Genetic and lifestyle-related determinants of circulatory amino acids and their effect on the risk of complex disorders

A thesis presented for the degree of Doctor of Philosophy by

Leila Abar

Department of Epidemiology and Biostatistics, School of Public Health,

Imperial College London

October 2022

Declaration of originality

I, Leila Abar, declare that this thesis represents my own work, and wherever I used materials that are not my own work, they have been referenced appropriately.

Copy right declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND).

Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

My first acknowledgement is that I have been very lucky to have a family that provided tremendous support and help. My father, Mr Mohammad-Reza Abar and my mother, Mrs Azam Estanbolchi, who taught me as a young girl to be brave, work hard and follow my dreams. They have done everything within their power to provide excellent education for me and have been the inspiration for any strength that I have learnt. My husband, Mr Alireza Beykpour and my little daughter Miss Hanna Beykpour, who have been my companion during the hard days of my PhD, and I owe them this work as they have sacrificed a lot. My siblings, Mrs Hanieh Abar, Miss Sara Abar, Miss Homeira Abar, and Dr Mohammad Abar, who have been a great team supporting my little daughter whenever she needed their help and attention while I have been very busy with this work, as well as supporting me with their love and kindness.

I would like to offer my heartfelt thanks and appreciation to my PhD supervisors, Dr Abbas Dehghan and Dr Verena Zuber. They have provided valuable support and feedback, also shown great patience and kindness in supervising me. This work could not be accomplished without their guidance!

Abbreviations

AD	Alzheimer's disease
Amino acids	AA's
ANNOVAR	Annotate variation
BCAA	Branch chain amino acids
COBLL1	Cordon-Bleu WH2 Repeat Protein Like 1
CPS1	Capcom Play System 1
CVD	Cardiovascular disease
DBP	Diastolic blood pressure
FUMA	Functional mapping and annotation
GCKR	Glucokinase regulatory protein
GLS2	Glutaminase 2
GWAS	Genome-wide association studies
GWGA	Genome-wide gene-based association analysis
HDL	High density lipoprotein
HWE	Hardy-Weinberg equilibrium
ICD	International Classification of Diseases
InSIDE	INstrument Strengths being Independent of Direct Effects
IR	Insulin resistance
IV's	Instrumental variables
IVW	Inverse variance weighted
LD	Linkage disequilibrium
LDL	Low density lipoprotein
LMO1	Rhombotin-1
MAF	Minor allele frequency
MAGMA	Multi-marker analysis of genomic Annotation
MIP	Major Intrinsic Protein of Lens Fibre
MR	Mendelian randomization
NHS	National Health Service
NMR	Nuclear magnetic resonance
OR	Odds ratio

PFAA	Plasma-free amino acid
PheWAS	Phenome wide association study
PRESSO	Pleiotropy and residual sum outlier
RCT	Randomized controlled trials
SBP	Systolic blood pressure
SLC	Solute-carrier gene
SNP	Single nucleotide polymorphisms
T2DM	Type 2 diabetes mellitus
TC	Total cholesterol
TG	Triglycerides
UKB	UK Biobank study

Abstract

Variations in serum amino acid (AA's) levels are associated with several diseases. However, the evidence from epidemiological studies remains conflicting and inconclusive. In this thesis, I aimed to identify genetic variants associated with serum AA's and investigate genetic and lifestyle-related determinants of circulatory AA's and their effect on the risk of complex disorders.

I conducted genome-wide association (GWAS) of circulating AA's including a sample of ~ 117 000 participants of the UK-Biobank (UKB) study, and the data were incorporated in a Mendelian Randomization- Phenome-wide association (MR-PheWAS) framework, using genotypic and phenotypic data of the UKB (n=488 377), to investigate the involvement of circulating AA's on a wide range of the clinical traits (n=617 with > 200 cases). To further validate the results, I performed MR analysis using the largest available GWAS studies on common complex disorders. I also performed an observational and MR analysis, exploring the effect of modifiable lifestyle risk factors on circulating AA's levels.

GWAS identified significant SNPs associated with serum AA's ranging from 48 to 138 SNPs for serum phenylalanine and glutamine, respectively. The total proportion of variance explained for serum AA's, ranged from 1% to 6%. MR analysis revealed that genetically elevated dietary protein and fat (% of energy) were associated with higher levels of BCAAs, while the carbohydrate showed an inverse association. Higher genetically determined accelerometer-measured physical activity was associated with higher valine levels. Higher genetically determined body fat (%) was nominally related to lower levels of glycine and glutamine. MR-PheWAS analysis and subsequent two-sample MR analysis provided evidence supporting potential causal associations between genetically elevated levels of serum valine and increased risk of type 2 diabetes, and elevated levels of serum alanine and BCAAs with higher levels of total cholesterol, triglyceride, and LDL, and lower HDL levels. These observed associations were successfully replicated using the largest available GWAS studies.

This study provides insight into the genomic and phenotypic profile of serum AA's concentrations, as well as novel findings that can be investigated further to understand the underlying mechanism of abnormal levels of serum AA's and their role in risk of complex disorders.

Table of Contents

Declaration of originality 2
Copy right declaration 2
Acknowledgements 3
Abbreviations 4
Abstract 6
Chapter One 11
Background and literature review 11
1.1 Introduction 12
1.1.1 Essential amino acids 13
1.1.2 non-Essential amino acids 13
1.2 Literature review on 9 proteogenic amino acids 15
1.2.1 Alanine 15
1.2.2 Branch chain amino acids (BCAA) 16
1.2.3 Glycine 20
1.2.4 Glutamine 21
1.2.5 Histidine 24
1.2.6 Aromatic amino acids 25
1.3 Epidemiology Methods to Investigate the effects of serum amino acids 27
1.3.1 Genome-Wide Association Studies 27
1.3.2 Phenome-Wide Association Studies 27
1.3.3 Mendelian Randomization 27
Chapter Two 31
GWAS and post-GWAS analysis 31
2.1 Introduction 32
2.2 Method 35
2.2.1 UK Biobank Data 35
2.2.2 UKB baseline data collection 35
2.2.3 UKB serum amino acids 36
2.2.4 GWAS on serum amino acid levels in UKB Population 37
2.2.5 Instrument selection using UKB GWAS on serum amino acids 37
2.2.6 Functional Annotation and Pathway Enrichment 38
2.3 Results 40
2.3.1 Functional mapping and annotation of serum amino acids 40
2.3.2 Genome-wide gene-based association analysis 43

2.3.3 Gene-set analysis 46
2.3.4 Gene-property analysis 47
2.4 Discussion 50
2.4.1 Strengths and limitations 52
Chapter Three 54
Modifiable lifestyle factors and their effect on serum amino acids 54
3.1 Introduction 55
3.2 Methods 58
3.2.1 Observational analysis 58
3.2.1.2 Data extraction 58
3.2.1.3 Statistical analysis 59
3.2.2 Genetic data on modifiable lifestyle factors 60
3.2.3 Statistical analysis 61
3.3 Results 61
3.3.1 Observational analysis 61
3.3.2 Mendelian Randomization analysis 62
3.4 Discussion 68
3.4.1 Strengths and limitations 72
Chapter Four 73
Phenome-wide association study on circulatory levels of serum amino acid and their potential causal role on common complex disorders 73
4.1 Introduction 74
4.2 Method 77
4.2.1 Study participants and recruitment 77
4.2.2 Genetic instruments selection 77
4.2.3 Phenotyping and mapping ICD-9 and ICD10 codes to phecode 78
4.2.4 Summary GWAS data 78
4.2.5 Statistical analysis 79
4.3 Results 82
4.3.1 Genetic loci associated with serum amino acids in UKB data 82
4.3.2 Genetically predicted serum amino acid levels and risk of diseases using one-sample MR-PheWAS analysis 82
4.3.3 MVMR analysis 86
4.3.4 Serum amino acids and type 2 diabetes 89
4.3.5 Serum amino acids and lipid levels 90

4.3.6 Serum amino acids and blood pressure 93
4.3.7 Serum amino acids and coronary artery disease 95
4.3.8 Serum amino acids and breast cancer 96
4.3.9 Serum amino acids and prostate cancer 97
4.3.10 Serum amino acids and Alzheimer's disease (AD) 98
4.4 Discussion 99
4.4.1 Strength and limitation 107
Chapter Five 108
Summary of findings & future perspectives 108
Summary of applied methods and findings 109
Conclusion 114
Future prospective 114
References 117
Appendix 159

Chapter One

Background and literature review

1.1 Introduction

Proteogenic amino acids are building blocks of protein synthesis in our body. The basic structure of all amino acids is the same, as each molecule has a central carbon atom linked together with a basic amino group, a carboxylic acid group, a hydrogen atom, an R-group, or a side-chain group (1). Each amino acid's chemical nature is determined by its R-group which interacts with other amino acids and their environment. The general formula of amino acids is R-CH(NH_2)-COOH (1).

Amino acids are involved in many biological and metabolic pathways and are important ingredients for human growth and development. Some of the proteogenic amino acids can be produced endogenously in our body (non-essential amino acids) and some should be obtained exogenously from diet or supplement (essential amino acids). A list of essential and non-essential proteogenic amino acids is presented in Table 1.

Table1. List of essential and non-essential proteogenic amino acids

Essential proteogenic amino acids	Non-essential proteogenic amino acids
Histidine	Arginine
Isoleucine	Alanine
Leucine	Asparagine
Lysine	Aspartic acid
Methionine	Cysteine
Phenylalanine	Glutamine
Threonine	Glycine
Tryptophan	Proline
Valine	Serine
-	Tyrosine

1.1.1 Essential amino acids

Histidine, a semi-essential amino acid (children should obtain it from diet), is needed for growth and tissue repair (2). It is metabolized to histamine, a neurotransmitter, that plays roles in immunity, gastric secretion, and sexual functions (2). BCAAs (isoleucine, leucine, and valine) have an important role in the modulation of glucose homeostasis, body weight, and cellular signalling and have beneficiary effects on decreasing exercise-induced fatigue, muscle recovery, and immune regulation (3). Lysine acts against herpes simplex viruses and is present in many forms of diet supplements (4). Methionine improves the tone and pliability of skin, hair, and strengthens nails, and is generally required for growth and tissue repair (5). Threonine is a component of tooth enamel, collagen, and elastin and thus is required for healthy teeth and skin (6). Threonine also plays role in fat metabolism and can alleviate anxiety and mild depression (6). Tryptophan as an essential amino acid is necessary for infants' proper growth and is a precursor of serotonin and melatonin. Serotonin regulates appetite, sleep, mood, and pain, and melatonin plays role in the regulation of sleep (7).

1.1.2 non-Essential amino acids

L-arginine is suggested to act as a vasodilator, opening (dilating) blood vessels and therefore has positive effects on atherosclerosis, a known risk factor for cardiovascular disease, redox stress, and the inflammatory process, regulation of synaptic plasticity and neurogenesis, and modulation of glucose metabolism and insulin activity (8, 9). Alanine plays as a source of energy for muscles and the central nervous system and boosts the immune system (10). Asparagine is a beta-amido derivative of aspartic acid which has a pivotal role in the biosynthesis of glycoproteins and many other proteins (11). Cysteine plays role in collagen production and the elasticity of skin (12). Glutamine, the most abundant amino acid in the body plays a pivotal role in immunity (13). Glycine is claimed to have an anti-inflammatory effect (14). Proline consists of collagen and has an essential role in the proper functioning of joints

and tendons (15). Serine is required for the metabolism of fats, fatty acids, and cell membranes and improves muscle growth and the immune system (16). Tyrosine, produced by phenylalanine, is claimed to improve alertness and increase focus (17).

In this chapter, I will review evidence on the biology, structure, and utilization of 9 proteogenic amino acids as they are the only amino acids with their serum levels data available from participants of UK Biobank study (UKB) (http://biobank.ndph.ox.ac.uk/ukb/label.cgi?id=220), including alanine, Branch chain amino acids (BCAA), glycine, glutamine, histidine, and aromatic amino acids (phenylalanine and tyrosine). In addition, I will review epidemiological evidence on the association of each previously mentioned amino acid with complex disorders. In the end, I will also discuss the molecular and genetic epidemiological methods to investigate serum amino acid association with the risk of complex disorders.

1.2 Literature review on 9 proteogenic amino acids

1.2.1 Alanine

Alanine is a non-essential proteogenic amino acid. It is also a glucogenic amino acid that has a critical role in glucose production and hence is involved in glucose metabolism and is a source of energy for muscle cells and the tissues of the brain and the central nervous system (18). When protein breakdown occurs in skeletal muscles during fasting or other stresses, a large quantity of alanine is released into the bloodstream and is carried to the liver, and is readily converted to glutamate and pyruvate by glutamate-pyruvate transaminase (alanine transaminase), which has a catalytic action, with α-ketoglutarate. Then the produced pyruvate is converted to glucose through the glucogenic pathway and then the glucose enters the bloodstream by glucose transporter 2 and can be readily available to be used as a source of energy for the muscles. Finally, the initially transported amino group of alanine will be converted to urea and will be excreted via the urea cycle.

Alanine supplement, such as L-alanine, is used in cases of hypoglycaemia by stimulating and prompting the release of glucose into the bloodstream and avoiding low blood sugar or insulin shocks (19). Other amino acids such as branched-chain amino acids including isoleucine, leucine, and valine can be converted to alanine (20).

A recent Mendelian randomization (MR) study found a decreased risk of type 2 diabetes with plasma alanine levels (21). This MR study used summary-level data from the DIAbetes Genetics Replication And Meta-analysis consortium including 74 124 type 2 diabetes cases and 824 006 controls from European ancestry and they also replicated the causal associations using the FinnGen consortium consisting of 11 006 type 2 diabetes cases and 82 655 controls of European ancestry (21).

1.2.2 Branch chain amino acids (BCAA)

Branch chain amino acids including isoleucine, leucine, and valine are essential amino acids that should be obtained exogenously from the diet, mainly from animal-sourced foods (22). A healthy adult requires around 10 g/day but this amount can vary between individuals (23). Rich foods in leucine are beef, raw salmon, eggs, and wheat products, and foods rich in isoleucine and valine are generally meat and animal products as well as dried peas and walnuts (23). BCAAs catabolism is different from the catabolism of other essential amino acids, as it catalysed in extrahepatic tissues via two enzymes (24):

1. Branched-chain amino acid aminotransferases (BCAT)
2. Branched-chain α-keto acid dehydrogenase enzyme complex (BCKDC).

After transamination of the BCAAs metabolites by BCAT and decarboxylation by BCKDC, they are further converted to acetyl-CoA and Succinyl-CoA through a series of enzymatic reactions, and at the final step they enter the metabolism of tricarboxylic acid (TCA) cycle (24-27).

BCAAs metabolism dysregulation is related to a range of diseases such as Maple Syrup Urine Disease (MSUD), an autosomal disorder that is caused by decreased activity of the BCKDC enzyme (28).

BCAAs have an important role not only in protein synthesis but also in nutrient-sensitive signalling pathways such as phosphoinositide 3-kinase-protein kinase B (PI3K-AKT) and mammalian target of rapamycin (mTOR)(24, 29-31). The mTOR is the main regulator of mammalian cell growth (24, 32) and cell proliferation and migration (24, 33), and also has a pivotal role in the regulation of hypertrophy (24, 34), which is an increase in cell mass without cellular proliferation, and insulin resistance (IR) (35). In addition, the mTOR pathway is the

most notorious amino acid sensor for maintaining the amino acids balance via amino acid transporters, and/or biosynthesis and degradation of amino acids and protein in cells (24).

BCAAs regulate glucose metabolism as well as lipid metabolism through the PI3K-AKT (also referred to as protein kinase B, PKB) pathway (24, 33). BCAAs act as a nitrogen donor for other amino acids including alanine and glutamine and act as a nitrogen shuttle for organs (24, 36). It also plays an important role in multiple metabolic functions through specific signalling pathways (24, 29, 31). For instance, increased gluconeogenesis and glucose intolerance through glutamate transamination to alanine can occur as a result of increased BCAAs catabolic flux (24, 37).

Isoleucine inhibits the increase of plasma glucose levels by stimulating the glucose uptake in skeletal muscle and prevents the development of visceral obesity and hyperinsulinemia (24, 38, 39). Leucine regulates the levels of three BCAAs concentrations in muscle tissues, and lower levels of isoleucine and valine may induce the activation of cardiac BCKDC in humans (24, 40). Moreover, excess dietary leucine resulting in increased levels of it in the bloodstream, has been shown to decrease growth performance and increase BCAAs catabolism in pigs (24, 41). Leucine also has shown to have a pivotal role in stimulating protein synthesis in skeletal muscles independently of increases in serum insulin (42) and also inhibiting protein degradation (24, 43).

There is evidence that shows BCAAs deficiency causes a significant change in the metabolism of lipids in white adipose tissue, for example, leucine deprivation can reduce fat mass by increasing energy expenditure through regulation of fat loss by the central nervous system (44), and isoleucine and valine deficiencies have the same effect on reducing fat mass (24, 45).

BCAAs are a popular supplement among gym-goers and athletes, particularly bodybuilders for accelerating muscle growth, as BCAAs supplements have beneficiary effects on decreasing

exercise-induced fatigue, muscle recovery, and immune regulation (3, 46) and additionally leucine supplementation may improve muscle protein synthesis among older individuals (47). Also, it has been reported in a cross-sectional study that a higher dietary intake of BCAAs is associated with a lower prevalence of overweight and obesity in healthy middle-aged adults from East Asian and Western countries (48).

BCAAs supplementation has also been shown to be an effective therapy for patients undergoing hepatic operation by reducing complication risk and hospital stay days and this evidence is from a meta-analysis study that included 11 randomized controlled trials (RCT) representing 510 patients (49). Moreover, another recent meta-analysis study that investigated the effect of oral administration of BCAAs on the mortality of patients with hepatocellular carcinoma (HCC) after treatment, including 14 studies covering 1 179 patients, reported that BCAA supplementation might improve liver reserve function and long-term prognosis of HCC patients and it can reduce the long-term mortality of radiofrequency ablation (RFA) patients, while no significant effect on hepatectomy patients (50).

Despite the reported beneficiary effects of BCAAs, epidemiological evidence have shown that high levels of circulating BCAAs are associated with type 2 diabetes mellitus (T2DM) and also higher levels of BCAAs in obese individuals than in normal-weight individuals which are associated with insulin resistance (IR)(24, 31, 51) as well as associating with increased risk of CVD, particularly stroke (52, 53).

Higher consumption of dietary BCAAs can accelerate metabolic abnormalities development including metabolic syndrome, and diabetes (24, 54). It has been suggested that concentrations of plasma BCAAs can be a good indicator of impaired IR in pre-diabetic patients than plasma glucose levels (24, 55) and in both human and animal studies it has been shown that the observed association between BCAAs and IR might be dependent on energy homeostasis, and

therefore circulating BCAAs can be used as potential indicators of insulin sensitivity (24, 56). Higher levels of BCAAs concentrations are significantly related to the development of T2DM as plasma-free amino acid (PFAA) profiles, especially levels of BCAAs concentrations are reported to be altered before the T2DM development (57).

A recent MR study conducted by Yuan and Larsson in 2020 (21), that mentioned previously showing the decreased risk of T2D with plasma alanine, showed an increased risk of T2D with serum BCAAs (isoleucine, leucine, and valine).

Another MR study, also reported that higher circulating levels of BCAAs were associated with a higher risk of type 2 diabetes and they concluded a causal role of BCAA metabolism in the aetiology of type 2 diabetes (58). This study used a meta-analysis of GWAS studies of Fenland, KORA and TwinsUK (total sample=16 596 individuals) for the amino acid traits and used a meta-analysis of GWAS studies of DIAGRAM, EPIC-InterAct, GoDARTs, and UK Biobank studies (47 877 type 2 diabetes cases and 267 694 controls)(58).

Alterations in the metabolism of BCAAs are found to be associated with cardiometabolic risk factors in cross-sectional studies and have been implicated in reduced cardiometabolic health including IR, elevated blood pressure, dyslipidaemia, and indicators of coronary artery disease (37, 59-63). In a prospective cohort of US Women's Health Study with 18.6 years of follow-up, circulating plasma levels of BCAAs were found to be positively associated with the incidence of coronary vascular disease (CVD) (59). This study concluded that impaired metabolism of BCAAs signifies a shared pathway and common aetiology for T2D and CVD, particularly in women who develop T2D before to a CVD event (59).

Apart from the evidence on the association between BCAAs and cardiometabolic health, there is epidemiological evidence suggesting a potential effect of BCAAs on cancer.

BCAAs are consumed by tumours in several biosynthetic pathways and therefore BCAAs act as an energy source for cancer cells (64, 65) that obtain it from circulation or surrounding tissues (66). Elevated plasma levels of BCAAs are found to be significantly associated with pancreatic cancer risk (66-68). Furthermore, studies have reported that plasma and tissue levels of BCAAs are increased in breast cancer, supplemented by the higher expression of cytosolic branched-chain aminotransferase 1 (BCAT1) that can result in activating the BCAAs catabolism in human breast cancer (24, 65) and in general epidemiological studies strongly point to a pivotal role of BCAT1 in multiple cancer types (29).

1.2.3 Glycine

Glycine is the simplest non-essential proteogenic amino acid (69) and therefore it is synthesised in the body from the amino acid serine, mostly catalysed by the serine hydroxy-methyl-transferase enzyme via cofactor pyridoxal phosphate (70). It can also be synthesised from other amino acids such as threonine in the liver and kidneys (71).

Glycine is a key amino acid in several metabolic reactions and is the major inhibitory neurotransmitter in the spinal cord and brain stem that facilitates the transmission of nerve impulses (14). It also has an anti-inflammatory effect and is a cytoprotective, and immune-modulating amino acid (14) and is involved in the production of DNA, phospholipids, and collagen and is the most abundant amino acid in collagen triple helices (72).

Glycine also regulates blood glucose levels by converting it into energy (73) and is essential for dietary fat digestion and the absorption of long-chain fatty acids (71).

Glycine supplementation in rodents have shown to lower blood pressure (74-77) and also a similar effect was observed in one small intervention study in humans (78). Glycine

supplements are also claimed to have anti-aging effects as 30% of collagen consists of glycine and studies in rodents showed glycine supplementation extended lifespan (79).

It has been reported in large-scale epidemiological studies that circulating levels of glycine are associated with lower incidence of T2DM (77, 80-83) and myocardial infarction (MI)(77, 84).

A recent Mendelian randomization (MR) study has shown that lower levels of glycine concentrations might be causally associated with the risk of T2D, mainly determined by *CPS1* gene (77, 82). A meta-analysis of genome-wide association studies that investigated the causality and potential mechanisms of the association between glycine and cardio-metabolic disease and included 80 003 participants of European ancestry have shown that glycine is genetically associated with a lower risk of coronary heart disease (CHD) and also implicated the role of blood pressure for their observed association (77). To investigate the possible mechanism, they examined the downstream effects of genetic differences in glycine levels on CHD risk factors such as systolic (SBP) and diastolic blood pressure (DBP), blood lipids (total cholesterol, triglycerides, LDL, and HDL) and 13 other potentially relevant blood cell traits (77). They found that genetically predicted glycine, using a 24 SNP score, was significantly associated with lower genetically predicted SBP and nominally with lower genetically predicted DBP and no association with blood lipids or other blood cell traits (77). Furthermore, this study reported that genetically lower levels of glycine were significantly associated with genetically higher IR (77).

1.2.4 Glutamine

Glutamine is the most abundant and versatile non-essential proteogenic amino acid in the body and a nitrogen transporter and NH3 carrier that transfer nitrogen among organs (13). It has a regulatory role in protein synthesis, cell defence and repair, cell integrity (apoptosis, cell

proliferation), degradation, and metabolism (oxidative fuel, gluconeogenic and lipogenic precursor)(85).

Glutamine also plays a role in IR and insulin secretion, as a study in vivo has shown that oral glutamine effectively increases circulating glucagon-like peptide 1 (GLP-1), glucose-dependent insulinotropic polypeptide (GIP), which are hormones that have a key role in meal-related insulin secretion and can stimulate insulin secretion in individuals with obesity and type 2 diabetes.

In most cells, glutamine can be used as a substrate for nucleotide synthesis including purines, pyrimidines, amino sugar, and nicotinamide adenine dinucleotide phosphate (NADPH), and antioxidants (13). In addition to its role in protein synthesis and metabolism, glutamine also regulates the expression of many metabolism-related genes, signal transduction, and activates intracellular signalling pathways (85).

Healthy individuals normally have a plentiful large amount of glutamine in their body as its rate of utilization by immune cells is comparable to glucose (13). Under catabolic conditions for example sepsis, recovery from burns or surgery, and malnutrition, and also during high intensity physical exercise, lymphocytes, neutrophils, and macrophages which are cells of the immune system, use glutamine (13, 86, 87). Therefore, glutamine is currently part of clinical nutrition supplementation recommendations for immune-suppressed individuals (13). Glutamine supplementation has shown promising effects in individuals with cardiometabolic and sickle cell disease, however, it has failed to show a significant reduction in mortality rate in epidemiological studies (88).

A recent meta-analysis study of 12 RCTs that investigated the effect of glutamine supplementation on cardio-metabolic risk factors and inflammatory markers, including Fasting plasma glucose (FPG), insulin, homeostasis model of assessment-estimated insulin resistance (HOMA-IR), triglyceride (TG), C-reactive protein (CRP), Interleukin 6 (IL-6), Glutathione (GSH), Interleukin 1(IL-1), and Tumour necrosis factor alpha (TNF-α), showed that glutamine supplementation might have a positive effect on FPG and CRP (89). However, they found no association between glutamine supplementation and other cardio-metabolic risk factors (89).

The role of glutamine in cancer has been investigated through several epidemiological studies (13, 90, 91). The accumulating evidence has found that cancer cells consume glutamine at the highest rate, and they also require glutamine for cell proliferation (92-94). This is because glutamine is a source of carbon and nitrogen supply for numerous biosynthetic reactions and its carbon in the form of α-ketoglutarate (αKG) is an important substrate for the TCA cycle (13).

Furthermore, studies have highlighted the possible role of serum glutamine on Alzheimer's disease. An MR study that utilized a GWAS study among 24 925 participants of 14 European cohorts for its metabolite traits, and International Genomics of Alzheimer's Project (IGAP) including 17 008 Alzheimer's disease cases and 37 154 controls, has reported an inverse association between circulating levels of glutamine and risk of Alzheimer's disease (95). Previous studies implicate the alterations in glutamine metabolism as possible early transformations in the pathogenesis of Alzheimer's disease (95, 96). Therefore, it has been suggested that exogenous glutamine may protect against the accumulation of amyloid plaques in early phase of the disease, as the amyloid plaques is associated with inflammation, oxidative stress and mitochondrial dysfunction (95, 97).

1.2.5 Histidine

Histidine is a semi-essential amino acid and particularly children should obtain it from food and supplement (2). Our body requires histidine for growth and tissue repair (2). Histidine plays a key role in the maintenance of myelin sheaths for protecting nerve cells (2). Histidine is metabolized to the neurotransmitter histamine which is essential for immunity, gastric secretion, and sexual functions as well as for blood cell manufacture and protecting tissues from damage by radiation and heavy metals (2).

In a systematic review and meta-analysis of randomized controlled trials (n=23) investigating the effect of supplementation with histidine-containing dipeptides on obesity and glycaemic outcome, it was reported that the supplementation improved glycaemic outcomes and central obesity reduced (98). Moreover, the effect of histidine-containing dipeptides on lipid profile was examined in another systematic review and meta-analysis of RCTs (n=21), and the study indicated that the supplementation has a role in improving dyslipidaemia (99).

A study that conducted whole exome sequencing on 1 152 African Americans in the Atherosclerosis Risk in Communities (ARIC) study and focused on loss-of-function (LoF) variants have found that three LoF mutations were related to elevated levels of histidine, which are shown to be inversely associated with chronic heart disease risk in individuals from both African Americans and European Americans (100).

Studies have shown that people have lower histidine levels when are under oxidative stress or have inflammatory diseases including rheumatoid arthritis (100-102). Inflammation has a crucial role in atherosclerosis and cardiovascular disease(100).

The possible mechanism, in which the serum histidine may reduce the risk of CHD, is suggested via reducing oxidative stress or inflammation suppression as pro-inflammatory cytokine is suppressed by histidine through pathways in adipocytes (100).

1.2.6 Aromatic amino acids

1.2.6.1 Phenylalanine and tyrosine

Phenylalanine is an essential amino acid that should be obtained from food or supplement. It plays a key role in the biosynthesis of other amino acids including tyrosine. Therefore, tyrosine is a non-essential proteogenic amino acid (103). However, in phenylketonuria patients who cannot convert phenylalanine into tyrosine due to a lack of phenylalanine hydroxylase, tyrosine is considered an essential amino acid that should be obtained from diet or supplement (103). While the L-form of Phenylalanine is involved in protein synthesis, the D-form acts as a painkiller (104).

Tyrosine, produced by phenylalanine, is an essential precursor of catecholamines, which is a group of neurotransmitters including dopamine, norepinephrine and epinephrine, and is also a precursor of thyroid hormones, and melanin (103).

In 1989 a study suggest that tyrosine administration, the precursor of catecholamines, can reduce the neurochemical deficits during acutely stressful situations as the brain depletes norepinephrine and dopamine (105). However, since then no consensus achieved among observational studies for the beneficiary effect of tyrosine supplementation to improve physical/cognitive function (106, 107).

Two recent MR studies showed a protective effect of serum tyrosine on type 2 diabetes but in contrast to tyrosine a positive significant association between phenylalanine and the risk of type 2 diabetes (21, 108). The first study was done by Yuan and Larsson in 2020, which was mentioned earlier reporting the results for serum alanine and BCAAs as well (21). The second study conducted by Jager et al., in 2020, reported that higher tyrosine levels might have a causal inverse association with type 2 diabetes, independent of other diabetes-associated amino acids. This study used genome-wide association studies (GWAS) of a sub-cohort of the European

Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study (up to 2265 participants) for amino acid traits and DIAGRAM (DIAbetes Genetics Replication And Meta-analysis, 74 124 cases and 824 006 controls) consortium for diabetes (108).

It has been suggested that tyrosine might affect the risk of type 2 diabetes via its critical role in the neurotransmitter dopamine and L-DOPA synthesis (108). In vivo studies, dopamine is suggested to influence glucose homeostasis and beta cell function but in human studies the research is incomplete (108-110).

Furthermore, there is evidence for the relationship between tyrosine and Non- Alcoholic Fatty Liver Disease (NAFLD)(111). A recent MR study using Estonian Biobank (10 809 individuals including 359 patients with NAFLD) data revealed a potential causal effect of NAFLD on blood concentrations of tyrosine and suggested that blood tyrosine levels can be a new biomarker for NAFLD (111).

1.3 Epidemiology Methods to Investigate the effects of serum amino acids

1.3.1 Genome-Wide Association Studies

Genome-wide association study (GWAS) is a method to identify genetic loci (most commonly as single nucleotide polymorphisms (SNPs)) associated with complex diseases or phenotypes of interest through the human genome of a population (112). In molecular genetic epidemiology, SNP analysis is a powerful tool to identify variations in DNA copy numbers and point mutations (113). In the past 10-15 years, thousands of GWAS have been conducted on a wide range of traits and disorders which provide a wealth of knowledge on genetic determinants of complex traits and disorders.

1.3.2 Phenome-Wide Association Studies

PheWAS is a method of testing comprehensive associations between genetic variants and a wide range of coded phenotypes and traits (114). PheWAS is designed to investigate which of many phenotypes are possibly associated with a given genetic variant and is a reverse genetics approach that begins with a genotype and then systematically queries a large number of phenotypes (115).

PheWAS utilizes electronic health records (EHR) and has been validated through replication of hundreds of known genotype-phenotype associations, and their utilization has distinguished between true pleiotropy and clinical comorbidity (114, 115).

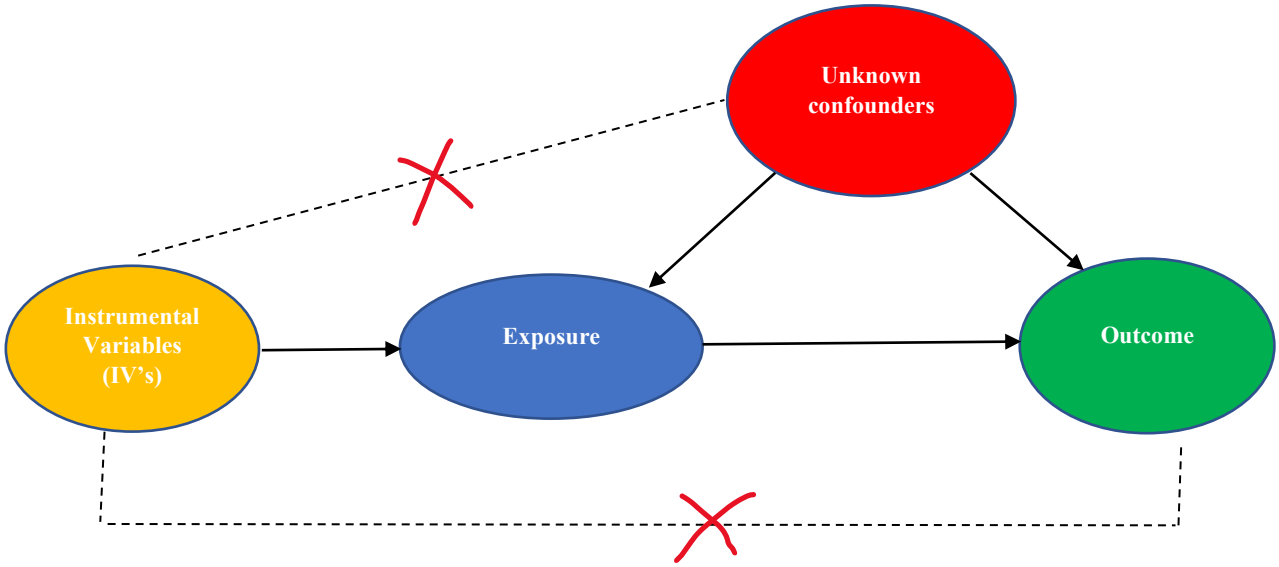
1.3.3 Mendelian Randomization

Mendelian Randomization (MR) uses genetic variants as instrumental variables to investigate the causal association between potentially modifiable risk factors and health outcomes in observational data (116). The MR method can overcome the limitation of confounding in

observational studies and other various biases that have proved to be unreliable indicators of the causal effects of modifiable exposures on disease outcomes (116). Therefore, the MR method generates more reliable evidence regarding which interventions should produce health benefits as well as enables the researchers to predict the related mediating pathways (117, 118) and adverse effects (119), and more importantly the direction of causality (116, 120).

The MR method reliability is based on assumptions that the instrument is only associated with the exposure of interest (relevance assumption), and not directly with the outcome of interest (exclusion restriction), nor any confounders of the association between the exposure and the outcome (independence assumption) (121)(**Figure 1**). Therefore, violation of these assumptions potentially leads to biased MR estimates (122).

Figure1. Schematic of instrumental variable analysis to generate causal estimates through MR



There is an increasing number of genetic variants which are demonstrated to be associated with exposures in GWAS, and the public availability of these summarized data on genetic associations with exposures and outcomes from large consortia made the MR analysis relatively simple to perform. However, using GWAS findings as instruments for MR analysis can be challenging due to the pleiotropic effect of genetic variants. GWAS method allows to

identification of many of these SNPs which would occur more frequently in individuals with disease than those without (123). Genetic human studies mimic the design of RCTs in observational studies with no drug and lifestyle factors intervention requirement (124). Therefore, disease-associated SNPs identified through GWAS studies might be interpreted as an under-used source of randomized human evidence to help with the identification of drug targets and validation (124). Furthermore, GWAS findings could point to genetic variants, randomly assigned at conception. This allows us to imitate not only the drug's effect and side effects of medications but also lifestyle factors such as physical activity or diet (125).

Pleiotropy is a phenomenon where a genetic variant associated with multiple traits either through a single biological pathway (vertical pleiotropy) or on discrete pathways (horizontal pleiotropy) (126). Horizontal pleiotropy is a challenge for MR analysis as a fundamental assumption of MR is the 'no horizontal pleiotropy' assumption (also called the 'exclusion-restriction criterion') (127) and consequently violation of this assumption results in inaccurate causal estimates, loss of statistical power and potential false-positive causal relationships (127). There are several methods that have been developed to allow for genetic pleiotropy including median estimator and MR Egger regression (128-130). These methods enable sensitivity analyses to be performed and investigate whether findings depend on the assumption that no pleiotropic effects exist among all the variants (128).

The median estimator method can provide reliable evidence if at least half the genetic variants have no pleiotropic effects (128, 130). MR Egger regression allows all variants to have pleiotropic effects, provided they are not proportional to the variants' effects on the risk factor of interest (128). However, the MR Egger regression method generates less precise estimates than other methods (128).

Another source of bias in MR approach is utilizing weak instruments, which can occur when using one or more genetic variants that only explain a small proportion of the variation in the risk factor combined with a small sample size (128). The F statistics can be applied to quantify the level of weak instrument bias and a value of 10 corresponds to an approximate 10% relative bias (128).

Chapter Two

GWAS and post-GWAS analysis

2.1 Introduction

In this chapter, I will explore the genetic architecture of 9 proteogenic amino acids including alanine, BCAAs (isoleucine, leucine, and valine), glycine, glutamine, histidine, phenylalanine, and tyrosine. By identifying genetic variants, I will further assess the biological pathways associated with each amino acid.

It has been previously discussed in chapter one, that the circulating levels of amino acids are associated with the development of diseases such as type 2 diabetes (21, 24, 51, 77, 131, 132) and cardiovascular disease (52, 53, 59) as well as being associated with several cardiometabolic risk factors including overweight and obesity, insulin resistance, blood pressure and lipid profile (37, 59-63, 77, 89), and with other diseases such as inflammatory disease (133, 134), Non-Alcoholic Fatty Liver disease (111), and Alzheimer's disease (95, 96). Many metabolomics studies also investigated the relationship between serum amino acids and gynaecological and breast cancers (135-149), and serum amino acids levels are suggested to be a useful marker for disease status in patients with head and neck cancer (150) as well as for plasma free amino acids (PFAAs) profile, which serves as one of the biological markers for cancer patients (140, 151). However, its exact biological pathways and mechanism are yet to be elucidated (140).

Most of these discussed observed associations are from observational studies, which are subject to bias, and confounding and the ones using the MR approach investigated the limited number of amino acids, and the evidence on the role of all serum amino acids on complex disorders is rare. In addition, observational studies could not investigate the causality and the direction of the effect of the studied exposure on their outcome of interest.

Therefore, identifying the genetic variants associated with the circulating serum amino acids concentrations can be utilised to investigate their effect on chronic conditions. The genetic

markers are inherited during conception and thus prevent the biases afflicted in traditional epidemiological studies. However, the genetic variants cannot generalize the effect of circulating serum amino acid levels in instances where they markedly rise, as they assess small effect size over a lifetime. GWAS approach unravels the genetic components of serum amino acids and may elucidate the role of serum amino acids levels in the pathophysiology of complex disorders.

In the past decade, thousands of genome-wide association studies (GWAS) have been conducted (152) aiming for identifying genotypes-phenotypes associations through investigating the allele frequency differences in genetic variants, single-nucleotide polymorphisms (SNPs)(153), among individuals who are from the same ancestry but are phenotypically different (153). More specifically, over 5 700 GWAS studies have now been conducted detecting more than 3 300 traits (152) consisting of GWAS studies with sample sizes of millions of participants (154). These well-powered GWAS studies (152) should translate the identified genetic loci into actionable variants, directing to functional genomics experimentation and drug target assessment (152, 155). Thus the main challenge in GWAS studies is pinpointing the potential causal variants (152, 153). However, identifying the causal variants from significant associated SNPs rarely happens as many of the GWAS hits are located in non-coding or intergenic regions (152, 156). Genomic regions include various correlated SNPs covering closely located genes, but not all of those genes necessarily are relevant to the disease (152). Therefore, because of the correlated nature of those closely located SNPs, distinguishing between disease related genes with non-related ones is unlikely, by just considering the association *P*-values (152). As a result, to identify the causal genes, integrating available information regarding regional linkage disequilibrium (LD) patterns, functional consequences of correlated SNPs, and their effects on gene expression is required (152). This

information increases our insight into the biological mechanisms underlying the diseases and traits (152).

FUMA, functional mapping and annotation of genetic associations web tool is developed by Watanabe et al in 2017 and is available at http://fuma.ctglab.nl. This online web tool functionally annotates GWAS findings and prioritizes the most likely causal SNPs and genes via evidence from repositories of biological data and tools (n=18)(152).

Therefore, this chapter aims to present a large-scale GWAS on serum amino acids (alanine, BCAAs, glycine, glutamine, phenylalanine, and tyrosine) of UK Biobank study participants to identify genetic determinants of previously mentioned serum amino acids levels and a post-GWAS analysis to explore the related biological pathways using FUMA web-tool.

2.2 Method

2.2.1 UK Biobank Data

For my project, I utilized data from UKB under the approved application **ID 52569**. The UKB was established by the Wellcome Trust medical charity, the Medical Research Council, the Department of Health, the Scottish Government, and the Northwest Regional Development Agency and is supported by the National Health Service (NHS) patient registers.

The UKB is a large longitudinal population-based cohort of ~500 000 participants who were recruited between 2006-2010 and were aged 40-69 at recruitment. This age group was selected because of being at risk of developing a wide range of chronic diseases and therefore enables investigators and researchers to examine a wide range of disease determinants including genetic, proteomic, and biochemical that attributes to a group of lifestyle and environmental factors.

The UKB population was recruited from different parts of the country including 89% in England, 7% in Scotland; and 4% in Wales.

2.2.2 UKB baseline data collection

The baseline assessment comprised of electronic signed consent, a two-part questionnaire including touch-screen health and lifestyle questionnaires following a brief interview to collect data on socio-demographics and occupation; lifestyle exposures diet, physical activity, and body fat percentage; early life exposures; psychological state; cognitive function; family history of illness; and medical history and general health.

In addition, physical measurements including height, blood pressure, and weight were measured as well as blood, urine, and saliva sample collection.

The personal medical history of participants was asked by a trained nurse, following the touch-screen questionnaire to confirm or correct reported illnesses or disabilities. The self-reported information was validated through the linkage to the participant's NHS medical records. Moreover, the participants were both asked via questionnaire and interviewed to confirm reported information about any common prescription, over the counter medication, and intake of dietary vitamins/supplements.

2.2.3 UKB serum amino acids

NMR metabolic biomarkers from EDTA plasma samples of around 117 000 European UK Biobank participants was generated by Nightingale Health. Metabolic biomarkers, including 9 serum amino acids, were measured from randomly selected EDTA plasma samples (aliquot 3) via a high-throughput NMR-based metabolic biomarker profiling platform developed by Nightingale Health Ltd (http://biobank.ndph.ox.ac.uk/ukb/label.cgi?id=220).

The available biomarkers for serum amino acids, that I used for my project, include serum levels of alanine (**Data-Field:23460**), glycine (**Data-Field:23462**), glutamine (**Data-Field:23461**), histidine (**Data-Field:23463**), isoleucine (**Data-Field:23465**), leucine (**Data-Field:23466**), phenylalanine (**Data-Field:23468**), tyrosine (**Data-Field:23469**), and valine (**Data-Field:23467**)(https://biobank.ndph.ox.ac.uk/ukb/label.cgi?id=220).

2.2.4 GWAS on serum amino acid levels in UKB Population

I performed a linear mixed model regression using BOLT-LMM version (2.3) in Linux, for the UKB GWAS. This BOLT-LMM method considers population stratification and relatedness in the sample population and therefore it can be applied with minimal sample exclusions (157). This approach boosts the power for detecting any potential association by modelling more realistic, non-infinitesimal genetic architectures using a Bayesian mixture prior to marker effect sizes (158). Therefore, this approach controls false positives while increasing the power (158). This method allows the assumption of complex genetic architecture for traits; thus, it does not assume that all variants are causal with small effect sizes to form a Gaussian distribution (158).

All serum amino acid levels of individuals from European ancestry of the UKB data were standardized and normalized via rank-based inverse normal transformation. I filtered variants with minor allele frequency (MAF) < 0.01, genotyping rate < 0.015 and Hardy-Weinberg equilibrium p-value < 0.0001. Age, sex and the first 10 genetic principal components were included in the model as covariates.

Furthermore, I calculated the total proportion of variance explained (R^2) and F-statistic (F) by regressing all serum amino acids on every single genetic instrument.

2.2.5 Instrument selection using UKB GWAS on serum amino acids

Through the identified SNPs significantly associated with each serum amino acids in the previous GWAS step, multi-allelic SNPs were excluded, and the remaining SNPs were clumped with r^2 <0.001 using TwoSampleMR package (0.5.6) in R software.

Those associated SNPs with P-value <1e-05 and F-statistics of > 10 (calculated by regressing serum levels per each SNPs) were selected as instruments for functional mapping, and pathway enrichment analysis in the next step.

2.2.6 Functional Annotation and Pathway Enrichment

I applied a range of Functional annotation analyses using FUMA web server (http://fuma.ctglab.nl)(152). The FUMA web tool contains information from 18 biological data repositories to characterize SNPs according to gene function, mapped genes and biological pathways and associations with other phenotypes (152).

All significant SNPs ($p < 1e\text{-}05$) and SNPs in LD with them ($r^2 \geq 0.6$) from the UKB GWAS summary statistics in the previous step, were annotated for functional consequences on gene functions, based on Ensemble genes (build 85) using ANNOVAR(152, 159). Independent significant SNPs and correlated SNPs were linked to the GWAS catalogue to postulate known associations of the SNPs in the risk loci that were reported previously with several phenotypes (152).

Annotated SNPs were subsequently mapped to the genes, using positional mapping (physical distance): based on annotations obtained from ANNOVAR (159), and the physical distance of SNP from known protein-coding genes is set to the window of 10kb (default), in the human reference assembly (hg19/GRCH37)(152).

2.2.6.1 Gene-Based Analysis

For genome-wide gene-based association analysis (GWGA), FUMA implements MAGMA gene-based analysis (152, 160) by mapping the input SNPs to protein-coding genes attained from Ensemble build 92.

2.2.6.2 Gene-Set Enrichment Analysis

Same as for gene-based analysis, FUMA implements MAGMA (152, 160) for gene-set enrichment analysis, and is conducted for curated gene sets and gene ontology terms attained from molecular signature database (MsigDB) v5.2. Gene-set enrichment analysis assesses if genes within the gene set are more strongly associated with any of the phenotypes than other genes.

2.2.6.3 Gene-Property Analysis

To test the association between tissue-specific gene expression and serum amino acids gene associations, MAGMA gene-property analysis conducted using FUMA (152, 160). This analysis is based on the computed gene-based p-values in the gene-based analysis.

2.3 Results

The number of identified independent (r^2 <0.001) significant SNPs ($p < 1 \times 10^{-5}$) associated with serum amino acids through the UKB GWAS analysis are shown in **Table1**. A full list of selected SNPs is provided in **Tables S1-S9**.

The total proportion of variance explained (R^2) for serum amino acids ranged from 1% for isoleucine with a combined F statistic of 24.35 for 60 SNPs, to 6% for serum glycine with a combined F statistic of 53.58 for 120 SNPs. The proportion of variance explained for serum BCAAs (isoleucine, leucine and valine) ranged from 1% to 3% with a F statistic of 24 to 34 (see **Table 1**).

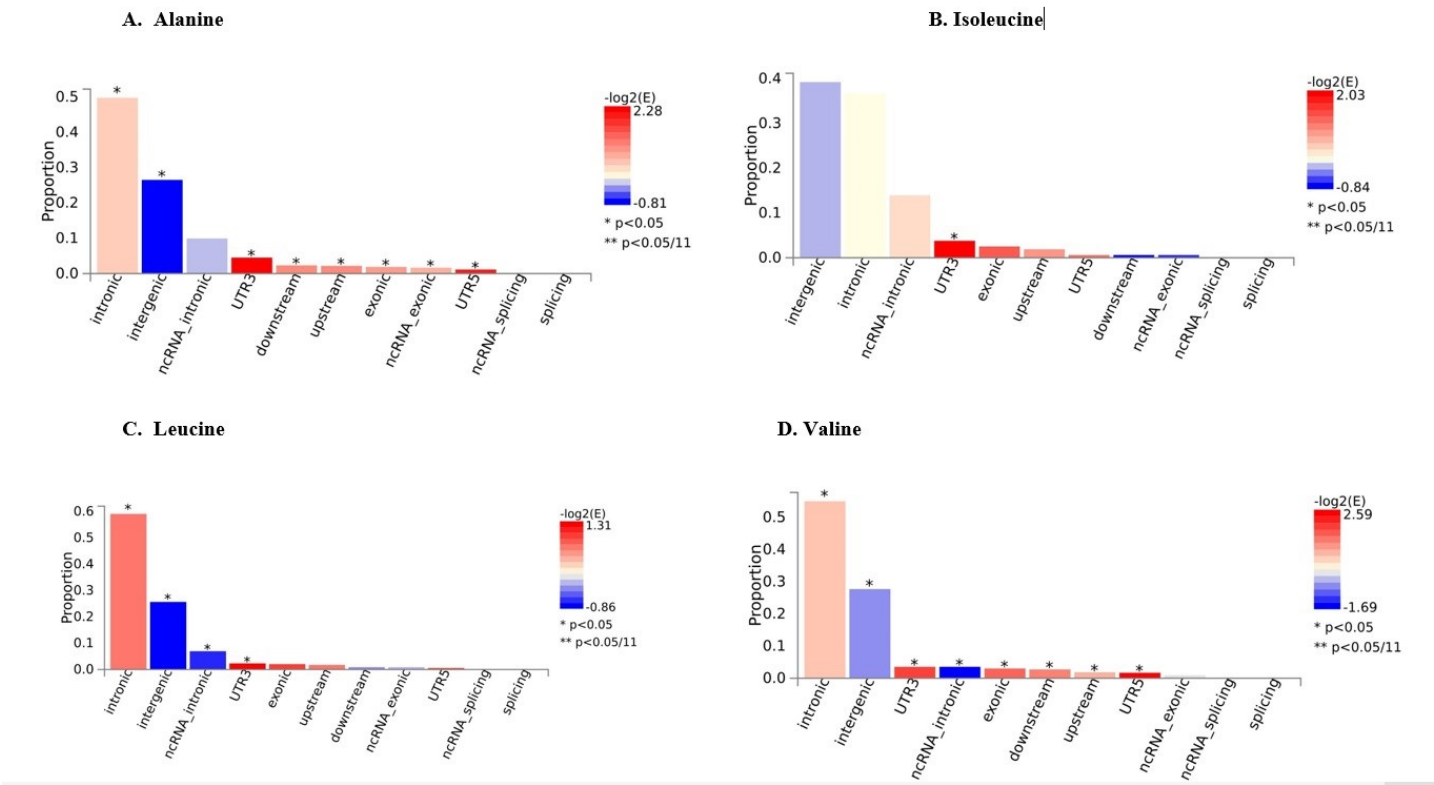
Table1. Total proportion of variance explained by instruments for each amino acid

	Number of SNPs	R^2	F statistic
Alanine	**101**	**0.03**	**35.02**
Glycine	**120**	**0.06**	**53.58**
Glutamine	**138**	**0.06**	**44.62**
Histidine	**62**	**0.02**	**28.45**
Isoleucine	**60**	**0.01**	**24.35**
Leucine	**67**	**0.02**	**28.66**
Phenylalanine	**48**	**0.01**	**31.86**
Tyrosine	**110**	**0.04**	**39.62**
Valine	**91**	**0.03**	**33.69**

2.3.1 Functional mapping and annotation of serum amino acids

Most of the significant SNPs and SNPs in LD with the significant SNPs lie within intronic followed by intergenic gene regions for all serum amino acids. However, there was an exception for serum tyrosine in which the most significant SNPs, and SNPs in LD with the significant SNPs, were in intronic following by ncRNA-intronic gene regions (**Figures 1-3**).

Figure 1. Functional consequence of SNPs on gens for serum alanine and BCCA's



Functional consequence of SNPs on gens for serum amino acids

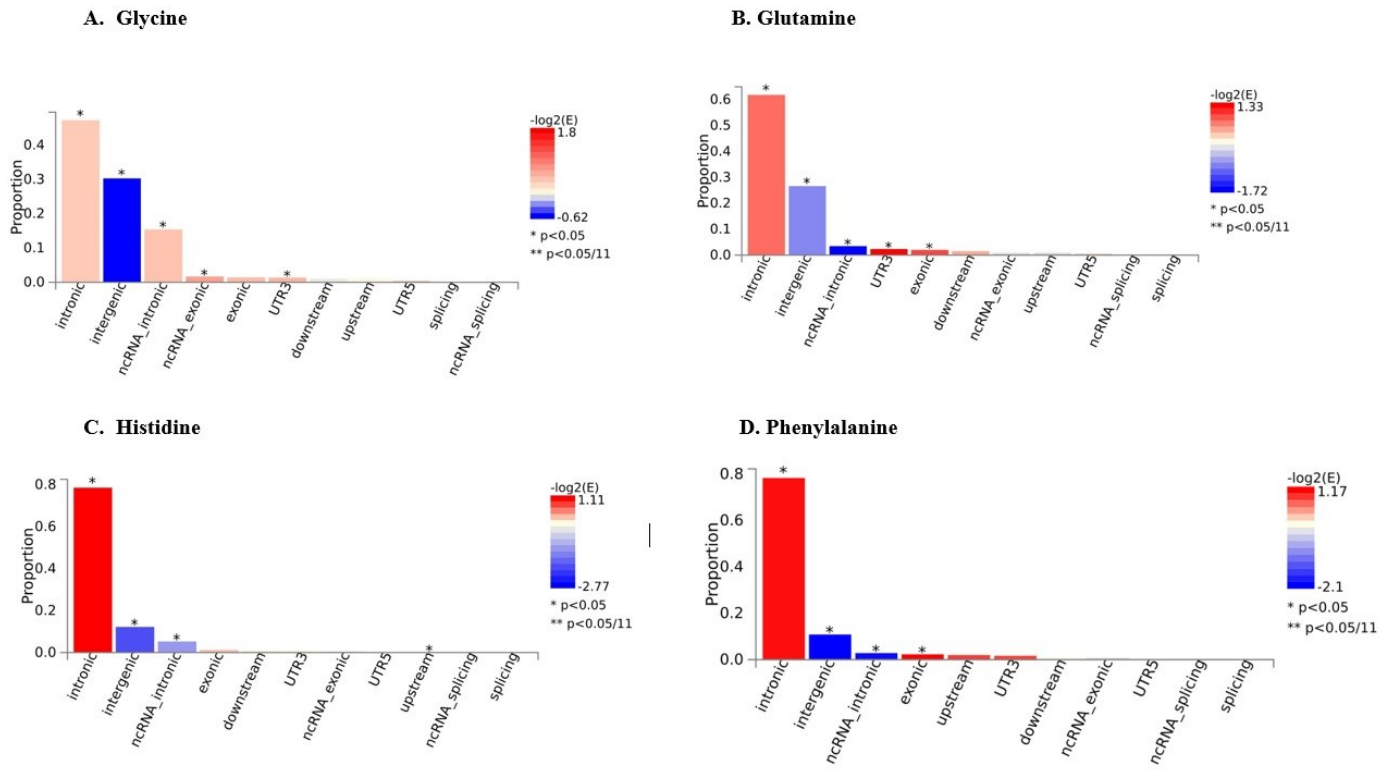
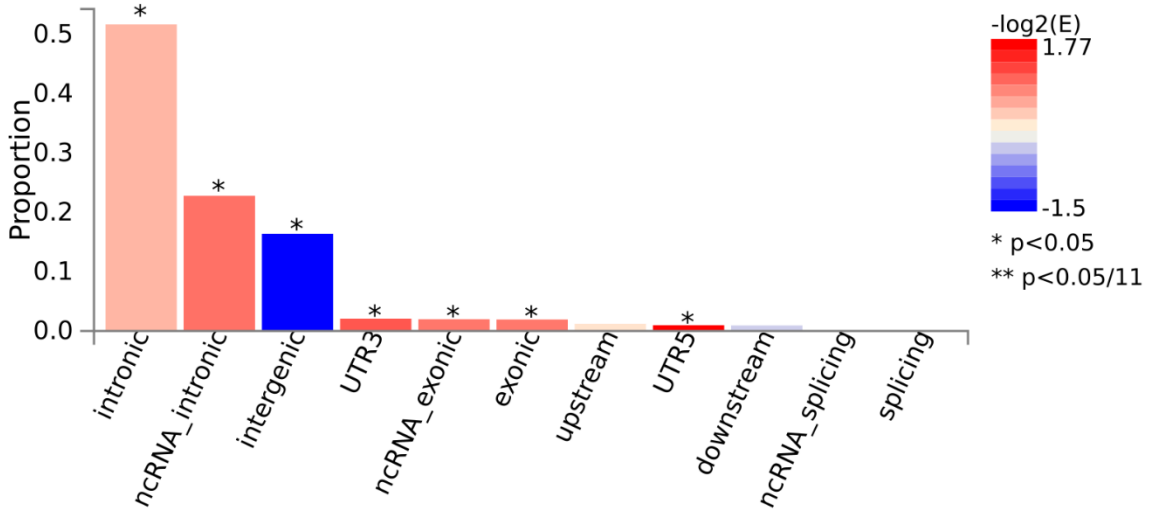


Figure3. Functional consequence of SNPs on gens for serum tyrosine



A summary of results including number of genomic risk loci, lead SNPs and tagged SNPs are presented in **Table 2**.

Table2. Summary of results

	Genomic risk loci	lead SNPs	Ind. Sig. SNPs	candidate SNPs	candidate GWAS tagged SNPs	mapped genes
Alanine	34	35	35	1559	35	84
Isoleucine	8	9	9	164	9	34
Leucine	15	17	17	357	17	61
Valine	18	21	21	776	21	62
Glycine	40	50	50	2603	50	116
Glutamine	45	46	46	2040	46	116
Histidine	15	17	17	652	17	36
Tyrosine	29	35	35	2069	35	87
Phenylalanine	8	9	9	353	9	23

2.3.2 Genome-wide gene-based association analysis

MAGMA-GWGAS mapped SNPs to 64 protein-coding genes for serum alanine, and 35, 43, and 54 protein-coding genes for serum isoleucine, leucine, and valine, respectively, followed by 70 genes for serum glycine, 95 for serum glutamine, 31 genes for serum histidine, 66 genes for serum tyrosine and 35 protein coding genes for serum phenylalanine using Genome Reference Consortium Human Build 37. All p-values were Bonferroni-corrected.

The top signal from gene-based analysis were ***GCKR, MIP, SLC*** *gene family*, ***COBLL1, LMO1, CPS1,*** and ***GLS2***. The common genes across the serum amino acids are demonstrated in **Table 3**.

Table3. Top common genes across amino acids

Gene	Chr	Alanine (pval)	Isoleucine (pval)	Leucine (pval)	Valine (pval)	Glycine (pval)	Glutamine (pval)	Histidine (pval)	Tyrosine (pval)	Phenylalanine (pval)
ABCG2	**4**	-	-	1.10E-10	1.30E-19	-	-	-	-	-
AC022431.2	**5**	-	-	-	3.40E-08	-	4.30E-08	-	-	-
BCAT2	**19**	-	-	2.40E-13	8.70E-35	-	-	-	-	-
CCDC137	**17**	1.30E-13	-	-	3.10E-09	-	-	-	-	-
CDKAL1	**6**	1.40E-15	-	-	-	-	2.90E-10	-	-	-
COBLL1	**2**	-	-	2.30E-15	1.50E-17	5.30E-11	-	-	-	-
CPS1	**2**	1.60E-10	-	-	-	1.80E-42	4.90E-09	4.00E-55	9.90E-12	4.20E-10
DDX19B	**16**		1.10E-31	-	1.10E-46	-	6.50E-09	-	3.70E-08	-
GCKR	**2**	1.90E-40	2.30E-29	3.60E-32	6.50E-36	-	1.60E-81	9.90E-16	-	-
GLIS3	**9**	9.10E-12	-	-	-	-	3.90E-15	-	-	-
GLS2	**12**	3.30E-56	3.60E-16	1.20E-15	2.20E-16	5.20E-22	-	1.10E-28	1.10E-36	-
GLUD1	**10**	-	-	-	1.40E-12	-	3.30E-09	-	-	-
HNF4G	**8**	-	9.10E-09	1.80E-08	-	-	-	-	-	-
HSD17B14	**19**	-	2.30E-19		8.70E-35	-	-	-	-	-
IL27RA	**19**	-	-	3.30E-09	4.80E-08	-	-	-	-	-
LMO1	**11**	9.00E-31	-	-	-	3.90E-11	7.60E-48	3.50E-09	1.60E-11	-
MIP	**12**	3.30E-56	3.60E-16	1.20E-15	2.20E-16	5.20E-22		1.10E-28	1.10E-36	-
MLXIP	**12**	2.00E-11	-	-	1.30E-13	-	-	-	-	-
NDRG2	**14**	-	-	-	-	-	1.60E-16	5.90E-20	-	-
OXLD1	**17**	1.30E-13	-	-	3.10E-09	-	1.00E+00	1.00E+00	-	-
PNPLA3	**22**	-	-	-		-	3.50E-11	-	4.80E-09	-
PPP2R3A	**3**	-	-	2.50E-08	6.00E-09	-	-	-	6.10E-15	-
PRODH	**22**	-	-	8.60E-09	5.80E-10	-	-	-	-	-
PROX1	**1**	4.70E-33	-	-	-	7.80E-11	1.30E-43	-	-	-
RBMS2	**12**	-	-	-	-	-	4.90E-09	1.19E-312	-	4.30E-09

RLN3	**19**	-	-	3.30E-09	4.80E-08	-	-	-	-	-
RP11-529K1.3	**16**	-	1.10E-31	-	1.10E-46	-	6.50E-09	-	3.70E-08	-
SERPINA1	**14**	-	-	-		-	6.60E-10	5.50E-12	-	-
SERPINA1	**14**	-	-	-	-	-	6.60E-10	5.50E-12	-	-
SLC1A4	**2**	5.90E-28	-	1.40E-09	2.90E-47	-	-	-	-	-
SLC25A47	**14**	-	-			8.20E-10	6.20E-15	-	-	-
SLC2A4	**17**	-	-	4.00E-21	6.30E-43		2.70E-08	-	-	-
SLC38A4	**12**	3.30E-27	-	-	-	7.70E-21	8.60E-33	1.40E-09	3.10E-21	-
SLC7A2	**8**	3.70E-14	-	-	-	-	4.90E-20	-	-	-
WARS	**14**	-	-	-	-	8.20E-10	6.20E-15	-	-	-
YBX2	**17**	-	-	4.00E-21	6.30E-43	-	2.70E-08	-	-	-

2.3.3 Gene-set analysis

Using the full distribution of SNP p-values, gene-set analysis tested the gene sets of each serum amino acids to represent their potential involvement in various pathways (**Figure S1-S9**).

The common highlighted pathways for all included serum amino acids indicated their involvement in the regulation of metabolites (small molecules and protein measures), with an exception for serum phenylalanine, and in the regulation of plasma free amino acid levels (adjusted for twenty other PFAAs), with an exception for serum tyrosine.

This enrichment analysis demonstrated the involvement of serum alanine gene-set in the regulation of fasting blood glucose, alcohol use disorder and mood instability, as well as in metabolites and PFAAs. The analysis also detected serum BCAAs related gene-set involvement in the regulation of for instance triglycerides levels for the serum valine related gene-set, and many other pathways in the regulation of blood pressure (both systolic and diastolic), blood lipid profile including triglycerides, cholesterol, HDL, and LDL, as well as in mean arterial pressure, coronary heart disease, CRP levels, Crohn's disease, inflammatory bowel disease, hypothyroidism, serum uric acid levels and oesophageal cancer for serum leucine related gene-set.

For serum glycine, gene-set analysis showed pathways related to regulation in glycine levels, ulcerative colitis, Crohn's disease, inflammatory bowel disease, extremely high intelligence, acylcarnitine levels, HDL cholesterol levels in drinkers, triglycerides, sleep duration (short sleep) and mood instability. It further showed the highlighted pathways for serum glutamine, such as fasting blood glucose, type 2 diabetes, bipolar disorder, mood instability, hyperglyceridaemia, autism spectrum disorder, or schizophrenia.

Serum histidine gene-set analysis demonstrated pathways involved in the regulation of triglycerides and glycine levels.

Lastly, for serum aromatic amino acids (phenylalanine, tyrosine), the enrichment analysis detected many pathways ranging from blood lipid profiles (cholesterol and triglycerides) to diseases such as Alzheimer's in hypertension, Laryngeal squamous cell carcinoma and gallstone disease (**Figure S8-S9)**.

2.3.4 Gene-property analysis

The gene-property analysis across 30 tissues, highlighted the expression of serum amino acids in muscles tissues for serum alanine and liver tissues for serum glycine (**Figure 4**), and serum tyrosine and in liver, kidney, and pancreas tissues for phenylalanine (**Figure 5**). This analysis did not highlight any further significant associations for the other amino acid levels including BCAAs, glutamine, histidine, and tyrosine.

Figure4. Gene-property analysis for serum alanine and glycine

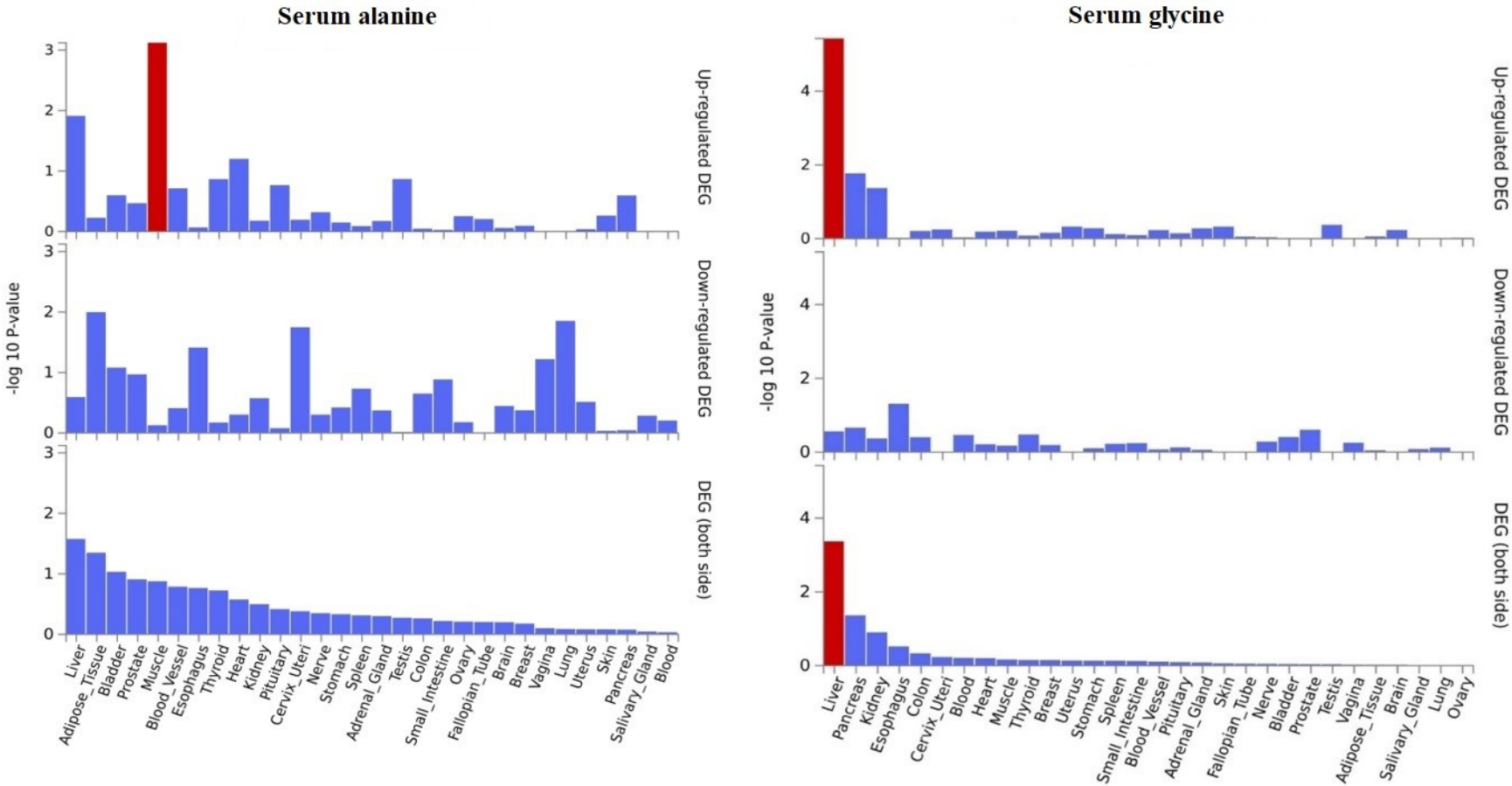
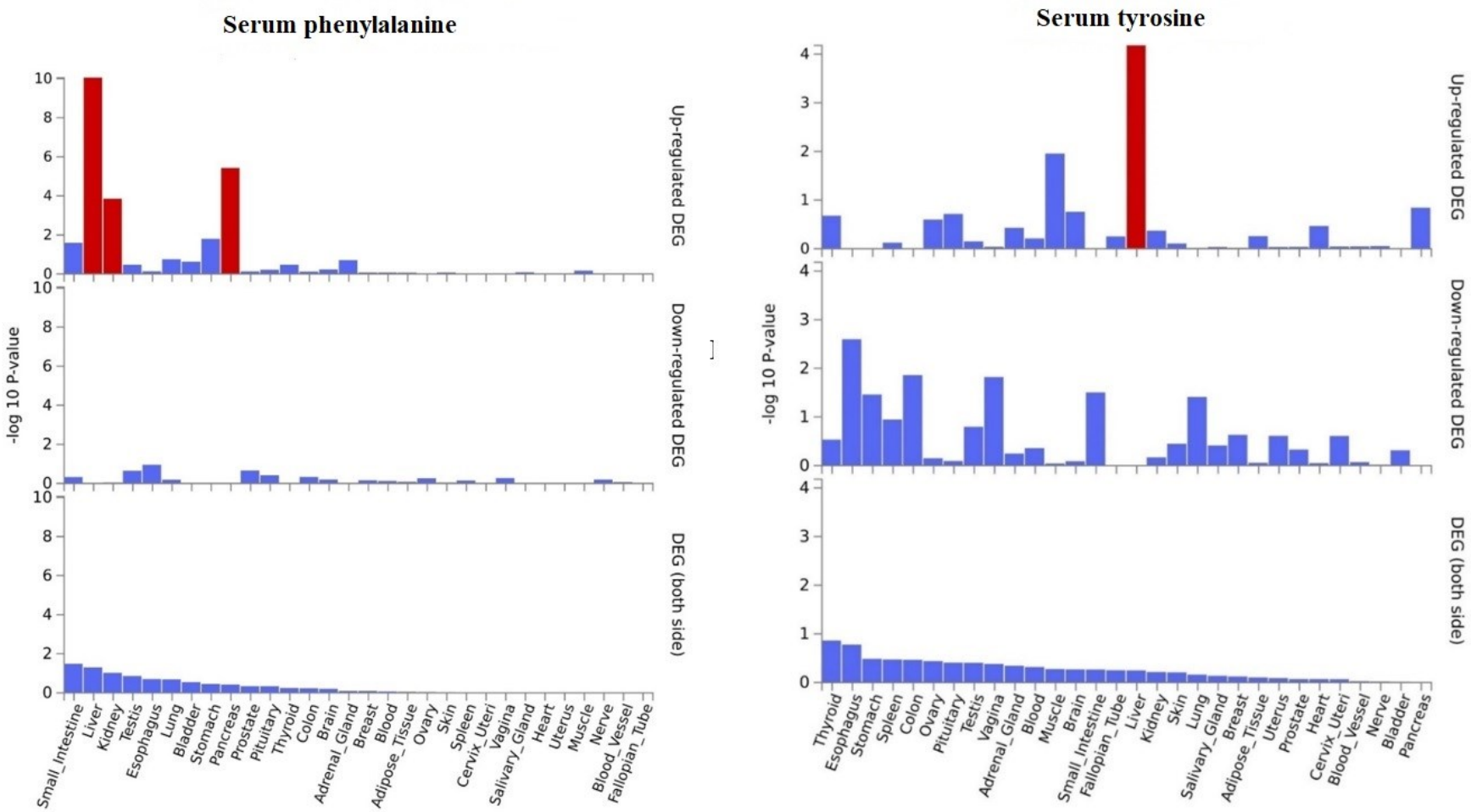


Figure5. Gene-property analysis for serum tyrosine and phenylalanine



2.4 Discussion

Utilising data from serum amino acid levels of around 117 000 individuals from European ancestry of UKB participants, detected a range of genetic variants that can be used as instrumental variables for assessing the related biological pathways, as well as for further analysis investing potential relationship between serum amino acids and risk of complex disorders, which will be examined in the next chapters.

The percentage of variance explained by genetics for serum amino acids ranged from 1-3% for serum BCAAs and 1% to 6% for serum phenylalanine, and both serum glutamine and glycine, respectively.

Functional annotation mapped SNPs to several significant protein-coding genes, ranging from 31 genes for serum histidine to 95 genes for serum glutamine. The MAGMA gene-based analysis within this study highlighted the most common Bonferroni significant genes associated with serum alanine and BCAAs as ***GCKR*** and ***SLC*** superfamily genes.

GCKR is a glucokinase regulator that is involved in the regulation of glucose metabolism and fasting plasma glucose levels, and the well-known disease associated with this gene is Maturity-Onset Diabetes Of The Young (161). As well as glycaemic control and diabetes, GCKR is implicated in dyslipidaemia (162). This is not surprising, as the existing literature has extensively reported that serum BCAAs levels are associated with insulin resistance and type 2 diabetes and are associated with cardiovascular risk (21, 24, 51, 59, 77, 131, 132).

The solute carrier ***SLC*** superfamily genes are responsible for the transmembrane transportation of many essential compounds including nutrients including amino acids, drugs, and osmolytes to neurons and glial cells (163) and SLC transporters are critical in the termination of synaptic transmission for amino acid neurotransmitters, that removes dopamine, serotonin, noradrenaline, glycine, and GABA (164). In epidemiological studies, SLC

transporters have been suggested to play a pivotal role in neurodegenerative disorders and recently, genome-wide association studies have detected 7 of these SLC transporters to be associated with Parkinson's disease (PD) (165). Moreover, the SLC transporter have been reported to be associated with epithelial permeability and barrier function in the intestine, thus SLC genes polymorphisms are shown to be related to inflammatory bowel disease (166).

In addition, ***MIP, CPS1, LMO1, COBLL1***, and ***GLS2*** genes were also highlighted as the most common genes significantly associated with serum amino acids. The related associated diseases with these genes are Cataract for ***MIP*** (167), hyper ammonia deficiency for ***CPS1*** (168), T-Cell Acute Lymphoblastic Leukaemia (169) and neuroblastoma (170, 171) for ***LMO1***, age-related macular degeneration and eye disease for ***COBLL1 (172)***, and Juvenile Dermatitis Herpetiformis and Sclerosing Keratitis for ***GLS2*** (173).

Gene-property analysis showed expression of serum alanine in muscle tissues, serum glycine and tyrosine in liver tissues, and serum phenylalanine in liver, kidney, and pancreas tissues. The expression of serum alanine, a glucogenic and proteogenic amino acid, in muscles tissues is known as when protein breakdown in skeletal muscles during fasting or other stresses, a large quantity of alanine is released into the bloodstream and is carried to the liver to be converted to glutamate and pyruvate, and finally to glucose and entering the blood stream to be available as a source of energy for the muscles (174). In the liver, phenylalanine is converted to tyrosine (175), and glycine is synthesised from other amino acids such as threonine in the liver and kidneys (71).

The MAGMA gene-set analysis showed that serum amino acids are involved in pathways related to regulations of serum metabolites and Plasma-free amino acid (PFAA) profile as well as fasting plasma glucose levels, and some cardiometabolic risk factors including blood pressure, lipid levels, and diabetes. Plasma-free amino acid (PFAA) profiles have been

associated with future risk of developing diabetes or cardiovascular disease in both nondiabetic and diabetic individuals (51, 176). In addition, the plasma-free amino acid (PFAA) profile has been shown to be correlated with Crohn's disease (177), which has also been highlighted in the gene-set analysis of this study for serum leucine and glycine. Crohn's disease, a type of inflammatory bowel disease (IBD), is characterized by epithelium damage and reduced immune response of the bowel, intestinal mucosal barrier function, and abnormal absorption and/or wasting in nutritional particles of the intestine including amino acids (177). Amino acid glycine is reported to have anti-inflammatory properties in intestinal epithelial cells (177).

2.4.1 Strengths and limitations

The strength of this investigation includes the large sample size of the UKB study. The main limitation of this study is that I included only participants of European ancestry; therefore, this must be considered when extending the findings to other ethnic groups. Moreover, rare variants (MAF < 0.01) were not investigated, thus only common variants influencing serum amino acids levels were studied. Lastly, as the UKB participants were younger and more health-oriented than the general population, there is a potential risk of selection bias that could not be avoided. However, as such differences between the UKB participants and the general population is minimal, any potential bias should be negligible.

This large-scale GWAS study investigated the genetic architecture of serum amino acids and has provided new insight into the biology of serum amino acids levels through pathway analysis and functional annotation. Gene targets implicated in levels of serum amino acids were identified that can be further investigated for potential therapeutic targets. To further understand the risk factors affecting the levels of serum amino acids and the possible

association of serum amino acids with diseases and explore potential causality, a wider scale study is required and will be tackled in the following chapters.

Chapter Three

Modifiable lifestyle factors and their effect on serum amino acids

3.1 Introduction

In this chapter, I will investigate the potential effect of modifiable lifestyle factors, including macronutrients intake (% energy from carbohydrate (CHO), protein (PRO) and fat), physical activity (moderate to vigorous, vigorous, strenuous, or other exercise, and accelerometer-measured physical activity), and bio-impedance measured body fat percentage on serum amino acid levels. To this end, firstly I will conduct an observational analysis using the UKB data on lifestyle factors and serum amino acid levels. Secondly, I will use the genetic variants associated with lifestyle factors and each of 9 serum proteogenic amino acid levels including serum alanine, histidine, glycine, glutamine, tyrosine, phenylalanine, and BCAA' s (isoleucine, leucine, and valine) and will examine the evidence for any potential causality.

Lifestyle factors are the modifiable habits and behaviours influencing the overall health and well-being of an individual (178). We know that diet plays a pivotal role in health outcomes as dietary choices and habits can contribute to the risk of developing cardiometabolic risk factors (179), resulting in an increased risk of morbidity and mortality through chronic diseases such as cardiovascular disease (CVD), diabetes, and cancer (179-181). Macronutrient intake, which is an important modifiable risk factor for disease prevention, refers to the calories amount consumed from carbohydrates, fats, and protein (182). In a large cohort study (3 sub-cohorts of Rotterdam Study), including 6 822 participants, it has been shown that higher intake of animal protein and not plant protein, is associated with higher longitudinal insulin resistance and risk of prediabetes and type 2 diabetes which highlights the importance of specific and habitual intake of protein from food sources in the development of type 2 diabetes (183). Eating behaviour and macronutrient intake is a result of an interaction between physiological, environmental, and genetic factors (184). Family studies have shown that the heritability of

macronutrient intake ranges between 20–40% (185), raising the interest for GWAS meta-analysis studies to identify the significant loci associated with macronutrient intake (182, 186).

Physical activity has also been shown to be associated with a wide range of diseases and protects against diseases such as CVD, cancer, osteoporosis, type 2 diabetes, and Alzheimer's disease (187-189). However, the health benefit of physical activity depends on the level of individual engagement, and it seems that most people are not as physically active as the recommended levels (187). There are various factors affecting the levels of physical activity of an individual including cultural, economic, environmental factors, and genetics (187, 190, 191). There is a couple of GWAS studies that reported the genetic basis for physical activity engagement (187, 192, 193).

Overweight and obesity, due to bad dietary habits and a sedentary lifestyle, are risk factor for many chronic diseases (194). The genetic basis of adiposity and its links to cardiometabolic disease risk has been assessed through large-scale meta-analyses of genome-wide association studies (GWAS) (195-199). Overall adiposity, measured by body mass index (BMI), has implicated genes showing the role of the central nervous system (CNS) in obesity susceptibility (197, 200-205), while genetic loci associated with body fat distribution, measured by waist-to-hip ratio (WHR), has shown genes involved in adipocyte metabolism (195, 196, 206, 207). However, people with the same BMI or WHR may have different percentages of body fat, resulting in differences in cardiometabolic risk (195, 208, 209).

Therefore, this chapter aims to explore the effect of modifiable lifestyle factors on levels of proteogenic amino acids, utilizing an observational analysis. In addition, the present study aimed to assess potentially causal associations between modifiable lifestyle risk factors and circulating amino acid levels using Mendelian randomization (MR) analysis. MR analysis

utilized genetic polymorphisms which are randomised naturally, as alleles are randomly allocated during conception, thereby overcoming the limitations of traditional epidemiological studies such as confounding and reverse causation.

3.2 Methods

3.2.1 Observational analysis

3.2.1.1 Study participants and characteristics

I used data from the UK Biobank study (UKB) under the approved application **ID 52569**. The details of the UKB study have been discussed in the previous chapter. Briefly, the UKB study is an open resource, all details are available online at https://www.ukbiobank.ac.uk/, and is a large longitudinal population-based cohort of ~500,000 participants aged 40-69 years at recruitment between 2006-2010. Participants provided self-reported baseline information on age, sex, ethnicity, BMI, physical activity, and diet. Furthermore, blood samples for biochemical tests and genotyping of participants were collected (193, 210).

3.2.1.2 Data extraction

3.2.1.2.1 Macronutrient intake (CHO, fat and PRO)

I extracted the data for carbohydrate intake (**Data-Field 100005**), fat intake (**Data-Field 100004**), and protein intake (**Data-Field 100003**) from the UKB study. This dietary data was collected using Oxford WebQ, which is an online 24-hour dietary recall, that presents 21 broad food groups, including more than 200 commonly consumed foods and drinks (211)(These data have now been superseded by new Data-Fields in 2021 which was not accessible under application **ID 52569**). These data are available for 210 977 participants. After data cleaning, macronutrients intake data were log-transformed as it had a skewed distribution.

3.2.1.2.2 Physical activity

Overall acceleration average (**Data-Field 90012**) data of physical activity was extracted from the UKB study. The average accelerations were measured during wearing of activity monitor

(the device truncated any values exceeding +/- 8 gravitites). The UKB study has data of accelerometer-measured physical activity for 103 687 participants. However, I excluded 6 977 participants whose data could not be calibrated (212) or those who did not wear the device for sufficiently long to get a stable measure of their physical activity status(212).

3.2.1.2.3 Body fat (%)

I extracted the body fat percentage data (**Data-Field 23099**), covering 492 787 participants. The body fat percentage data of the UKB is a subset of whole-body bio-impedance measures using the Tanita BC418MA body composition analyser.

3.2.1.2.4 Serum amino acids levels

I extracted the data on circulating levels of 9 nuclear magnetic resonance (NMR) of 9 proteogenic serum amino acids, available for EDTA plasma samples from approximately 117 000 UK Biobank participants, generated by Nightingale Health. The details have been mentioned in the previous chapter. The included amino acids are serum alanine (**Data-Field 23460**), serum BCAAs (**Data-Fields: 23465-23467**), histidine (**Data-Field 23463**), glycine (**Data-Field 23462**), glutamine (**Data-Field 23461**), phenylalanine (**Data-Field 23468**), and tyrosine (**Data-Field 23469**). All the serum amino acids levels were normalized prior to analyses using rank-based inverse normal transformation.

3.2.1.3 Statistical analysis

The data of macronutrient intake, accelerometer-measured physical activity, and body fat percentage of the UKB participants were merged with the serum amino acids data of the same individuals for the analysis.

I conducted linear regression analysis for assessing the effect of accelerometer-measured physical activity on all serum amino acids. I included 2 models, unadjusted and adjusted for age and sex.

I also performed linear mix-model regression analysis with a random-effects structure including a random intercept and slope (for time of repeated measurements) to investigate the effect of macronutrients intake and body fat percentage on serum amino acids levels. For macronutrient intake, I included both unadjusted and adjusted model including age, sex, weight, and total energy intake as covariates. For body fat percentage analysis, age and sex were adjusted. As included lifestyle factors are correlated, I have also included a model, in which they were further adjusted for each other. All analysis performed using R software (version 4.1.3) in Linux environment.

3.2.2 Genetic data on modifiable lifestyle factors

3.2.2.1 Genetic association estimates for macronutrients (% energy from CHO, PRO and fat)

I drew summary statistics of GWASs associated with diet compositions, including percentage of energy from carbohydrate intake, protein intake and fat intake, from the CHARGE Consortium Nutrition Working Group (182), which their discovery included 91 114 European ancestry participants from 24 epidemiologic cohorts from the CHARGE Consortium Nutrition Working Group (182).

3.2.2.2 Genetic association estimates for physical activity

A GWAS study on UK Biobank (UKB) study reported physical activity measures based on wrist-worn accelerometery data (N=91,084, of which there is 24,000 overlaps with serum amino acids data) (187). I used the reported significant loci of this study in my analysis.

3.2.2.3 Genetic association estimates for body fat (%)

I used the summary data of a genome-wide association meta-analysis of body fat percentage (BF%) in individual from 43 GWAS (*n* up to 76,137) and 13 MetaboChip studies (*n* up to 24,582), predominantly of European ancestry (*n* up to 89,297) (195).

3.2.2.4 Genetic data on serum amino acids

3.2.2.5.1 Genetic instruments selection

I used the list of genetic instruments for macronutrients, physical activity and body fat percentage from the available GWAS studies, mentioned in the previous section, and extracted the same sets of SNPs from the GWAS I conducted and fully explained in chapter one.

3.2.3 Statistical analysis

3.2.3.1 Mendelian Randomisation

A two-sample MR approach using summary data and IVW method, as the main analysis (213), was implemented. I further utilised sensitivity analyses, that provide consistent estimates under weaker assumptions than standard instrumental variable assumptions (214), including MR-Egger, weighted median, simple mode, weighted mode and MR-PRESSO.

All results correspond to an OR per 1SD increment in the genetically predicted modifiable lifestyle factors including macronutrients, physical activity, and body fat percentage.

3.3 Results

3.3.1 Observational analysis

3.3.1.1 Association between macronutrient intakes and serum amino acids

The mean intakes were 82.41 g/day for protein, 78.18 g/day for fat, and 254.70 g/day for carbohydrate. The observational analysis for macronutrient intakes and serum BCAAs (isoleucine, leucine and valine) revealed no significant (p-value <0.05) relationship between the carbohydrate, fat, and protein intake with serum amino acids levels (**Table S10**). Further adjustment for physical activity and body fat percentage did not change results (**Table S10).**

3.3.1.2 Association between measured physical activity and serum amino acids

The analysis showed an inverse significant (p <0.05) association between average accelerometer-measured physical activity and serum isoleucine and valine, showing per 1 unit increase in average acceleration of physical activity there is a 1% unit decrease in levels of serum isoleucine and valine. On the other hand, a positive association for serum glycine and glutamine were observed (**Table S11**). Further adjustment for total energy intake and body fat percentage, demonstrated per 1 unit increase in average acceleration of physical activity there is 1% unit decrease in levels of serum alanine, isoleucine, and valine levels (**Table S11**).

3.3.1.3 Association between measured body fat percentage and serum amino acids

Results from the observational analysis, showed that per 1% unit increase in bio-impedance measured body fat percentage there is 1% unit increase in serum alanine to 3% unit increase in serum BCAAs and tyrosine (**Table S12**). On the other hand, the analysis indicated an inverse significant association between body fat (%) and serum glycine and glutamine by 3% unit decrease in serum levels per 1 unit increase in body fat percentage (**Table S12**). Further adjustment for physical activity and total energy intake didn't change the results (**Table S12**).

3.3.2 Mendelian Randomization analysis

3.3.2.1 Genetically predicted macronutrients intake (% energy) and serum amino acid levels

MR IVW analysis showed that higher genetically predicted protein and fat intakes (% energy) were associated with higher levels of serum BCAAs (isoleucine, leucine and valine) (**Figure1 and Table S13 & S15**), with consistent results for IVW and weighted median across all analysis. MR-Egger intercept analysis found no evidence of directional pleiotropy (P-value > 0.05). The modified Q statistic indicated no notable heterogeneity (**Table S13 & S15).**

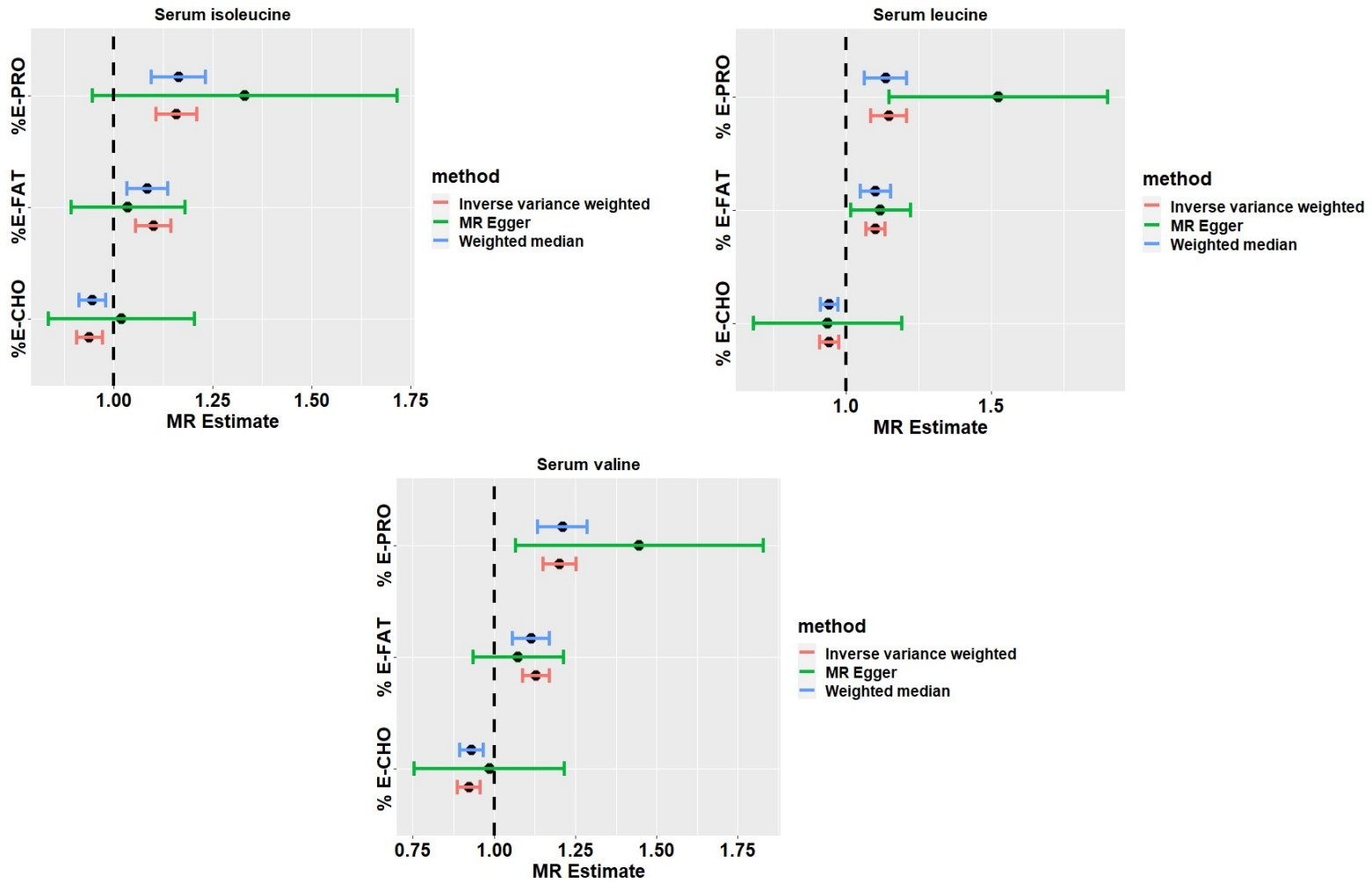
The OR estimates showed in **Supplementary Tables S13 & S15** are per 1 standard deviation (SD) increase in macronutrients intake (% of energy).

I have also reported that higher genetically predicted carbohydrate intake (% energy) was associated with lower levels of BCAAs concentration, showing a protective effect (**Figure 1 and Table S14**) and the finding was consistent across both IVW and weighted median methods. MR-Egger intercept analysis found no evidence of directional pleiotropy (P-value > 0.05). The modified Q statistic indicated no notable heterogeneity (**Table S14**).

The OR estimates reported in **Table S14** per 1 standard deviation (SD) increase in macronutrients intake (% of energy).

Due to limited number of included genetic instruments (n=3) for all these analysis, MR-PRESSO analysis could not be conducted to detect any potential outlier. For the remaining 5 serum amino acids including serum histidine, glycine, glutamine, tyrosine, and phenylalanine, I didn't find any significant association using MR analysis (**Tables S13-S15**).

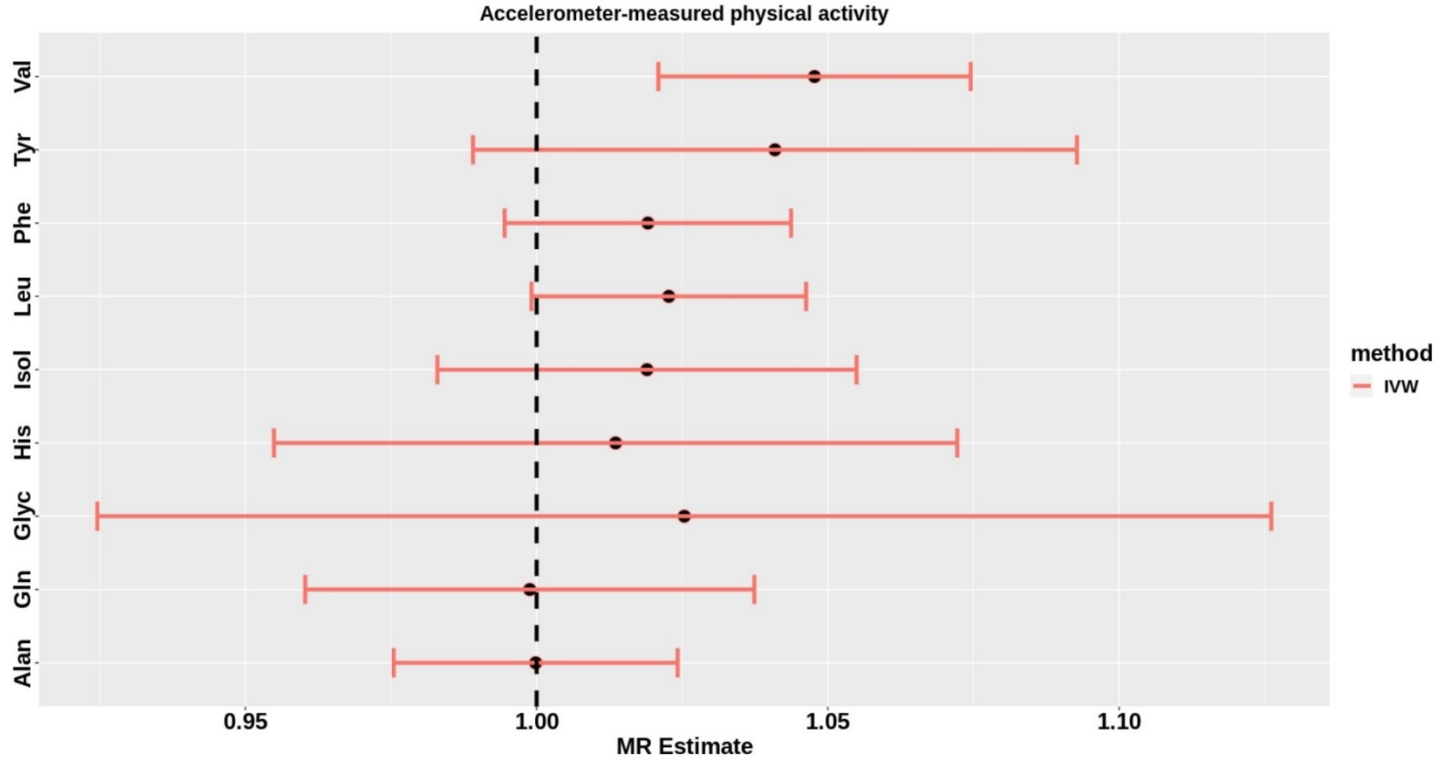
Figure1. Mendelian randomisation analysis showing the effect of % energy from macronutrients on serum BCAAs



3.3.2.2 Genetically predicted physical activity and serum amino acid levels

I estimated that a 1 standard deviation (SD) increment in the genetically predicted levels of accelerometer-measured physical activity was associated with a 5% higher levels of serum valine, using the 2 genetic instruments (odds ratio [OR]:1.05, 95% confidence interval [CI]:1.02-1.07, P-value=<0.01) (**Figure 5 and Table S16**). No heterogeneity was observed using Q-statistic.

Figure2. Plot of MR IVW results for the effect of accelerometer-measured physical activity on serum amino acids levels



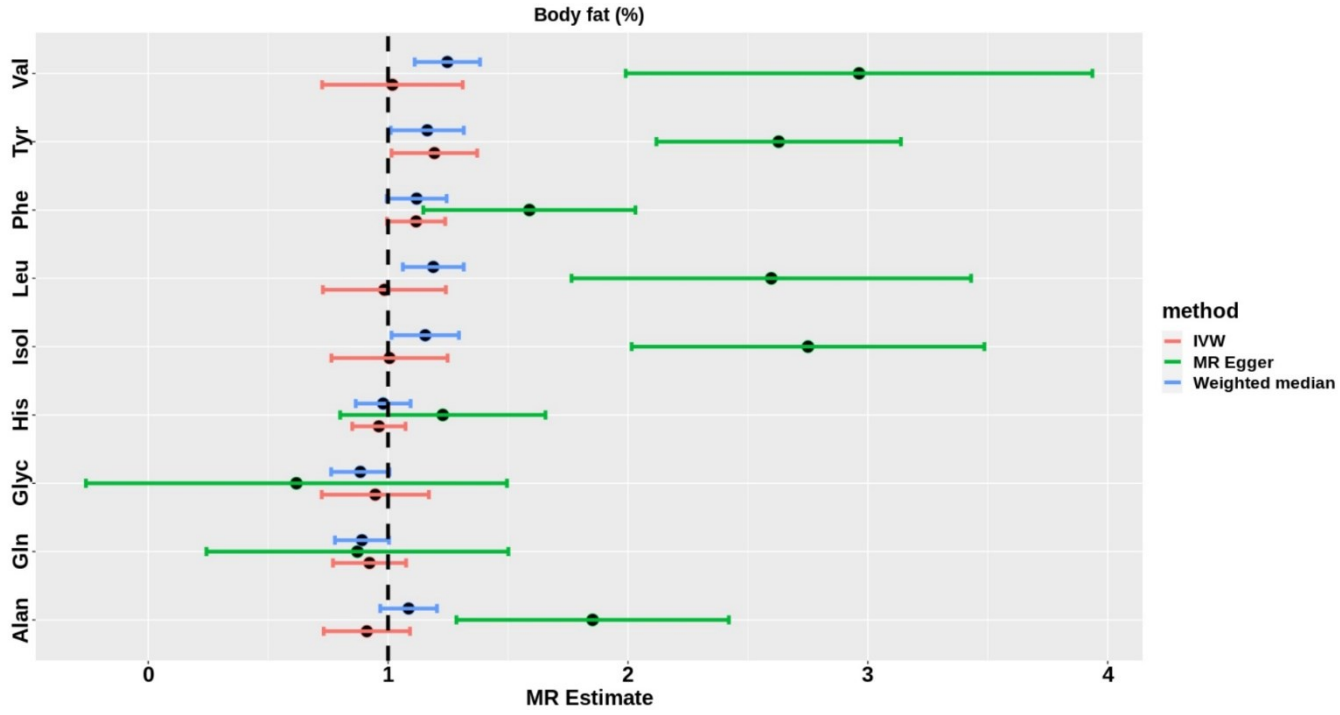
3.3.2.3 Genetically predicted body fat (%) and serum amino acid levels

The results from IVW method showed a nominally suggestive significant inverse association between the body fat percentage and serum glutamine (OR:90, 95% CI:0.82-0.97, P-value=0.02, **Figure 3 and Table S17**) and serum glycine levels (OR:87, 95% CI:0.77-0.96, P-value=0.02, **Figure 3 and Table S17**), after removing outliers using MR-PRESSO. These findings were consistent across sensitivity analysis using weighted median for serum glutamine and both weighted median and weighted mode for serum glycine.

In addition, body fat percentage showed a significant association with serum tyrosine using IVW method (OR:1.19, 95% CI: 1.01-13.7, P-value= 0.05). This result was in concordant using both weighted median and weighted mode. However, this result did not remain significant after correcting for multiple testing (P-value=0.14) and after removing the outlier using MR-PRESSO method (P-value=0.15) (**Figure 3 and Supplementary Table S17**). MR-egger was also significant showing a potential pleiotropy (**Figure 3 and Supplementary Table S17**).

After removing 1 SNPs (rs1558902) that were related to the FTO gene, no significant result was observed between genetically predicted body fat (%) and serum amino acid levels.

Figure3. Plot of MR results for the effect of body fat percentage on serum amino acids levels



3.4 Discussion

In the current study, I explored the association between modifiable lifestyle risk factors including macronutrients (CHO, Fat, and PRO), physical activity and body fat (%), and the 9 serum proteogenic amino acids. To this end, I utilized an observational analysis followed by an MR analysis to investigate the causality of genetically determined lifestyle factors and serum amino acid levels.

Through the observational analysis, I found no association between macronutrient intake and serum amino acids. However, using an MR approach, I identified suggestive evidence for the association between genetically predicted macronutrients intake (% energy from CHO, Fat, and PRO) and serum BCAAs levels, an inverse association between CHO intake and serum BCAAs and a positive association between PRO and fat intake with serum BCAAs. Therefore, MR analysis showed that higher genetically determined protein and fat intakes is associated with higher levels of serum BCAAs while higher genetically predicted carbohydrate intake is protective against higher levels of serum BCAAs.

The disagreement for no relationship between macronutrient intakes with levels of serum amino acids in observational analysis and a significant relationship between the percentage of energy from all macronutrient intake in the MR analysis could be related to the effect of an unknown confounder or measurement error bias in the observational analysis. However, the MR analysis, which can overcome the confounding limitation of observational studies, showed a significant increased association between both genetically elevated PRO and fat intake with serum BCAAs levels. Additionally, a study comparing the performance of the used Oxford WebQ 24-hour dietary recall in the UKB against some tested biomarkers, showed that the Oxford WebQ performs well across key nutrients as compared to more administratively

burdensome interviewer-based 24-hour recalls and this study suggest that using average values helps to reduce the related measurement error bias (211).

Moreover, the contribution of diet to serum BCAAs levels is still unclear and contradictory (215). A study reported that 80% of dietary intake of BCAAs absorbed (216) while another study reported the absorption rate is as low as 40% (215, 217-219). On the other hand, observational studies showed a weak correlation between dietary and serum BCAAs (215, 220, 221).

There is convincing evidence that high protein intake in short term might raise the rate of gluconeogenesis, resulting in a high ketogenic state and subsequently contribute to increased satiety and energy expenditure (183, 222, 223). Therefore, short term high protein intake can cause lower energy intake and a negative energy balance, and in that way stimulate weight loss and support weight maintenance (224). That is the rationale behind the popularity of BCAAs supplement intake among gym-goers and athletes. However, there is still a lack of evidence to support whether serum concentrations of BCAAs are a good biomarker of its dietary/supplement intake or not.

Some observational studies showed a dietary pattern high in protein (mainly animal protein) and fat intake (a westernised diet) modestly increases the levels of serum BCAAs levels (215, 225), as dietary branched-chain and aromatic amino acids are mainly derived from animal-sourced protein (183), which their results are in line with the observed significant association of genetically predicted percentage of energy from protein and fat intake with serum BCAAs levels in my MR analysis. On the other hand, other observational studies found no relationship between higher dietary intake of protein and fat with higher levels of BCAAs (215, 226). Therefore, the reason behind the elevation of serum BCAAs is still unknown (227).

In this study, in an observational design, I also found that every 1 unit increase in accelerometer-measured physical activity resulted in 1% unit decrease in serum BCAAs levels. However, the observational analysis could be affected by unknown confounders as in the MR analysis higher genetically predicted accelerometer-measured physical activity was associated with higher levels of serum valine concentrations. Though only 2 genetic instruments were used, thus sensitivity analysis could not be performed to validate the observed association. A recent cross-sectional and interventional study that used a 12-week supervised combined aerobic resistance-exercise program showed that the exercise training programme did not decrease the serum BCAAs levels (227).

Furthermore, the observational analysis of this study showed that body fat percentage is positively associated with serum concentrations of alanine, BCAAs, and aromatic amino acids, also inversely associated with serum histidine, glycine, and glutamine. However, utilizing an MR approach only a nominally significant inverse association between body fat percentage and serum glutamine and glycine levels was detected. This observed protective effect ranged between 10 to 13 % in levels of serum glutamine and glycine, respectively.

It has been previously reported from a pilot study that glutamine supplementation favours weight loss (228). Glutamine supplementation at a dose not exceeding 0.5 g/KgBW/day has also been consistently shown to improve glucose homeostasis and reduce exogenous insulin requirements in critically ill patients (228). Moreover, glutamine supplementation in rats fed a high-fat diet was demonstrated to inhibit the development of pro-inflammatory markers, obesity, and diabetes (229, 230). Glycine supplementation also showed benefits like glutamine

supplementation in stimulating the loss of adipose tissue and preserving muscle mass in mice (231).

There is no observational study, to my knowledge, that has previously looked at any possible effect of body fat percentage on all 9 proteogenic serum amino acid concentrations. However, studies are looking at the effect of serum amino acids, particularly BCAAs, on body fat related exposures (overweight and obesity) as higher serum BCAAs have shown to be associated with cardiometabolic risk factors including higher risk of diabetes (24, 31, 51) and cardiovascular disease (52, 53).

BCAAs (leucine, isoleucine, and valine) are essential amino acids that can't be synthesized in the body and should be obtained exogenously from the diet, mainly from animal-sourced foods (22). BCAAs supplements are popular among gym-goers and athletes as it believed to have possible beneficiary effects on decreasing exercise-induced fatigue, muscle recovery, and immune regulation (232-234). They also have an important role in the modulation of glucose homeostasis, body weight, and cellular signalling (3, 46). A randomized controlled trial study (RCT) reported that greater decrease in body fat percentage and an increase in lean mass after ingestion of a supplement containing BCAAs, following an 8-week resistance training program (235). This can be interpreted as BCAAs intake should help with weight loss, reduce the body fat percentage, and should be higher among those individuals with higher physical activity but, as discussed previously, serum amino acid levels are not a good biomarker of its dietary or supplement intake. There is epidemiological evidence that has reported a higher level of circulating BCAAs in obese individuals than in normal-weight individuals (24, 31, 51). Recent epidemiological studies have shown that elevated serum BCAAs levels are associated with metabolic disorders, independent of body weight (236, 237), waist circumference (238) and BMI (238). On the other hand, there are studies suggesting that BCAAs promote intrahepatic fat accumulation and are higher in human individuals with non-alcoholic fatty liver disease

(NAFLD)(239-245), which indicates a possible association between impaired amino acid metabolism and liver fat accumulation (239, 243). The underlying mechanism and pathways for increased levels of serum BCAAs in obesity are still unclear (239), but chronic low-grade inflammation is suggested to have a role in inducing pro-inflammatory gene expression in adipose tissue (246, 247).

3.4.1 Strengths and limitations

The strength of this study includes the hypothesis-searching approach to the assessment of the causality of genetically determined modifiable lifestyle factors and the proteogenic amino acids. This allowed a wide exploratory base to assess the potential causality of macronutrient intake, physical activity, and body fat percentage with 9 proteogenic amino acid serum levels which was unprecedented. Moreover, using summary genetic associations from the largest available GWAS meta-analysis studies on lifestyle factors and the GWAS of the UKB participants for serum amino acid levels allowed for having the largest possible strong instruments (genome-wide significant with F-statistic >10) for the MR analysis, resulting in increased statistical power. Population stratification was not a source of potential bias in our analysis, as European ancestry was predominant in all the included data in the MR analysis.

The limitation of this study is that the MR model does not examine potential interactions between proteogenic serum amino acids. Another limitation is that we had a limited number of instruments for macronutrient intake (n=3), and physical activity (n=2), thus it was not possible to assess the existence of any outliers using MR-PRESSO analysis.

Chapter Four

Phenome-wide association study on circulatory levels of serum amino acid and their potential causal role on common complex disorders

4.1 Introduction

In this chapter, I will investigate the effect of genetically determined circulating serum levels of amino acids and the risk of complex disorder using data from the UK Biobank (UKB) within an MR-PheWAS framework. Moreover, I will examine any potential association and causality between serum amino acids and common complex disorders in independent populations, using the largest available GWAS studies on complex disorders.

As discussed previously, serum amino acids are found to be associated with chronic diseases. Serum histidine is suggested to have anti-inflammatory properties (248) and its low levels are associated with chronic diseases such as rheumatoid arthritis (134) and increased mortality in patients (249) with chronic kidney disease (133).

Serum Glycine levels have been demonstrated to be inversely associated with cardiovascular factors including hypertension (250), type 2 diabetes (251), obesity (252), and in general with metabolic syndrome (MetS) particularly in the elderly population (253). It has been shown that serum glycine levels in patients with MetS are significantly lower than in normal subjects (254, 255). Also, its levels in diabetic patients and obese are significantly lower than those in healthy individuals (253, 256-258).

It is believed that glutamine is an essential amino acid, particularly during critical illness for recovery as it attenuates oxidant damage and inflammatory cytokine production, it also reduces gut bacterial translocation and improves N balance (259). Therefore, glutamine supplementation is suggested for reducing mortality, the development of infection, and organ failure in critical illness (259). Glutamine levels have been reported to be associated with tumour cell proliferation and cancer. In addition, lower serum glutamine levels have been associated with respiratory diseases (260) such as chronic obstructive pulmonary disease (261).

Glutamine plays role in immunity and its high levels have been shown to have protective effects on patients with coronary heart disease (262).

Phenylalanine is catalysed into tyrosine which is a precursor for neurotransmitters such as dopamine, epinephrine and norepinephrine (17) and there's epidemiological evidence showing that a lack of these neurotransmitters is a causal factor for the development of neurological disorders such as schizophrenia (263) as schizophrenic patients have higher levels of phenylalanine and lower levels of tyrosine levels, affecting the balance of neurotransmitter downstream (264). This may be due to phenylalanine hydroxylase enzyme impairment that catalyses phenylalanine to tyrosine (265).

Furthermore, amino acids are building blocks of proteins in the body and have an important role in many metabolic pathways (266-268). In observational studies, variations in serum amino acid concentrations have been shown to be associated with a higher risk of chronic diseases such as cancer (269), Alzheimer's disease (270), fatty liver (271), muscle dysfunction (272, 273), liver failure (274) and renal failure (275), type 2 diabetes mellitus (T2DM) (24, 31, 51) and cardiovascular disease (CVD) (53, 88). However, observational studies are prone to well-characterised biases such as confounding, and therefore it is difficult to measure the true causal effect of serum amino acids on complex disorders.

Mendelian Randomization (MR) is a method that uses genetic variants as instrumental variables to investigate the causal association between potentially modifiable risk factors and health outcomes in observational data and is believed to overcome the limitations of observational studies such as confounding and reverse causation (116). MR could be used in the framework of a Phenome wide association study (PheWAS) where the association between genetic variants and a wide range of phenotypes is studied (114, 115). PheWAS utilizes electronic health records (EHR) and has been validated through replication of hundreds of

known genotype-phenotype associations, and their utilization have distinguished between true pleiotropy and clinical comorbidity (114, 115).

Using the MR-PheWAS approach facilitates both a hypothesis-searching approach and investigation of a potential causal association between an exposure (using genetic IVs) and a panel of PheWAS outcomes in a high-throughput manner. Furthermore, the MR-PheWAS method allows to replicate of the observed associations reported in observational studies, also the discover any unknown associations.

In this chapter, I aim to utilise serum amino acids associated variants identified in chapter two, and rich phenotypic data from the UKB study, to elucidate the potential relationship between serum amino acids concentrations and diseases as well as assess any pleiotropic effects and potential causality through an MR-PheWAS approach. Furthermore, I will use the largest available GWAS studies on cardiometabolic outcomes including type 2 diabetes, blood pressure, lipid levels, and coronary artery disease, and on Alzheimer's disease, and breast and prostate cancer to elucidate any potential association and causality observed in the previous chapter of MR-PheWAS analysis but this time in independent populations.

4.2 Method

4.2.1 Study participants and recruitment

I used data from the the UK Biobank (UKB) study, which is previously discussed in chapter2. Briefly, the UKB is a population-based prospective cohort study including more than half a million participants, aged 40–69 years who were recruited between 2006 and 2010 (193, 210). The genetic data of participants were linked to Hospital Episode Statistics (HES) using the International Classification of Diseases (ICD) codes (210). In this study, I used the genotypic and phenotypic UK Biobank data for 488,378 participants under an approved data request application.

I also used metabolic biomarkers data of the UK Biobank study measured by Nightingale Health NMR platform (276), available for around 120,000 individuals. The metabolic biomarkers included in my study are serum alanine, histidine, glutamine, glycine, tyrosine, phenylalanine, and BCAAs (isoleucine, leucine and valine).

4.2.2 Genetic instruments selection

I obtained genetic instruments using GWAS of amino acids I performed and explained in chapter two. I selected single-nucleotide polymorphisms (SNPs) associated with serum amino acids at a threshold of $p < 1 \times 10^{-5}$ and selected those which were independently (Linkage Disequilibrium (LD) $r^2 < 0.001$) (277) and robustly (F-statistics > 10) (278) associated with each of the serum biomarkers. The variance explained for all serum amino acids were calculated for the pooled variants, using linear regression analysis.

4.2.3 Phenotyping and mapping ICD-9 and ICD10 codes to phecode

I used ICD9 and ICD10 codes in the HES data (including a wide range of the clinical traits) and converted them into phenotype codes (phecodes) (114). Cases were selected as individuals with a given phecode, who had at least one documented diagnosis, and controls as individuals without the phecode of interest and any other phecode within the same disease category.

4.2.4 Summary GWAS data

4.2.4.1 Genetic association estimates for type 2 diabetes

Genetic association estimates for type 2 diabetes were obtained from a meta-analysis GWAS study (279) that included 62,892 T2D cases and 596,424 controls from participants with European ancestry, by combining 3 GWAS data sets: DIAbetes Genetics Replication and Meta-analysis (DIAGRAM)(12,171 cases vs. 56,862 controls in stage 1 and 22,669 cases vs. 58,119 controls in stage 2)(280), Genetic Epidemiology Research on Aging (GERA) (6905 cases and 46,983 controls)(281), and the UK Biobank (UKB) (21,147 cases and 434,460 controls) (282).

4.2.4.2 Genetic association estimates for lipid levels

Low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides and total cholesterol genetic associations estimates were obtained from a GWAS study (283) consisting of 188,577 individuals of European, East Asian, South Asian and African ancestries.

4.2.4.3 Genetic association estimates for blood pressure

For blood pressure traits (systolic, diastolic and pulse pressure), genetic association estimates were taken from a GWAS study in over 1 million individuals of European ancestry (284).

4.2.4.4 Genetic association estimates for coronary artery disease

CARDIoGRAMplusC4D 1000 Genomes GWAS (CARDIoGRAMplusC4D 1000G) study was used to derive genetic association estimates (285). This GWAS study included 60,801 CAD cases and 123,504 controls, with adjustment made for population stratification using the genomic control method. Cases and controls were of European (approximately 75%), East and South Asian, Hispanic and African American ancestry (285).

4.2.4.5 Genetic association estimates for breast cancer

For breast cancer, genetic association estimates were attained from a GWAS study including 133,384 breast cancer cases and 113,789 controls, plus 18,908 BRCA1 mutation carriers (9,414 with breast cancer) of European ancestry (286).

4.2.4.6 Genetic association estimates for prostate cancer

Genetic association estimates for prostate cancer were obtained from a GWAS study that included 46,939 prostate cancer cases and 27,910 controls from European ancestry (287).

4.2.4.7 Genetic association estimates for Alzheimer's disease

Genetic association estimates were attained from a meta-analysis GWAS study on 94,437 individuals of European ancestry who were clinically diagnosis with Risk for late-onset Alzheimer's disease (LOAD)(288).

4.2.5 Statistical analysis

4.2.5.1 One-sample MR-PheWAS

Initially, I conducted a one-Sample MR analysis (289) using the two-stage least-squares (2SLS) method (289). First, I regressed each serum amino acids on their genetic instruments and saved the residual values. In the second stage, I conducted a PheWAS analysis using a logistic regression model on each phecode against a wide range of clinical diagnoses in the UKB study.

The PheWAS analysis was restricted to participants from European ancestry and a kinship coefficient of > 0.088 was used to mark the related individuals. My PheWAS analysis was adjusted for age, sex and the first 10 genetic principal components. I only included 617 phecodes with more than 200 cases to avoid testing phecodes with very small statistical power. I also used the false discovery rate (FDR) method to account for multiple testing as the Bonferroni procedure is suggested to be too conservative when the number of tests is large (290).

4.2.5.2 Two-sample MR

To investigate the potential overall causal effect for the observed association in the MR-PheWAS analysis for serum alanine and BCAAs (isoleucine, leucine and valine), I applied the two-sample MR analysis using the Inverse-variance weighted (IVW) method (289)(**Table1**). The estimates were calculated per 1 SD increase in each of serum amino acids levels.

To further investigate the robustness of the findings to possible pleiotropy, the weighted median MR (130) and MR-Egger regression (289, 291) sensitivity analysis was performed. The weighted median MR method assumes that at least half of the are valid instruments and makes a less weak assumption about invalid instruments. MR-Egger regression was applied to identify any potential bias resulting from horizontal pleiotropy. The intercept of MR-Egger regression will be different from zero in the presence of directional pleiotropy, and the slope of that regression will be a consistent estimate of the causal effect of exposure on the outcome of interest (292).

Additionally, the MR-PRESSO method was applied to detect and remove outlying genetic variants that substantially affect the IVW estimates (127, 289).

4.2.5.3 Multivariable MR (MVMR)

As the serum amino acids were highly correlated, I conducted an MVMR analysis to estimate the direct causal effect of each serum amino acids in a single analysis model (289, 293-295). MVMR analysis uses genetic variants associated with two or more exposures to estimate the causal effect of each of exposures on the outcome (296). Therefore, MVMR analysis can aid to clarify whether the observed causal effect on the outcome is due to all included exposures or one of the exposures mediates the effect of others on the outcome (295, 296).

4.2.5.3.1 Genetic Instruments for MVMR

Through the pool of our previously selected genetic instruments for each of the serum biomarkers, I used the same SNPs for all amino acids (alanine, isoleucine, leucine and valine), which were at least associated significantly with one of them. Therefore, 52 independent ($r^2 <$ 0.001) SNPs which were common among all 4 exposures were included in the MVMR analysis. I also conducted an MVMR-Egger analysis to investigate the potential pleiotropy (297).

All statistical analysis was performed using the software R (version 4.3.4).

4.3 Results

4.3.1 Genetic loci associated with serum amino acids in UKB data

I selected genetic instruments for the analysis that each had F-statistics >10, which were strong IVs to be used for the MR analysis **(Chapter Two, Table S1-S9)**.

4.3.2 Genetically predicted serum amino acid levels and risk of diseases using one-sample MR-PheWAS analysis

I tested associations between all the genetically determined serum amino acids levels and 617 outcomes (>200 cases). **Figures 1**, shows the Manhattan plot of significant associations observed for serum alanine and BCAAs (leucine, isoleucine, and valine) with the phenotypes. The OR estimates are per 1 standard deviation (SD) increase in serum amino acids levels.

No significant association was observed for serum histidine, glycine, glutamine, tyrosine, and phenylalanine across phenome.

Figure1. PheWAS Manhattan plot for serum alanine and BCAA

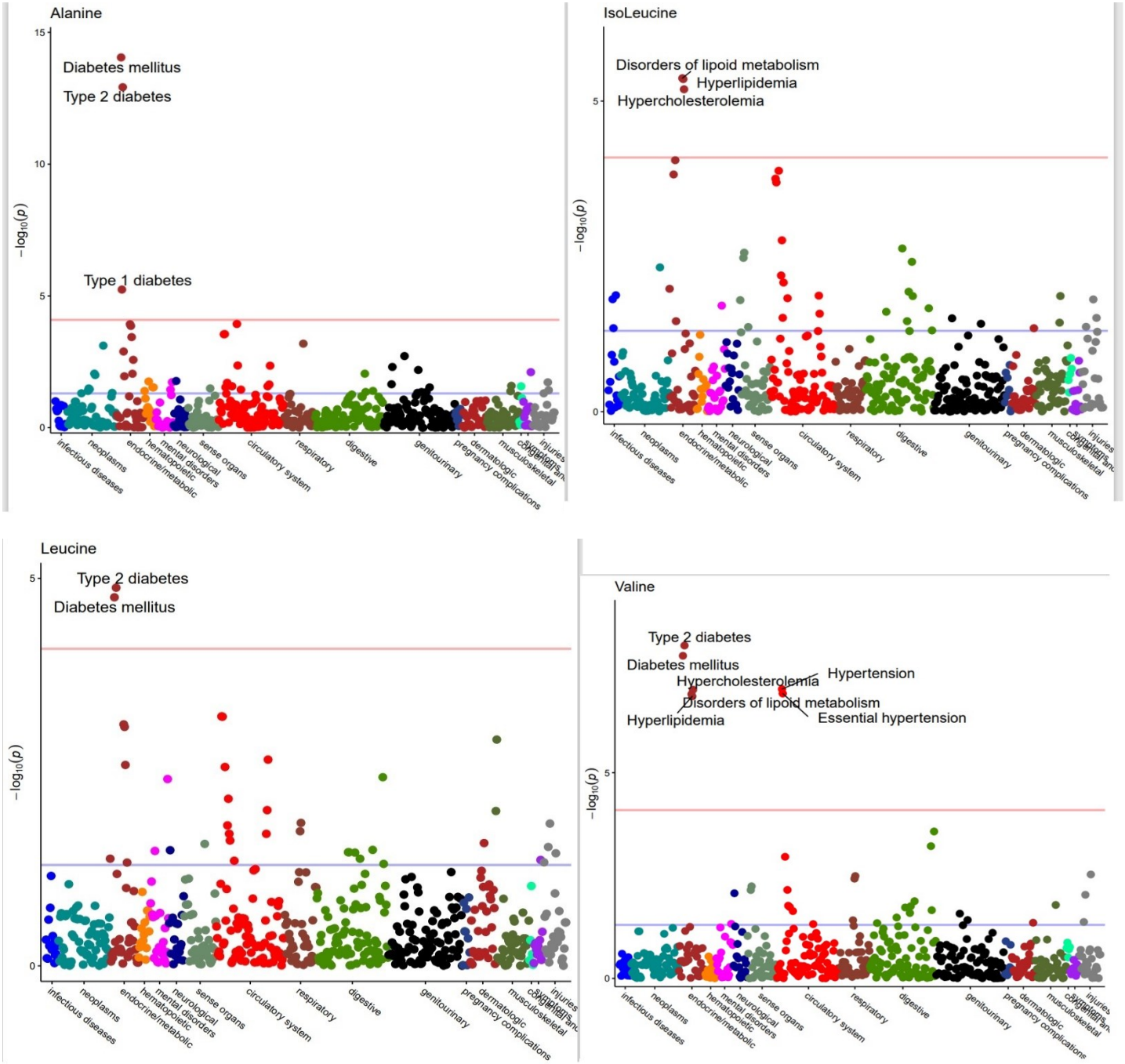


Table 1. Univariate MR analysis for serum alanine and branch chain amino acids and risk of specific diseases

		IVW			MR Egger			Weighted Median			MR-PRESSO		
Exposure	**Outcome**	**OR**	**95% CI**	**p**	**OR**	**95% CI**	**p**	**OR**	**95% CI**	**p**	**OR**	**95% CI**	**p**
Alanine (N SNPs=101)	Diabetes mellitus	1.43	1.18-1.74	**<0.001**	0.98	0.60-1.60	0.95	1.15	1.00-1.32	**0.04**	1.21	1.08-1.24	**<0.001**
	Type 2 Diabetes	1.43	1.15-1.80	**0.001**	0.99	0.60-1.62	0.95	1.14	0.99-1.31	0.07	1.19	1.06-1.33	**<0.01**
	Type 1 Diabetes	1.58	1.20-2.08	**0.001**	0.95	0.49-1.85	0.89	1.80	1.32-2.47	**<0.001**	1.60	1.26-2.02	**<0.001**
Isoleucine (N SNPs=60)	Disorders of lipid metabolism	1.30	1.14-1.50	**<0.001**	1.17	0.86-1.60	0.32	1.14	0.97-1.33	0.10	1.25	1.13-1.37	**<0.001**
	Hypercholesterolemia	1.30	1.13-1.49	**<0.001**	1.16	0.83-1.62	0.37	1.06	0.91-1.24	0.46	1.22	1.11-1.35	**<0.001**
	Hyperlipidaemia	1.31	1.14-1.50	**<0.001**	1.17	0.84-1.64	0.35	1.12	0.97-1.28	0.12	1.25	1.13-1.37	**<0.001**
	Ischemic Heart Disease	1.20	1.06-1.35	**0.005**	1.07	0.80-1.44	0.61	1.20	1.02-1.40	**0.03**	1.19	1.05-1.33	**0.004**
	Type 2 Diabetes	1.32	1.02-1.71	**0.03**	1.31	0.39-1.13	0.13	1.22	1.00-1.49	**0.04**	1.26	1.08-1.46	**<0.01**
Leucine (N SNPs=72)	Diabetes mellitus	1.22	1.02-1.46	**0.03**	1.20	0.69-1.39	0.90	1.30	1.11-1.52	**0.002**	1.20	1.06-1.37	**<0.01**
	Type 2 Diabetes	1.22	1.00-1.48	**0.04**	1.21	0.66-1.38	0.79	1.25	1.04-1.49	**0.01**	1.20	1.06-1.36	**<0.01**
Valine (N SNPs=91)	Diabetes mellitus	1.35	1.10-1.64	**0.002**	1.20	0.72-1.47	0.87	1.26	1.10-1.44	**<0.001**	1.19	1.08-1.31	**<0.001**
	Disorders of lipid metabolism	1.25	1.09-1.43	**0.001**	1.14	0.83-1.38	0.58	1.06	0.96-1.17	0.27	1.13	1.04-1.22	**<0.01**
	Essential Hypertension	1.25	1.15-1.35	**1.11E-06**	1.08	0.95-1.31	0.19	1.15	1.06-1.24	**0.001**	1.20	1.12-1.29	**1.98E-06**
	Hypercholesterolemia	1.22	1.06-1.40	**0.002**	1.14	0.82-1.37	0.64	1.04	0.94-1.15	0.50	1.12	1.04-1.22	**<0.01**

	Hyperlipidaemia	1.24	1.09-1.43	**0.001**	1.14	0.83-1.38	0.61	1.08	0.98-1.19	0.13	1.14	1.05-1.24	**0.001**
	Hypertension	1.25	1.15-1.35	**1.10E-06**	1.08	0.95-1.31	0.20	1.15	1.04-1.27	**0.003**	1.20	1.12-1.30	**1.95E-06**
	Type 2 Diabetes	1.35	1.11-1.64	**0.003**	1.20	0.68-1.38	0.86	1.21	1.05-1.39	**0.008**	1.22	1.11-1.35	**<0.001**

4.3.3 MVMR analysis

Figure3, demonstrates the overlap between genetic IVs used for serum alanine, isoleucine, leucine and valine. The percentage of instruments that were in common across two amino acids ranged from 9.3% for serum alanine and isoleucine up to 17.1% for serum leucine and valine (**Figure 3**).

The MVMR analysis results are shown in **Table 2**. Isoleucine was significantly associated with disorders of lipid metabolism (OR, 2.61;95% CI, 1.21-5.62; *p*=0.01, IVW), hypercholesterolemia (OR, 2.37;95% CI, 1.09-5.13; *p*=0.03, IVW), and hyperlipidaemia (OR, 2.63;95% CI, 1.22-5.68; *p*=0.01, IVW). However, after correcting for multiple testing no associations remained significant (**Table 2**). Therefore, the MVMR analysis indicated that the observed significant associations are not related to the direct effect of isoleucine on those outcomes (**Table 2**).

Figure3. Venn diagram representing the degree of overlap among SNPs of the named metabolites

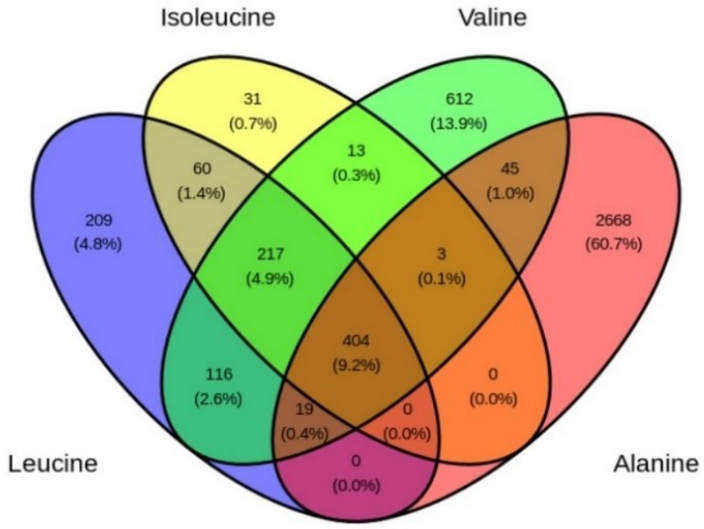


Table2. Multivariable MR results for serum alanine and branch chain amino acids and specific diseases

		IVW				MR-Egger			
Exposure	**Outcome**	**OR**	**Lower 95% CI**	**Upper 95% CI**	**p**	**OR**	**Lower 95% CI**	**Upper 95% CI**	**p**
Valine	Diabetes mellitus (N SNPs=52)	0.76	0.31	1.93	0.56	0.74	0.31	1.74	0.49
Isoleucine		0.75	0.12	4.74	0.76	0.96	0.16	5.59	0.97
Leucine		1.97	0.44	8.77	0.37	1.28	0.30	5.44	0.74
Alanine		1.45	0.95	2.21	0.08	0.84	0.47	1.49	0.55
Valine	Disorders of lipid metabolism (N SNPs=52)	0.89	0.61	1.29	0.53	0.78	0.60	1.28	0.50
Isoleucine		**2.61**	**1.21**	**5.62**	**0.01**	2.78	**1.30**	**5.96**	**0.008**
Leucine		0.64	0.34	1.19	0.16	0.58	0.31	1.08	0.09
Alanine		1.13	0.95	1.35	0.17	0.99	0.77	1.27	0.93
Valine	Essential Hypertension (N SNPs=52)	1.04	0.73	1.49	0.81	1.03	0.74	1.42	0.87
Isoleucine		0.84	0.41	1.73	0.63	0.96	0.49	1.87	0.91
Leucine		1.47	0.82	2.63	0.20	1.17	5.34	2.57	0.57
Alanine		1.04	0.88	1.23	0.62	0.79	**0.64**	**0.99**	**0.04**
Valine	Hypercholesterolemia (N SNPs=52)	0.89	0.61	1.30	0.55	0.88	0.60	1.29	0.52
Isoleucine		**2.37**	**1.09**	**5.13**	**0.03**	2.52	**1.17**	**5.45**	**0.02**
Leucine		0.67	0.36	1.26	0.21	0.60	0.32	1.45	0.12
Alanine		1.16	0.97	1.70	0.10	1.02	0.79	1.31	0.90
Valine	Hyperlipidaemia (N SNPs=52)	0.88	0.60	1.30	0.51	0.87	60	1.27	0.48
Isoleucine		**2.63**	**1.22**	**5.68**	**0.01**	2.80	**1.31**	**6.03**	**0.008**
Leucine		0.64	0.34	1.19	0.16	0.57	0.30	1.08	0.09
Alanine		1.13	0.95	1.35	0.15	0.99	0.77	1.27	0.94
Valine	Hypertension (N SNPs=52)	1.04	0.73	1.49	0.81	1.03	0.74	1.42	0.87
Isoleucine		0.85	0.41	1.75	0.65	0.96	0.69	1.33	0.91
Leucine		1.45	0.81	2.61	0.21	1.17	0.68	2.03	0.57
Alanine		1.04	0.88	1.23	0.63	0.79	**0.64**	**0.99**	**0.04**
Valine	Ischemic Heart Disease (N SNPs=52)	0.92	0.60	1.41	0.70	0.90	0.61	1.31	0.57
Isoleucine		0.84	0.35	2.00	0.69	0.99	0.46	2.14	0.99
Leucine		1.63	0.81	3.3	0.17	1.62	0.25	2.29	0.54

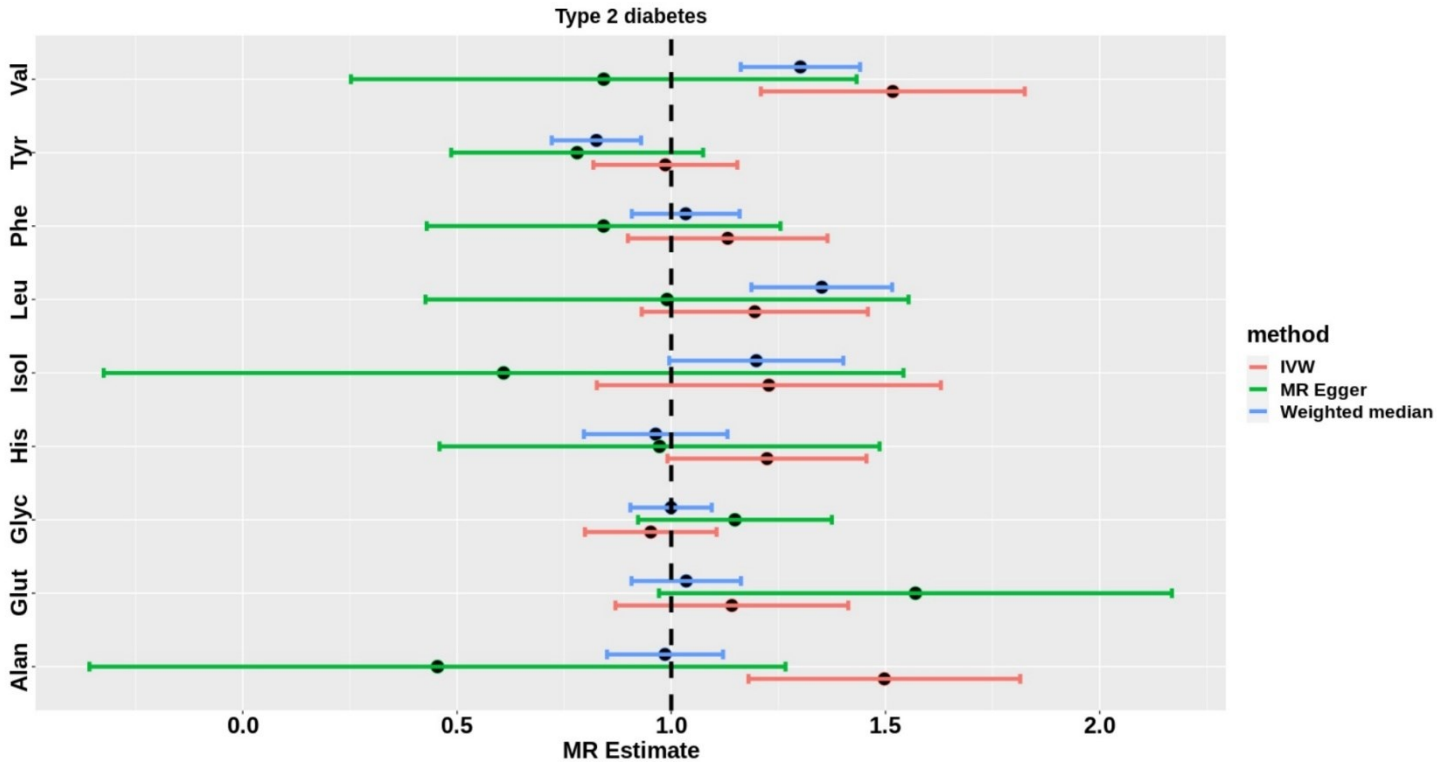
Alanine		1.10	0.90	1.34	0.35	0.76	**0.59**	**0.98**	**0.03**
Valine	Type 2 Diabetes (N SNPs=52)	0.85	0.33	2.17	0.73	0.82	0.33	1.99	0.65
Isoleucine		0.61	0.09	4.10	0.62	0.79	0.33	1.94	0.80
Leucine		2.03	0.44	9.33	0.36	1.31	0.30	5.78	0.72
Alanine		1.46	0.95	2.24	0.09	0.84	0.46	1.51	0.56

4.3.4 Serum amino acids and type 2 diabetes

Genetically elevated levels of serum alanine and valine were significantly associated with higher risk of type 2 diabetes (**Figure 1 and Table S18**). However, the observed relationship for serum alanine did not remain significant after correcting for multiple testing using FDR and removing outliers using MR-PRESSO (**Table S18**).

The odds ratio (OR) for serum valine was 1.51 (95%CI:1.11-1.82, p =0.04). After removing the outliers using MR-PRESSO, the observed effect estimate attenuated but remained significant (p <0.001) and this finding was also consistent with the results of the weighted median method (**Table S18**).

Figure1. Plot of MR results for 9 serum amino acids



4.3.5 Serum amino acids and lipid levels

Genetically elevated serum levels of alanine, isoleucine and valine were associated with higher levels of total cholesterol, triglycerides, and LDL (**Figure2 and Table S19**). Genetically elevated levels of serum leucine were associated with TC and TG and not with LDL levels. All genetically determined serum BCAAs were associated with lower levels of HDL (**Figure2 and Table S19**). The observed significant associations were consistent using sensitivity analysis including weighted median and after removing outliers using the MR-PRESSO method (**Table S19**). In addition, the associations remained significant after correcting for multiple testing using FDR (**Table S19**). No further associations for the rest of 5 amino acids were observed (**Table S19**).

Figure2a. Plot of MR results for serum amino acids and total cholesterol levels

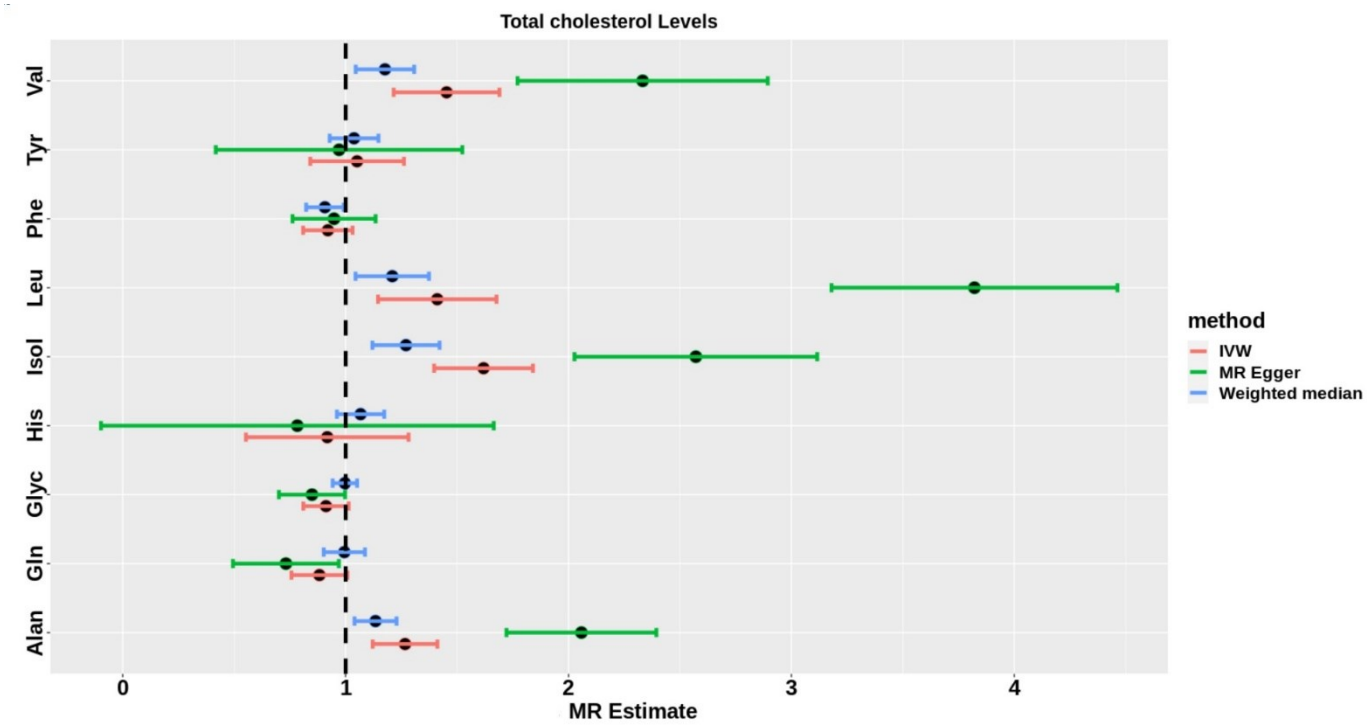


Figure2b. Plot of MR results for serum amino acids and triglycerides levels

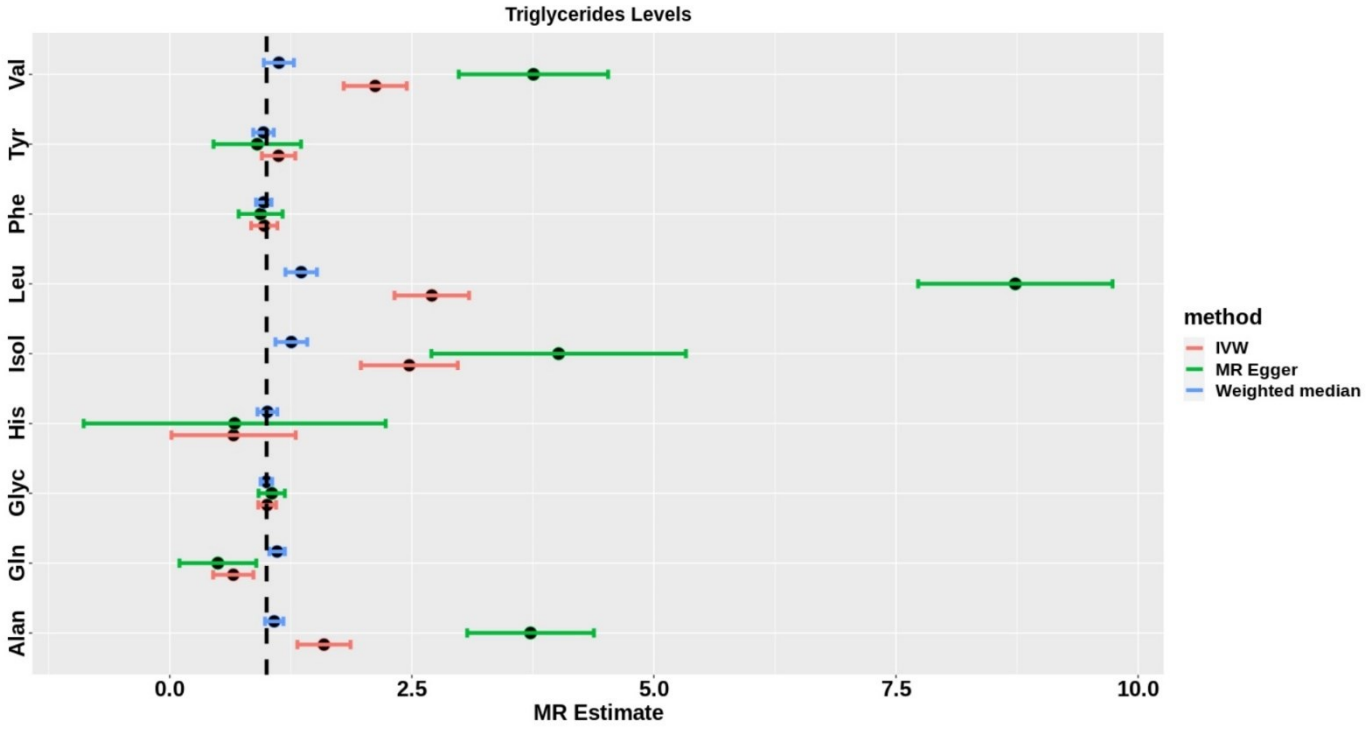


Figure2c. Plot of MR results for serum amino acids and LDL levels

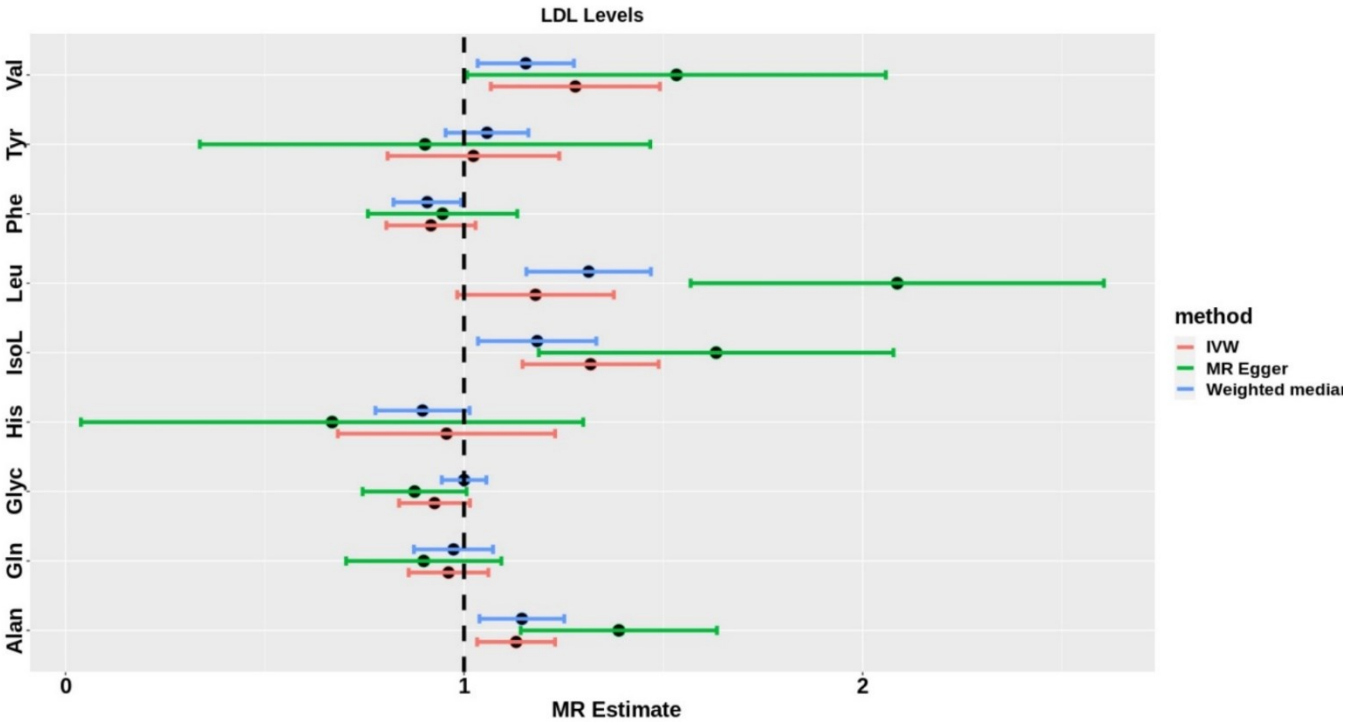
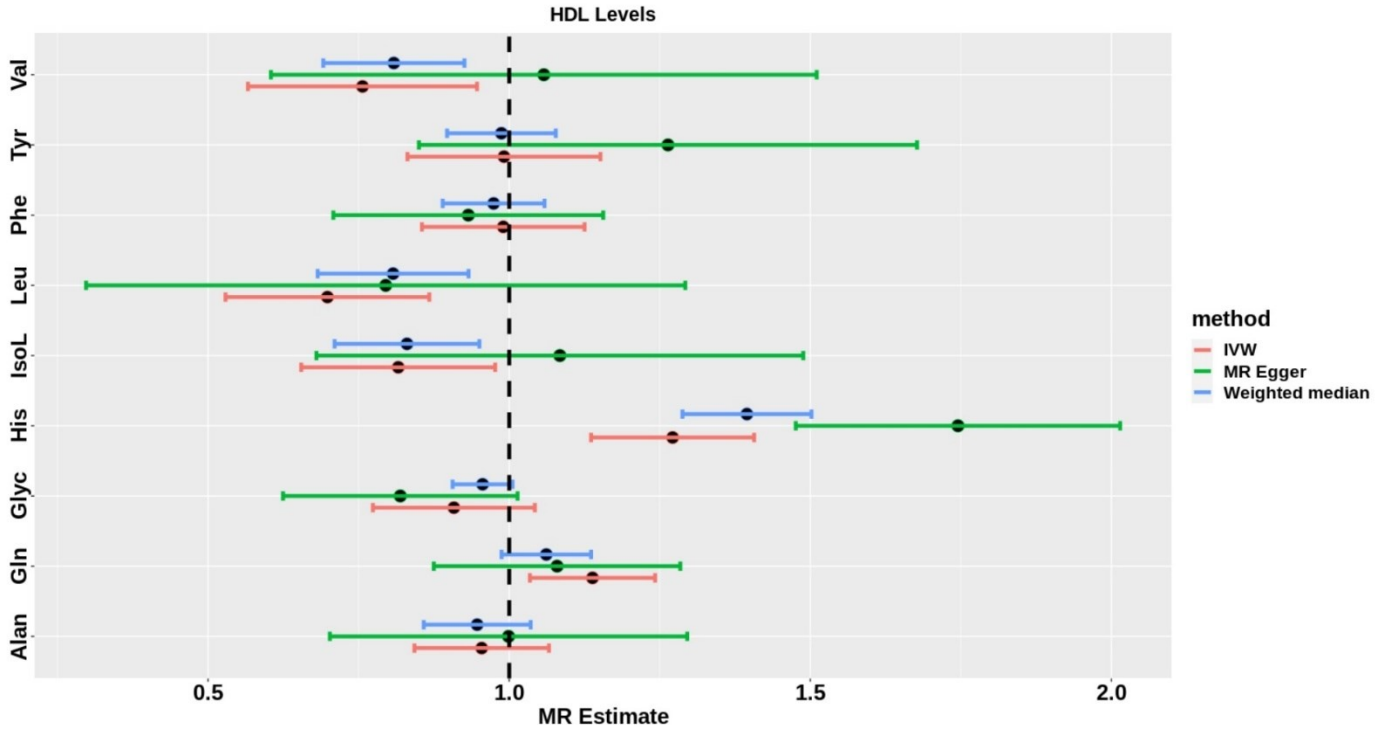


Figure2d. Plot of MR results for serum amino acids and HDL levels



4.3.6 Serum amino acids and blood pressure

Genetically elevated serum levels of alanine and BCAAs were significantly associated with higher systolic blood pressure, after correction for multiple testing using FDR (**Figure 3 and Table S20**). However, the findings were not consistent using sensitivity analysis, except for serum alanine (**Table S20**).

Figure 4 illustrated the result of IVW method for all 9 serum amino acids, showing significant associations between serum alanine and BCAAs with diastolic blood pressure (**Table S21**). However, the significant association for genetically determined levels of serum alanine was the only association that was in line with the result of weighted median method and excluding outliers using MR-PRESSO and after correcting for multiple testing using FDR (**Table S21**).

Figure3. Plot of MR results for serum amino acids and systolic blood pressure (SBP)

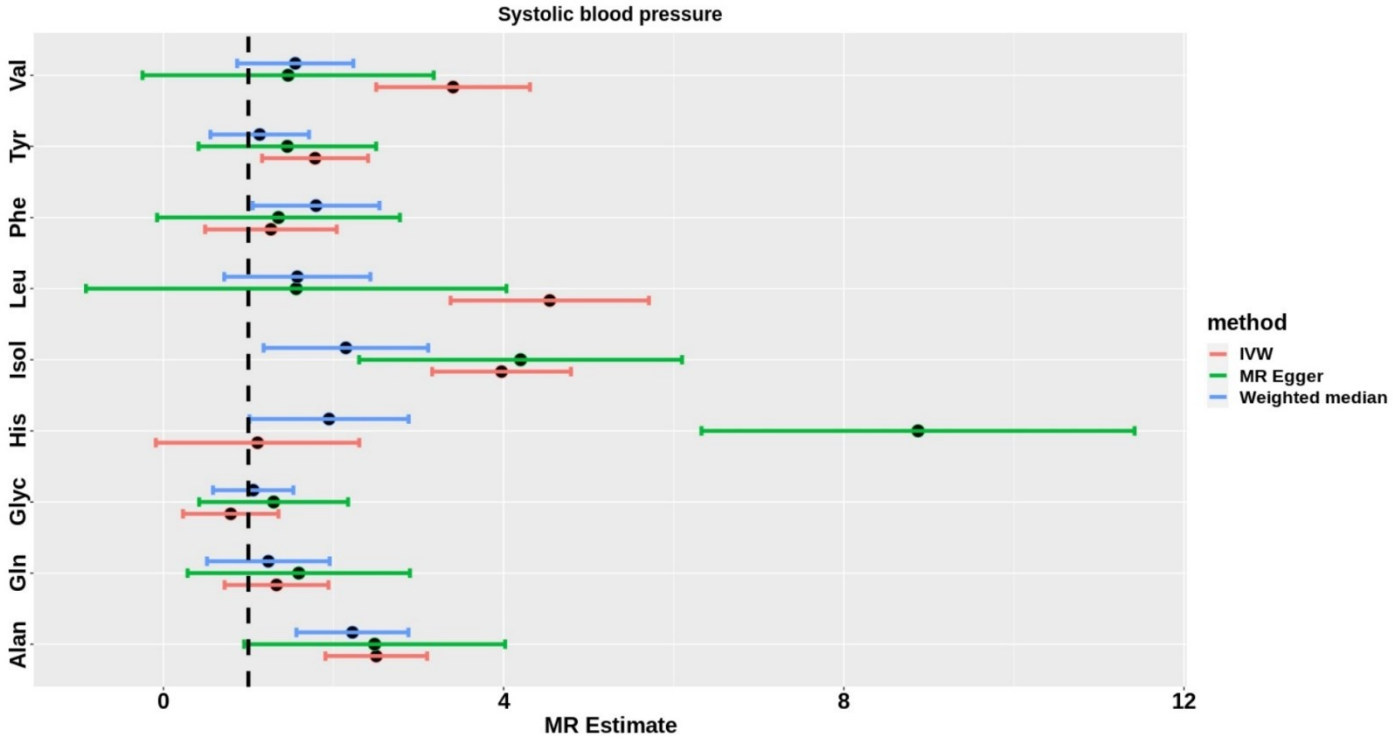
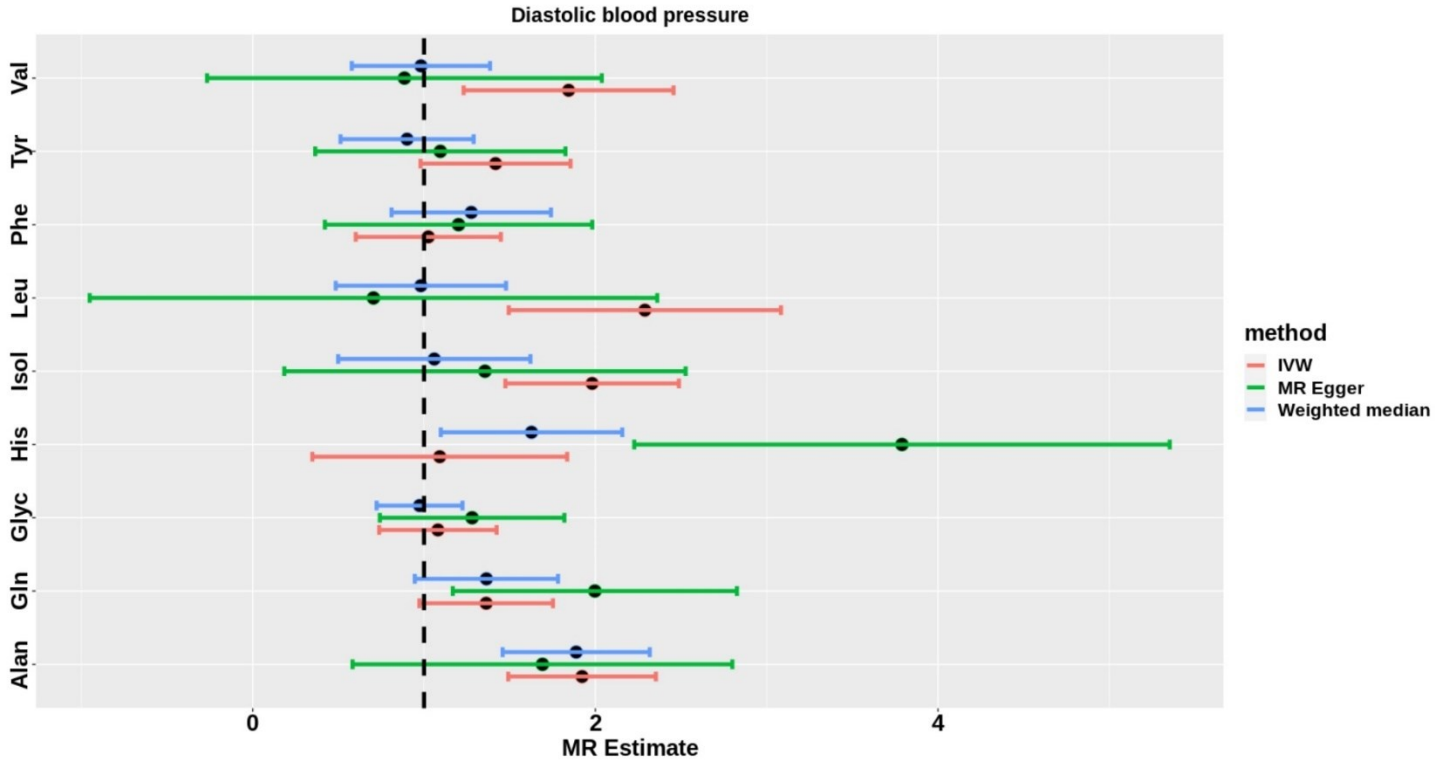


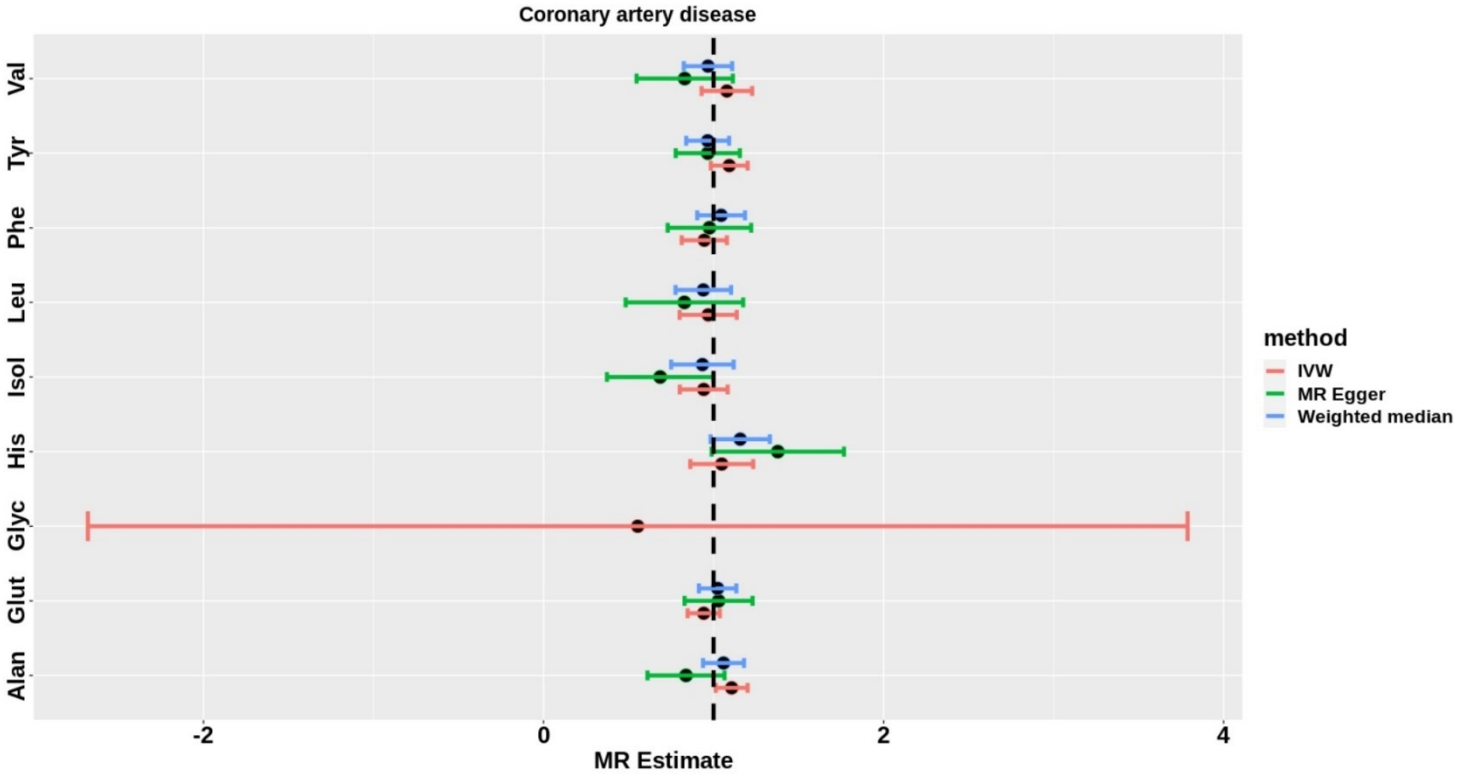
Figure4. Plot of MR results for serum amino acids and diastolic blood pressure (DBP)



4.3.7 Serum amino acids and coronary artery disease

In the current study, I didn't detect any significant association between genetically elevated levels of any serum amino acids with risk of coronary artery disease **(Figure 5 and Table S22).**

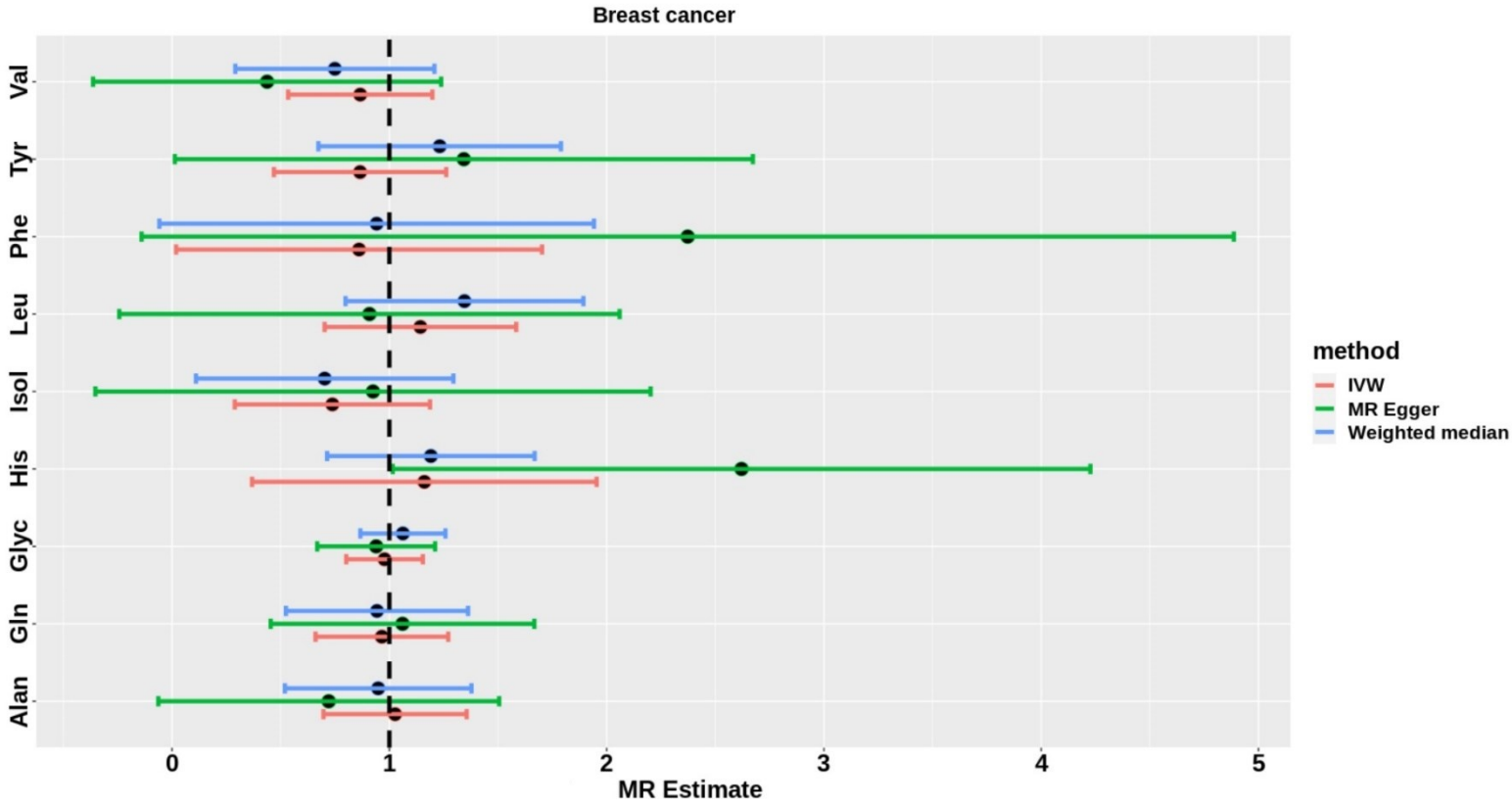
Figure5. Plot of MR results for serum amino acids and risk of CAD



4.3.8 Serum amino acids and breast cancer

No association was observed for genetically elevated levels of serum amino acids and risk of breast cancer **(Figure 6 and Table S23).**

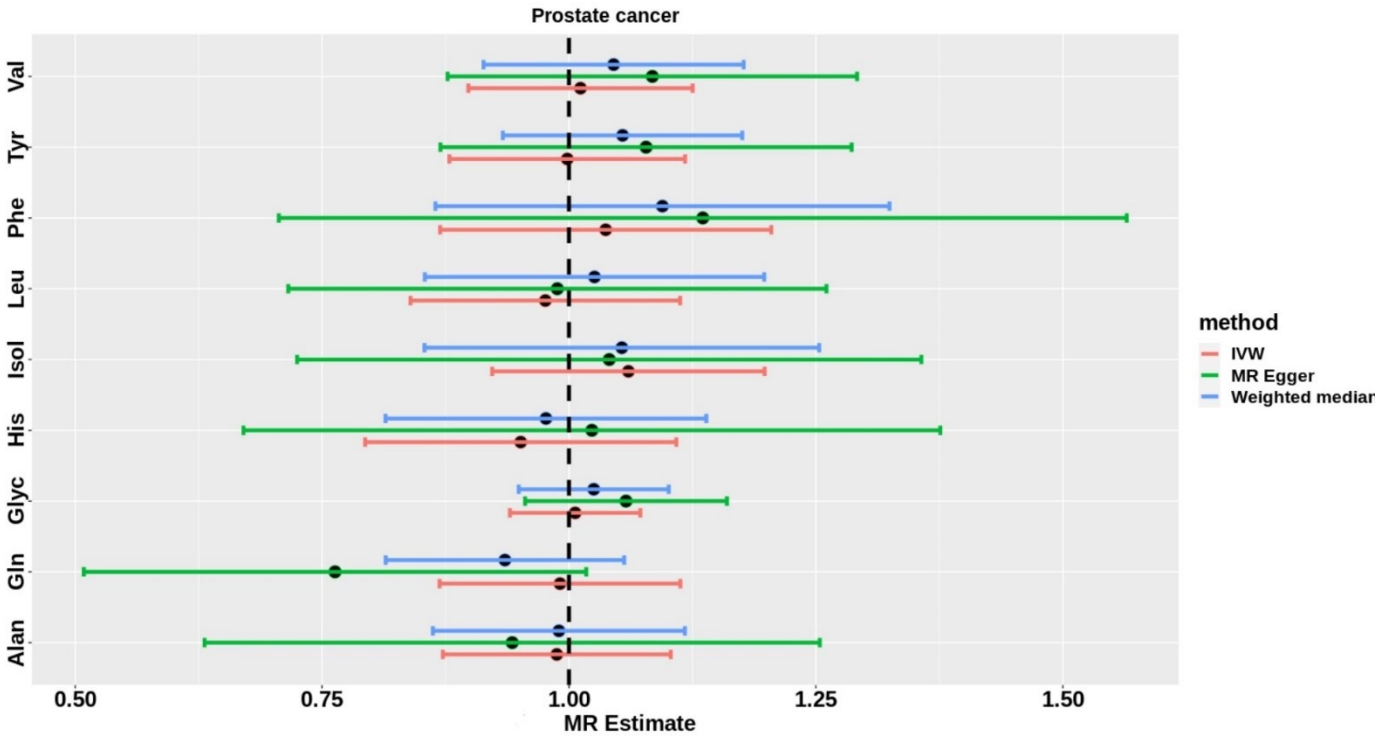
Figure6. Plot of MR results for serum amino acids and risk of breast cancer



4.3.9 Serum amino acids and prostate cancer

No significant associations observed between genetically elevated levels of serum amino acids and risk of prostate cancer **(Figure 7 and Table S24).**

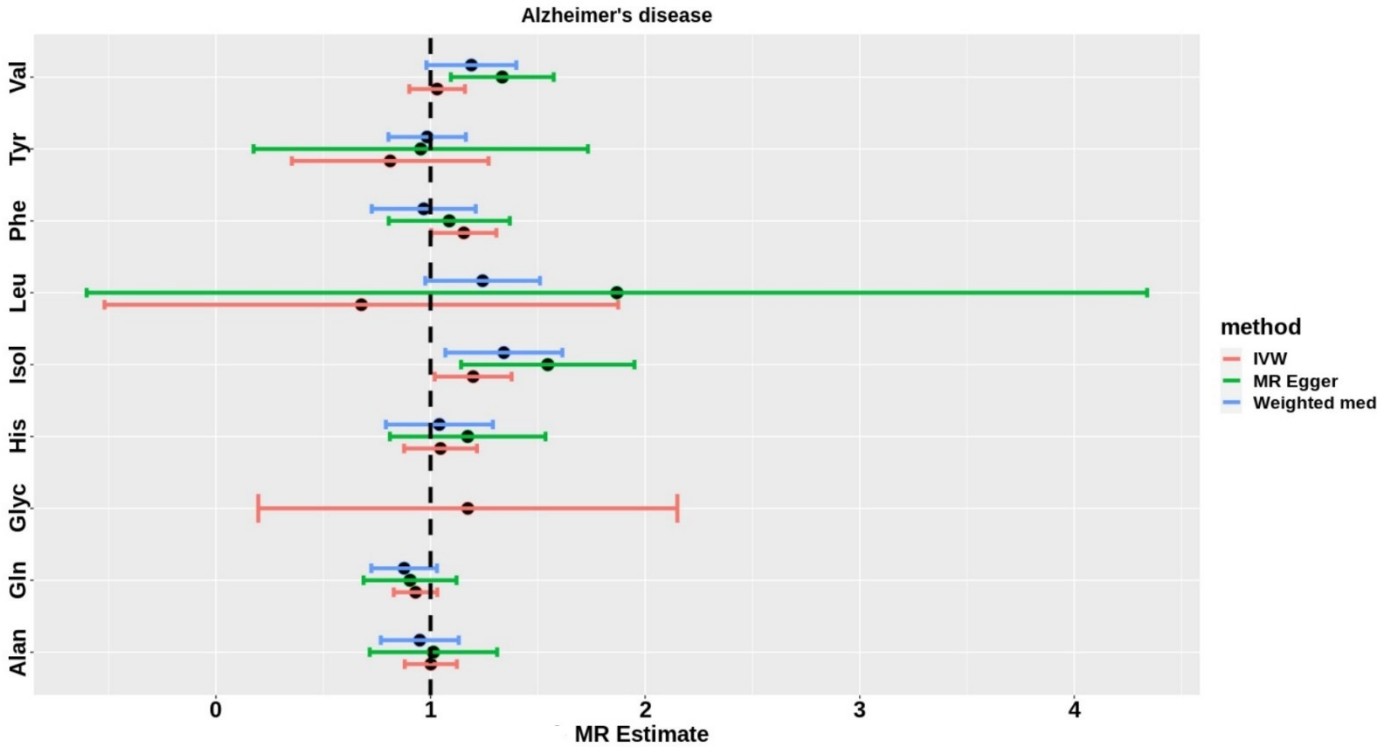
Figure7. Plot of MR results for serum amino acids and risk of prostate cancer



4.3.10 Serum amino acids and Alzheimer's disease (AD)

The only significant association was detected for the relationship between genetically elevated levels of serum isoleucine and risk of AD, however, the association didn't remain significant after correcting for multiple testing using FDR **(Figure 8 and Table S25).**

Figure8. Plot of MR results for serum amino acids and risk of Alzheimer's disease



4.4 Discussion

This chapter, utilizing the large-scale exploratory PheWAS approach, highlighted the outcomes associated with genetically determined serum alanine and BCAAs across the phenome, nevertheless, the association's causality could not be determined. The PheWAS analysis elucidated significant associations between serum alanine and diabetes and significant associations between serum BCAAs and several cardiometabolic risk factors including diabetes, hyperlipidaemia, hypercholesterolemia, disorders of lipid metabolism, and hypertension.

The two-sample MR analysis using the largest available GWAS studies on common complex disorders, showed that genetically elevated levels of serum valine is associated with higher risk of type 2 diabetes and higher genetically determined levels of serum alanine and BCAAs are associated with higher levels of total cholesterol, triglycerides and LDL levels, and with lower levels of HDL concentrations.

The association between serum BCAAs with diabetes have been reported in previous observational studies that reported an association between serum BCAAs with insulin resistance (57, 298, 299). The observational studies demonstrated that serum levels of BCAAs and aromatic amino acids (tyrosine and phenylalanine) are significantly different among individual who are lean, obese, and those who have diabetes, that is highly related to insulin resistance (254, 300-303). The significant associations of amino acids with future diabetes were identified in different populations including Americans (37, 57, 236), young Finns (304), and Chinese (300). A mendelian randomization study investigated the association between genetically determined serum amino acids (glycine, phenylalanine, tryptophan, tyrosine, methionine, valine, leucine, serine) using summary level data

of a sub-cohort of European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam, KORA (Kooperative Gesundheitsforschung in der Region Augsburg) and TwinsUK studies and type 2 diabetes using GWAS study of DIAGRAM (108). This MR study reported that higher tyrosine levels might be causally associated with increased the risk of type 2 diabetes and the association is independent of other diabetes-associated amino acids (108).

Observational studies also suggest that circulating BCAAs are associated with several cardiometabolic risk factors including dyslipidaemia and are indicators of coronary artery disease (59, 61, 305). Tobias et al conducted a prospective cohort study among 27 041 healthy US women in the Women's Health Study to investigate the association between plasma BCAA and incident CVD and T2D. After 18.6 years of follow up, they found that BCAAs concentrations were strongly associated with incident CVD events (n=2207) (HR per SD, 1.13; 95%CI's, 1.08–1.18), adjusted for age, BMI, smoking, diet, physical activity, and other established CVD risk factors (59). They also reported that BCAA–CVD association was greater among women who developed T2D before CVD (HR, 1.20; 95% CI, 1.08–1.32) compared to women without T2D (HR, 1.08; 95% CI, 1.03–1.14). They concluded that any dysfunction in BCAA metabolism may capture the long-term risk of the common cause underlying T2D and CVD (59).

Moreover, a recent meta-analysis including 21 non-prospective and prospective studies assessing the effects of circulating BCAAs on cardiovascular disease (CVD) risk reported that elevated concentrations of circulating Isoleucine were associated with increased risks of CVD. In their subgroup analysis among adults <=60 years of age, they observed a 15%, 13%, and 9% increased risk of developing CVD per study-specific SD difference for isoleucine (RR 1.15 [1.11–1.19]), valine (RR 1.13 [1.09–1.17]), and leucine (RR 1.09 [1.03–1.16]), respectively but no such associations reported for adults older than 60 years (306).

Moreover, two MR studies reported an increased association of serum BCAAs with the risk of T2D (21, 131). The first study used a single instrument for BCAA levels and summary-level data from the DIAbetes Genetics Replication And Meta-analysis consortium and included 74 124 type 2 diabetes cases and 824 006 controls from European ancestry (21). This study also replicated their findings by utilizing the FinnGen consortium study which included 11 006 type 2 diabetes cases and 82 655 controls, both from European ancestry (21). The second study, which also reported a causal role of BCAA metabolism in the aetiology of type 2 diabetes, used meta-analysis of GWAS studies of Fenland, KORA and TwinsUK (total sample=16 596 individuals) for the amino acid traits and for the diabetes outcome they used meta-analysis of GWAS studies of DIAGRAM, EPIC-InterAct, GoDARTs, and UK Biobank studies (47 877 type 2 diabetes cases and 267 694 controls)(131). Therefore, compared to those two MR studies that used a single instrument in the first study (21) and GWAS study of serum BCAAs on around 16 000 individuals in the second study (131), my study was conducted in a much larger setting using the GWAS study of serum amino acids on more than 117 000 individuals from European ancestry.

The mechanism in which higher circulating levels of BCAAs increase the risk of T2D is not fully elucidated. However, as the levels of BCAAs concentrations are reported to be altered before the development of T2DM (57), epidemiological studies have implicated the alterations in metabolism of BCAAs for reduced cardiometabolic health including IR, elevated blood pressure, dyslipidaemia, which are indicators of coronary artery disease (59-63). Insulin plays a role in catabolism of BCAAs and has been long recognized as the regulator of BCKD complex (branched-chain alpha-keto acid dehydrogenase) (237, 307, 308). Insulin suppresses the catabolism of BCAAs by reducing the levels of BCKD enzyme which is implicated as the potential aetiology of BCAAs increased levels in individuals with obesity and type 2 diabetes (237, 308). Insulin

resistance was also shown to be correlated with alterations of other amino acids including alanine and glycine (309).

Accumulating evidence suggests that increased levels of BCAAs concentrations might adversely affect the regulation of glucose homeostasis, as the oxidation of BCAAs releases glucose consumption in the skeletal muscle, and this abnormal glucose homeostasis may lead to future risk of developing type 2 diabetes (57, 307, 310).

Moreover, higher circulation of BCAAs is reported to activate mTORC (mammalian target of rapamycin complex that controls protein synthesis) and its downstream effecter S6K1 (Ribosomal protein S6 kinase beta-1) in the muscle, adipose tissue, and liver, and this persistent activation results in serine phosphorylation of Insulin receptor substrate 1 (IRS-1) and therefore the inhibition of IRS-1, leads to insulin resistance.

There is increasing evidence suggesting a potential effect of free plasma free amino acids amino on cancer (67, 307), particularly BCAAs (311). A study has shown that elevated levels of plasma BCAAs concentration increase the risk of a future diagnosis of pancreatic cancer two-fold and they concluded that the levels of BCAAs have the potential to be used as a general marker to diagnose pancreatic ductal adenocarcinoma (67, 307). Another study suggested enrichment of BCAAs levels can suppress the growth of cultured human hepatocellular carcinoma cells (312). BCAAs are consumed by tumours in many biosynthetic pathways (64, 65), hence BCAAs are an energy source for cancer cells that obtain their BCAAs supply from circulation or surrounding tissues(65).

However, in my study, I didn't observe any association between circulating levels of BCAAs and cancer using the MR-PheWAS approach. This could be related to the lack of power in my study as the UKB study has not included a large number of cancer cases.

Finally, in this study using the summary GWAS data of the UKB study on over half a million European participants, I didn't identify any potential causal association for genetically elevated levels of serum histidine, glycine, glutamine, tyrosine, and phenylalanine with any traits or diseases.

Higher levels of triglycerides and lower levels of HDL are shown to be related to metabolic syndrome and insulin resistance which are all risk factors for cardiometabolic diseases (313). There is evidence reporting the involvement of serum BCAAs in the development of dyslipidaemia (314) as studies have revealed that higher serum BCAA were associated with an increased risk of TG levels (5,6) and another cross-sectional study reported higher total BCAA levels were associated with higher levels of TG and lower level of HDL (7). However, the evidence for the association between serum BCAAs with LDL levels are inconclusive (314) .

For the possible mechanism, it has been suggested that serum leucine is involved in the Rag GTPase-dependent activation of the mammalian target of rapamycin (mTOR) signalling pathway (314, 315) which promotes lipid synthesis through activating the sterol regulatory element-binding transcription factor-1 (SREBP1) (314). This SREBP1 is a transcriptional regulator of insulin-stimulated fatty acid synthesis (314, 316). Furthermore, mTOR activation by BCAAs is different compared to the activation by insulin or by the insulin-like growth factor (314, 317) and therefore serum BCAA might be directly involved in the process of lipid metabolism and causing dyslipidaemia (314).

In my study, I also detected that genetically elevated serum levels of alanine were associated with both higher systolic and diastolic blood pressure. However, no further significant association was detected between genetically elevated levels of serum amino acids including BCAAs, glycine, glutamine, histidine, tyrosine or phenylalanine with blood pressure.

The observed positive association in my study for genetically elevated levels of serum alanine and higher blood pressure is in agreement with a cohort study that assessed the effect of dietary amino acids intake, not serum levels, on blood pressure in patients with prevalent cardiovascular disease and concluded that intakes of methionine and alanine were associated with higher blood pressure, while intakes of threonine and histidine showed protective effect and were inversely associated with blood pressure (318). The evidence from observational studies on the relationship between dietary or serum BCAAs with blood pressure is inconclusive (319) and in this study, I didn't detect any association between genetically determined serum BCAAs with neither systolic nor diastolic blood pressure.

Using the largest available GWAS studies on coronary artery disease, I didn't observe any significant associations between genetically elevated levels of serum amino acids and the risk of coronary artery disease. However, the findings from observational studies for the relationship between serum amino acids and CAD show significant associations. A case-control study in the Chinese population (320) reported that BCAAs are significantly related to the development of CAD. Also, a prospective cohort study with 12.3 years of follow up including 9 584 Finnish men reported alanine, phenylalanine, and tyrosine were associated with increased risk of CAD (321).

Furthermore, I didn't identify any significant associations between genetically elevated serum levels of amino acids and the risk of breast or prostate cancer. These findings on breast cancer, is line with a result of a nested case-control study that indicated no consistent association between serum BCAAs (leucine, isoleucine, and valine) with breast cancer risk (322). This nested case-control study (322) included participants of the Nurses' Health Study (NHS) and NHSII (1997 cases and 1997 controls) and further validated their analysis by including the participants of the Women's Health Study (WHS) and reported that the elevated serum BCAA levels were associated with lower risk of breast cancer among premenopausal NHSII women while associated with higher risk of breast cancer among postmenopausal women in NHS but not in the WHS. Furthermore, another cohort study that looked at the association between dietary intake of BCAAs and breast cancer risk among the 196 161 women participants of the NHS and NHSII study found no association after 20.8 years of follow up and concluded that dietary intakes of BCAAs are not likely a risk factor for breast cancer (323). However, a case-control study including 350 breast cancer cases and 700 controls found that higher dietary intake of BCAAs was significantly associated with a reduced risk of postmenopausal breast cancer (324).

Cancer is a complex disease that survives through cell proliferation which requires energy from protein biosynthesis (325). There are cumulating research evidence showing that dietary restriction of amino acids plays role in cancer interventions including glycine restriction (325, 326), leucine deprivation (325, 327), and glutamine blockade (325, 328, 329). However, in my study, I didn't detect any association between genetically determined serum amino acid levels and the risk of breast or prostate cancer.

In addition, in my study I investigated the relationship between genetically elevated serum amino acid concentrations and risk of Alzheimer's disease, and I found no significant association. This finding is not in agreement with the findings of observational studies that reported, the serum and brains of patients diagnosed with Alzheimer's' disease have altered levels of serum amino acids (330). As serum BCAAs compete to enter into the brain with aromatic amino acids including phenylalanine, tyrosine, and tryptophan thus any alteration in the levels of serum BCAAs can affect the levels of the neurotransmitters such as serotonin, dopamine, epinephrine, and norepinephrine in the brain (330). Increasing the ratio of BCAAs to tryptophan and threonine results in hyperphagia and is associated with central serotonin depletion (331). Higher levels of serum BCAAs through dietary supplementation with BCAAs in rats has shown to cause a decrease in neural growth factor (NGF) in the hippocampus (330, 332) is involved in the formation of memory in the brain and is believed to suffer from neuronal loss among AD patients (330). Another study reported decreased levels of serum valine levels (as a BCAA amino acid) in AD patients (330, 333). However, the role of serum BCAAs in AD is still unclear. Aromatic amino acids including tyrosine and phenylalanine, as metabolized through the same metabolic pathway, have been shown to have a beneficiary effect on memory and cognitive function (17, 334). A research study reported decreased levels of aromatic serum amino acid levels in AD patients (330, 333) while another study showed increased levels of phenylalanine in AD patients (330, 335), and also a study in rats showed increased levels of phenylalanine in the brain (330, 336). These findings on the role of aromatic amino acids in AD are also conflicting and require further investigations.

4.4.1 Strength and limitation

The strength of this study is that I investigated a causal effect of proteogenic serum amino acids, particularly BCAAs, in a large sample of ~ 117 000 individuals across the whole spectrum of disease outcomes. The PheWAS design enabled us to investigate the association between serum amino acids status with a wide range of phenotypes (617 phenotypes) and further explore a causal, independent association between the selected serum amino acids status and a wide spectrum of disease outcomes using multiple methods including both univariable and multivariable MR methods and additionally testing the robustness of our findings by applying IVW, MR-Egger and MR-PRESSO tests.

Therefore, the main strength of this study is related to its design and the use of MR analysis with its known ability to overcome the environmental confounding that exist in observational studies through using genetic instruments that are inherited randomly at conception and are considered unbiased lifetime exposures.

This study has some limitations. Firstly, our analysis merely included individuals of European ancestry and the included population of the UKB study are young and health-oriented thus, caution is needed when generalizing the results to other populations (337). Secondly, some of the included diseases in this study have a very low prevalence in the UKB, so although I had a large number of controls, I may have had very few cases. It is consequently possible that some of my Null results are false and only due to a lack of power. Thirdly, for the diabetes analysis I used the UKB GWAS study for serum amino acids and also used DIAGRAM study for diabetes outcome which includes participants from the UKB study as well, so there is an overlap that might have resulted in an overestimation of the observed association.

Chapter Five

Summary of findings & future perspectives

Summary of applied methods and findings

Through this thesis, I have applied observational, molecular, and genetic epidemiological approaches to comprehensively investigate, the genetic architecture of 9 proteogenic amino acids levels (alanine, BCAAs, glycine, glutamine, histidine, and aromatic amino acids) and their related metabolic pathways as well as the role of modifiable lifestyle factors in levels of aforementioned serum amino acids and lastly the genetically predicted effect of those serum amino acids concentrations on the risk of common complex disorder and assess any potential causality.

To this end, I have utilized the data from the UKB study which consists of data on half a million participants of European ancestry. The UKB provides data on lifestyle factors including macronutrient intake (CHO, fat, and PRO) (n=210 977), accelerometer-measured physical activity (n=103 669), and bio-impedance-measured body fat percentage (n=492 787. The UKB also has data on serum amino acids levels of ~117 000 participants as well as genotypic data and phenotypic data (n=487 442).

In the first chapter of this thesis, I have reviewed evidence on the biology and structure of the proteogenic amino acids, and further reviewed epidemiological evidence on the association of each amino acids with complex disorders.

In the second chapter, I explored the genetic architecture of serum amino acids data of the UKB study and performed a linear mixed model regression using BOLT-LMM in Linux environment. I used this method as it considers population stratification and relatedness in the sample population, therefore it can be applied with minimal sample exclusions, as well as boosting the power while controlling false positives (338, 339). This method also allows the assumption of complex genetic architecture for traits; thus, it does not assume that all variants are causal with small effect sizes to

form a Gaussian distribution (338, 339). Through the BOLT-LMM method, I identified significant ($p < 1 \times 10^{-5}$) independent SNPs ($r^2 < 0.001$) associated with each of serum amino acids, ranging from 48 IVs for serum phenylalanine to 138 IVs for serum glutamine, all with F-statistic > 10, representing strong instruments. I also calculated total proportion of variance explained (R^2) for serum amino acids, which ranged from 1% for isoleucine (combined F statistic of 24.35) to 6% for serum glycine (combined F statistic of 53.58). The proportion of variance explained for serum BCAAs (isoleucine, leucine and valine) ranged from 1% to 3% (with a F statistic of 24 to 34).

In addition, in the second chapter I used FUMA web tool (340) to assess the related metabolic pathways and related important genes to the serum amino acids. The analysis showed that most of the significant SNPs associated with the serum amino acids and SNPs in LD with the significant SNPs lie within intronic followed by intergenic gene regions for all serum amino acids with an exception for serum tyrosine, in which the most significant SNPs, and SNPs in LD with the significant SNPs, were in intronic following by ncRNA-intronic gene regions. The analysis highlighted several genetic risk loci associated with serum amino acids ranging from 8 genomic risk loci for serum isoleucine and phenylalanine to 45 genomic risk loci for serum glutamine. The top signal from gene-based analysis revealed the most common genes among all serum amino acids as ***GCKR, MIP, SLC*** gene family***, COBLL1, LMO1, CPS1, GLS2*** and ***SPRYD4.*** The gene-set analysis highlighted pathways in the regulation of metabolites (small molecules and protein measures) for all serum amino acids, except for serum phenylalanine, and regulation in plasma free amino acid levels (adjusted for twenty other PFAAs) for all serum amino acids, except serum tyrosine. Many other pathways were highlighted including the pathways related to type 2 diabetes for serum glutamine, blood pressure (both systolic and diastolic) for serum leucine, blood lipid profile including triglycerides, cholesterol, HDL, and LDL for serum glycine, leucine and tyrosine,

Crohn's disease and inflammatory bowel disease for serum glycine and leucine, mood instability for serum leucine and glutamine, bipolar disorder, autism spectrum disorder or schizophrenia for serum glutamine, and Alzheimer's disease for serum tyrosine.

In the third chapter, I assessed the potential effect of modifiable risk factors on levels of 9 serum proteogenic amino acids and used Linear Mixed-Effects Models in an observational setting. This method avoids p-value inflation and considers variations in exposure measurements when its measured/reported multiple times (341). The modifiable risk factors include dietary intakes of CHO, fat, and PRO, the total average of accelerometer-measured physical activity, and bio-impedance-measured body fat (%).

Through the observational analysis, I did not find any association between macronutrient intake and serum amino acids. I revealed that accelerometer-measure physical activity was inversely associated with serum isoleucine and valine levels. I also showed that higher measured body fat (%) was significantly associated with higher levels of serum alanine, BCAAs, and aromatic amino acids while inversely associated with serum glycine and glutamine.

In this chapter, I used Mendelian randomization approach to investigate the effect of genetically determined modifiable lifestyle factors on serum amino acids concentration. The MR method proxies risk factors by using genetic variants to reinforce causal inferences of observational studies and evaluate whether a modifiable risk factor has a causal role in the aetiology of the disease (342). Therefore, using the genetic instruments proxies avoids biases such as confounding and reverse causation which are the main limitations of observational studies.

Using the MR method, I detected PRO and fat intake are positively and significantly (FDR <0.05) associated with serum BCAAs levels, while the CHO intake was inversely related to the levels of

serum BCAAs. These results were consistent using sensitivity analysis methods and therefore the observed associations were not due to a pleiotropy effect. I also found a positive significant association between genetically determined accelerometer-measured physical activity and serum valine. However, only 2 instruments were available for this analysis and sensitivity analysis could not be conducted due to the limited number of IVs. For body fat (%), I found a nominal significant positive association between genetically determined measured body fat (%) and serum glycine and glutamine.

In the fourth chapter, I investigated the association between genetically determined serum amino acids concentrations and the risk of disease and applied an MR-PheWAS analysis. PheWAS is a method of assessing comprehensive associations between wide ranges of coded phenotypes and genetic variants and is a tool that allows unbiased assessment of target genotypes across multiple phenotypes (343).

Under the MR-PheWAS paradigm, I revealed that higher levels of serum alanine, leucine and valine were associated with the higher risk of diabetes, and higher levels of serum isoleucine and valine are related to disorders of lipid metabolism, hyperlipidaemia, and hypercholesterolemia. In addition, higher levels of serum valine showed to be associated with a higher risk of hypertension. The MVMR analysis (using 52 common SNPS among all serum BCAAs) showed that the observed associations in the previous step, were not related to serum isoleucine only.

In addition, the Two-sample MR analysis replicated the observed significant associations for genetically predicted serum alanine and leucine levels with an increased risk of diabetes, and the significant association between serum valine and an increased risk of diabetes and hypertension. These results were consistent across both IVW and weighted median methods and remained

significant after removing the outliers, using the MR-PRESSO method that showed the absence of pleiotropic effect.

In the fifth chapter, I replicated the findings of this thesis, under the MR paradigm, and used the largest available GWAS studies in independent populations for common complex disorders including type 2 diabetes, coronary artery disease, Alzheimer's disease, breast cancer, prostate cancer as well as cardiometabolic risk factors including blood pressure and lipid profile (total cholesterol, triglycerides, LDL and HDL levels).

The findings successfully replicated the significant associations for a 51% increased risk of type 2 diabetes with higher levels of genetically predicted serum valine. Genetically elevated serum levels of alanine, isoleucine, and valine were associated with higher levels of total cholesterol, triglycerides, and LDL and genetically elevated levels of serum leucine were associated with higher TC, and TG levels. All genetically determined serum BCAAs were associated with lower levels of HDL. These observed significant associations were consistent using sensitivity analysis including weighted median and after removing outliers using the MR-PRESSO method. In addition, the associations remained significant after correcting for multiple testing using FDR.

Conclusion

In conclusion, the comprehensive analyses conducted in this thesis provide a thorough map of the genetic architecture of 9 proteogenic amino acids (serum alanine, BCAAs, histidine, glycine, glutamine, and aromatic amino acids). This thesis has provided evidence that higher genetically determined PRO and fat intake (% of energy) increases the levels of serum BCAAs, while the genetically predicted levels of CHO intake decrease the serum BCAAs levels. Using MR analysis, I have also shown a potential causal association between higher levels of serum BCAAs and an increased risk of type 2 diabetes and higher levels of lipid total cholesterol, triglycerides, LDL, and HDL levels, which are all important risk factors in developing the risk cardiovascular disease. Therefore, serum BCAAs are a relevant player in disease development however, the complex network and pathways require further detailed investigations.

Future prospective

BCAAs are popular supplements among gymgoers and athletes. In this study, I have shown that higher levels of serum BCAAs increase the risk of diabetes and cardiometabolic risk factors. However, as there is no established biomarker to represent the exact levels of serum BCAAs, whether this higher intake of BCAAs among those with higher levels of physical activity increases their risk of diabetes and cardiovascular disease is still yet to be elucidated. To my knowledge, no cohort study has collected large-scale data on BCAAs or amino acids supplements intake. A large study comprising such a data would shed light on the association between the BCAAs supplements intake and the risk of complex disorders.

Funding research studies through governmental bodies would make more certainty regarding the health safety and dosage of intake, particularly for professional athletes who use these supplements to reduce fatigue and muscle alleviation. This is a crucial step towards generating recommendations by public health organisations, to clarify whether the BCAA's supplements use is safe for all subjects with any different levels of physical activity or perhaps there is an exception regarding the target population for the intake of such supplements.

In terms of establishing a valid biomarker for serum amino acids, new evidence has emerged suggesting a robust role of the human gut microbiota in its metabolites of the host (344). It has been reported that the gut microbiota can synthesize essential amino acids (344, 345) including BCAAs, and a recent study reported a positive correlation between the biosynthesised BCAAs in the gut microbial and circulating BCAAs levels in a healthy population (346). Moreover, this recent study identified a positive correlation between microbial BCAA biosynthesis genes and plasma levels of BCAA, and further reported decreased abundance of microbial genes involved in the biosynthesis of essential amino acids attributable to the depletion of 2 genes (Eubacterium and Prevotella), among patients with heart failure (347). Therefore, new studies are required to provide novel insight into the potential role of the gut microbiota in proteogenic amino acids metabolism, particularly BCAAs, and their subsequent effect on the risk of complex disorders.

Diabetes mellitus has been long regarded as a risk factor for developing cardiovascular disease (348). Along with diabetes, high cholesterol levels and high blood pressure lead to an increased risk of cardiovascular complications (349). In this thesis, I have provided evidence for supporting

the potential role of higher BCAAs levels with risk of type 2 diabetes. Serum BCAAs levels might have potential as an emerging biomarker in risk prediction for cardiovascular disease and its complications including angina, coronary artery diseases (CADs), myocardial infarction, stroke, peripheral artery disease (PAD), and congestive heart failure. However, further larger studies are required to support and confirm this.

Implementing strategies to apply new potential novel biomarkers in a cost-effective manner to detect patients who are at higher risk of coronary artery disease, is an efficient way to reduce the burden of such diseases. It was not in the scope of this thesis to examine predictive ability of using serum BCAA's levels in early detection of coronary artery disease, though this could be of interest for future work.

Furthermore, large accessible biobank studies with comprehensive follow up data as well as genotypic and phenotypic data have enabled researchers to identify genetic risk factors associated with a large number of diseases. These biobank studies have also aided to understanding of biological pathways that have been implicated for non-communicable diseases and helps to target novel therapies for identified diseases. Therefore, improving disease prediction by refining genetic risk scores for a wider range of disease would also be of an interest for future work.

References

1. BYJUS.com. Amino acid structure [Available from: https://byjus.com/chemistry/amino-acid-structure/.

2. Information NCfB. PubChem Compound Summary for CID 6274, Histidine. 2022 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/histidine.

3. Negro M, Giardina S, Marzani B, Marzatico F. Branched-chain amino acid supplementation does not enhance athletic performance but affects muscle recovery and the immune system. J Sports Med Phys Fitness. 2008;48(3):347-51.

4. Information NCfB. PubChem Compound Summary for CID 5962, Lysine 2022 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/Lysine.

5. Information NCfB. PubChem Compound Summary for CID 6137, Methionine. 2022.

6. Information NCfB. PubChem Compound Summary for CID 6288, L-threonine 2022 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/L-threonine.

7. Information NCfB. PubChem Compound Summary for CID 6305, Tryptophan. 2022.

8. Yi J, et al. L-arginine and Alzheimer's disease. Int J Clin Exp Pathol,. 2009;2(3):211-38.

9. Dong JY, et al 162(6): p. 959-65. Effect of oral L-arginine supplementation on blood pressure: a meta-analysis of randomized, double-blind, placebo-controlled trials. Am Heart 2011;162(61):959-65.

10. Information NCfB. PubChem Compound Summary for CID 5950, Alanine. 2022.

11. Information NCfB. PubChem Compound Summary for CID 6267, Asparagine 2022 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/Asparagine.

12. Information NCfB. PubChem Compound Summary for CID 5862, Cysteine 2022 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/Cysteine.

13. Cruzat V, Macedo Rogero M, Noel Keane K, Curi R, Newsholme P. Glutamine: Metabolism and Immune Function, Supplementation and Clinical Translation. Nutrients. 2018;10(11).

14. Gundersen RY, Vaagenes P, Breivik T, Fonnum F, Opstad PK. Glycine--an important neurotransmitter and cytoprotective agent. Acta Anaesthesiol Scand. 2005;49(8):1108-16.

15. Information NCfB. PubChem Compound Summary for CID 145742, Proline 2022 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/Proline.

16. Information NCfB. PubChem Compound Summary for CID 5951, Serine 2022 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/Serine.

17. Fernstrom JD, Fernstrom MH. Tyrosine, phenylalanine, and catecholamine synthesis and function in the brain. J Nutr. 2007;137(6 Suppl 1):1539S-47S; discussion 48S.

18. Information NCfB. PubChem Compound Summary for CID 5950, Alanine. [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/alanine.

19. aminoacidstudies.org. Alanine 2022 [cited 2022. Available from: https://aminoacidstudies.org/l-alanine/.

20. Litwack G. Human Biochemistry, Second Edition. 2 ed2021.

21. Yuan S, Larsson SC. An atlas on risk factors for type 2 diabetes: a wide-angled Mendelian randomisation study. Diabetologia. 2020;63(11):2359-71.

22. Zhang S, Zeng X, Ren M, Mao X, Qiao S. Novel metabolic and physiological functions of branched chain amino acids: a review. J Anim Sci Biotechnol. 2017;8:10.

23. aminoacidstudies.org. Branch Chain Amino Acid (BCAAs) 2022 [Available from: https://aminoacidstudies.org/bcaa/#Dietary_sources.

24. Nie C, He T, Zhang W, Zhang G, Ma X. Branched Chain Amino Acids: Beyond Nutrition Metabolism. Int J Mol Sci. 2018;19(4).

25. Adeva-Andany MM, López-Maside L, Donapetry-García C, Fernández-Fernández C, Sixto-Leal C. Enzymes involved in branched-chain amino acid metabolism in humans. Amino Acids. 2017;49(6):1005-28.

26. Fan P, Li L, Rezaei A, Eslamfam S, Che D, Ma X. Metabolites of Dietary Protein and Peptides by Intestinal Microbes and their Impacts on Gut. Curr Protein Pept Sci. 2015;16(7):646-54.

27. Sperringer JE, Addington A, Hutson SM. Branched-Chain Amino Acids and Brain Metabolism. Neurochem Res. 2017;42(6):1697-709.

28. Xu J, Jakher Y, Ahrens-Nicklas RC. Brain Branched-Chain Amino Acids in Maple Syrup Urine Disease: Implications for Neurological Disorders. Int J Mol Sci. 2020;21(20).

29. Neinast M, Murashige D, Arany Z. Branched Chain Amino Acids. Annu Rev Physiol. 2019;81:139-64.

30. Jewell JL, Russell RC, Guan KL. Amino acid signalling upstream of mTOR. Nat Rev Mol Cell Biol. 2013;14(3):133-9.

31. Lynch CJ, Adams SH. Branched-chain amino acids in metabolic signalling and insulin resistance. Nat Rev Endocrinol. 2014;10(12):723-36.

32. Zhenyukh O, Civantos E, Ruiz-Ortega M, Sánchez MS, Vázquez C, Peiró C, et al. High concentration of branched-chain amino acids promotes oxidative stress, inflammation and migration of human peripheral blood mononuclear cells via mTORC1 activation. Free Radic Biol Med. 2017;104:165-77.

33. Liu Y, Dong W, Shao J, Wang Y, Zhou M, Sun H. Branched-Chain Amino Acid Negatively Regulates KLF15 Expression via PI3K-AKT Pathway. Front Physiol. 2017;8:853.

34. Neishabouri SH, Hutson SM, Davoodi J. Chronic activation of mTOR complex 1 by branched chain amino acids and organ hypertrophy. Amino Acids. 2015;47(6):1167-82.

35. Jang C, Oh SF, Wada S, Rowe GC, Liu L, Chan MC, et al. A branched-chain amino acid metabolite drives vascular fatty acid transport and causes insulin resistance. Nat Med. 2016;22(4):421-6.

36. van Zanten AR. Glutamine and antioxidants: status of their use in critical illness. Curr Opin Clin Nutr Metab Care. 2015;18(2):179-86.

37. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell Metab. 2009;9(4):311-26.

38. Doi M, Yamaoka I, Nakayama M, Mochizuki S, Sugahara K, Yoshizawa F. Isoleucine, a blood glucose-lowering amino acid, increases glucose uptake in rat skeletal muscle in the absence of increases in AMP-activated protein kinase activity. J Nutr. 2005;135(9):2103-8.

39. Nishimura J, Masaki T, Arakawa M, Seike M, Yoshimatsu H. Isoleucine prevents the accumulation of tissue triglycerides and upregulates the expression of PPARalpha and uncoupling protein in diet-induced obese mice. J Nutr. 2010;140(3):496-500.

40. Zhen H, Kitaura Y, Kadota Y, Ishikawa T, Kondo Y, Xu M, et al. mTORC1 is involved in the regulation of branched-chain amino acid catabolism in mouse heart. FEBS Open Bio. 2016;6(1):43-9.

41. Wiltafsky MK, Pfaffl MW, Roth FX. The effects of branched-chain amino acid interactions on growth performance, blood metabolites, enzyme kinetics and transcriptomics in weaned pigs. Br J Nutr. 2010;103(7):964-76.

42. Anthony JC, Anthony TG, Kimball SR, Vary TC, Jefferson LS. Orally administered leucine stimulates protein synthesis in skeletal muscle of postabsorptive rats in association with increased eIF4F formation. J Nutr. 2000;130(2):139-45.

43. Davis TA, Fiorotto ML. Regulation of muscle growth in neonates. Curr Opin Clin Nutr Metab Care. 2009;12(1):78-85.

44. Cheng Y, Zhang Q, Meng Q, Xia T, Huang Z, Wang C, et al. Leucine deprivation stimulates fat loss via increasing CRH expression in the hypothalamus and activating the sympathetic nervous system. Mol Endocrinol. 2011;25(9):1624-35.

45. Du Y, Meng Q, Zhang Q, Guo F. Isoleucine or valine deprivation stimulates fat loss via increasing energy expenditure and regulating lipid metabolism in WAT. Amino Acids. 2012;43(2):725-34.

46. Maclure M, Travis LB, Willett W, MacMahon B. A prospective cohort study of nutrient intake and age at menarche. Am J Clin Nutr. 1991;54(4):649-56.

47. Casperson SL, Sheffield-Moore M, Hewlings SJ, Paddon-Jones D. Leucine supplementation chronically improves muscle protein synthesis in older adults consuming the RDA for protein. Clin Nutr. 2012;31(4):512-9.

48. Qin LQ, Xun P, Bujnowski D, Daviglus ML, Van Horn L, Stamler J, et al. Higher branched-chain amino acid intake is associated with a lower prevalence of being overweight or obese in middle-aged East Asian and Western adults. J Nutr. 2011;141(2):249-54.

49. Shu X, Kang K, Zhong J, Ji S, Zhang Y, Hu H, et al. [Meta-analysis of branched chain amino acid-enriched nutrition to improve hepatic function in patients undergoing hepatic operation]. Zhonghua Gan Zang Bing Za Zhi. 2014;22(1):43-7.

50. Xue WX, Yang F, Duan JJ, Sun JB, Peng H. [Effect of oral administration of branched-chain amino acids supplementation on the mortality of patients with hepatocellular carcinoma: a meta-analysis]. Zhonghua Yi Xue Za Zhi. 2020;100(47):3793-800.

51. Nakamura H, Jinzu H, Nagao K, Noguchi Y, Shimba N, Miyano H, et al. Plasma amino acid profiles are associated with insulin, C-peptide and adiponectin levels in type 2 diabetic patients. Nutr Diabetes. 2014;4(9):e133.

52. Ruiz-Canela M, Toledo E, Clish CB, Hruby A, Liang L, Salas-Salvadó J, et al. Plasma Branched-Chain Amino Acids and Incident Cardiovascular Disease in the PREDIMED Trial. Clin Chem. 2016;62(4):582-92.

53. Grajeda-Iglesias C, Aviram M. Specific Amino Acids Affect Cardiovascular Diseases and Atherogenesis via Protection against Macrophage Foam Cell Formation: Review Article. Rambam Maimonides Med J. 2018;9(3).

54. Asghari G. FH, Teymoori F., Mirmiran P., Tohidi M., Azizi F. . High dietary intakes of branched-hain amino acids is associated with increased risk of insulin resistance in adults. J Diabetes. 2017.

55. Allam-Ndoul B. GF, Garneau V., Barbier O., Pérusse L., Vohl M. Associations between branched chain amino acid levels, obesity and cardiometabolic complications. Integr Obes Diabetes. 2015(1):157–62.

56. Gannon NP, Schnuck JK, Vaughan RA. BCAA Metabolism and Insulin Sensitivity - Dysregulated by Metabolic Status? Mol Nutr Food Res. 2018;62(6):e1700756.

57. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nat Med. 2011;17(4):448-53.

58. Luca A. Lotta RAS, Stephen J. Sharp,Stephen Burgess,Jian'an Luan,Therese Tillin,Amand F. Schmidt,Fumiaki Imamura,Isobel D. Stewart,John R. B. Perry,Luke Marney,Albert Koulman,Edward D. Karoly,Nita G. Forouhi,Rasmus J. O. Sjögren,Erik Näslund,Juleen R. Zierath,Anna Krook,David B. Savage,Julian L. Griffin,Nishi Chaturvedi,Aroon D. Hingorani,Kay-Tee Khaw,Inês Barroso,Mark I. McCarthy,Stephen O'Rahilly,Nicholas J. Wareham ,Claudia Langenberg. Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. PLoS Med. 2016.

59. Tobias DK, Lawler PR, Harada PH, Demler OV, Ridker PM, Manson JE, et al. Circulating Branched-Chain Amino Acids and Incident Cardiovascular Disease in a Prospective Cohort of US Women. Circ Genom Precis Med. 2018;11(4):e002157.

60. Yang R, Dong J, Zhao H, Li H, Guo H, Wang S, et al. Association of branched-chain amino acids with carotid intima-media thickness and coronary artery disease risk factors. PLoS One. 2014;9(6):e99598.

61. Yang P, Hu W, Fu Z, Sun L, Zhou Y, Gong Y, et al. The positive association of branched-chain amino acids and metabolic dyslipidemia in Chinese Han population. Lipids Health Dis. 2016;15:120.

62. Mels CM, Schutte AE, Schutte R, Huisman HW, Smith W, Fourie CM, et al. The link between vascular deterioration and branched chain amino acids in a population with high glycated haemoglobin: the SABPA study. Amino Acids. 2013;45(6):1405-13.

63. Magnusson M, Lewis GD, Ericson U, Orho-Melander M, Hedblad B, Engström G, et al. A diabetes-predictive amino acid score and future cardiovascular disease. Eur Heart J. 2013;34(26):1982-9.

64. DeBerardinis RJ, Chandel NS. Fundamentals of cancer metabolism. Sci Adv. 2016;2(5):e1600200.

65. Ananieva EA, Wilkinson AC. Branched-chain amino acid metabolism in cancer. Curr Opin Clin Nutr Metab Care. 2018;21(1):64-70.

66. Lei M, Li, XX., Zhang, Y. et al. . Acetylation promotes BCAT2 degradation to suppress BCAA catabolism and pancreatic cancer growth. . Sig Transduct Target Ther. 2020(5):70.

67. Mayers JR, Wu C, Clish CB, Kraft P, Torrence ME, Fiske BP, et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. Nat Med. 2014;20(10):1193-8.

68. Katagiri R, Goto A, Nakagawa T, Nishiumi S, Kobayashi T, Hidaka A, et al. Increased Levels of Branched-Chain Amino Acid Associated With Increased Risk of Pancreatic Cancer in a Prospective Case-Control Study of a Large Cohort. Gastroenterology. 2018;155(5):1474-82.e1.

69. Information NCfB. PubChem Compound Summary for CID 750, Glycine. 2022.

70. Nelson DLC, Michael M. (2005). . New York: W. H. Freeman. Principles of Biochemistry 4th ed2005.

71. Wang W, Wu Z, Dai Z, Yang Y, Wang J, Wu G. Glycine metabolism in animals and humans: implications for nutrition and health. Amino Acids. 2013;45(3):463-77.

72. Ubuka T. Handbook of Hormones (Second Edition);Comparative Endocrinology for Basic and Clinical Research: Academic Press; 2021.

73. aminoacids.org. Glycine 2022 [Available from: https://aminoacidstudies.org/glycine/#fn-1167-6.

74. Jackson AA, Dunn RL, Marchand MC, Langley-Evans SC. Increased systolic blood pressure in rats induced by a maternal low-protein diet is reversed by dietary supplementation with glycine. Clin Sci (Lond). 2002;103(6):633-9.

75. El Hafidi M, Pérez I, Zamora J, Soto V, Carvajal-Sandoval G, Baños G. Glycine intake decreases plasma free fatty acids, adipose cell size, and blood pressure in sucrose-fed rats. Am J Physiol Regul Integr Comp Physiol. 2004;287(6):R1387-93.

76. Quan H, Athirakul K, Wetsel WC, Torres GE, Stevens R, Chen YT, et al. Hypertension and impaired glycine handling in mice lacking the orphan transporter XT2. Mol Cell Biol. 2004;24(10):4166-73.

77. Wittemans LBL, Lotta LA, Oliver-Williams C, Stewart ID, Surendran P, Karthikeyan S, et al. Assessing the causal association of glycine with risk of cardio-metabolic diseases. Nat Commun. 2019;10(1):1060.

78. Díaz-Flores M, Cruz M, Duran-Reyes G, Munguia-Miranda C, Loza-Rodríguez H, Pulido-Casas E, et al. Oral supplementation with glycine reduces oxidative stress in patients with metabolic syndrome, improving their systolic blood pressure. Can J Physiol Pharmacol. 2013;91(10):855-60.

79. Miller RA, Harrison DE, Astle CM, Bogue MA, Brind J, Fernandez E, et al. Glycine supplementation extends lifespan of male and female mice. Aging Cell. 2019;18(3):e12953.

80. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, et al. Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol. 2012;8:615.

81. Floegel A, Stefan N, Yu Z, Mühlenbruch K, Drogan D, Joost HG, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes. 2013;62(2):639-48.

82. Merino J, Leong A, Liu CT, Porneala B, Walford GA, von Grotthuss M, et al. Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose. Diabetologia. 2018;61(6):1315-24.

83. Palmer ND, Stevens RD, Antinozzi PA, Anderson A, Bergman RN, Wagenknecht LE, et al. Metabolomic profile associated with insulin resistance and conversion to diabetes in the Insulin Resistance Atherosclerosis Study. J Clin Endocrinol Metab. 2015;100(3):E463-8.

84. Ding Y, Svingen GF, Pedersen ER, Gregory JF, Ueland PM, Tell GS, et al. Plasma Glycine and Risk of Acute Myocardial Infarction in Patients With Suspected Stable Angina Pectoris. J Am Heart Assoc. 2015;5(1).

85. Curi R, Lagranha CJ, Doi SQ, Sellitti DF, Procopio J, Pithon-Curi TC, et al. Molecular mechanisms of glutamine action. J Cell Physiol. 2005;204(2):392-401.

86. Cruzat VF, Krause M, Newsholme P. Amino acid supplementation and impact on immune function in the context of exercise. J Int Soc Sports Nutr. 2014;11(1):61.

87. Newsholme P. Why is L-glutamine metabolism important to cells of the immune system in health, postinjury, surgery or infection? J Nutr. 2001;131(9 Suppl):2515S-22S; discussion 23S-4S.

88. Durante W. The Emerging Role of l-Glutamine in Cardiovascular Health and Disease. Nutrients. 2019;11(9).

89. Hasani M, Mansour A, Asayesh H, Djalalinia S, Mahdavi Gorabi A, Ochi F, et al. Effect of glutamine supplementation on cardiometabolic risk factors and inflammatory markers: a systematic review and meta-analysis. BMC Cardiovasc Disord. 2021;21(1):190.

90. Ahluwalia GS, Grem JL, Hao Z, Cooney DA. Metabolism and action of amino acid analog anti-cancer agents. Pharmacol Ther. 1990;46(2):243-71.

91. Griffiths M, Keast D, Patrick G, Crawford M, Palmer TN. The role of glutamine and glucose analogues in metabolic inhibition of human myeloid leukaemia in vitro. Int J Biochem. 1993;25(12):1749-55.

92. Choi BH, Coloff JL. The Diverse Functions of Non-Essential Amino Acids in Cancer. Cancers (Basel). 2019;11(5).

93. Hosios AM, Hecht VC, Danai LV, Johnson MO, Rathmell JC, Steinhauser ML, et al. Amino Acids Rather than Glucose Account for the Majority of Cell Mass in Proliferating Mammalian Cells. Dev Cell. 2016;36(5):540-9.

94. Eagle H. Nutrition needs of mammalian cells in tissue culture. Science. 1955;122(3168):501-14.

95. Adams CD. Circulating Glutamine and Alzheimer's Disease: A Mendelian Randomization Study. Clin Interv Aging. 2020;15:185-93.

96. Andersen JV, Christensen SK, Aldana BI, Nissen JD, Tanila H, Waagepetersen HS. Alterations in Cerebral Cortical Glucose and Glutamine Metabolism Precedes Amyloid Plaques in the APPswe/PSEN1dE9 Mouse Model of Alzheimer's Disease. Neurochem Res. 2017;42(6):1589-98.

97. De Strooper B, Karran E. The Cellular Phase of Alzheimer's Disease. Cell. 2016;164(4):603-15.

98. Menon K, Marquina C, Liew D, Mousa A, de Courten B. Histidine-containing dipeptides reduce central obesity and improve glycaemic outcomes: A systematic review and meta-analysis of randomized controlled trials. Obes Rev. 2020;21(3):e12975.

99. Menon K, Marquina C, Hoj P, Liew D, Mousa A, de Courten B. Carnosine and histidine-containing dipeptides improve dyslipidemia: a systematic review and meta-analysis of randomized controlled trials. Nutr Rev. 2020;78(11):939-51.

100. Yu B, Li AH, Muzny D, Veeraraghavan N, de Vries PS, Bis JC, et al. Association of Rare Loss-Of-Function Alleles in HAL, Serum Histidine: Levels and Incident Coronary Heart Disease. Circ Cardiovasc Genet. 2015;8(2):351-5.

101. Watanabe M SM, Qureshi AR, Garcia-Lopez E, Barany P, Heimburger O, et al. , . Consequences of low plasma histidine in chronic kidney disease patients: Associations with inflammation, oxidative stress, and mortality. Am J Clin Nutr 2008: 87:1860–6.

102. DA. G. Low free serum histidine concentration in rheumatoid arthritis. A measure of disease activity. J Clin Invest. 1975:1975; 55:164–173.

103. Information NCfB. PubChem Compound Summary for CID 6057, Tyrosine. 2022 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/tyrosine.

104. (2022). NCfBI. PubChem Compound Summary for CID 6140 2022 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/phenylalanin.

105. Banderet LE, Lieberman HR. Treatment with tyrosine, a neurotransmitter precursor, reduces environmental stress in humans. Brain Res Bull. 1989;22(4):759-62.

106. Attipoe S, Zeno SA, Lee C, Crawford C, Khorsan R, Walter AR, et al. Tyrosine for Mitigating Stress and Enhancing Performance in Healthy Adult Humans, a Rapid Evidence Assessment of the Literature. Mil Med. 2015;180(7):754-65.

107. Jongkees BJ, Hommel B, Kühn S, Colzato LS. Effect of tyrosine supplementation on clinical and healthy populations under stress or cognitive demands--A review. J Psychiatr Res. 2015;70:50-7.

108. Jäger S, Cuadrat R, Wittenbecher C, Floegel A, Hoffmann P, Prehn C, et al. Mendelian Randomization Study on Amino Acid Metabolism Suggests Tyrosine as Causal Trait for Type 2 Diabetes. Nutrients. 2020;12(12).

109. Korner J, Cline GW, Slifstein M, Barba P, Rayat GR, Febres G, et al. A role for foregut tyrosine metabolism in glucose tolerance. Mol Metab. 2019;23:37-50.

110. Lopez Vicchi F, Luque, G.M., Brie, B., Nogueira, J.P., Garcia Tornadu, I. and Becu-Villalobos, D. . Dopaminergic Drugs in Type 2 Diabetes and Glucose Homeostasis. . Pharmacological Research. 2016:109, 74-80.

111. Gobeil É, Maltais-Payette I, Taba N, Brière F, Ghodsian N, Abner E, et al. Mendelian Randomization Analysis Identifies Blood Tyrosine Levels as a Biomarker of Non-Alcoholic Fatty Liver Disease. Metabolites. 2022;12(5).

112. Genome-Wide Association Studies Factsheet [Internet]. 2017 [cited 2019]. Available from: https://www.genome.gov/about-genomics/fact-sheets/Genome-Wide-Association-Studies-Fact-Sheet.

113. Ali C. Ravanpay CEaEH. Molecular Genetics and the Development of Targets for Glioma Therapy Youmans and Winn Neurological Surgery. 7th ed: Elsevier; 2017.

114. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol. 2013;31(12):1102-10.

115. Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. Annu Rev Genomics Hum Genet. 2016;17:353-73.

116. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23(R1):R89-98.

117. Burgess S, Daniel RM, Butterworth AS, Thompson SG. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. Int J Epidemiol. 2015;44(2):484-95.

118. Burgess S, Thompson DJ, Rees JMB, Day FR, Perry JR, Ong KK. Dissecting Causal Pathways Using Mendelian Randomization with Summarized Genetic Data: Application to Age at Menarche and Risk of Breast Cancer. Genetics. 2017;207(2):481-7.

119. Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol. 2017;46(6):2078-89.

120. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet. 2017;13(11):e1007081.

121. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1-22.

122. Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. PLoS Med. 2008;5(8):e177.

123. National Institute of Health- genetics home reference. What are genome-wide association studies? - Genetics Home Reference. (2018) 2018 [updated 03 March 2020. Available from: https://ghr.nlm.nih.gov/primer/genomicresearch/gwastudies.

124. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med. 2017;9(383).

125. Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes MV, Engmann JE, Shah T, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. Lancet. 2015;385(9965):351-61.

126. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Hum Mol Genet. 2018;27(R2):R195-r208.

127. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50(5):693-8.

128. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. Bmj. 2018;362:k601.

129. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985-98.

130. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14.

131. Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan J, Tillin T, et al. Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. PLoS Med. 2016;13(11):e1002179.

132. Greenfield JR, Farooqi IS, Keogh JM, Henning E, Habib AM, Blackwood A, et al. Oral glutamine increases circulating glucagon-like peptide 1, glucagon, and insulin concentrations in lean, obese, and type 2 diabetic subjects. Am J Clin Nutr. 2009;89(1):106-13.

133. Watanabe M, Suliman ME, Qureshi AR, Garcia-Lopez E, Bárány P, Heimbürger O, et al. Consequences of low plasma histidine in chronic kidney disease patients: associations with inflammation, oxidative stress, and mortality. Am J Clin Nutr. 2008;87(6):1860-6.

134. Gerber DA. Low free serum histidine concentration in rheumatoid arthritis. A measure of disease activity. J Clin Invest. 1975;55(6):1164-73.

135. Gaudet MM, Falk RT, Stevens RD, Gunter MJ, Bain JR, Pfeiffer RM, et al. Analysis of serum metabolic profiles in women with endometrial cancer and controls in a population-based case-control study. J Clin Endocrinol Metab. 2012;97(9):3216-23.

136. Ihata Y, Miyagi E, Numazaki R, Muramatsu T, Imaizumi A, Yamamoto H, et al. Amino acid profile index for early detection of endometrial cancer: verification as a novel diagnostic marker. Int J Clin Oncol. 2014;19(2):364-72.

137. Shi K, Wang Q, Su Y, Xuan X, Liu Y, Chen W, et al. Identification and functional analyses of differentially expressed metabolites in early stage endometrial carcinoma. Cancer Sci. 2018;109(4):1032-43.

138. Suzuki Y, Tokinaga-Uchiyama A, Mizushima T, Maruyama Y, Mogami T, Shikata N, et al. Normalization of abnormal plasma amino acid profile-based indexes in patients with gynecological malignant tumors after curative treatment. BMC Cancer. 2018;18(1):973.

139. Audet-Delage Y, Villeneuve L, Grégoire J, Plante M, Guillemette C. Identification of Metabolomic Biomarkers for Endometrial Cancer and Its Recurrence after Surgery in Postmenopausal Women. Front Endocrinol (Lausanne). 2018;9:87.

140. Pietkiewicz D, Klupczynska-Gabryszak A, Plewa S, Misiura M, Horala A, Miltyk W, et al. Free Amino Acid Alterations in Patients with Gynecological and Breast Cancer: A Review. Pharmaceuticals (Basel). 2021;14(8).

141. Trousil S, Lee P, Pinato DJ, Ellis JK, Dina R, Aboagye EO, et al. Alterations of choline phospholipid metabolism in endometrial cancer are caused by choline kinase alpha overexpression and a hyperactivated deacylation pathway. Cancer Res. 2014;74(23):6867-77.

142. Cheng SC, Chen K, Chiu CY, Lu KY, Lu HY, Chiang MH, et al. Metabolomic biomarkers in cervicovaginal fluid for detecting endometrial cancer through nuclear magnetic resonance spectroscopy. Metabolomics. 2019;15(11):146.

143. Zhou M, Guan W, Walker LD, Mezencev R, Benigno BB, Gray A, et al. Rapid mass spectrometric metabolic profiling of blood sera detects ovarian cancer with high accuracy. Cancer Epidemiol Biomarkers Prev. 2010;19(9):2262-71.

144. Hilvo M, de Santiago I, Gopalacharyulu P, Schmitt WD, Budczies J, Kuhberg M, et al. Accumulated Metabolites of Hydroxybutyric Acid Serve as Diagnostic and Prognostic Biomarkers of Ovarian High-Grade Serous Carcinomas. Cancer Res. 2016;76(4):796-804.

145. Horala A, Plewa S, Derezinski P, Klupczynska A, Matysiak J, Nowak-Markwitz E, et al. Serum Free Amino Acid Profiling in Differential Diagnosis of Ovarian Tumors-A Comparative Study with Review of the Literature. Int J Environ Res Public Health. 2021;18(4).

146. Miyagi E, Maruyama Y, Mogami T, Numazaki R, Ikeda A, Yamamoto H, et al. Comparison of plasma amino acid profile-based index and CA125 in the diagnosis of epithelial ovarian cancers and borderline malignant tumors. Int J Clin Oncol. 2017;22(1):118-25.

147. Cala MP, Aldana J, Medina J, Sánchez J, Guio J, Wist J, et al. Multiplatform plasma metabolic and lipid fingerprinting of breast cancer: A pilot control-case study in Colombian Hispanic women. PLoS One. 2018;13(2):e0190958.

148. Eniu DT, Romanciuc F, Moraru C, Goidescu I, Eniu D, Staicu A, et al. The decrease of some serum free amino acids can predict breast cancer diagnosis and progression. Scand J Clin Lab Invest. 2019;79(1-2):17-24.

149. Shen J, Yan L, Liu S, Ambrosone CB, Zhao H. Plasma metabolomic profiles in breast cancer patients and healthy controls: by race and tumor receptor subtypes. Transl Oncol. 2013;6(6):757-65.

150. Scioscia KA, Snyderman CH, Wagner R. Altered serum amino acid profiles in head and neck cancer. Nutrition and Cancer. 1998;30(2):144-7.

151. Lai H-S, Lee J-C, Lee P-H, Wang S-T, Chen W-J. Plasma free amino acid profile in cancer patients. Seminars in Cancer Biology. 2005;15(4):267-76.

152. Watanabe K TE, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA

Nat Commun. 2017;8(1):1826.

153. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nature Reviews Methods Primers. 2021;1(1):59.

154. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature Genetics. 2018;50(8):1112-21.

155. Breen G, Li Q, Roth BL, O'Donnell P, Didriksen M, Dolmetsch R, et al. Translating genome-wide association findings into new therapeutics for psychiatry. Nat Neurosci. 2016;19(11):1392-6.

156. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190-5.

157. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. Nature Genetics. 2018;50(7):906-8.

158. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nature Genetics. 2015;47(3):284-90.

159. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

160. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11(4):e1004219.

161. Medicine NLo. GCKR glucokinase regulator 2022 2022 [Available from: GCKR glucokinase regulator [Homo sapiens (human)] - Gene - NCBI (nih.gov).

162. Onuma H, Tabara Y, Kawamoto R, Shimizu I, Kawamura R, Takata Y, et al. The GCKR rs780094 polymorphism is associated with susceptibility of type 2 diabetes, reduced fasting

plasma glucose levels, increased triglycerides levels and lower HOMA-IR in Japanese population. J Hum Genet. 2010;55(9):600-4.

163. Ayka A, Şehirli A. The Role of the SLC Transporters Protein in the Neurodegenerative Disorders. Clin Psychopharmacol Neurosci. 2020;18(2):174-87.

164. Fang X, Liu Y, Xiao W, Zhao N, Zhu C, Yu D, et al. Prognostic SLC family genes promote cell proliferation, migration, and invasion in hepatocellular carcinoma. Acta Biochim Biophys Sin (Shanghai). 2021;53(8):1065-75.

165. Li C, Ou R, Chen Y, Gu X, Wei Q, Cao B, et al. Mutation analysis of seven SLC family transporters for early-onset Parkinson's disease in Chinese population. Neurobiol Aging. 2021;103:152.e1-.e6.

166. Kotka M, Lieden A, Pettersson S, Trinchieri V, Masci A, D'Amato M. Solute carriers (SLC) in inflammatory bowel disease: a potential target of probiotics? J Clin Gastroenterol. 2008;42 Suppl 3 Pt 1:S133-5.

167. Yu Y, Yu Y, Chen P, Li J, Zhu Y, Zhai Y, et al. A novel MIP gene mutation associated with autosomal dominant congenital cataracts in a Chinese family. BMC Med Genet. 2014;15:6.

168. Chen L, Tian Q, Zhang M, Chen D, Gao X, Yang H, et al. CPS1 T1405N polymorphism, HDL cholesterol, homocysteine and renal function are risk factors of VPA induced hyperammonemia among epilepsy patients. Epilepsy Res. 2019;154:139-43.

169. Bardelli V, Arniani S, Pierini V, Di Giacomo D, Pierini T, Gorello P, et al. T-Cell Acute Lymphoblastic Leukemia: Biomarkers and Their Clinical Usefulness. Genes (Basel). 2021;12(8).

170. Oldridge DA, Wood AC, Weichert-Leahey N, Crimmins I, Sussman R, Winter C, et al. Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. Nature. 2015;528(7582):418-21.

171. Hashemi M, Sarabandi S, Karami S, Śmieja J, Moazeni-Roodi A, Ghavami S, et al. LMO1 polymorphisms and the risk of neuroblastoma: Assessment of meta-analysis of case-control studies. J Cell Mol Med. 2020;24(2):1160-8.

172. Genecards.org. COBLL1 2022 [Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=COBLL1.

173. Genecards.org. GLS2 2022 [Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=Gls2.

174. G. L. Human Biochemistry. 2nd ed2021.

175. Tessari P, Vettore M, Millioni R, Puricelli L, Orlando R. Effect of liver cirrhosis on phenylalanine and tyrosine metabolism. Curr Opin Clin Nutr Metab Care. 2010;13(1):81-6.

176. Li T, Luo HH, Feng XF, Bai Y, Fang ZZ, Wu GG, et al. Plasma Free Amino Acids and Risk of Cardiovascular Disease in Chinese Patients With Type 2 Diabetes. Front Endocrinol (Lausanne). 2020;11:519923.

177. Chiba T, Suzuki K, Matsumoto T. Plasma-Free Amino Acid Profiles in Crohn's Disease: Relationship With the Crohn Disease Activity Index. Clin Med Insights Gastroenterol. 2018;11:1179552218791173.

178. Sharma R, Biedenharn KR, Fedor JM, Agarwal A. Lifestyle factors and reproductive health: taking control of your fertility. Reprod Biol Endocrinol. 2013;11:66.

179. Cena H, Calder PC. Defining a Healthy Diet: Evidence for The Role of Contemporary Dietary Patterns in Health and Disease. Nutrients. 2020;12(2).

180. Mozaffarian D. Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity: A Comprehensive Review. Circulation. 2016;133(2):187-225.

181. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared Risk Factors in Cardiovascular Disease and Cancer. Circulation. 2016;133(11):1104-14.

182. Merino J, Dashti HS, Li SX, Sarnowski C, Justice AE, Graff M, et al. Genome-wide meta-analysis of macronutrient intake of 91,114 European ancestry participants from the cohorts for heart and aging research in genomic epidemiology consortium. Mol Psychiatry. 2019;24(12):1920-32.

183. Chen Z, Franco OH, Lamballais S, Ikram MA, Schoufour JD, Muka T, et al. Associations of specific dietary protein with longitudinal insulin resistance, prediabetes and type 2 diabetes: The Rotterdam Study. Clin Nutr. 2020;39(1):242-9.

184. de Castro JM. The control of food intake of free-living humans: putting the pieces back together. Physiol Behav. 2010;100(5):446-53.

185. Rankinen T, Bouchard C. Genetics of food intake and eating behavior phenotypes in humans. Annu Rev Nutr. 2006;26:413-34.

186. Teucher B, Skinner J, Skidmore PM, Cassidy A, Fairweather-Tait SJ, Hooper L, et al. Dietary patterns and heritability of food choice in a UK female twin cohort. Twin Res Hum Genet. 2007;10(5):734-48.

187. Klimentidis YC, Raichlen DA, Bea J, Garcia DO, Wineinger NE, Mandarino LJ, et al. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. Int J Obes (Lond). 2018;42(6):1161-76.

188. Fiuza-Luces C, Garatachea N, Berger NA, Lucia A. Exercise is the real polypill. Physiology (Bethesda). 2013;28(5):330-58.

189. Blair SN. Physical inactivity: the biggest public health problem of the 21st century. Br J Sports Med. 2009;43(1):1-2.

190. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW. Correlates of physical activity: why are some people physically active and others not? Lancet. 2012;380(9838):258-71.

191. Trost SG, Owen N, Bauman AE, Sallis JF, Brown W. Correlates of adults' participation in physical activity: review and update. Med Sci Sports Exerc. 2002;34(12):1996-2001.

192. Kim J, Kim J, Min H, Oh S, Kim Y, Lee AH, et al. Joint identification of genetic variants for physical activity in Korean population. Int J Mol Sci. 2014;15(7):12407-21.

193. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.

194. Jezewska-Zychowicz M, Gębski J, Guzek D, Świątkowska M, Stangierska D, Plichta M, et al. The Associations between Dietary Patterns and Sedentary Behaviors in Polish Adults (LifeStyle Study). Nutrients. 2018;10(8).

195. Lu Y, Day FR, Gustafsson S, Buchkovich ML, Na J, Bataille V, et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. Nat Commun. 2016;7:10495.

196. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518(7538):187-96.

197. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197-206.

198. Monda KL, Chen GK, Taylor KC, Palmer C, Edwards TL, Lange LA, et al. A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. Nat Genet. 2013;45(6):690-6.

199. Scherag A, Dina C, Hinney A, Vatin V, Scherag S, Vogel CI, et al. Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. PLoS Genet. 2010;6(4):e1000916.

200. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007;316(5826):889-94.

201. Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J, et al. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS Genet. 2007;3(7):e115.

202. Chambers JC, Elliott P, Zabaneh D, Zhang W, Li Y, Froguel P, et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. Nat Genet. 2008;40(6):716-8.

203. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet. 2009;41(1):25-34.

204. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet. 2009;41(1):18-24.

205. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010;42(11):937-48.

206. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet. 2010;42(11):949-60.

207. Heard-Costa NL, Zillikens MC, Monda KL, Johansson A, Harris TB, Fu M, et al. NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. PLoS Genet. 2009;5(6):e1000539.

208. Heo M, Faith MS, Pietrobelli A, Heymsfield SB. Percentage of body fat cutoffs by sex, age, and race-ethnicity in the US adult population from NHANES 1999-2004. Am J Clin Nutr. 2012;95(3):594-602.

209. Després JP, Lemieux I, Prud'homme D. Treatment of obesity: need to focus on high risk abdominally obese patients. Bmj. 2001;322(7288):716-20.

210. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.

211. Greenwood DC, Hardie LJ, Frost GS, Alwan NA, Bradbury KE, Carter M, et al. Validation of the Oxford WebQ Online 24-Hour Dietary Questionnaire Using Biomarkers. Am J Epidemiol. 2019;188(10):1858-67.

212. Data-Field 90012 - University of Oxford 2022 [Available from: https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=90012.

213. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65.

214. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology. 2017;28(1):30-42.

215. Merz B, Frommherz L, Rist MJ, Kulling SE, Bub A, Watzl B. Dietary Pattern and Plasma BCAA-Variations in Healthy Men and Women-Results from the KarMeN Study. Nutrients. 2018;10(5).

216. Rietman A, Schwarz J, Tomé D, Kok FJ, Mensink M. High dietary protein intake, reducing or eliciting insulin resistance? Eur J Clin Nutr. 2014;68(9):973-9.

217. Cortiella J, Matthews DE, Hoerr RA, Bier DM, Young VR. Leucine kinetics at graded intakes in young men: quantitative fate of dietary leucine. Am J Clin Nutr. 1988;48(4):998-1009.

218. Biolo G, Tessari P, Inchiostro S, Bruttomesso D, Fongher C, Sabadin L, et al. Leucine and phenylalanine kinetics during mixed meal ingestion: a multiple tracer approach. Am J Physiol. 1992;262(4 Pt 1):E455-63.

219. Meguid MM, Matthews DE, Bier DM, Meredith CN, Young VR. Valine kinetics at graded valine intakes in young men. Am J Clin Nutr. 1986;43(5):781-6.

220. Shah SH, Crosslin DR, Haynes CS, Nelson S, Turer CB, Stevens RD, et al. Branched-chain amino acid levels are associated with improvement in insulin resistance with weight loss. Diabetologia. 2012;55(2):321-30.

221. Schmidt JA, Rinaldi S, Scalbert A, Ferrari P, Achaintre D, Gunter MJ, et al. Plasma concentrations and intakes of amino acids in male meat-eaters, fish-eaters, vegetarians and vegans: a cross-sectional analysis in the EPIC-Oxford cohort. Eur J Clin Nutr. 2016;70(3):306-12.

222. Veldhorst MA, Westerterp-Plantenga MS, Westerterp KR. Gluconeogenesis and energy expenditure after a high-protein, carbohydrate-free diet. Am J Clin Nutr. 2009;90(3):519-26.

223. Johnstone AM, Horgan GW, Murison SD, Bremner DM, Lobley GE. Effects of a high-protein ketogenic diet on hunger, appetite, and weight loss in obese men feeding ad libitum. Am J Clin Nutr. 2008;87(1):44-55.

224. Goran MI. ENERGY METABOLISM AND OBESITY. Medical Clinics of North America. 2000;84(2):347-62.

225. Rousseau M, Guénard F, Garneau V, Allam-Ndoul B, Lemieux S, Pérusse L, et al. Associations Between Dietary Protein Sources, Plasma BCAA and Short-Chain Acylcarnitine Levels in Adults. Nutrients. 2019;11(1).

226. López AM, Noriega LG, Diaz M, Torres N, Tovar AR. Plasma branched-chain and aromatic amino acid concentration after ingestion of an urban or rural diet in rural Mexican women. BMC Obes. 2015;2:8.

227. Vanweert F, Boone SC, Brouwers B, Mook-Kanamori DO, de Mutsert R, Rosendaal FR, et al. The effect of physical activity level and exercise training on the association between plasma branched-chain amino acids and intrahepatic lipid content in participants with obesity. Int J Obes (Lond). 2021;45(7):1510-20.

228. Laviano A, Molfino A, Lacaria MT, Canelli A, De Leo S, Preziosa I, et al. Glutamine supplementation favors weight loss in nondieting obese female patients. A pilot study. Eur J Clin Nutr. 2014;68(11):1264-6.

229. Opara EC, Petro A, Tevrizian A, Feinglos MN, Surwit RS. L-Glutamine Supplementation of a High Fat Diet Reduces Body Weight and Attenuates Hyperglycemia and Hyperinsulinemia in C57BL/6J Mice. The Journal of Nutrition. 1996;126(1):273-9.

230. Abboud KY, Reis SK, Martelli ME, Zordão OP, Tannihão F, de Souza AZZ, et al. Oral Glutamine Supplementation Reduces Obesity, Pro-Inflammatory Markers, and Improves Insulin Sensitivity in DIO Wistar Rats and Reduces Waist Circumference in Overweight and Obese Humans. Nutrients. 2019;11(3).

231. Caldow MK, Ham DJ, Godeassi DP, Chee A, Lynch GS, Koopman R. Glycine supplementation during calorie restriction accelerates fat loss and protects against further muscle loss in obese mice. Clin Nutr. 2016;35(5):1118-26.

232. Panjwani AA, Ji Y, Fahey JW, Palmer A, Wang G, Hong X, et al. Maternal Dyslipidemia, Plasma Branched-Chain Amino Acids, and the Risk of Child Autism Spectrum Disorder: Evidence of Sex Difference. J Autism Dev Disord. 2020;50(2):540-50.

233. Okekunle AP, Zhang M, Wang Z, Onwuka JU, Wu X, Feng R, et al. Dietary branched-chain amino acids intake exhibited a different relationship with type 2 diabetes and obesity risk: a meta-analysis. Acta Diabetol. 2019;56(2):187-95.

234. Lu J, Xie G, Jia W, Jia W. Insulin resistance and the metabolism of branched-chain amino acids. Front Med. 2013;7(1):53-9.

235. Stoppani J, Scheett T, Pena J, Rudolph C, Charlebois D. 2009 international society of sports nutrition conference and expo new orleans, la, USA. 14-15 june 2009. Abstracts. J Int Soc Sports Nutr. 2009;6 Suppl 1(Suppl 1):P1-p19.

236. Batch BC, Shah SH, Newgard CB, Turer CB, Haynes C, Bain JR, et al. Branched chain amino acids are novel biomarkers for discrimination of metabolic wellness. Metabolism. 2013;62(7):961-9.

237. Lackey DE, Lynch CJ, Olson KC, Mostaedi R, Ali M, Smith WH, et al. Regulation of adipose branched-chain amino acid catabolism enzyme expression and cross-adipose amino acid flux in human obesity. Am J Physiol Endocrinol Metab. 2013;304(11):E1175-87.

238. Wiklund PK, Pekkala S, Autio R, Munukka E, Xu L, Saltevo J, et al. Serum metabolic profiles in overweight and obese women with and without metabolic syndrome. Diabetol Metab Syndr. 2014;6(1):40.

239. Lischka J, Schanzer A, Hojreh A, Ba Ssalamah A, Item CB, de Gier C, et al. A branched-chain amino acid-based metabolic score can predict liver fat in children and adolescents with severe obesity. Pediatr Obes. 2021;16(4):e12739.

240. Gaggini M, Carli F, Rosso C, Buzzigoli E, Marietti M, Della Latta V, et al. Altered amino acid concentrations in NAFLD: Impact of obesity and insulin resistance. Hepatology. 2018;67(1):145-58.

241. Haufe S, Witt H, Engeli S, Kaminski J, Utz W, Fuhrmann JC, et al. Branched-chain and aromatic amino acids, insulin resistance and liver specific ectopic fat storage in overweight to obese subjects. Nutr Metab Cardiovasc Dis. 2016;26(7):637-42.

242. Kaikkonen JE, Würtz P, Suomela E, Lehtovirta M, Kangas AJ, Jula A, et al. Metabolic profiling of fatty liver in young and middle-aged adults: Cross-sectional and prospective analyses of the Young Finns Study. Hepatology. 2017;65(2):491-500.

243. Cheng S, Wiklund P, Autio R, Borra R, Ojanen X, Xu L, et al. Adipose Tissue Dysfunction and Altered Systemic Amino Acid Metabolism Are Associated with Non-Alcoholic Fatty Liver Disease. PLoS One. 2015;10(10):e0138889.

244. Goffredo M, Santoro N, Tricò D, Giannini C, D'Adamo E, Zhao H, et al. A Branched-Chain Amino Acid-Related Metabolic Signature Characterizes Obese Adolescents with Non-Alcoholic Fatty Liver Disease. Nutrients. 2017;9(7).

245. Koch M, Freitag-Wolf S, Schlesinger S, Borggrefe J, Hov JR, Jensen MK, et al. Serum metabolomic profiling highlights pathways associated with liver fat content in a general population sample. Eur J Clin Nutr. 2017;71(8):995-1001.

246. Breitman I, Saraf N, Kakade M, Yellumahanthi K, White M, Hackett JA, et al. The effects of an amino acid supplement on glucose homeostasis, inflammatory markers, and incretins after laparoscopic gastric bypass. J Am Coll Surg. 2011;212(4):617-25; discussion 25-7.

247. Mu WC, VanHoosier E, Elks CM, Grant RW. Long-Term Effects of Dietary Protein and Branched-Chain Amino Acids on Metabolism and Inflammation in Mice. Nutrients. 2018;10(7).

248. Son DO, Satsu H, Shimizu M. Histidine inhibits oxidative stress- and TNF-alpha-induced interleukin-8 secretion in intestinal epithelial cells. FEBS Lett. 2005;579(21):4671-7.

249. Li C, Chen B, Fang Z, Leng YF, Wang DW, Chen FQ, et al. Metabolomics in the development and progression of rheumatoid arthritis: A systematic review. Joint Bone Spine. 2020;87(5):425-30.

250. El Hafidi M, Pérez I, Baños G. Is glycine effective against elevated blood pressure? Curr Opin Clin Nutr Metab Care. 2006;9(1):26-31.

251. De Luca G, Calpona PR, Caponetti A, Macaione V, Di Benedetto A, Cucinotta D, et al. Preliminary report: Amino acid profile in platelets of diabetic patients. Metabolism. 2001;50(7):739-41.

252. Oberbach A, Blüher M, Wirth H, Till H, Kovacs P, Kullnick Y, et al. Combined proteomic and metabolomic profiling of serum reveals association of the complement system with obesity and identifies novel markers of body fat mass changes. J Proteome Res. 2011;10(10):4769-88.

253. Li X, Sun L, Zhang W, Li H, Wang S, Mu H, et al. Association of serum glycine levels with metabolic syndrome in an elderly Chinese population. Nutr Metab (Lond). 2018;15:89.

254. Badoud F, Lam KP, DiBattista A, Perreault M, Zulyniak MA, Cattrysse B, et al. Serum and adipose tissue amino acid homeostasis in the metabolically healthy obese. J Proteome Res. 2014;13(7):3455-66.

255. Yamakado M, Nagao K, Imaizumi A, Tani M, Toda A, Tanaka T, et al. Plasma Free Amino Acid Profiles Predict Four-Year Risk of Developing Diabetes, Metabolic Syndrome, Dyslipidemia, and Hypertension in Japanese Population. Sci Rep. 2015;5:11918.

256. Adeva-Andany M, Souto-Adeva G, Ameneiros-Rodríguez E, Fernández-Fernández C, Donapetry-García C, Domínguez-Montero A. Insulin resistance and glycine metabolism in humans. Amino Acids. 2018;50(1):11-27.

257. Takashina C, Tsujino I, Watanabe T, Sakaue S, Ikeda D, Yamada A, et al. Associations among the plasma amino acid profile, obesity, and glucose metabolism in Japanese adults with normal glucose tolerance. Nutr Metab (Lond). 2016;13:5.

258. Drábková P, Šanderová J, Kovařík J, kanďár R. An Assay of Selected Serum Amino Acids in Patients with Type 2 Diabetes Mellitus. Adv Clin Exp Med. 2015;24(3):447-51.

259. Avenell A. Glutamine in critical care: current evidence from systematic reviews. Proc Nutr Soc. 2006;65(3):236-41.

260. Oliveira GP, de Abreu MG, Pelosi P, Rocco PR. Exogenous Glutamine in Respiratory Diseases: Myth or Reality? Nutrients. 2016;8(2):76.

261. Pouw EM, Schols AM, Deutz NE, Wouters EF. Plasma and muscle amino acid levels in relation to resting energy expenditure and inflammation in stable chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1998;158(3):797-801.

262. Khogali SE, Pringle SD, Weryk BV, Rennie MJ. Is glutamine beneficial in ischemic heart disease? Nutrition. 2002;18(2):123-6.

263. Poisner AM. Serum phenylalanine in schizophrenia: biochemical genetic aspects. J Nerv Ment Dis. 1960;131:74-6.

264. Rao ML, Gross G, Strebel B, Bräunig P, Huber G, Klosterkötter J. Serum amino acids, central monoamines, and hormones in drug-naive, drug-free, and neuroleptic-treated schizophrenic patients and healthy subjects. Psychiatry Res. 1990;34(3):243-57.

265. Okusaga O, Muravitskaja O, Fuchs D, Ashraf A, Hinman S, Giegling I, et al. Elevated levels of plasma phenylalanine in schizophrenia: a guanosine triphosphate cyclohydrolase-1 metabolic pathway abnormality? PLoS One. 2014;9(1):e85945.

266. Bröer S. Amino Acid Transporters as Disease Modifiers and Drug Targets. SLAS Discov. 2018;23(4):303-20.

267. Soory M. Nutritional antioxidants and their applications in cardiometabolic diseases. Infect Disord Drug Targets. 2012;12(5):388-401.

268. Nagao K, Kimura T. Use of plasma-free amino acids as biomarkers for detecting and predicting disease risk. Nutrition Reviews. 2020;78(Supplement_3):79-85.

269. Miyagi Y, Higashiyama M, Gochi A, Akaike M, Ishikawa T, Miura T, et al. Plasma free amino acid profiling of five types of cancer patients and its application for early detection. PLoS One. 2011;6(9):e24143.

270. Ravaglia G, Forti P, Maioli F, Bianchi G, Martelli M, Talerico T, et al. Plasma amino acid concentrations in patients with amnestic mild cognitive impairment or Alzheimer disease. Am J Clin Nutr. 2004;80(2):483-8.

271. Yamakado M, Tanaka T, Nagao K, Imaizumi A, Komatsu M, Daimon T, et al. Plasma amino acid profile associated with fatty liver disease and co-occurrence of metabolic risk factors. Sci Rep. 2017;7(1):14485.

272. Lustgarten MS, Price LL, Chale A, Phillips EM, Fielding RA. Branched chain amino acids are associated with muscle mass in functionally limited older adults. J Gerontol A Biol Sci Med Sci. 2014;69(6):717-24.

273. Ter Borg S, Luiking YC, van Helvoort A, Boirie Y, Schols J, de Groot C. Low Levels of Branched Chain Amino Acids, Eicosapentaenoic Acid and Micronutrients Are Associated with Low Muscle Mass, Strength and Function in Community-Dwelling Older Adults. The journal of nutrition, health & aging. 2019;23(1):27-34.

274. Holm E, Sedlaczek O, Grips E. Amino acid metabolism in liver disease. Curr Opin Clin Nutr Metab Care. 1999;2(1):47-53.

275. Laidlaw SA, Berg RL, Kopple JD, Naito H, Walker WG, Walser M. Patterns of fasting plasma amino acid levels in chronic renal insufficiency: results from the feasibility phase of the Modification of Diet in Renal Disease Study. Am J Kidney Dis. 1994;23(4):504-13.

276. UK-Biobank. Nightingale Health Metabolic Biomarkers: Phase 1 Release. 2021.

277. Panagiotou OA, Ioannidis JP. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. Int J Epidemiol. 2012;41(1):273-86.

278. Liu YS, Wu PE, Chou WC, Vikram R, Chen WT, Yang SL, et al. Body mass index and type 2 diabetes and breast cancer survival: a Mendelian randomization study. Am J Cancer Res. 2021;11(8):3921-34.

279. Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun. 2018;9(1):2941.

280. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44(9):981-90.

281. Banda Y, Kvale MN, Hoffmann TJ, Hesselson SE, Ranatunga D, Tang H, et al. Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. Genetics. 2015;200(4):1285-95.

282. Clare Bycroft CF, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, Adrian Cortes, Samantha Welsh, Gil McVean, Stephen Leslie, Peter Donnelly. Genome-wide genetic data on ~500,000 UK Biobank participants: reprint at bioRxiv; 2017 [Available from: https://www.biorxiv.org/content/10.1101/166298v1.

283. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45(11):1274-83.

284. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet. 2018;50(10):1412-25.

285. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121-30.

286. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat Genet. 2020;52(6):572-81.

287. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018;50(7):928-36.

288. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019;51(3):414-30.

289. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations. Wellcome Open Res. 2019;4:186.

290. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. Bmj. 1995;310(6973):170.

291. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512-25.

292. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017;32(5):377-89.

293. Gormley M, Dudding T, Sanderson E, Martin RM, Thomas S, Tyrrell J, et al. A multivariable Mendelian randomization analysis investigating smoking and alcohol consumption in oral and oropharyngeal cancer. Nat Commun. 2020;11(1):6071.

294. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181(4):251-60.

295. Sanderson E. Multivariable Mendelian Randomization and Mediation. Cold Spring Harb Perspect Med. 2021;11(2).

296. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. Int J Epidemiol. 2019;48(3):713-27.

297. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. Stat Med. 2017;36(29):4705-18.

298. Saleem T, Dahpy M, Ezzat G, Abdelrahman G, Abdel-Aziz E, Farghaly R. The Profile of Plasma Free Amino Acids in Type 2 Diabetes Mellitus with Insulin Resistance: Association with Microalbuminuria and Macroalbuminuria. Appl Biochem Biotechnol. 2019;188(3):854-67.

299. Guasch-Ferré M, Hruby A, Toledo E, Clish CB, Martínez-González MA, Salas-Salvadó J, et al. Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. Diabetes Care. 2016;39(5):833-46.

300. Chen T, Ni Y, Ma X, Bao Y, Liu J, Huang F, et al. Branched-chain and aromatic amino acid profiles and diabetes risk in Chinese populations. Sci Rep. 2016;6:20594.

301. van Doorn M, Vogels J, Tas A, van Hoogdalem EJ, Burggraaf J, Cohen A, et al. Evaluation of metabolite profiles as biomarkers for the pharmacological effects of thiazolidinediones in Type 2 diabetes mellitus patients and healthy volunteers. Br J Clin Pharmacol. 2007;63(5):562-74.

302. Suhre K, Meisinger C, Döring A, Altmaier E, Belcredi P, Gieger C, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PLoS One. 2010;5(11):e13953.

303. Würtz P, Tiainen M, Mäkinen VP, Kangas AJ, Soininen P, Saltevo J, et al. Circulating metabolite predictors of glycemia in middle-aged men and women. Diabetes Care. 2012;35(8):1749-56.

304. Würtz P, Soininen P, Kangas AJ, Rönnemaa T, Lehtimäki T, Kähönen M, et al. Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. Diabetes Care. 2013;36(3):648-55.

305. Le Couteur DG, Ribeiro R, Senior A, Hsu B, Hirani V, Blyth FM, et al. Branched Chain Amino Acids, Cardiometabolic Risk Factors and Outcomes in Older Men: The Concord Health and Ageing in Men Project. J Gerontol A Biol Sci Med Sci. 2020;75(10):1805-10.

306. Wang Y, Huang K, Liu F, Lu X, Huang J, Gu D. Association of circulating branched-chain amino acids with risk of cardiovascular disease: A systematic review and meta-analysis. Atherosclerosis. 2022;350:90-6.

307. Bi X, Henry CJ. Plasma-free amino acid profiles are predictors of cancer and diabetes development. Nutrition & Diabetes. 2017;7(3):e249-e.

308. Costeas PA, Chinsky JM. Effects of insulin on the regulation of branched-chain alpha-keto acid dehydrogenase E1 alpha subunit gene expression. Biochem J. 1996;318 (Pt 1)(Pt 1):85-92.

309. Tai ES, Tan ML, Stevens RD, Low YL, Muehlbauer MJ, Goh DL, et al. Insulin resistance is associated with a metabolic profile of altered protein metabolism in Chinese and Asian-Indian men. Diabetologia. 2010;53(4):757-67.

310. Buse MG, Biggers JF, Friderici KH, Buse JF. Oxidation of branched chain amino acids by isolated hearts and diaphragms of the rat. The effect of fatty acids, glucose, and pyruvate respiration. J Biol Chem. 1972;247(24):8085-96.

311. Lei MZ, Li XX, Zhang Y, Li JT, Zhang F, Wang YP, et al. Acetylation promotes BCAT2 degradation to suppress BCAA catabolism and pancreatic cancer growth. Signal Transduct Target Ther. 2020;5(1):70.

312. Sugiyama K, Yu L, Nagasue N. Direct effect of branched-chain amino acids on the growth and metabolism of cultured human hepatocellular carcinoma cells. Nutr Cancer. 1998;31(1):62-8.

313. Rana JS, Visser ME, Arsenault BJ, Després JP, Stroes ES, Kastelein JJ, et al. Metabolic dyslipidemia and risk of future coronary heart disease in apparently healthy men and women: the EPIC-Norfolk prospective population study. Int J Cardiol. 2010;143(3):399-404.

314. Fukushima K, Harada S, Takeuchi A, Kurihara A, Iida M, Fukai K, et al. Association between dyslipidemia and plasma levels of branched-chain amino acids in the Japanese population without diabetes mellitus. J Clin Lipidol. 2019;13(6):932-9.e2.

315. Shimobayashi M, Hall MN. Multiple amino acid sensing inputs to mTORC1. Cell Res. 2016;26(1):7-20.

316. Han J, Wang Y. mTORC1 signaling in hepatic lipid metabolism. Protein Cell. 2018;9(2):145-51.

317. Laplante M, Sabatini DM. mTOR signaling at a glance. J Cell Sci. 2009;122(Pt 20):3589-94.

318. Tuttle KR, Milton JE, Packard DP, Shuler LA, Short RA. Dietary amino acids and blood pressure: a cohort study of patients with cardiovascular disease. Am J Kidney Dis. 2012;59(6):803-9.

319. Poggiogalle E, Fontana M, Giusti AM, Pinto A, Iannucci G, Lenzi A, et al. Amino Acids and Hypertension in Adults. Nutrients. 2019;11(7).

320. Yang RY, Wang SM, Sun L, Liu JM, Li HX, Sui XF, et al. Association of branched-chain amino acids with coronary artery disease: A matched-pair case-control study. Nutr Metab Cardiovasc Dis. 2015;25(10):937-42.

321. Jauhiainen R, Vangipurapu J, Laakso A, Kuulasmaa T, Kuusisto J, Laakso M. The Association of 9 Amino Acids With Cardiovascular Events in Finnish Men in a 12-Year Follow-up Study. J Clin Endocrinol Metab. 2021;106(12):3448-54.

322. Zeleznik OA, Balasubramanian R, Ren Y, Tobias DK, Rosner BA, Peng C, et al. Branched-Chain Amino Acids and Risk of Breast Cancer. JNCI Cancer Spectr. 2021;5(5).

323. Tobias DK, Chai B, Tamimi RM, Manson JE, Hu FB, Willett WC, et al. Dietary Intake of Branched Chain Amino Acids and Breast Cancer Risk in the NHS and NHS II Prospective Cohorts. JNCI Cancer Spectr. 2021;5(3).

324. Nouri-Majd S, Salari-Moghaddam A, Benisi-Kohansal S, Azadbakht L, Esmaillzadeh A. Dietary intake of branched-chain amino acids in relation to the risk of breast cancer. Breast Cancer. 2022.

325. Kang JS. Dietary restriction of amino acids for Cancer therapy. Nutr Metab (Lond). 2020;17:20.

326. Jain M, Nilsson R, Sharma S, Madhusudhan N, Kitami T, Souza AL, et al. Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. Science. 2012;336(6084):1040-4.

327. Xiao F, Wang C, Yin H, Yu J, Chen S, Fang J, et al. Leucine deprivation inhibits proliferation and induces apoptosis of human breast cancer cells via fatty acid synthase. Oncotarget. 2016;7(39):63679-89.

328. Chen L, Cui H. Targeting Glutamine Induces Apoptosis: A Cancer Therapy Approach. Int J Mol Sci. 2015;16(9):22830-55.

329. Leone RD, Zhao L, Englert JM, Sun IM, Oh MH, Sun IH, et al. Glutamine blockade induces divergent metabolic programs to overcome tumor immune evasion. Science. 2019;366(6468):1013-21.

330. Griffin JWD, Bradshaw PC. Amino Acid Catabolism in Alzheimer's Disease Brain: Friend or Foe? Oxidative Medicine and Cellular Longevity. 2017;2017:5472792.

331. Solon-Biet SM, Cogger VC, Pulpitel T, Wahl D, Clark X, Bagley E, et al. Branched chain amino acids impact health and lifespan indirectly via amino acid balance and appetite control. Nat Metab. 2019;1(5):532-45.

332. Scaini G, Mello-Santos LM, Furlanetto CB, Jeremias IC, Mina F, Schuck PF, et al. Acute and chronic administration of the branched-chain amino acids decreases nerve growth factor in rat hippocampus. Mol Neurobiol. 2013;48(3):581-9.

333. González-Domínguez R, García-Barrera T, Gómez-Ariza JL. Metabolite profiling for the identification of altered metabolic pathways in Alzheimer's disease. J Pharm Biomed Anal. 2015;107:75-81.

334. van de Rest O, van der Zwaluw NL, de Groot LC. Literature review on the role of dietary protein and amino acids in cognitive functioning and cognitive decline. Amino Acids. 2013;45(5):1035-45.

335. Xu J, Begley P, Church SJ, Patassini S, Hollywood KA, Jüllig M, et al. Graded perturbations of metabolism in multiple regions of human brain in Alzheimer's disease: Snapshot of a pervasive metabolic disorder. Biochim Biophys Acta. 2016;1862(6):1084-92.

336. Nilsen LH, Witter MP, Sonnewald U. Neuronal and astrocytic metabolism in a transgenic rat model of Alzheimer's disease. J Cereb Blood Flow Metab. 2014;34(5):906-14.

337. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol. 2017;186(9):1026-34.

338. Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. Nat Genet. 2018;50(7):906-8.

339. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015;47(3):284-90.

340. Watanabe K TE vBA, Posthuma D. . Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8(1):1826.

341. Yuelin Li JB. Linear Mixed-Effects Models in Analyzing Repeated-Measures Data. Behavioral Research Data Analysis with R: Springer; 2012.

342. Davey Smith G HG. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23.

343. Denny JC BL, Roden DM. . Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. Annu Rev Genomics Hum Genet. 2016;17:353-73.

344. Lin R, Liu W, Piao M, Zhu H. A review of the relationship between the gut microbiota and amino acid metabolism. Amino Acids. 2017;49(12):2083-90.

345. Wu L, Tang Z, Chen H, Ren Z, Ding Q, Liang K, et al. Mutual interaction between gut microbiota and protein/amino acid metabolism for host mucosal immunity and health. Anim Nutr. 2021;7(1):11-6.

346. Dhakan DB, Maji A, Sharma AK, Saxena R, Pulikkan J, Grace T, et al. The unique composition of Indian gut microbiome, gene catalogue, and associated fecal metabolome deciphered using multi-omics approaches. GigaScience. 2019;8(3).

347. Hayashi T, Yamashita T, Takahashi T, Tabata T, Watanabe H, Gotoh Y, et al. Uncovering the Role of Gut Microbiota in Amino Acid Metabolic Disturbances in Heart Failure Through Metagenomic Analysis. Front Cardiovasc Med. 2021;8:789325.

348. Association AH. Diabetes 2022 [Available from: https://www.heart.org/en/health-topics/diabetes.

349. Federation ID. Diabetes and cardiovascular disease 2021 [Available from: https://idf.org/our-activities/care-prevention/cardiovascular-disease.html.

Appendix

Table S1. Identified variants associated with serum alanine in the current UKB GWAS

SNP	p-value	Chr	BP	EA	NEA	EAF	Beta	SE
rs2933243	3.30E-56	12	56860577	A	G	0.183189	0.083539	0.005289
rs1260326	1.90E-40	2	27730940	C	T	0.603981	-0.05574	0.004186
rs8061221	4.00E-35	16	70078960	A	G	0.745547	0.058203	0.004707
rs79687284	4.70E-33	1	2.14E+08	C	G	0.034641	0.134338	0.011217
rs11183581	1.30E-31	12	47122774	A	C	0.212669	0.058477	0.005
rs2168101	9.00E-31	11	8255408	A	C	0.308213	-0.05278	0.004576
rs2160387	5.90E-28	2	65220910	C	T	0.427786	-0.04515	0.004119
rs4554975	3.30E-27	12	47201814	G	A	0.562773	-0.04502	0.004167
rs2808454	1.70E-25	1	2.07E+08	T	A	0.543121	-0.043	0.00412
rs7624902	4.10E-21	3	1.6E+08	G	A	0.537534	-0.03877	0.004111
rs4985532	2.10E-20	16	70575214	G	A	0.070422	-0.0746	0.008057
rs6931514	1.40E-15	6	20703952	G	A	0.262271	0.03726	0.004663
rs4693210	1.80E-15	4	89208565	G	A	0.440215	-0.03287	0.004134
rs74093304	2.30E-15	12	59923805	A	G	0.11906	0.050278	0.006344
rs13277542	4.90E-15	8	1.46E+08	G	T	0.488895	-0.03258	0.004162
rs56335308	3.70E-14	8	17419461	A	G	0.027352	0.094964	0.012541
rs7503139	1.30E-13	17	79641534	T	C	0.167225	-0.04069	0.005493
rs7925445	5.60E-13	11	18398958	G	A	0.556188	0.029806	0.004134
rs138373837	1.30E-12	5	36219710	T	C	0.023615	-0.09572	0.013497
rs11558471	3.10E-12	8	1.18E+08	G	A	0.314671	-0.03076	0.004413
rs4237150	9.10E-12	9	4290085	C	G	0.403366	0.028459	0.004173
rs2385127	1.80E-11	8	49975746	A	G	0.616547	-0.02831	0.004211
rs7978353	2.00E-11	12	1.23E+08	G	A	0.404674	0.027906	0.004162
rs1047891	1.60E-10	2	2.12E+08	A	C	0.315002	-0.02821	0.004411
rs11715633	3.30E-10	3	1.23E+08	A	G	0.270529	-0.02898	0.004611
rs807672	3.90E-10	22	19157819	G	C	0.531442	0.02565	0.004098

rs6752053	5.70E-10	2	65666674	C	T	0.408902	-0.02582	0.004166
rs4766214	8.30E-10	12	4290249	G	A	0.456752	-0.02543	0.004142
rs12379111	2.50E-09	9	22128180	G	C	0.094813	-0.04224	0.007089
rs5023078	2.80E-09	12	1.33E+08	C	T	0.531306	-0.02441	0.004108
rs339969	7.30E-09	15	60883281	A	C	0.61583	0.024364	0.004212
rs73201506	8.60E-09	21	33044408	G	A	0.046008	0.056456	0.009808
rs144262262	2.10E-08	7	95222201	C	T	0.003511	-0.19421	0.03464
rs4903264	2.60E-08	14	75343829	T	A	0.559472	0.02346	0.004215
rs58116791	4.20E-08	14	56680822	C	T	0.2346	0.026587	0.00485
rs6484738	5.60E-08	11	34860858	A	G	0.3984	-0.02274	0.004187
rs4846922	6.50E-08	1	2.3E+08	C	T	0.654684	0.023412	0.004332
rs112846708	7.20E-08	14	86618600	A	G	0.769585	0.032712	0.006074
rs1880000	8.60E-08	2	1.03E+08	C	T	0.143438	0.031553	0.005893
rs561931	8.70E-08	1	1.2E+08	G	A	0.581003	-0.02225	0.004158
rs12673978	9.80E-08	7	44717523	T	C	0.069052	0.043051	0.008077
rs7959621	1.00E-07	12	3353636	T	C	0.343187	-0.02325	0.004364
rs35635959	2.00E-07	17	40772288	C	T	0.284239	0.023597	0.004539
rs6499167	2.10E-07	16	68337798	A	T	0.725901	-0.02383	0.004592
rs66464442	2.30E-07	1	1.18E+08	A	C	0.315866	0.022776	0.004404
rs76895963	2.80E-07	12	4384844	G	T	0.020556	-0.0818	0.015921
rs2957933	2.80E-07	17	46335579	A	T	0.761336	0.024715	0.004811
rs3785477	3.20E-07	17	55164142	T	C	0.401264	-0.02134	0.004178
rs58470815	3.30E-07	19	36132922	C	T	0.344476	0.022142	0.004337
rs12964812	4.00E-07	18	31414323	T	A	0.399965	0.02174	0.00429
rs2704797	4.20E-07	3	71408090	C	T	0.31358	0.022541	0.004456
rs8107974	4.20E-07	19	19388500	T	A	0.075553	-0.03928	0.007764
rs61676179	4.50E-07	22	32176825	A	C	0.051004	-0.04717	0.009348
rs4789698	4.60E-07	17	80179786	T	C	0.43583	-0.02087	0.004141
rs727331	5.20E-07	6	1.27E+08	C	T	0.29093	-0.02262	0.004508
rs113223489	7.60E-07	15	64262688	T	C	0.025814	-0.0638	0.012902
rs113830923	7.70E-07	12	21397321	G	A	0.109612	0.032882	0.006653
rs12692735	8.10E-07	2	1.66E+08	T	G	0.360742	-0.02102	0.004261

rs73158708	8.10E-07	7	1.37E+08	C	A	0.316984	0.0217	0.004399
rs2495729	8.20E-07	10	1.02E+08	C	T	0.156099	-0.0279	0.005658
rs145460781	9.20E-07	4	1.42E+08	T	G	0.068364	-0.04004	0.008156
rs10736885	1.20E-06	10	1.29E+08	T	C	0.952674	-0.0469	0.009643
rs9883916	1.30E-06	3	1.36E+08	T	C	0.775001	0.02374	0.004905
rs28690260	1.30E-06	8	17365162	T	A	0.403672	0.022535	0.004655
rs7732481	1.40E-06	5	1.2E+08	T	C	0.169361	-0.02654	0.005503
rs9414767	1.40E-06	10	63893687	C	T	0.142591	-0.02838	0.005886
rs4647736	1.40E-06	11	47253137	T	C	0.101582	-0.03291	0.006824
rs12725301	1.60E-06	1	2.2E+08	T	C	0.211946	0.024064	0.005018
rs4470984	1.70E-06	7	56033558	G	A	0.759116	-0.02298	0.004797
rs1076572	1.80E-06	16	3230588	A	G	0.240791	0.023395	0.004901
rs116256520	1.90E-06	3	1.23E+08	T	C	0.014256	-0.08227	0.01727
rs62623385	1.90E-06	17	7847837	T	A	0.034143	0.053579	0.011239
rs304829	2.00E-06	3	30809058	A	G	0.398096	0.020013	0.004214
rs11769922	2.00E-06	7	1.51E+08	T	C	0.231617	-0.02325	0.004889
rs4399146	2.60E-06	1	1.57E+08	A	G	0.335089	0.020366	0.004334
rs7973823	2.60E-06	12	1.21E+08	C	T	0.506326	0.019768	0.004209
rs62042562	2.80E-06	16	58404923	C	T	0.095114	-0.03284	0.007004
rs34768796	3.20E-06	1	3625807	T	C	0.092313	-0.03342	0.007173
rs11187815	3.20E-06	10	95971321	A	G	0.544541	-0.01941	0.004164
rs149486185	3.20E-06	20	42006230	T	C	0.062716	-0.03983	0.008548
rs6888884	3.60E-06	5	1.37E+08	A	G	0.015799	-0.07621	0.016454
rs25789	3.70E-06	5	55800724	T	C	0.743867	0.021722	0.004693
rs3135041	3.70E-06	6	31555657	G	A	0.340588	0.020002	0.004323
rs13108587	3.80E-06	4	39588892	A	G	0.336379	0.020097	0.004347
rs34308099	4.00E-06	3	1.95E+08	A	G	0.400995	-0.01936	0.004198
rs3770687	4.10E-06	2	2.11E+08	G	A	0.487232	-0.01891	0.004108
rs62320528	5.00E-06	4	1.4E+08	T	C	0.40627	-0.01923	0.004212
rs75226856	5.10E-06	2	2.35E+08	A	C	0.019535	0.067633	0.014831
rs6875539	5.30E-06	5	1.7E+08	G	T	0.925627	-0.03583	0.007866
rs146184369	5.30E-06	15	97658188	A	G	0.021684	-0.07112	0.015617

rs9394215	6.20E-06	6	34399872	A	G	0.811983	-0.02378	0.005262
rs196377	6.30E-06	3	38421098	G	C	0.697122	0.020239	0.004481
rs111670693	6.60E-06	10	1369372	G	A	0.199701	0.025717	0.005707
rs11187149	6.80E-06	10	94485763	A	G	0.379299	-0.01897	0.004215
rs4819296	6.90E-06	21	45133868	T	C	0.631989	-0.01945	0.004325
rs1176790	7.10E-06	2	77439602	T	G	0.024665	-0.06997	0.015585
rs2066577	7.50E-06	13	50665444	T	C	0.014065	0.077815	0.017371
rs62576130	7.60E-06	9	1.19E+08	G	T	0.129837	0.028076	0.00627
rs148647914	7.80E-06	13	26738138	G	A	0.030074	0.056111	0.012552
rs34179517	8.60E-06	6	39281456	C	A	0.121682	0.028048	0.006305
rs1202384	9.80E-06	11	91972287	G	T	0.182143	0.023464	0.005308
rs28544889	1.00E-05	7	1.3E+08	A	C	0.324702	0.019316	0.00438
rs2663854	1.00E-05	18	55458827	T	C	0.302238	0.019693	0.004464

Chr: Chromosome, BP: Position, EA: Effect Allele, NEA: Non-Effect Allele, EAF: Effect Allele Frequency, SE: Standard Error

Table S2. Identified variants associated with serum glycine in the current UKB GWAS

SNP	p-value	Chr	BP	EA	NEA	EAF	Beta	SE
rs12991443	1.60E-315	2	2.12E+08	G	C	0.309118	-0.15738	0.004198
rs4461202	1.10E-246	2	2.12E+08	T	C	0.471863	0.129962	0.003888
rs2026972	2.20E-87	9	6538279	C	G	0.308331	-0.08216	0.004232
rs4240624	2.00E-82	8	9184231	A	G	0.90923	-0.12686	0.006793
rs10934753	3.40E-74	3	1.26E+08	A	G	0.417116	0.069003	0.003962
rs9922910	1.10E-64	16	81126662	T	G	0.915043	0.116686	0.007039
rs4948102	3.30E-61	7	56097265	G	C	0.751228	0.071556	0.004525
rs28601761	1.40E-55	8	1.27E+08	G	C	0.420161	0.060914	0.003998
rs74561492	6.40E-50	2	2.11E+08	T	C	0.043845	0.138455	0.009478
rs112247225	5.20E-45	16	81154900	T	C	0.046621	-0.12823	0.009285
rs550674059	1.80E-42	2	2.11E+08	A	C	0.016772	0.213325	0.015573
rs62178226	1.30E-35	2	2.12E+08	A	T	0.366194	0.050927	0.004047
rs12125269	1.90E-26	1	76125211	A	C	0.291212	0.043695	0.004352
rs62073965	1.00E-24	17	45599591	C	T	0.513199	0.03971	0.003918
rs115811144	2.50E-22	2	2.12E+08	G	C	0.010593	-0.1987	0.020184
rs2657879	5.20E-22	12	56865338	G	A	0.182536	0.047889	0.005038
rs12307895	7.70E-21	12	47150176	T	C	0.207617	0.04303	0.00481
rs11666281	1.30E-17	19	18234588	T	C	0.25424	-0.03708	0.004479
rs149181595	3.90E-17	15	43685807	C	A	0.027601	0.094259	0.011912
rs34945403	1.70E-15	15	58430763	G	A	0.067631	-0.06347	0.008036
rs561931	4.50E-14	1	1.2E+08	G	A	0.580975	0.029099	0.003959
rs13107325	3.00E-13	4	1.03E+08	T	C	0.074309	-0.04944	0.007433
rs7704653	3.70E-13	5	90255685	G	A	0.724033	-0.03165	0.004409
rs2711697	1.10E-11	12	47265729	C	A	0.369926	0.026521	0.004034
rs3782860	3.20E-11	12	361996	T	C	0.546309	0.026265	0.003923
rs2168101	3.90E-11	11	8255408	A	C	0.308174	-0.02944	0.00436
rs192322963	4.10E-11	8	17445955	A	G	0.025449	0.082944	0.012538
rs36105243	5.30E-11	2	1.66E+08	G	C	0.347118	0.025833	0.004074

rs79687284	7.80E-11	1	2.14E+08	C	G	0.034645	0.067982	0.010678
rs11130192	1.20E-10	3	49467206	G	C	0.433069	-0.02397	0.003933
rs11172187	1.50E-10	12	57764193	G	C	0.469409	-0.02401	0.003981
rs6601302	2.10E-10	8	9239458	G	T	0.750342	0.027685	0.004516
rs3736951	8.20E-10	14	1.01E+08	T	C	0.513478	0.022027	0.003939
rs28435239	8.60E-10	9	5989087	A	G	0.769224	-0.02896	0.004649
rs555754	1.20E-09	6	1.61E+08	A	G	0.468479	0.023294	0.003916
rs11045866	7.60E-09	12	21367382	C	A	0.152288	0.031328	0.00548
rs9532939	8.70E-09	13	42440496	A	T	0.345473	0.025203	0.004264
rs1965869	9.00E-09	4	89677537	C	T	0.715726	-0.02421	0.004332
rs7188156	1.20E-08	16	79938114	G	T	0.14474	0.031719	0.005559
rs12566232	1.30E-08	1	1.52E+08	C	A	0.284491	0.022419	0.004333
rs1801133	1.80E-08	1	11856378	A	G	0.337391	0.023308	0.004113
rs11242109	2.00E-08	5	1.32E+08	T	G	0.47905	0.022605	0.003901
rs553564605	2.10E-08	19	7197931	A	T	0.281857	0.024373	0.004544
rs2608913	2.30E-08	6	1.32E+08	C	T	0.217028	-0.02435	0.004737
rs56113850	2.40E-08	19	41353107	C	T	0.579158	-0.0219	0.003954
rs17036342	2.50E-08	3	12393682	G	A	0.121047	0.033538	0.005966
rs114400025	3.00E-08	2	2.1E+08	G	T	0.01885	-0.07851	0.014674
rs2972154	3.10E-08	2	2.27E+08	T	C	0.538836	-0.02248	0.003896
rs116747369	3.50E-08	2	2.1E+08	T	C	0.018305	-0.07991	0.014929
rs78312207	3.70E-08	3	1.26E+08	A	G	0.068982	0.041417	0.007725
rs7078003	6.00E-08	10	99359412	T	C	0.178432	-0.02579	0.005105
rs499600	6.40E-08	1	46039077	G	T	0.841387	-0.02839	0.005345
rs3936511	6.90E-08	5	55860781	G	A	0.192401	-0.0254	0.004946
rs34439264	8.50E-08	19	49103243	T	C	0.136659	0.028119	0.005682
rs72675794	1.00E-07	8	1.21E+08	A	G	0.213361	-0.0242	0.004744
rs45446698	1.20E-07	7	99332948	G	T	0.041915	0.045207	0.009732
rs74273659	1.30E-07	20	5510794	T	C	0.0275	-0.06048	0.012206
rs61789561	1.40E-07	3	1.36E+08	A	G	0.195931	0.026621	0.004918
rs72776460	2.20E-07	10	819685	C	G	0.096433	-0.03305	0.006609
rs6125987	3.80E-07	20	48987950	T	C	0.192219	-0.02608	0.005029

rs7252042	3.90E-07	19	33720811	T	C	0.601576	0.018929	0.004019
rs35484034	4.00E-07	7	9090687	C	T	0.202676	-0.02559	0.004963
rs2967	4.20E-07	6	33689534	G	A	0.700949	0.019644	0.004271
rs34505073	4.40E-07	13	50237255	A	G	0.021697	0.065904	0.013488
rs34536774	5.90E-07	3	1.25E+08	C	G	0.136495	0.029084	0.005775
rs7608700	8.80E-07	2	1.36E+08	A	C	0.33343	0.021407	0.004228
rs6963105	9.30E-07	7	75097488	G	A	0.563173	-0.01873	0.003926
rs4254836	1.10E-06	4	55938914	T	C	0.881724	0.029503	0.006118
rs111504781	1.20E-06	2	31085106	T	C	0.033207	-0.05266	0.010993
rs145678014	1.20E-06	11	32927778	T	G	0.0444	0.044791	0.009492
rs117249010	1.20E-06	11	1.14E+08	A	G	0.009518	-0.10794	0.021788
rs4493564	1.30E-06	4	77420407	G	A	0.645572	0.018967	0.00408
rs56311083	1.30E-06	16	30557392	A	C	0.104656	-0.03063	0.006354
rs2293889	1.50E-06	8	1.17E+08	G	T	0.567099	-0.01864	0.003949
rs144132875	1.70E-06	6	20592286	A	G	0.045029	0.047868	0.009458
rs79325880	1.90E-06	7	8659322	A	C	0.010182	0.091283	0.020495
rs59668107	1.90E-06	9	1.12E+08	A	G	0.032975	0.051448	0.011037
rs28549287	2.10E-06	1	1.1E+08	A	G	0.773806	-0.02318	0.005074
rs73144782	2.10E-06	12	1.29E+08	T	C	0.187137	-0.02514	0.005003
rs12573237	2.20E-06	10	21417606	G	C	0.301693	-0.01978	0.004254
rs6879874	2.50E-06	5	1.77E+08	T	A	0.725202	0.021598	0.004375
rs76796332	2.50E-06	12	97925226	T	C	0.028515	-0.05533	0.012329
rs12811045	2.50E-06	12	1.16E+08	G	A	0.216027	-0.02118	0.004777
rs2123003	2.60E-06	3	1.6E+08	G	C	0.476116	-0.01951	0.003907
rs11098665	2.70E-06	4	75985263	T	C	0.801043	-0.02431	0.005061
rs2009509	2.70E-06	16	2824930	G	C	0.750222	-0.0208	0.004514
rs62357251	3.00E-06	5	8482499	A	G	0.044063	-0.03995	0.009469
rs72805967	3.00E-06	5	1.5E+08	T	C	0.120138	-0.02686	0.005987
rs1812599	3.00E-06	6	37852115	G	A	0.426761	-0.01656	0.003982
rs76013262	3.20E-06	7	56685225	G	A	0.02569	0.052292	0.012368
rs55727466	3.40E-06	2	12387457	T	C	0.572429	0.018454	0.003934
rs35626482	3.60E-06	9	6661410	G	A	0.37868	0.01659	0.004036

rs1936789	3.70E-06	6	1.27E+08	A	G	0.532884	-0.01806	0.00394
rs17114722	3.70E-06	15	25032802	G	A	0.094293	-0.03065	0.006765
rs35603048	3.90E-06	15	40391965	T	C	0.253969	-0.01916	0.004515
rs6449313	4.80E-06	4	17588320	G	A	0.623333	0.018061	0.004019
rs140695755	5.00E-06	2	61664762	A	G	0.022313	-0.05966	0.013263
rs118057622	5.00E-06	9	20074464	C	T	0.054405	0.03372	0.008569
rs188716412	5.70E-06	2	2.1E+08	C	T	0.015398	-0.07494	0.016185
rs537715477	6.00E-06	20	45207546	T	C	0.010439	0.093135	0.022383
rs77515689	6.20E-06	5	78311037	T	C	0.053835	0.040868	0.008677
rs34453616	7.30E-06	11	1.35E+08	G	C	0.199581	0.021542	0.004892
rs8127409	7.40E-06	21	21802487	A	G	0.175031	0.023911	0.005213
rs1209623	7.50E-06	2	1.83E+08	C	T	0.94664	0.039669	0.008828
rs34656869	7.70E-06	16	89709516	T	C	0.137589	-0.02668	0.005678
rs633911	8.00E-06	18	40944720	G	A	0.252682	0.018551	0.004526
rs1243473	8.40E-06	14	21544274	C	T	0.748832	-0.01793	0.004517
rs2133899	8.50E-06	10	57882308	T	C	0.397605	0.016964	0.003993
rs2246199	8.90E-06	6	29945620	T	C	0.091949	-0.03318	0.006753
rs7120967	8.90E-06	11	73108685	A	G	0.009507	-0.09004	0.021245
rs147504633	9.00E-06	5	28096709	C	A	0.039745	-0.04889	0.010047
rs489009	9.00E-06	11	89896457	T	C	0.507906	-0.01639	0.003941
rs12005290	9.10E-06	9	82705241	T	C	0.097522	0.026121	0.006624
rs150388628	1.00E-05	9	1.41E+08	A	G	0.020108	-0.06772	0.015161
rs4886484	1.00E-05	15	76527955	C	G	0.621676	0.016958	0.004018

Chr: Chromosome, BP: Position, EA: Effect Allele, NEA: Non-Effect Allele, EAF: Effect Allele Frequency, SE: Standard Error

Table S3. Identified variants associated with serum glutamine in the current UKB GWAS

SNP	p-value	Chr	BP	EA	NEA	EAF	Beta	SE
rs58995108	1.20e-312	12	56913616	A	T	0.791826	0.199502	0.00531
rs7078003	6.50E-82	10	99359412	T	C	0.17848	0.103948	0.005408
rs1260326	1.60E-81	2	27730940	C	T	0.604031	0.079257	0.004224
rs62182473	8.10E-52	2	1.92E+08	T	C	0.265297	-0.07073	0.004697
rs2168101	7.60E-48	11	8255408	A	C	0.308265	-0.06717	0.004617
rs79687284	1.30E-43	1	2.14E+08	C	G	0.034608	0.154204	0.011323
rs12306007	8.60E-33	12	47184042	T	C	0.141684	0.071369	0.005931
rs13234131	2.60E-32	7	73025975	G	A	0.129494	0.072124	0.00615
rs74093341	7.20E-32	12	59945201	A	G	0.126097	0.074998	0.006405
rs56156703	8.60E-25	3	1.6E+08	T	A	0.497917	-0.04168	0.004137
rs56335308	4.90E-20	8	17419461	A	G	0.027373	0.115579	0.012652
rs1998848	1.60E-16	14	21492229	A	G	0.010532	0.167169	0.020962
rs9482770	2.20E-15	6	1.27E+08	C	T	0.448707	-0.03312	0.004173
rs4237150	3.90E-15	9	4290085	C	G	0.403428	0.03342	0.004213
rs180426	6.10E-15	12	47173662	G	A	0.760684	-0.03766	0.004853
rs3736951	6.20E-15	14	1.01E+08	T	C	0.513473	0.032382	0.004173
rs28601761	1.40E-13	8	1.27E+08	G	C	0.420128	0.031584	0.004239
rs78431863	1.90E-13	11	74109553	T	C	0.041225	0.075814	0.010443
rs2608977	3.80E-13	6	1.32E+08	C	T	0.222857	-0.03565	0.004982
rs10849008	4.40E-12	12	4301876	C	T	0.381178	0.029345	0.004271
rs7925445	1.50E-11	11	18398958	G	A	0.556145	0.028324	0.004172
rs738408	3.50E-11	22	44324730	T	C	0.216546	0.033256	0.005016
rs8074974	1.30E-10	17	25595818	C	T	0.45004	0.027174	0.004158
rs35261542	2.90E-10	6	20675792	A	C	0.260951	0.029651	0.004713
rs11171756	3.60E-10	12	56539567	C	T	0.188823	0.03365	0.0054

rs28929474	6.60E-10	14	94844947	T	C	0.019978	0.090502	0.014798
rs7147721	7.60E-10	14	75186010	G	A	0.462072	0.025473	0.004157
rs58673065	1.90E-09	7	1885600	G	A	0.229249	0.029338	0.00491
rs17096421	3.30E-09	10	88820592	T	A	0.056411	-0.05237	0.009085
rs1274961	4.10E-09	3	39191335	C	T	0.782876	-0.02921	0.00501
rs2039098	4.50E-09	20	56112882	T	C	0.597146	-0.02426	0.004202
rs13396296	4.90E-09	2	2.11E+08	C	A	0.436659	-0.02454	0.004349
rs4309084	4.90E-09	10	1.15E+08	G	A	0.466486	-0.02408	0.004135
rs12325419	6.50E-09	16	70368909	A	G	0.11532	-0.03653	0.006469
rs10757283	8.40E-09	9	22134172	T	C	0.423362	-0.02401	0.004189
rs7966322	1.10E-08	12	1.21E+08	T	C	0.539917	-0.02526	0.004458
rs1750768	1.40E-08	10	22926227	T	C	0.041937	0.058293	0.010362
rs3988	1.50E-08	4	18241903	G	A	0.396083	-0.02235	0.004225
rs838737	2.10E-08	2	2.34E+08	A	G	0.565468	-0.02364	0.004169
rs7240148	2.50E-08	18	31557481	C	T	0.433385	-0.02276	0.004178
rs117643180	2.70E-08	17	7185779	A	C	0.025849	0.069617	0.0131
rs73186030	3.10E-08	3	1.22E+08	T	C	0.130234	0.032395	0.006137
rs147139486	3.40E-08	5	1.28E+08	A	G	0.054064	-0.05134	0.00928
rs2395779	4.30E-08	6	41511950	T	C	0.285284	0.024825	0.004617
rs9687846	4.30E-08	5	55861894	A	G	0.201368	-0.02888	0.005146
rs2083180	4.50E-08	3	52668119	T	G	0.40395	0.022926	0.004212
rs1323320	5.00E-08	6	56287985	A	G	0.263338	-0.02563	0.004694
rs11993225	6.30E-08	8	1.34E+08	C	A	0.269895	0.026094	0.004644
rs35146081	7.70E-08	11	30675441	C	T	0.31012	-0.02369	0.004486
rs28362590	1.10E-07	5	1.77E+08	T	G	0.754396	-0.02583	0.004806
rs9903592	1.40E-07	17	58854993	A	G	0.092189	0.037542	0.007243
rs139097404	1.60E-07	15	43933941	C	T	0.023504	0.074731	0.013823
rs4722551	1.60E-07	7	25991826	C	T	0.158113	0.029883	0.005657
rs2291444	1.70E-07	11	47179143	T	C	0.555313	0.021194	0.004161
rs13386066	1.90E-07	2	1.98E+08	G	A	0.193342	-0.02693	0.005232
rs13107325	2.10E-07	4	1.03E+08	T	C	0.074292	-0.04151	0.007879
rs11693327	2.20E-07	2	1.12E+08	C	T	0.138514	0.030969	0.006003

rs1744418	2.20E-07	6	39222982	G	A	0.293817	-0.02323	0.004577
rs2713550	2.20E-07	2	2.27E+08	C	T	0.341312	-0.02263	0.004359
rs115271198	2.70E-07	17	46319476	T	C	0.070161	0.040976	0.008119
rs4946936	3.00E-07	6	1.09E+08	C	T	0.710737	-0.02462	0.004602
rs568299	3.20E-07	5	80313786	A	G	0.661617	0.022325	0.004382
rs10916877	3.80E-07	1	21172024	A	G	0.564297	-0.02092	0.004202
rs35381887	3.80E-07	5	90191468	A	G	0.276636	-0.02315	0.004621
rs499600	4.50E-07	1	46039077	G	T	0.841372	-0.02774	0.005666
rs10019104	4.70E-07	4	1.62E+08	G	A	0.368705	0.020929	0.00429
rs11568438	4.80E-07	14	23282336	A	G	0.018148	0.07861	0.015442
rs147115029	5.10E-07	3	12321331	T	C	0.013056	0.09132	0.018339
rs7200772	5.60E-07	16	68446811	C	T	0.574994	0.020932	0.004266
rs79339930	6.20E-07	3	1.68E+08	T	C	0.190526	-0.02603	0.005407
rs1975283	6.80E-07	1	1.18E+08	C	A	0.687493	-0.02131	0.004482
rs10062238	6.90E-07	5	46396424	T	G	0.523346	0.02093	0.004187
rs71370489	7.70E-07	17	7182187	C	T	0.026325	0.063379	0.013098
rs145457999	8.20E-07	1	1.16E+08	G	A	0.0138	-0.08846	0.01851
rs10492229	8.30E-07	12	1.11E+08	T	C	0.235129	-0.02404	0.004868
rs72555385	8.40E-07	7	73123473	G	A	0.049316	-0.04542	0.009592
rs149539065	9.00E-07	12	1.33E+08	T	C	0.358349	-0.02114	0.004382
rs10005413	9.10E-07	4	1.49E+08	T	C	0.013284	-0.08764	0.01813
rs149374882	9.50E-07	4	1841663	C	T	0.009322	-0.10675	0.022263
rs878521	9.60E-07	7	44255643	A	G	0.254746	-0.02321	0.004775
rs206966	9.70E-07	12	1.21E+08	T	C	0.155383	0.02918	0.005915
rs115926813	1.10E-06	5	1.32E+08	A	G	0.038511	-0.05244	0.010762
rs157923	1.20E-06	7	1.31E+08	C	G	0.922011	-0.03488	0.007761
rs56336142	1.20E-06	6	39134099	C	T	0.212081	0.025984	0.005092
rs59385441	1.20E-06	14	89943244	A	G	0.323249	-0.0225	0.004652
rs72637216	1.20E-06	1	1.65E+08	G	T	0.050214	0.045332	0.009471
rs1229984	1.40E-06	4	1E+08	C	T	0.972519	0.060002	0.012499
rs2024947	1.40E-06	20	62205498	T	C	0.213985	-0.0249	0.005211
rs7502539	1.50E-06	17	38219005	G	A	0.665241	0.022333	0.004421

rs56015043	1.60E-06	4	69440672	T	C	0.397304	-0.02522	0.005114
rs7849271	1.60E-06	9	75597830	A	G	0.305964	0.022039	0.004482
rs167769	1.70E-06	12	57503775	T	C	0.398575	0.020123	0.004222
rs12526848	1.80E-06	6	7232776	G	C	0.098314	0.032372	0.006934
rs1991880	1.80E-06	1	2.14E+08	C	A	0.43102	-0.02046	0.004173
rs8129303	2.00E-06	21	40711959	A	G	0.356953	0.020568	0.004351
rs13136642	2.10E-06	4	1.32E+08	T	A	0.468759	-0.01845	0.00418
rs61909515	2.20E-06	11	1.28E+08	G	C	0.131138	0.02905	0.006318
rs75394182	2.30E-06	12	3849973	T	C	0.063114	0.041682	0.008613
rs2858980	2.40E-06	13	33554587	A	G	0.830832	-0.02566	0.005509
rs115478735	2.50E-06	9	1.36E+08	T	A	0.183277	0.023422	0.005346
rs4302820	2.50E-06	8	56969683	C	T	0.211002	-0.02302	0.005109
rs12889267	2.70E-06	14	21542766	G	A	0.166898	0.026247	0.005532
rs12409343	3.00E-06	1	56369195	C	T	0.02236	0.065885	0.014911
rs117747082	3.20E-06	11	62989133	T	C	0.011516	-0.08853	0.019348
rs761437	3.40E-06	20	24611585	G	A	0.633839	-0.01987	0.004315
rs2006227	3.50E-06	22	24995756	A	C	0.343651	-0.02047	0.004348
rs4984916	3.60E-06	16	749057	A	G	0.220462	-0.02367	0.004987
rs236658	3.80E-06	7	75049460	G	A	0.551529	-0.01902	0.004303
rs9966683	3.80E-06	18	55339942	C	T	0.138139	0.027462	0.00601
rs9923988	4.20E-06	16	88840554	G	C	0.383482	-0.02006	0.004299
rs149078666	4.40E-06	18	12426732	A	G	0.021847	0.063535	0.014305
rs143504274	5.00E-06	15	65414923	G	A	0.009294	0.098335	0.021503
rs79170412	5.10E-06	14	56714637	C	A	0.213095	0.022824	0.005075
rs117555550	5.40E-06	9	1.2E+08	T	C	0.027765	0.054868	0.0126
rs115155629	5.70E-06	6	1.35E+08	C	T	0.027471	-0.06071	0.01282
rs138502078	5.70E-06	14	21427022	T	C	0.006564	-0.12113	0.027012
rs11751024	5.80E-06	6	32586236	A	C	0.39853	0.020299	0.004221
rs1556026	5.80E-06	9	98808736	A	G	0.326745	0.019794	0.004452
rs28690260	5.80E-06	8	17365162	T	A	0.403588	0.020472	0.004699
rs9449390	5.80E-06	6	82753144	T	A	0.079346	0.034516	0.007793
rs192492501	6.00E-06	16	402901	A	G	0.175092	0.025891	0.005468

rs8075185	6.00E-06	17	36060216	C	T	0.605098	0.019411	0.004262
rs17703797	6.70E-06	4	88080971	C	T	0.120322	0.030039	0.006355
rs143384	6.80E-06	20	34025756	G	A	0.404031	0.018454	0.004208
rs67705163	7.00E-06	7	27954738	T	C	0.142327	0.026969	0.005912
rs9647335	7.10E-06	3	1.36E+08	T	A	0.192047	0.02458	0.005264
rs11881511	7.30E-06	19	2543476	A	G	0.483841	-0.01889	0.004155
rs77033759	8.00E-06	8	1.16E+08	T	C	0.025391	0.061259	0.013193
rs116656747	8.20E-06	2	56179018	T	C	0.011215	-0.09106	0.021291
rs77010315	8.30E-06	5	1.51E+08	A	C	0.011935	-0.08303	0.019087
rs7805168	8.30E-06	7	29603280	C	G	0.660038	0.019693	0.004365
rs147767080	8.60E-06	5	1.41E+08	G	C	0.027517	-0.05849	0.01304
rs741190	9.40E-06	2	20452868	A	G	0.035224	0.049487	0.011198
rs55825790	9.50E-06	10	1.02E+08	A	G	0.060977	-0.04018	0.008777
rs4845406	9.60E-06	1	1.55E+08	G	A	0.547728	0.018461	0.00415
rs72844596	9.70E-06	17	73268373	A	G	0.01349	0.076498	0.01837
rs71635591	9.80E-06	1	2.03E+08	T	C	0.066402	-0.03812	0.008354
rs7699214	1.00E-05	4	1.21E+08	G	A	0.519757	0.017912	0.004145

Chr: Chromosome, BP: Position, EA: Effect Allele, NEA: Non-Effect Allele, EAF: Effect Allele Frequency, SE: Standard Error

Table S4. Identified variants associated with serum histidine in the current UKB GWAS

SNP	p-value	Chr	BP	EA	NEA	EAF	Beta	SE
rs1047891	4.00E-55	2	2.12E+08	A	C	0.314892	-0.06982	0.004465
rs75341400	4.00E-36	12	96315168	T	G	0.098269	-0.08816	0.007025
rs2939302	1.10E-28	12	56860020	A	G	0.183177	0.059485	0.005355
rs34004251	3.50E-23	3	1.49E+08	A	T	0.176425	-0.05426	0.005471
rs1958393	5.90E-20	14	21516447	G	C	0.109029	0.061491	0.006723
rs780094	9.90E-16	2	27741237	C	T	0.615114	0.034181	0.004258
rs12422818	3.00E-15	12	96478835	A	G	0.138836	0.047511	0.006021
rs61938133	7.80E-14	12	96120605	A	G	0.040056	0.079641	0.010657
rs1883711	7.70E-13	20	39179822	C	G	0.031167	0.087242	0.012174
rs2272662	3.70E-12	8	1.46E+08	C	T	0.576631	0.029189	0.004201
rs150844304	3.90E-12	15	43726625	C	A	0.025691	0.090879	0.013096
rs709932	5.50E-12	14	94849201	T	C	0.155773	0.039402	0.005717
rs11621792	5.50E-11	14	24871926	T	C	0.453809	-0.02742	0.004182
rs117591301	1.00E-09	6	1.35E+08	T	A	0.027531	-0.07835	0.012835
rs76191003	1.40E-09	12	47182531	T	C	0.068358	0.049709	0.008216
rs2168101	3.50E-09	11	8255408	A	C	0.308204	-0.02735	0.00463
rs12140995	4.10E-09	1	1.69E+08	C	T	0.410996	-0.02479	0.004217
rs8007176	1.40E-07	14	21502345	G	A	0.360648	-0.02284	0.004341
rs9385710	2.00E-07	6	1.35E+08	G	A	0.710715	-0.02394	0.004601
rs554877699	2.90E-07	6	32487677	C	T	0.103363	-0.04415	0.008603
rs2105941	3.50E-07	20	42183219	A	G	0.314941	-0.02295	0.004506
rs799157	4.90E-07	7	73020301	C	T	0.956469	0.051007	0.010143
rs35848234	5.10E-07	16	53367192	T	C	0.460914	-0.02092	0.004166
rs532281808	5.50E-07	5	1.78E+08	T	C	0.492052	0.023939	0.004779
rs139749729	6.30E-07	13	93890862	C	T	0.053366	-0.04728	0.009491
rs17368188	8.00E-07	2	1.24E+08	C	T	0.008965	0.110168	0.022324

rs62518081	8.90E-07	8	1.34E+08	A	G	0.107546	0.033202	0.006757
rs7086	9.00E-07	15	42451108	C	G	0.136889	0.029725	0.006051
rs60122847	9.40E-07	12	96363081	G	A	0.041077	-0.05271	0.010747
rs4646248	1.30E-06	8	18260355	T	C	0.714996	-0.0222	0.004588
rs575146	1.50E-06	22	24295074	A	G	0.601916	-0.02125	0.00442
rs4706920	2.20E-06	6	82742017	A	T	0.126453	0.029582	0.006252
rs13222398	2.70E-06	7	1.52E+08	T	G	0.166147	-0.02634	0.005615
rs6853509	2.80E-06	4	1.89E+08	T	C	0.4613	-0.01964	0.004195
rs139190506	2.80E-06	10	88981118	G	A	0.023016	-0.06961	0.014861
rs11264237	2.90E-06	1	1.55E+08	A	G	0.272112	0.02178	0.004659
rs72999466	3.00E-06	19	18248499	T	C	0.258332	0.022122	0.004733
rs11513729	3.20E-06	12	1.12E+08	T	C	0.413347	-0.0199	0.004272
rs11650918	3.70E-06	17	27969220	T	C	0.037852	0.05017	0.010836
rs12497438	5.20E-06	3	1.48E+08	A	G	0.194517	-0.02407	0.005284
rs28601761	5.80E-06	8	1.27E+08	G	C	0.420102	0.019269	0.004249
rs72763302	6.10E-06	9	1.4E+08	G	C	0.081867	0.035369	0.007819
rs2556569	6.30E-06	15	44895326	T	C	0.049271	-0.0449	0.009939
rs7012637	6.50E-06	8	9173209	A	G	0.474112	-0.01885	0.00418
rs4311466	6.70E-06	5	40896837	C	T	0.988547	-0.08807	0.019561
rs35836665	6.80E-06	13	42921882	C	T	0.039743	-0.04809	0.01069
rs11179852	7.20E-06	12	74366332	T	C	0.204686	0.023064	0.005138
rs117913959	7.30E-06	12	96935491	A	G	0.034599	0.051197	0.011418
rs9958619	7.30E-06	18	74312273	A	G	0.020249	0.067491	0.015048
rs17800760	7.50E-06	19	45013959	A	G	0.156629	-0.02554	0.005702
rs902840	8.10E-06	7	1.29E+08	T	C	0.282244	0.021015	0.004709
rs539245	8.20E-06	6	7019417	C	A	0.391776	0.019042	0.004271
rs1924071	8.20E-06	13	28055737	A	G	0.043172	0.045566	0.010216
rs561139158	8.20E-06	16	64837511	T	A	0.036968	0.056267	0.012617
rs3774276	8.30E-06	3	1.87E+08	T	C	0.119762	-0.02849	0.006391
rs72881240	8.50E-06	18	22666810	T	C	0.209565	0.022722	0.005105
rs114932870	9.20E-06	1	1.15E+08	T	A	0.038578	-0.04915	0.011081
rs150183237	9.20E-06	19	691322	G	A	0.020061	-0.06708	0.015127

rs4846235	9.80E-06	1	10632235	T	C	0.327114	0.019567	0.004425

Chr: Chromosome, BP: Position, EA: Effect Allele, NEA: Non-Effect Allele, EAF: Effect Allele Frequency, SE: Standard Error

Table S5. Identified variants associated with serum isoleucine in the current UKB GWAS

SNP	p-value	Chr	BP	EA	NEA	EAF	Beta	SE
rs10018448	4.10E-60	4	89225171	G	A	0.535322	0.066556	0.00407
rs12325419	1.10E-31	16	70368909	A	G	0.115325	-0.0743	0.006343
rs1260326	2.30E-29	2	27730940	C	T	0.60398	-0.04662	0.004144
rs545587	2.30E-19	19	49319664	C	A	0.514774	0.036642	0.004072
rs7302925	3.60E-16	12	56861458	G	A	0.801117	-0.04133	0.005071
rs117643180	2.90E-15	17	7185779	A	C	0.025874	-0.10139	0.012841
rs10184004	4.00E-13	2	1.66E+08	T	C	0.406102	-0.02995	0.004128
rs2127864	5.00E-09	4	89147993	G	A	0.175281	-0.03135	0.005361
rs2941456	9.10E-09	8	76443463	A	G	0.198451	0.029259	0.005092
rs1471740	6.10E-08	3	1.36E+08	C	T	0.74018	0.025116	0.004637
rs9899432	8.90E-08	17	8204370	C	G	0.157987	-0.0298	0.005572
rs2422358	1.00E-07	2	65231806	T	A	0.435743	0.022618	0.004251
rs117164273	1.70E-07	16	68292616	A	G	0.045425	-0.05103	0.009757
rs116431586	1.90E-07	4	1.43E+08	T	C	0.02676	-0.06681	0.012836
rs2326458	3.80E-07	16	84987679	A	C	0.745449	-0.02363	0.004653
rs2973444	4.90E-07	5	90417593	C	T	0.085986	-0.03656	0.007266
rs1605749	5.70E-07	17	68475145	G	A	0.510911	-0.02024	0.004049
rs79259321	5.90E-07	12	47056356	A	C	0.072144	0.039124	0.007834
rs2616272	6.60E-07	11	97448214	A	G	0.49593	-0.02027	0.004076
rs941952	7.50E-07	14	1.05E+08	C	T	0.718723	0.022585	0.004564
rs73343569	1.50E-06	8	1.33E+08	A	G	0.009641	-0.10664	0.022182
rs30351	1.70E-06	5	55794632	A	G	0.744178	0.022361	0.004671
rs199607859	2.30E-06	6	1.4E+08	T	G	0.594537	-0.01959	0.004144
rs36216559	2.40E-06	2	262335	G	A	0.350063	-0.02	0.004245
rs1682628	2.50E-06	7	95283613	A	G	0.642143	-0.01999	0.004246
rs7201861	2.70E-06	16	79532675	C	G	0.710928	0.020983	0.004473
rs150277164	3.00E-06	4	89258580	A	G	0.014177	0.080314	0.017185

rs11166420	3.10E-06	1	1.01E+08	A	T	0.904411	-0.03222	0.006903
rs3771600	3.30E-06	2	1.59E+08	G	C	0.275449	-0.02111	0.004541
rs146774705	3.30E-06	3	96082587	A	G	0.030486	0.056081	0.012052
rs111423330	3.40E-06	15	62418810	C	T	0.327073	-0.02089	0.004499
rs7649917	3.60E-06	3	1.93E+08	C	T	0.227384	0.022421	0.004842
rs1924699	3.80E-06	10	95565254	T	G	0.939678	0.03933	0.008513
rs4805474	3.80E-06	19	30250970	G	T	0.712419	0.020761	0.004493
rs113821719	4.10E-06	10	11824902	C	G	0.018322	0.075767	0.016443
rs589625	4.10E-06	18	524173	C	G	0.812033	0.024067	0.005224
rs35509506	4.20E-06	6	44595179	T	C	0.13489	0.027663	0.006014
rs7380516	4.30E-06	5	1.35E+08	T	A	0.643668	-0.02253	0.004899
rs565181303	4.60E-06	6	1.6E+08	T	G	0.203236	-0.02468	0.00539
rs2168101	4.90E-06	11	8255408	A	C	0.308213	-0.0207	0.004529
rs11583974	5.40E-06	1	55551718	A	G	0.042438	0.045783	0.010062
rs17540179	6.00E-06	6	1.12E+08	C	T	0.014755	0.077078	0.017035
rs148772605	6.00E-06	12	1.08E+08	A	G	0.022434	0.065525	0.01448
rs139190404	6.10E-06	12	1.33E+08	A	G	0.037597	-0.04848	0.010722
rs7206790	6.20E-06	16	53797908	G	C	0.468079	0.018507	0.004095
rs147000073	6.50E-06	5	1.51E+08	G	C	0.011637	-0.08606	0.019083
rs34958196	6.50E-06	7	73043101	A	G	0.130936	-0.02704	0.005999
rs34238610	7.60E-06	3	53026480	T	A	0.101068	-0.03012	0.00673
rs12219397	7.70E-06	10	5146578	C	T	0.217671	0.021897	0.004894
rs28435138	7.70E-06	13	20694866	A	G	0.32666	-0.01958	0.004377
rs72702062	7.90E-06	4	1.87E+08	T	C	0.336018	0.019184	0.004294
rs112356680	8.00E-06	7	41623268	A	G	0.020793	-0.06518	0.014592
rs28457808	8.10E-06	7	99023246	G	C	0.163639	-0.02451	0.005493
rs6542674	9.30E-06	2	3632442	T	C	0.866926	0.026606	0.006003
rs113986322	9.60E-06	21	43167550	C	T	0.09468	-0.03074	0.006947
rs17703223	9.70E-06	8	88207033	A	G	0.033531	-0.05105	0.011539
rs138331001	9.80E-06	1	2.45E+08	A	G	0.007243	-0.10993	0.024863
rs200115642	1.00E-05	2	2.27E+08	A	T	0.637116	0.01927	0.004372
rs1660613	1.00E-05	10	34567543	C	T	0.578571	0.01841	0.004173

Chr: Chromosome, BP: Position, EA: Effect Allele, NEA: Non-Effect Allele, EAF: Effect Allele Frequency, SE: Standard Error

Table S6. Identified variants associated with serum leucine in the current UKB GWAS

SNP	p-value	Chr	BP	EA	NEA	EAF	Beta	SE
rs10018448	4.30E-180	4	89225171	G	A	0.535318	0.087817	0.003976
rs35452938	4.00E-50	16	70270907	C	T	0.114284	-0.09473	0.006363
rs1260326	3.60E-32	2	27730940	C	T	0.603977	-0.0478	0.004048
rs9930957	2.10E-22	16	72149923	T	C	0.159836	0.052545	0.005398
rs117643180	4.00E-21	17	7185779	A	C	0.02587	-0.11838	0.012549
rs2638315	1.20E-15	12	56865056	C	G	0.182491	0.040947	0.005116
rs13389219	2.30E-15	2	1.66E+08	T	C	0.392533	-0.03214	0.004057
rs4801776	2.40E-13	19	49304215	T	C	0.299241	-0.03237	0.004422
rs150277164	5.20E-12	4	89258580	A	G	0.014177	0.115856	0.01679
rs7655059	1.10E-10	4	89149144	G	C	0.175216	-0.03378	0.005239
rs2422358	1.40E-09	2	65231806	T	A	0.435742	0.025136	0.004153
rs12974412	3.30E-09	19	14151809	G	A	0.779952	-0.02911	0.00492
rs10774625	5.30E-09	12	1.12E+08	G	A	0.502842	-0.02314	0.003964
rs5747934	8.60E-09	22	18915282	T	C	0.042294	-0.05684	0.009873
rs11166420	1.70E-08	1	1.01E+08	A	T	0.90441	-0.03801	0.006744
rs2977929	1.80E-08	8	76454025	T	C	0.196785	0.028202	0.005006
rs34894639	2.50E-08	3	1.36E+08	T	C	0.231589	-0.0264	0.004737
rs1601935	7.10E-08	15	58671765	T	G	0.654661	0.022603	0.004195
rs200115642	7.80E-08	2	2.27E+08	A	T	0.637113	0.022945	0.004271
rs145815601	1.60E-07	1	1.55E+08	T	G	0.0367	-0.05591	0.01067
rs13100474	1.90E-07	3	1.22E+08	A	G	0.205777	0.025701	0.004934
rs79020793	2.30E-07	12	47199549	A	G	0.068228	0.040728	0.007869
rs17096421	4.70E-07	10	88820592	T	A	0.056395	0.043874	0.008707
rs61587941	5.90E-07	5	55814890	A	G	0.051232	-0.04479	0.008965
rs10822186	7.20E-07	10	65350383	G	A	0.494277	-0.01964	0.003963
rs182222149	8.00E-07	7	12609043	G	T	0.016427	-0.08044	0.016295
rs592423	8.70E-07	6	1.4E+08	C	A	0.553909	-0.01958	0.003981

rs2326458	9.30E-07	16	84987679	A	C	0.745446	-0.0223	0.004546
rs2950851	1.40E-06	5	90399012	T	A	0.095627	-0.03258	0.006756
rs7649917	1.50E-06	3	1.93E+08	C	T	0.227381	0.022762	0.004731
rs36181536	1.50E-06	12	1.23E+08	C	T	0.673499	-0.02037	0.004233
rs3796352	2.10E-06	3	52913279	T	C	0.099619	-0.03136	0.006605
rs62111417	2.10E-06	19	7193981	T	C	0.712803	-0.02083	0.004393
rs17703747	2.40E-06	8	88255713	A	G	0.040431	-0.04807	0.010196
rs17310469	2.60E-06	5	1.16E+08	G	A	0.159063	0.026002	0.005536
rs79327462	2.60E-06	16	67958058	A	G	0.045151	-0.04479	0.009535
rs9940976	2.80E-06	16	72311826	C	A	0.201957	-0.0232	0.004949
rs549751842	3.00E-06	2	296410	C	T	0.319952	-0.02047	0.004384
rs429358	3.20E-06	19	45411941	C	T	0.154826	-0.02555	0.005484
rs11948950	3.60E-06	5	55876283	A	G	0.105355	0.030452	0.00657
rs1564369	3.70E-06	12	21335290	G	T	0.178218	0.024027	0.005191
rs2168101	3.80E-06	11	8255408	A	C	0.308207	-0.02045	0.004426
rs79545275	3.90E-06	4	1.42E+08	A	G	0.027539	-0.05572	0.012074
rs3848453	4.00E-06	17	68362094	C	T	0.583643	0.018529	0.004018
rs182491129	4.20E-06	1	1E+08	A	C	0.011988	-0.08381	0.01822
rs4584484	4.40E-06	10	5148669	A	G	0.217697	0.021967	0.004783
rs10848097	4.40E-06	12	1.31E+08	C	A	0.607488	0.018679	0.004067
rs13402475	4.70E-06	2	3639909	G	C	0.815966	0.023842	0.005208
rs1438506	5.10E-06	10	1.33E+08	A	T	0.352719	-0.01893	0.00415
rs146774705	5.20E-06	3	96082587	A	G	0.030486	0.053652	0.011776
rs45620134	5.30E-06	11	1.12E+08	T	C	0.041692	0.045234	0.009938
rs138331001	6.20E-06	1	2.45E+08	A	G	0.007243	-0.10982	0.024291
rs16833240	6.50E-06	2	1.37E+08	A	G	0.063704	-0.03686	0.008173
rs2023494	6.90E-06	14	72560539	T	C	0.646671	0.018719	0.004164
rs139683582	7.90E-06	2	57893329	A	G	0.017845	-0.06734	0.015075
rs62175484	8.10E-06	2	1.61E+08	C	A	0.529437	-0.01837	0.004116
rs10734344	8.10E-06	11	23779713	A	T	0.960959	-0.04899	0.010979
rs115640614	8.20E-06	4	1.62E+08	C	A	0.032584	0.050726	0.011377
rs11765000	8.20E-06	7	98888291	C	T	0.069204	-0.03498	0.007843

rs73805830	8.70E-06	5	1.72E+08	T	C	0.120596	-0.02734	0.006147
rs72835458	8.70E-06	6	17545255	G	C	0.070297	-0.03511	0.007893
rs7691693	8.80E-06	4	1.41E+08	C	T	0.331736	0.018823	0.004236
rs1398079	8.80E-06	15	57707706	T	C	0.287429	-0.01953	0.004395
rs9751871	9.00E-06	2	2.29E+08	C	T	0.535294	-0.01766	0.003976
rs12318565	9.10E-06	12	92736142	C	A	0.348125	0.019833	0.00447
rs743714	9.50E-06	4	3419986	G	A	0.620317	-0.0181	0.004087
rs958648	9.90E-06	8	11103895	G	A	0.556904	0.017609	0.003985

Chr: Chromosome, BP: Position, EA: Effect Allele, NEA: Non-Effect Allele, EAF: Effect Allele Frequency, SE: Standard Error

TableS7. Identified variants associated with serum phenylalanine in the current UKB GWAS

SNP	p-value	Chr	BP	EA	NEA	EAF	Beta	SE
rs1498694	1.60E-146	12	1.03E+08	G	A	0.610845	-0.1093	0.00424
rs4646986	6.00E-51	12	1.03E+08	A	G	0.313218	-0.06718	0.004475
rs10750864	5.00E-34	11	57259694	A	T	0.741891	0.058049	0.004773
rs17253619	1.30E-26	14	55411744	C	T	0.118765	0.068032	0.006372
rs142025056	2.00E-22	12	1.03E+08	C	T	0.012969	0.188268	0.019322
rs116868345	2.50E-12	12	1.02E+08	T	C	0.006217	0.189848	0.027107
rs9467618	1.70E-10	6	25824809	T	C	0.255343	-0.03028	0.004742
rs715	4.20E-10	2	2.12E+08	C	T	0.310916	-0.02801	0.004486
rs150772783	3.30E-09	6	18117955	G	C	0.049698	-0.05638	0.00953
rs2933235	4.30E-09	12	56921418	G	A	0.166225	0.032544	0.005542
rs1533403	4.90E-09	12	1.03E+08	C	T	0.896425	0.039924	0.006823
rs12299423	5.80E-09	12	1.04E+08	G	A	0.048351	0.057995	0.009959
rs4781730	1.90E-08	16	16254626	T	C	0.284469	-0.02575	0.004583
rs4135105	2.80E-08	12	1.04E+08	T	C	0.010662	0.113913	0.020501
rs2180314	4.10E-08	6	52617731	G	C	0.596097	0.023109	0.004211

Chr: Chromosome, BP: Position, EA: Effect Allele, NEA: Non-Effect Allele, EAF: Effect Allele Frequency, SE: Standard Error

Table S8. Identified variants associated with serum tyrosine in the current UKB GWAS

SNP	p-value	Chr	BP	EA	NEA	EAF	Beta	SE
rs12206654	3.00E-225	6	1.12E+08	C	T	0.146605	0.188204	0.005874
rs150851429	1.90E-120	16	71625831	C	G	0.01598	0.399298	0.017111
rs7979473	1.80E-47	12	1.21E+08	G	A	0.613011	-0.06176	0.004268
rs2638315	1.10E-36	12	56865056	C	G	0.182446	0.067396	0.005328
rs9940990	2.40E-30	16	71831343	A	T	0.358944	-0.04924	0.004301
rs4149056	2.70E-25	12	21331549	C	T	0.150912	-0.05967	0.005743
rs35614968	1.20E-21	6	1.12E+08	A	G	0.028196	-0.12188	0.012749
rs79020793	3.10E-21	12	47199549	A	G	0.068215	0.077516	0.008194
rs1126671	1.70E-19	4	1E+08	C	T	0.697944	-0.04059	0.004496
rs28601761	7.30E-19	8	1.27E+08	G	C	0.420115	-0.03752	0.00423
rs9414801	5.30E-18	10	65149089	A	G	0.528632	-0.03585	0.004146
rs174547	5.10E-15	11	61570783	C	T	0.342661	0.033926	0.004336
rs34894639	6.10E-15	3	1.36E+08	T	C	0.23159	-0.03849	0.004933
rs740122	1.10E-13	7	28166442	A	G	0.217464	-0.03714	0.005002
rs535097680	1.70E-13	6	1.12E+08	G	T	0.050246	-0.08857	0.012019
rs114232169	1.90E-13	3	1.86E+08	G	T	0.667814	-0.03223	0.004381
rs715	9.90E-12	2	2.12E+08	C	T	0.310821	-0.03049	0.004478
rs2230681	1.40E-11	12	1.22E+08	C	T	0.874664	-0.04208	0.00623
rs2168101	1.60E-11	11	8255408	A	C	0.308215	-0.03106	0.00461
rs62415384	1.50E-10	6	43261767	C	G	0.453285	0.027476	0.00429
rs6575900	2.00E-10	14	1.03E+08	C	G	0.204047	-0.03268	0.005138
rs814573	2.30E-10	19	45424351	T	A	0.186313	-0.03452	0.005442
rs75661418	9.70E-10	6	1.11E+08	G	A	0.024336	-0.08165	0.013354
rs1800961	1.30E-09	20	43042364	T	C	0.030193	-0.07314	0.012043
rs7019778	2.50E-09	9	22134651	C	A	0.424391	-0.02495	0.004186
rs6868618	3.00E-09	5	90191256	G	A	0.380117	0.025292	0.004265
rs3747207	4.80E-09	22	44324855	A	G	0.21503	0.029401	0.005021

rs7786102	8.10E-09	7	25965759	A	G	0.715423	-0.02635	0.004569
rs56236906	9.30E-09	6	1.12E+08	A	G	0.020553	0.083634	0.014563
rs116862171	1.00E-08	6	1.12E+08	G	A	0.01745	0.092549	0.01615
rs36055409	2.10E-08	12	59990603	T	G	0.14239	0.033308	0.005944
rs61676179	2.10E-08	22	32176825	A	C	0.050985	-0.05277	0.009418
rs1433210	3.00E-08	4	1.25E+08	C	A	0.246119	-0.02658	0.004794
rs9267947	3.60E-08	6	32211218	G	A	0.478636	0.022718	0.004122
rs28733806	3.70E-08	16	70335323	G	C	0.422878	0.02298	0.004173
rs194749	4.10E-08	14	69273905	C	T	0.24019	0.026561	0.00484
rs4665972	6.00E-08	2	27598097	C	T	0.604592	0.022994	0.004243
rs7570775	7.60E-08	2	1.92E+08	G	T	0.339554	-0.02343	0.004358
rs67549392	1.00E-07	12	1.03E+08	A	T	0.311646	0.023793	0.004471
rs17050272	1.10E-07	2	1.21E+08	A	G	0.409325	0.022287	0.004193
rs148165862	1.10E-07	7	1.28E+08	G	A	0.258824	0.025379	0.004786
rs17817288	1.20E-07	16	53807764	G	A	0.484892	0.021831	0.004124
rs570539343	1.30E-07	2	87141670	A	G	0.4331	0.026162	0.004953
rs1550335	1.30E-07	15	51405511	A	G	0.391253	-0.02241	0.004246
rs115281911	1.50E-07	3	27935353	C	T	0.010271	0.112724	0.021455
rs10027275	1.60E-07	4	1.49E+08	C	G	0.740868	-0.02489	0.004754
rs367992425	2.90E-07	10	23758985	C	T	0.152077	0.033332	0.006499
rs11024636	3.00E-07	11	18405491	T	C	0.550549	0.021289	0.004154
rs13135092	3.10E-07	4	1.03E+08	G	A	0.082897	-0.0384	0.007505
rs4698232	3.70E-07	4	18413926	C	T	0.16046	0.030426	0.005986
rs140584594	3.80E-07	1	1.1E+08	G	A	0.730043	-0.02347	0.004621
rs35207259	5.50E-07	2	1.36E+08	T	C	0.095016	-0.03712	0.007414
rs6129629	5.80E-07	20	39185278	A	G	0.323135	-0.02229	0.004461
rs62106258	5.90E-07	2	417167	C	T	0.048172	-0.04806	0.009625
rs78787298	6.20E-07	14	34660824	A	G	0.043749	0.050523	0.010133
rs7005453	9.00E-07	8	1.27E+08	A	G	0.211271	0.024897	0.005067
rs3799350	9.60E-07	6	25799929	T	G	0.266118	0.022877	0.00467
rs951506	1.20E-06	1	81341758	C	T	0.357738	0.020886	0.004299
rs10025494	1.20E-06	4	77189380	G	A	0.218511	0.024265	0.004989

rs10493013	1.30E-06	1	22703035	C	T	0.176337	-0.0262	0.005413
rs17232826	1.50E-06	16	89842784	T	C	0.104214	0.032525	0.006766
rs74950025	1.70E-06	16	16245125	A	G	0.108439	0.031898	0.006667
rs7624902	1.80E-06	3	1.6E+08	G	A	0.537566	-0.01979	0.004141
rs635634	1.80E-06	9	1.36E+08	T	C	0.183525	0.025494	0.005345
rs61926203	1.80E-06	12	32902712	T	C	0.01527	-0.08365	0.017537
rs7209951	1.80E-06	17	30234986	A	T	0.025293	0.063022	0.013185
rs141731602	1.90E-06	5	13197576	C	A	0.018996	0.075355	0.015804
rs753381	2.00E-06	20	39797465	C	T	0.551892	0.019663	0.004133
rs11263465	2.50E-06	11	69157360	G	A	0.135905	0.028473	0.006047
rs145989179	2.60E-06	3	1.22E+08	T	G	0.031867	-0.05557	0.011827
rs1248666	2.80E-06	10	79594504	C	T	0.665213	0.02051	0.004381
rs111608960	2.90E-06	3	87410009	C	A	0.083414	0.039122	0.008369
rs55747707	2.90E-06	7	73037366	A	G	0.20359	-0.02402	0.005137
rs987951	3.00E-06	5	76221198	A	G	0.29408	0.021081	0.004516
rs330095	3.00E-06	8	9172718	G	A	0.243162	0.022779	0.004877
rs58158339	3.00E-06	1	16596542	C	G	0.339736	0.020328	0.004354
rs12435428	3.10E-06	14	61684338	T	C	0.066603	-0.03992	0.008562
rs79687284	3.30E-06	1	2.14E+08	C	G	0.034655	0.052533	0.011297
rs10169825	3.60E-06	2	1.58E+08	G	T	0.059417	-0.04085	0.008812
rs117975574	3.60E-06	8	15459047	C	G	0.023533	-0.06488	0.013994
rs8100204	3.60E-06	19	19393714	A	G	0.145255	0.027806	0.005999
rs2434068	3.70E-06	10	83974607	T	C	0.042297	-0.04764	0.01029
rs9311400	3.90E-06	3	4764171	G	A	0.482504	-0.01909	0.004137
rs8081539	4.10E-06	17	73792372	T	G	0.646348	-0.01986	0.004309
rs9892018	4.30E-06	17	47958908	A	G	0.884595	0.029988	0.006521
rs1426962	4.40E-06	3	1.41E+08	T	C	0.146581	0.027033	0.005887
rs1345901	4.40E-06	16	79728049	C	T	0.353167	0.019883	0.004331
rs75460349	4.60E-06	1	27180088	C	A	0.023281	-0.06352	0.013867
rs1447449	5.60E-06	3	63103449	A	G	0.611523	0.019225	0.004234
rs17690703	5.80E-06	17	43925297	T	C	0.261781	-0.02125	0.004688
rs143205225	6.10E-06	9	1.27E+08	G	C	0.007604	-0.11021	0.024371

rs2729363	6.10E-06	11	57226854	C	A	0.777121	0.022666	0.005011
rs5402	6.30E-06	3	1.71E+08	A	T	0.121579	0.028685	0.006353
rs138050285	6.40E-06	3	50122660	C	T	0.195954	-0.0241	0.005342
rs1861179	6.80E-06	12	4303181	C	G	0.378705	0.019166	0.004259
rs2053701	6.80E-06	15	1.02E+08	T	C	0.321116	-0.01994	0.00443
rs56291006	7.10E-06	2	16197214	A	G	0.211357	-0.02279	0.005075
rs1626109	7.40E-06	21	17605238	C	T	0.928978	0.036179	0.008073
rs12811045	7.60E-06	12	1.16E+08	G	A	0.216024	0.022617	0.005052
rs7034933	7.70E-06	9	25896367	G	A	0.057573	-0.0399	0.008919
rs2424641	8.30E-06	20	24717867	T	G	0.371342	-0.01925	0.00432
rs199659388	8.40E-06	16	29000806	C	G	0.012059	0.091531	0.020543
rs73198247	9.30E-06	7	85002315	A	C	0.031956	-0.05215	0.011765
rs76089176	9.60E-06	17	17995728	G	C	0.08507	0.033009	0.007459
rs112811496	9.80E-06	2	2.4E+08	T	C	0.008408	-0.10829	0.024487
rs1660613	9.90E-06	10	34567543	C	T	0.5786	0.01877	0.004246
rs1886690	1.00E-05	1	39285553	T	G	0.749055	0.02123	0.004813
rs75539574	1.00E-05	2	58871658	C	A	0.085125	-0.03284	0.007444
rs11078716	1.00E-05	17	7820183	T	G	0.407461	-0.01859	0.004216

Chr: Chromosome, BP: Position, EA: Effect Allele, NEA: Non-Effect Allele, EAF: Effect Allele Frequency, SE: Standard Error

Table S9. Identified variants associated with serum valine in the current UKB GWAS

SNP	p-value	Chr	BP	EA	NEA	EAF	Beta	SE
rs10018448	2.30E-148	4	89225171	G	A	0.535321	0.1046	0.004032
rs2422358	2.90E-47	2	65231806	T	A	0.435721	0.060813	0.004212
rs12325419	1.10E-46	16	70368909	A	G	0.115339	-0.09018	0.006286
rs117643180	6.30E-43	17	7185779	A	C	0.025862	-0.17484	0.01273
rs1260326	6.50E-36	2	27730940	C	T	0.603978	-0.05136	0.004105
rs117048185	8.70E-35	19	49309776	C	G	0.017258	0.189858	0.015431
rs35230038	1.10E-23	19	49300431	A	G	0.0454	-0.09842	0.009811
rs2127864	1.30E-19	4	89147993	G	A	0.175269	-0.04812	0.005312
rs1128249	1.50E-17	2	1.66E+08	T	G	0.39201	-0.03508	0.004115
rs2638315	2.20E-16	12	56865056	C	G	0.182454	0.042605	0.00519
rs36181536	1.30E-13	12	1.23E+08	C	T	0.673521	-0.03179	0.004294
rs17096421	1.40E-12	10	88820592	T	A	0.056393	0.062542	0.008831
rs8071084	3.10E-09	17	79634162	G	T	0.161732	-0.03234	0.005457
rs34894639	6.00E-09	3	1.36E+08	T	C	0.23158	-0.02795	0.004805
rs2943652	1.20E-08	2	2.27E+08	T	C	0.644806	0.023866	0.004189
rs837616	1.60E-08	19	49365588	G	A	0.291527	-0.02504	0.00443
rs2274815	1.80E-08	14	1.03E+08	A	G	0.205044	0.02804	0.004982
rs61587941	3.40E-08	5	55814890	A	G	0.051235	-0.05018	0.009094
rs6941263	3.40E-08	6	1.09E+08	A	T	0.187845	-0.0285	0.005164
rs12974412	4.80E-08	19	14151809	G	A	0.779952	-0.02725	0.00499
rs66921136	7.10E-08	19	7197880	C	T	0.71211	-0.02399	0.004452
rs2394861	1.10E-07	10	74078627	C	T	0.657615	0.02245	0.00423
rs2977929	1.20E-07	8	76454025	T	C	0.196778	0.026874	0.005078
rs17817288	2.20E-07	16	53807764	G	A	0.484862	0.020808	0.004017
rs4660293	2.60E-07	1	40028180	G	A	0.23342	0.024434	0.004747
rs6953476	4.90E-07	7	98921600	C	G	0.127532	-0.03036	0.006034

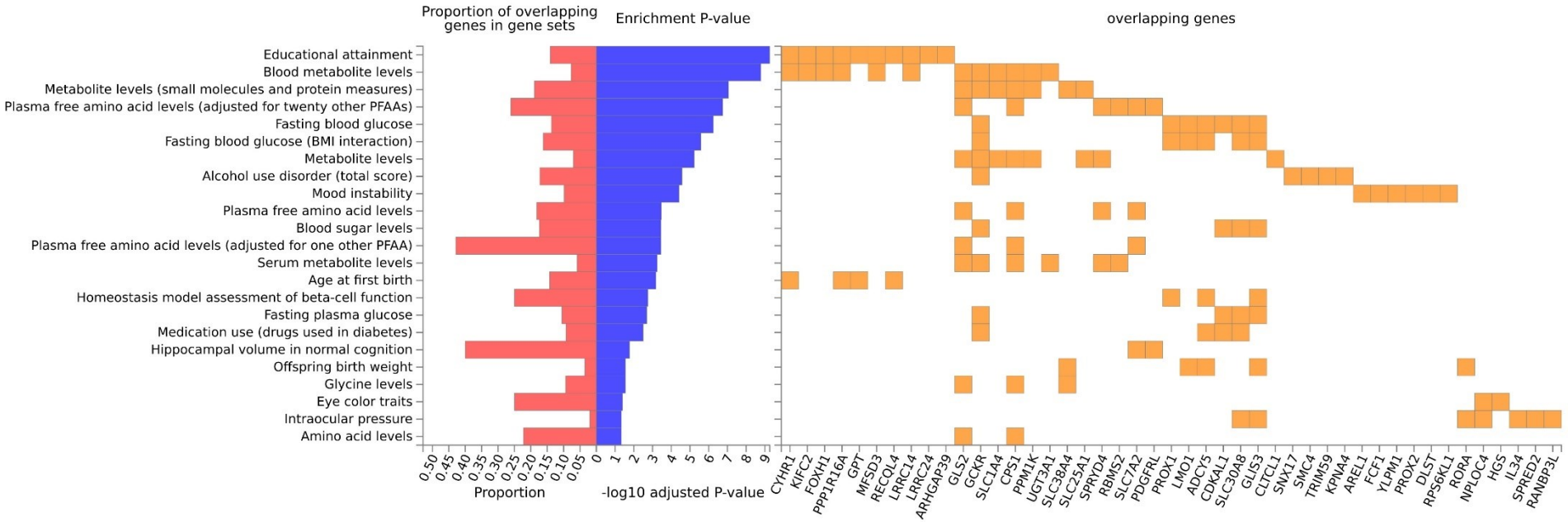
rs71624137	5.20E-07	5	55859146	C	T	0.12388	0.030606	0.006099
rs9941239	5.50E-07	16	84987947	G	A	0.756127	-0.02352	0.004696
rs12174589	6.50E-07	6	1.61E+08	T	A	0.206571	0.025059	0.005036
rs12318565	7.30E-07	12	92736142	C	A	0.348111	0.02246	0.004534
rs13402475	8.00E-07	2	3639909	G	C	0.81597	0.026067	0.005281
rs74823953	8.00E-07	5	1.51E+08	T	C	0.013274	-0.08815	0.01786
rs6427304	8.10E-07	1	1.56E+08	G	A	0.354924	0.020875	0.004231
rs2497349	8.20E-07	10	94521565	T	G	0.307901	0.021494	0.00436
rs7650482	1.10E-06	3	12841804	G	A	0.65107	-0.02057	0.004216
rs1028883	1.20E-06	13	74108587	G	T	0.572183	-0.01995	0.004101
rs12936169	1.20E-06	17	40824823	A	G	0.28637	-0.02167	0.00446
rs41294810	1.30E-06	1	55063953	G	T	0.260522	-0.02209	0.004565
rs3112484	1.30E-06	5	90401221	A	G	0.095616	-0.03316	0.006854
rs222856	1.30E-06	17	7110413	A	G	0.019961	-0.07144	0.014776
rs67555182	1.30E-06	21	26686210	A	G	0.106294	0.031575	0.006533
rs2616272	1.40E-06	11	97448214	A	G	0.495911	-0.01948	0.00404
rs117164273	1.40E-06	16	68292616	A	G	0.045436	-0.04664	0.009669
rs13100474	1.60E-06	3	1.22E+08	A	G	0.205784	0.024035	0.005004
rs28435138	1.60E-06	13	20694866	A	G	0.326663	-0.02083	0.004338
rs75663391	1.60E-06	15	63910433	T	G	0.679805	-0.02086	0.004344
rs7649917	1.70E-06	3	1.93E+08	C	T	0.227404	0.022967	0.004799
rs76895963	1.70E-06	12	4384844	G	T	0.020553	-0.07478	0.015622
rs1333051	1.80E-06	9	22136489	T	A	0.136465	-0.02792	0.00585
rs911268	1.80E-06	11	30809653	C	T	0.330666	0.020529	0.0043
rs2168101	1.90E-06	11	8255408	A	C	0.308191	-0.02137	0.004489
rs1682628	2.10E-06	7	95283613	A	G	0.642147	-0.01994	0.004209
rs62088469	2.10E-06	17	73519589	C	T	0.166001	-0.02559	0.005393
rs75041599	2.20E-06	6	1.44E+08	A	G	0.016686	0.076662	0.016201
rs579526	2.40E-06	12	373120	C	T	0.231453	0.022486	0.004771
rs10469763	2.80E-06	2	2.12E+08	C	T	0.450525	0.019304	0.004118
rs71562509	3.00E-06	6	1.4E+08	T	G	0.593337	-0.01913	0.004099
rs56001710	3.00E-06	7	25983400	T	A	0.580558	-0.01961	0.004196

rs62271373	3.20E-06	3	1.5E+08	A	T	0.059955	0.040127	0.008621
rs11045819	3.50E-06	12	21329813	A	C	0.161285	0.025352	0.005464
rs35509506	3.90E-06	6	44595179	T	C	0.134882	0.027531	0.00596
rs6501179	3.90E-06	16	3610642	C	T	0.439763	0.018731	0.004059
rs7659144	4.10E-06	4	3098321	G	C	0.344452	-0.01952	0.004238
rs140900031	4.10E-06	11	18489528	A	G	0.478583	0.019054	0.004135
rs1438506	4.50E-06	10	1.33E+08	A	T	0.352751	-0.01931	0.004209
rs875908	5.00E-06	14	23907608	G	C	0.641537	-0.01912	0.004191
rs115111747	5.70E-06	1	2139916	T	C	0.021426	0.062798	0.013841
rs72713107	5.80E-06	15	45871720	G	A	0.127227	-0.02738	0.006038
rs9984420	5.90E-06	21	45199112	A	T	0.111883	0.028983	0.006396
rs73547946	6.00E-06	9	1.33E+08	A	T	0.252538	-0.02091	0.004621
rs77475232	6.10E-06	12	1.25E+08	T	C	0.111075	0.029385	0.006498
rs11169172	6.20E-06	12	50232169	T	A	0.053964	-0.04013	0.008881
rs17018345	6.80E-06	4	92244300	T	C	0.277603	-0.02033	0.004518
rs12219397	7.00E-06	10	5146578	C	T	0.217674	0.021786	0.00485
rs200375975	7.10E-06	11	70483181	T	C	0.179632	-0.02566	0.005714
rs73121772	7.50E-06	12	66424265	T	C	0.092192	-0.03113	0.006949
rs2920900	7.70E-06	2	55280083	C	T	0.395114	0.018443	0.004122
rs7308448	7.70E-06	12	1.08E+08	T	G	0.214549	0.021885	0.004892
rs589625	7.90E-06	18	524173	C	G	0.812024	0.023132	0.005179
rs12093903	8.00E-06	1	2.27E+08	A	C	0.302259	-0.022	0.004927
rs2181083	8.20E-06	14	51293706	T	C	0.186916	-0.02303	0.005163
rs4766578	8.60E-06	12	1.12E+08	A	T	0.503247	-0.0179	0.004023
rs112875651	8.80E-06	8	1.27E+08	A	G	0.392344	-0.01854	0.004172
rs150817269	9.10E-06	12	3803291	T	C	0.00865	0.100107	0.022557
rs147302846	9.20E-06	3	58356116	A	C	0.025324	-0.05791	0.013054
rs116687155	9.40E-06	4	1.45E+08	T	C	0.024082	-0.0583	0.013159
rs142570322	9.50E-06	20	34003073	C	T	0.597185	-0.01862	0.004204
rs12916085	9.80E-06	15	79474419	C	T	0.773213	0.022628	0.005117
rs9809727	9.90E-06	3	24759872	A	G	0.115973	0.028395	0.006425
rs139683582	1.00E-05	2	57893329	A	G	0.017845	-0.06742	0.015287

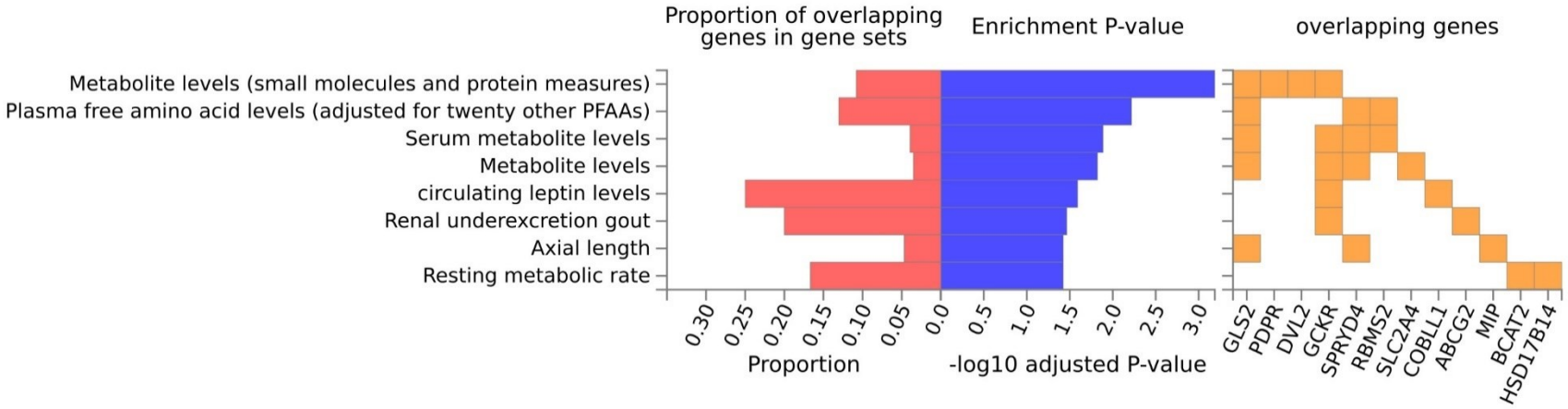
rs635634	1.00E-05	9	1.36E+08	T	C	0.183563	0.022956	0.005206

Chr: Chromosome, BP: Position, EA: Effect Allele, NEA: Non-Effect Allele, EAF: Effect Allele Frequency, SE: Standard Error

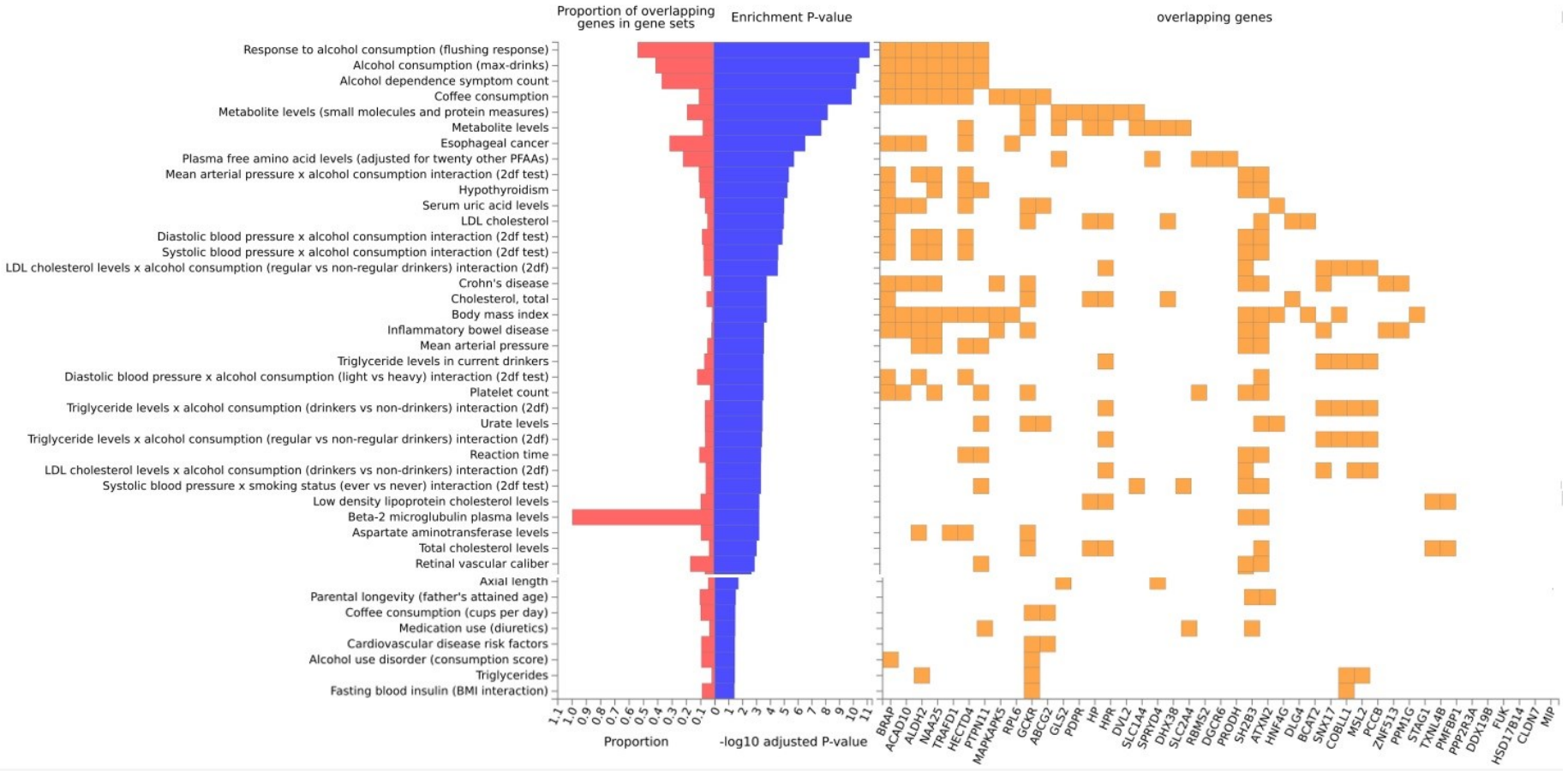
FigureS1. Gene-set analysis for serum alanine



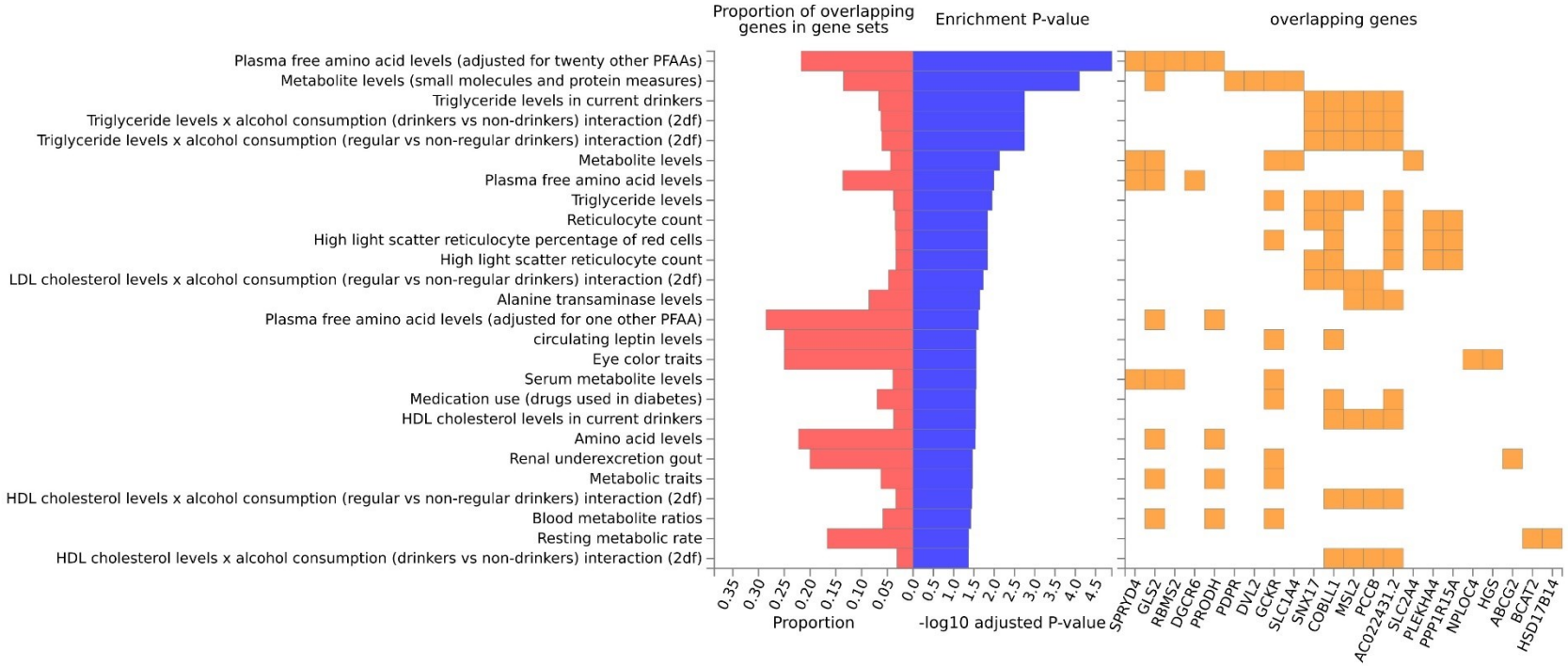
FigureS2. Gene-set analysis for serum isoleucine



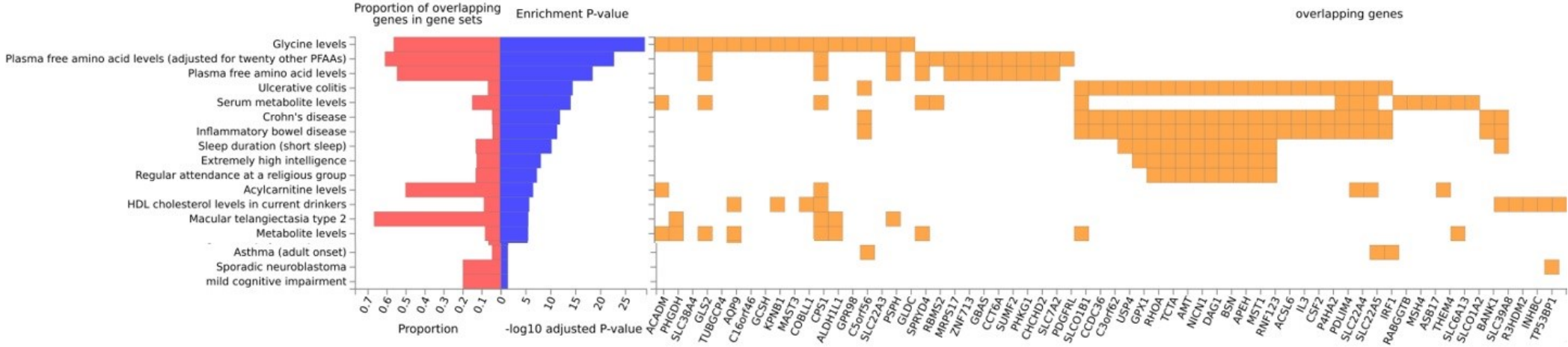
FigureS3. Gene-set analysis for serum leucine



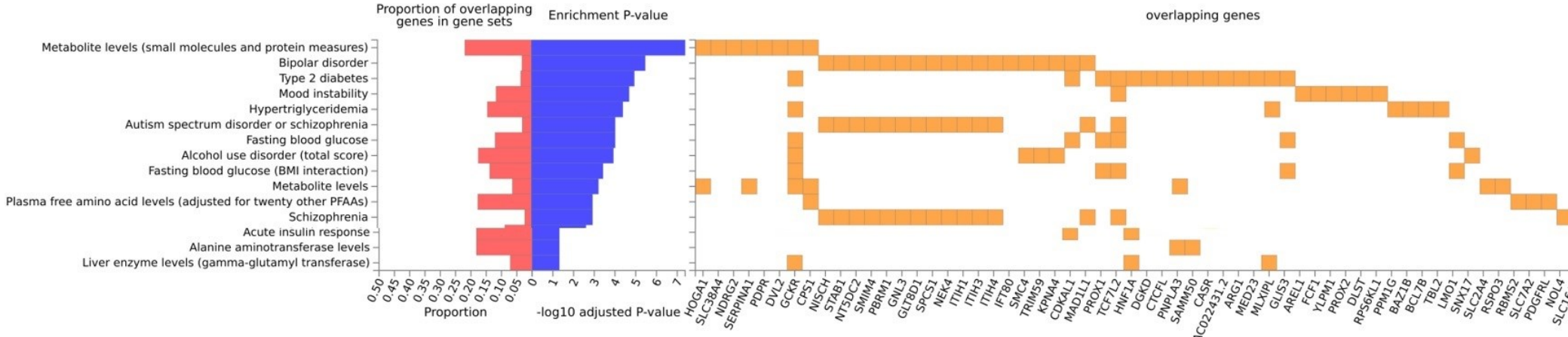
FigureS4. Gene-set analysis for serum valine



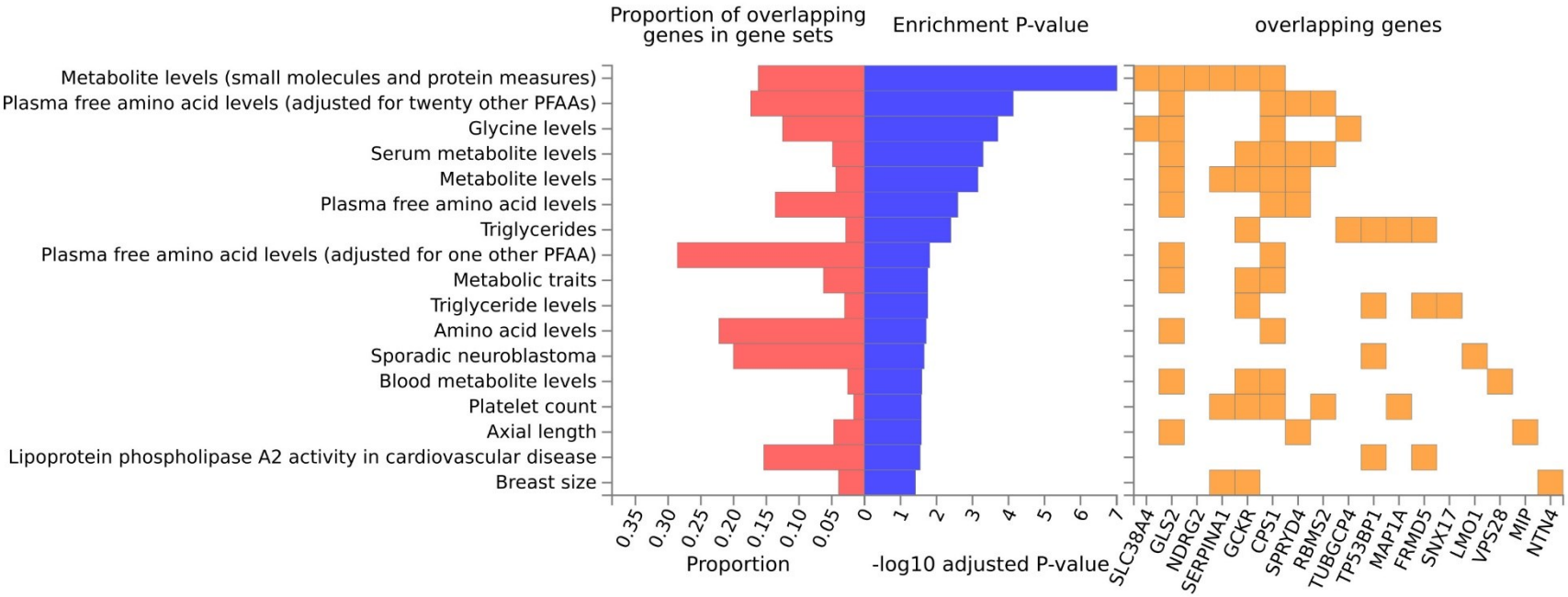
FigureS5. Gene-set analysis for serum glycine



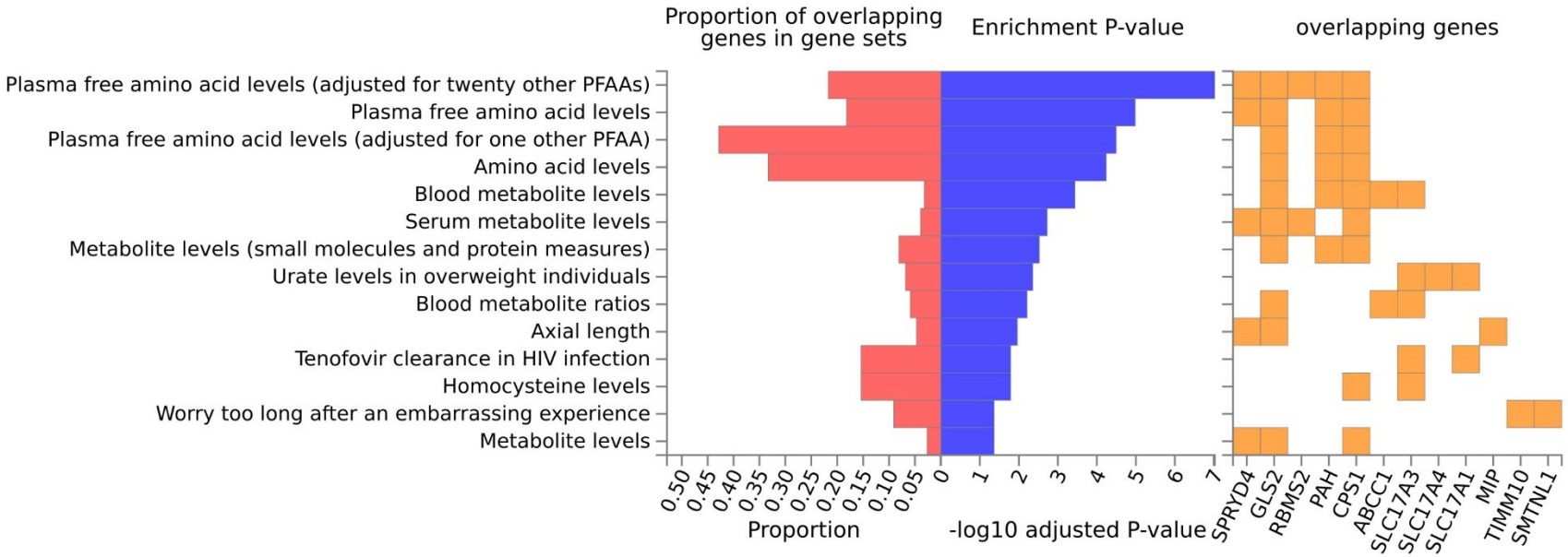
FigureS6. Gene-set analysis for serum glutamine



FigureS7. Serum histidine



FigureS8. Serum phenylalanine



FigureS9. Gene-set analysis for serum tyrosine

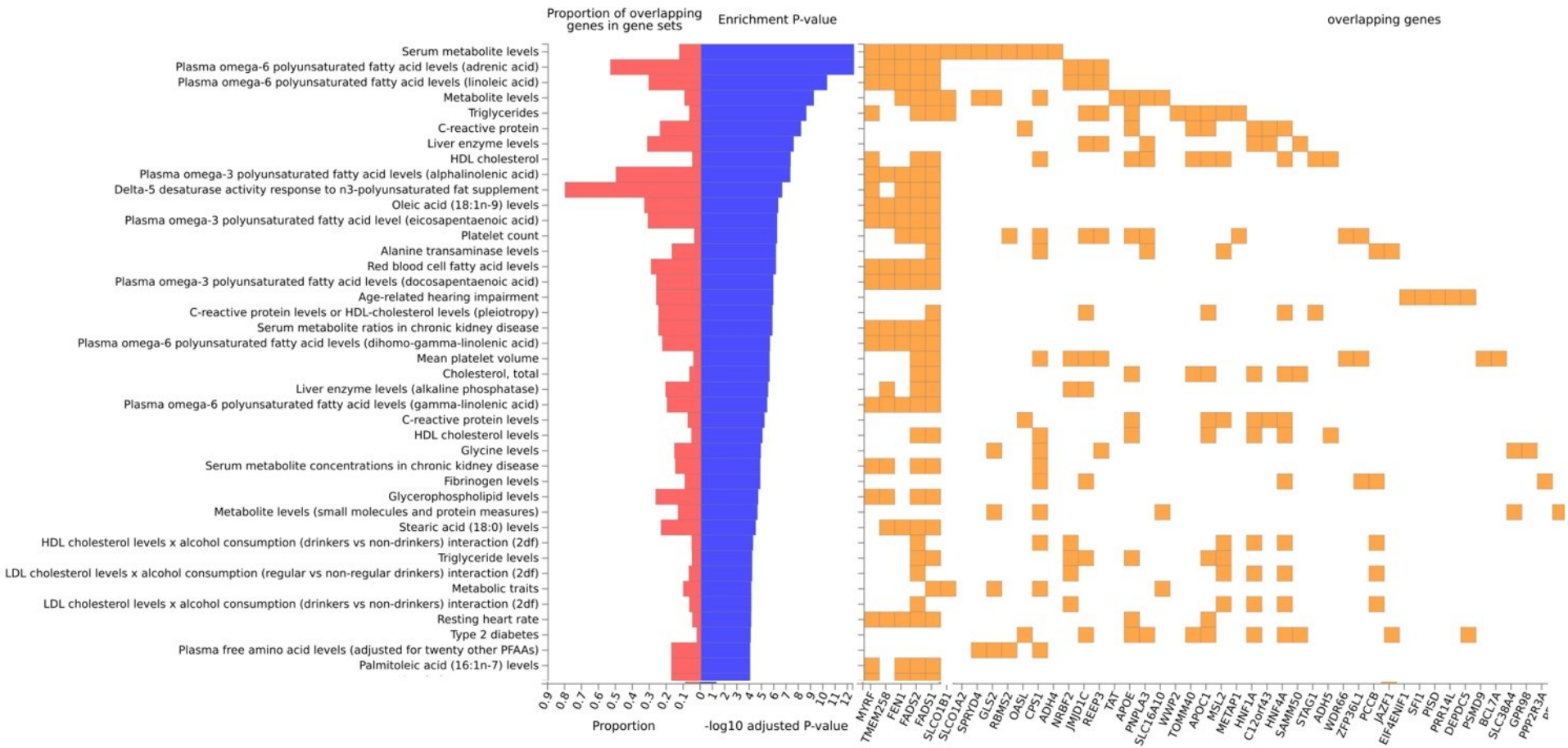


Table S10. The effect of macronutrient intakes on serum amino acids

			Carbohydrate intake			Fat intake			Protein intake		
	N	**Models**	**beta**	**SE**	**P-value**	**beta**	**SE**	**P-value**	**beta**	**SE**	**P-value**
Alanine	108 341	unadjusted	0.0003	0.04	**< 0.001**	0.0004	0.02	**< 0.001**	0.0002	7.97e-03	**0.01**
		Model 1	0.0001	0.0001	**0.003**	-0.0002	-0.0002	**0.001**	-0.0006	-6.05e-04	**< 0.001**
	14 191	Model2	0.0001	0.00003	**< 0.001**	-0.00004	0.00008	**< 0.001**	-0.0001	0.0001	**0.07**
Isoleucine	108 342	unadjusted	0.0002	0.03	**< 0.001**	0.001	0.05	**<0.001**	0.002	0.06	**< 0.001**
		Model 1	-0.00007	-0.007	**0.03**	-0.0002	-0.0002	**0.001**	0.0007	0.02	**< 0.001**
	14 191	Model2	-0.00003	0.00003	**0.24**	0.0001	0.00008	**0.13**	0.00007	0.0001	**0.50**
Leucine	108 342	unadjusted	0.0004	0.04	**< 0.001**	0.001	0.06	**< 0.001**	0.003	0.09	**< 0.001**
		Model 1	-0.0001	-0.01	**0.001**	-0.0002	-0.0002	**0.001**	0.001	0.03	**< 0.001**
	14 191	Model2	-0.00007	0.00003	**0.02**	0.0002	0.00008	**0.003**	0.0005	0.0001	**< 0.001**
Valine	108 326	unadjusted	0.0002	0.02	**< 0.001**	0.00103	0.04	**< 0.001**	0.003	0.10	**< 0.001**
		Model 1	-0.0001	-0.0001	**< 0.001**	-0.0002	-0.0002	**0.001**	0.001	1.71e-03	**< 0.001**
	14 189	Model2	-0.00005	0.0002	**0.08**	0.00003	0.00007	**0.96**	0.001	0.00009	**< 0.001**
Glycine	108 253	unadjusted	-0.0001	-0.01	**< 0.001**	-0.0007	-0.03	**< 0.001**	-0.002	-0.09	**< 0.001**
		Model 1	0.0003	0.04	**< 0.001**	0.0007	0.03	**< 0.001**	-0.001	-0.04	**< 0.001**
	14 183	Model2	0.0004	0.00003	**< 0.001**	0.0007	0.00008	**< 0.001**	-0.001	0.0001	**< 0.001**
Glutamine	108 009	unadjusted	0.0005	0.05	**< 0.001**	0.00007	0.002	0.40	-0.0006	-0.02	**< 0.001**
		Model 1	0.0004	0.04	**< 0.001**	-0.0002	-0.01	**0.003**	-0.001	-0.03	**< 0.001**
	14 010	Model2	0.0004	0.00003	**< 0.001**	-0.0003	0.00009	**< 0.001**	-0.0001	0.0001	**< 0.001**
Histidine	108 146	unadjusted	0.0001	0.02	**< 0.001**	0.0002	0.007	**0.01**	0.0005	0.02	**< 0.001**
		Model 1	-0.000006	-0.0006	0.85	-0.0003	< 0.001	**-0.01**	0.00007	2.27e-03	0.51
	14 010	Model2	-0.00001	0.00003	0.80	-0.0002	0.00009	**0.01**	-0.0001	0.0001	< 0.001
Phenylalanine	108 276	unadjusted	0.0003	0.03	**< 0.001**	0.0005	0.02	**< 0.001**	0.001	0.04	**< 0.001**
		Model 1	0.0001	0.01	**< 0.001**	-0.0001	-0.004	0.18	0.0007	0.02	**< 0.001**
	14 132	Model2	0.00007	0.00003	**0.03**	-0.0001	0.00009	0.16	-0.0003	0.0001	**0.005**

Tyrosine	108 162	unadjusted	0.0001	0.01	**< 0.001**	0.0002	0.008	**0.007**	0.001	0.06	**< 0.001**
		Model 1	0.00002	0.002	0.53	-0.0004	-0.01	**< 0.001**	0.001	0.04	**< 0.001**
	14 132	Model2	0.00001	0.00003	0.003	-0.0002	0.00009	**0.005**	0.0006	0.0001	**< 0.001**

N is number of participants

Model 1: adjusted for age, sex, weight, and total energy intake. Model 2: adjusted for age, sex, weight, total energy intake, physical activity, and body fat percentage.

Table S11. The effect of accelerometer-measured physical activity on serum amino acids

	Number of participants	Models	beta	SE	P-value
Alanine	24 182	unadjusted	-0.005	0.0007	2.32e-15
		Model 1	-0.004	0.0007	< 2.0e-16
		Model2	-0.006	0.0009	< 0.001
Isoleucine	24 182	unadjusted	-0.008	0.0007	< 2.2e-16
		Model 1	-0.006	0.0007	< 2.2e-16
		Model2	-0.006	0.0009	< 0.001
Leucine	24 182	unadjusted	-0.006	0.0007	< 2.2e-16
		Model 1	-0.005	0.0007	< 2.2e-16
		Model2	-0.003	0.0008	< 0.001
Valine	24 179	unadjusted	-0.01	0.0007	< 2.2e-16
		Model 1	-0.008	0.0007	< 2.2e-16
		Model2	-0.007	0.0008	< 0.001
Glycine	24 162	unadjusted	0.008	0.0007	< 2.2e-16
		Model 1	0.005	0.0007	< 2.2e-16
		Model2	-0.001	0.0008	0.01
Glutamine	24 126	unadjusted	0.004	0.0007	8.72e-10
		Model 1	0.006	0.0007	< 2.2e-16
		Model2	-0.0007	0.0009	0.04
Histidine	24 142	unadjusted	0.002	0.0007	0.002
		Model 1	0.001	0.0007	< 2.2e-16
		Model2	-0.003	0.0009	0.001
Phenylalanine	24 173	unadjusted	-0.002	0.0007	0.0003
		Model 1	-0.001	0.0007	< 2.2e-16
		Model2	-0.003	0.0009	0.003
Tyrosine	24 144	unadjusted	-0.006	0.0007	< 2.2e-16
		Model 1	-0.002	0.0007	< 2.2e-16
		Model2	-0.002	0.0009	0.008

Model 1: adjusted for age and sex. Model 2: adjusted for age, sex, total energy intake and body fat percentage.

Table S12. The effect of body fat percentage on serum amino acids

	Number of participants	Models	beta	SE	P-value
Alanine	117 977	unadjusted	0.0007	0.0003	0.06
		Model 1	0.01	0.12	<0 .001
	14 191	Model2	0.01	0.10	<0.01
Isoleucine	117 978	unadjusted	-0.003	-0.03	<0 .001
		Model 1	0.03	0.03	< 0.001
	14 191	Model 2	0.03	0.03	<0.01
Leucine	117 977	unadjusted	-0.008	-0.08	< 0.001
		Model 1	0.03	0.23	< 0.001
	14 191	Model 2	0.03	0.23	<0.01
Valine	117 950	unadjusted	0.001	0.0003	<0 .001
		Model 1	0.03	0.0004	<0 .001
	14 189	Model 2	0.03	0.0003	<0.01
Glycine	117 874	unadjusted	0.008	0.07	< 0.001
		Model 1	-0.03	-0.25	< 0.001
	14 183	Model 2	-0.03	-0.23	<0.01
Glutamine	117 625	unadjusted	-0.02	-0.16	<0.001
		Model 1	-0.03	-0.22	<0.001
	14 010	Model 2	-0.03	-0.20	<0.01
Histidine	117 763	unadjusted	-0.01	-0.10	<0.001
		Model 1	-0.004	-0.04	< 0.001
	14 010	Model 2	-0.004	-0.05	<0.01
Phenylalanine	117 904	unadjusted	0.004	0.04	< 0.001
		Model 1	0.02	0.14	<0.001
	14 132	Model 2	0.02	0.14	< 0.01
Tyrosine	117 781	unadjusted	0.01	0.12	<0.001
		Model 1	0.03	0.23	< 0.001
	14 132	Model 2	0.03	0.21	< 0.01

Model 1: adjusted for age and sex were included as covariates. Model 2: adjusted for age, sex, total energy intake and physical activity.

Table S13. MR estimates for associations of percentage of energy from protein intake with serum amino acids

Outcome	Method	N SNPs	OR	SE	95% Upper	95% Lower	P-value	FDR-corrected P-value	Heterogeneity P-value
Serum alanine									
	IVW	3	0.92	0.06	0.82	1.03	0.16	0.39	**0.01**
	Weighted median		0.90	0.04	0.83	0.97	**0.03**		
	MR Egger		1.03	0.58	-0.11	2.18	0.97		
	Weighted mode		0.87	0.08	0.71	1.03	0.23		
	Simple mode		0.87	0.06	0.76	0.98	0.13		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum glutamine		3							0.21
	IVW		0.95	0.03	0.88	1.01	0.11	0.37	
	Weighted median		0.96	0.03	0.9	1.02	0.21		
	MR Egger		1.32	0.2	0.93	1.72	0.39		
	Weighted mode		0.97	0.04	0.89	1.05	0.51		
	Simple mode		0.96	0.05	0.87	1.05	0.47		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum glycine	IVW	3	0.97	0.03	0.91	1.02	0.23	0.46	0.25
	Weighted median		0.95	0.03	0.88	1.01	0.09		
	MR Egger		1.12	0.28	0.58	1.66	0.75		
	Weighted mode		0.93	0.05	0.83	1.03	0.3		

	Simple mode		0.93	0.04	0.85	1.02	0.27		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum histidine		3							0.66
	IVW		1.01	0.03	0.96	1.07	0.61	0.79	
			1	0.03	0.94	1.06	0.97		
	Weighted median		1.11	0.2	0.71	1.5	0.7		
	MR Egger		0.99	0.04	0.92	1.07	0.88		
	Weighted mode		0.99	0.04	0.92	1.07	0.9		
	Simple mode		1.01	0.03	0.96	1.07	0.61		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum isoleucine		3							0.70
	IVW		1.16	0.03	1.11	1.21	**1.61E-08**	**3.62E-07**	
	Weighted median		1.16	0.04	1.09	1.23	**2.31E-05**		
	Weighted mode		1.33	0.2	0.95	1.72	0.38		
	Simple mode		1.17	0.04	1.09	1.25	0.07		
	MR Egger		1.17	0.04	1.08	1.25	0.07		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum leucine	IVW	3	1.15	0.03	1.09	1.21	**1.24E-05**	**<0.001**	0.32
	Weighted median		1.14	0.04	1.06	1.21	**<0.001**		
	MR Egger		1.52	0.19	1.15	1.9	0.27		
	Weighted mode		1.13	0.05	1.03	1.22	0.13		

	Simple mode		1.12	0.05	1.02	1.21	0.15		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum phenylalanine	IVW	3	1.08	0.04	1.01	1.16	0.03	0.20	0.12
	Weighted median		1.05	0.04	0.98	1.13	0.14		
	MR Egger		1.11	0.41	0.31	1.91	0.84		
	Simple mode		1.05	0.05	0.94	1.15	0.48		
	Weighted mode		1.04	0.05	0.95	1.13	0.49		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum tyrosine	IVW	3	1.08	0.06	0.96	1.21	0.19	0.42	0.45
	Weighted median		1.04	0.04	0.96	1.12	0.34		
	MR Egger		1.11	0.65	-0.18	2.39	0.9		
	Simple mode		1.03	0.06	0.92	1.13	0.69		
	Weighted mode		1.01	0.04	0.93	1.10	0.78		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum valine	IVW	3	1.20	0.03	1.15	1.25	**1.41E-12**	**6.36E-11**	0.63
	Weighted median		1.21	0.04	1.14	1.28	**3.32E-07**		
	MR Egger		1.45	0.19	1.07	1.83	0.31		
	Weighted mode		1.22	0.04	1.13	1.3	**0.05**		
	Simple mode		1.22	0.05	1.12	1.31	**0.05**		

	MR-PRESSO	***NA***	***NA***	***NA***	***NA***	***NA***	

Table S14. MR estimates for associations of percentage of energy from carbohydrate intake with serum amino acids

Outcome	Method	N SNPs	OR	SE	95% Upper	95% Lower	P-value	FDR-corrected P-value	Heterogeneity P-value
Serum alanine									
	IVW	3	1.04	0.02	1.00	1.08	0.03	0.19	0.06
	Weighted median		1.05	0.02	1.02	1.08	**0.001**		
	MR Egger		1.20	0.06	1.08	1.33	0.21		
	Weighted mode		1.06	0.02	1.02	1.09	0.09		
	Simple mode		1.06	0.02	1.02	1.09	0.11		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum glutamine									
	IVW	3	1.02	0.01	1.00	1.05	0.10	0.30	0.23
	Weighted median		1.03	0.01	1.00	1.05	0.08		
	MR Egger		0.96	0.09	0.79	1.13	0.71		
	Weighted mode		1	0.02	0.97	1.04	0.83		
	Simple mode		1.04	0.02	0.99	1.09	0.24		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum glycine									
	IVW	3	1.01	0.01	0.98	1.04	0.36	0.59	0.19
	Weighted median		1.01	0.01	0.98	1.04	0.41		
	MR Egger		0.91	0.06	0.79	1.03	0.37		

	Weighted mode		1	0.01	0.98	1.03	0.79		
	Simple mode		1.03	0.02	0.98	1.07	0.36		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum histidine									
	IVW	3	0.99	0.01	0.97	1.02	0.55	0.73	0.69
	Weighted median		0.99	0.01	0.97	1.02	0.69		
	MR Egger		1.00	0.02	0.97	1.04	0.90		
	Weighted mode		0.95	0.06	0.82	1.07	0.55		
	Simple mode		1.00	0.02	0.97	1.04	0.87		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum isoleucine									
	IVW	3	0.94	0.02	0.91	0.97	**<0.001**	**<0.01**	0.13
	Weighted median		0.95	0.02	0.91	0.98	**<0.001**		
	MR Egger		1.02	0.09	0.84	1.2	0.87		
	Weighted mode		0.95	0.02	0.91	0.98	0.10		
	Simple mode		0.95	0.02	0.91	0.99	0.12		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum leucine									
	IVW	3	0.94	0.02	0.91	0.98	**<0.001**	**<0.01**	0.11

	Weighted median		0.94	0.02	0.91	0.97	**<0.001**		
	MR Egger		0.94	0.13	0.68	1.19	0.70		
	Weighted mode		0.92	0.03	0.87	0.98	0.12		
	Simple mode		0.92	0.03	0.87	0.98	0.09		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum phenylalanine									
	IVW	3	0.97	0.02	0.93	1.01	0.16	0.38	**0.03**
	Weighted median		0.98	0.01	0.95	1.01	0.23		
	MR Egger		1.14	0.06	1.01	1.26	0.30		
	Weighted mode		0.98	0.01	0.96	1.01	0.37		
	Simple mode		0.98	0.02	0.95	1.01	0.40		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum tyrosine									
	IVW	3	0.97	0.03	0.91	1.04	0.38	0.59	**<0.001**
	Weighted median		0.99	0.01	0.96	1.02	0.55		
	MR Egger		1.23	0.06	1.1	1.36	0.19		
	Weighted mode		0.99	0.01	0.97	1.02	0.72		
	Simple mode		0.99	0.02	0.96	1.03	0.67		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum valine									

	IVW	3	0.92	0.02	0.89	0.96	**8.43E-06**	**<0.001**	0.09
	Weighted median		0.93	0.02	0.90	0.96	**4.39E-05**		
	MR Egger		0.99	0.12	0.75	1.22	0.92		
	Weighted mode		0.93	0.02	0.89	0.98	0.09		
	Simple mode		0.92	0.02	0.87	0.97	0.08		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		

Table S15. MR estimates for associations of percentage of energy from fat intake with serum amino acids

Outcome	Method	N SNPs	OR	SE	95% Upper	95% Lower	P-value	FDR-corrected P-value	Heterogeneity P-value
Serum alanine									
	IVW	3	0.94	0.03	0.88	1.00	0.03	0.18	0.07
	Weighted median		0.94	0.03	0.88	0.99	0.01		
	MR Egger		0.9	0.11	0.67	1.12	0.51		
	Weighted mode		0.92	0.03	0.87	0.98	0.1		
	Simple mode		0.91	0.03	0.85	0.98	0.12		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum glutamine									
	IVW	3	0.97	0.03	0.92	1.02	0.28	0.66	0.11
	Weighted median		0.98	0.02	0.94	1.02	0.37		
	MR Egger		1.08	0.05	0.97	1.18	0.41		
	Weighted mode		0.99	0.02	0.95	1.03	0.71		
	Simple mode		0.95	0.04	0.87	1.03	0.35		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum glycine									
	IVW	3	0.98	0.02	0.94	1.03	0.51	0.69	0.14
	Weighted median		0.99	0.02	0.95	1.03	0.55		
	MR Egger		1.08	0.05	0.98	1.18	0.38		

	Weighted mode		1.01	0.02	0.97	1.05	0.77		
	Simple mode		0.94	0.04	0.85	1.03	0.31		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum histidine									
	IVW	3	1.01	0.02	0.98	1.05	0.47	0.69	0.75
	Weighted median		1.01	0.02	0.97	1.05	0.55		
	MR Egger		1.05	0.05	0.94	1.16	0.52		
	Weighted mode		1.02	0.02	0.98	1.07	0.41		
	Simple mode		0.99	0.03	0.94	1.05	0.86		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum isoleucine									
	IVW	3	1.10	0.02	1.06	1.15	**2.50E-05**	**<0.001**	0.19
	Weighted median		1.09	0.03	1.02	1.15	**0.01**		
	MR Egger		1.04	0.07	0.89	1.18	0.71		
	Weighted mode		1.08	0.02	1.04	1.13	0.08		
	Simple mode		1.09	0.04	1.02	1.16	0.14		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum leucine									
	IVW	3	1.10	0.02	1.07	1.13	**2.98E-08**	**6.71E-07**	0.57

	Weighted median		1.10	0.03	1.05	1.15	**<0.001**		
	MR Egger		1.12	0.05	1.02	1.22	0.28		
	Weighted mode		1.11	0.03	1.05	1.16	0.07		
	Simple mode		1.11	0.03	1.05	1.18	0.08		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum phenylalanine									
	IVW	3	1.05	0.03	0.98	1.11	0.20	0.53	**0.02**
	Weighted median		1.03	0.03	0.98	1.07	0.33		
	MR Egger		0.94	0.1	0.75	1.13	0.64		
	Weighted mode		1.02	0.02	0.98	1.06	0.47		
	Simple mode		1.03	0.04	0.96	1.1	0.49		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum tyrosine									
	IVW	3	1.04	0.05	0.94	1.14	0.44	0.69	**<0.001**
	Weighted median		1.01	0.02	0.97	1.05	0.68		
	MR Egger		0.89	0.14	0.62	1.17	0.57		
	Weighted mode		1.00	0.02	0.96	1.05	0.88		
	Simple mode		1.02	0.04	0.94	1.1	0.66		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		
Serum valine									

	IVW	3	1.13	0.02	1.09	1.17	**5.57E-09**	**2.51E-07**	0.24
	Weighted median		1.11	0.04	1.04	1.19	**<0.01**		
	MR Egger		1.07	0.07	0.93	1.21	0.50		
	Weighted mode		1.11	0.03	1.06	1.17	0.06		
	Simple mode		1.11	0.04	1.04	1.19	0.10		
	MR-PRESSO		***NA***	***NA***	***NA***	***NA***	***NA***		

Table S16. MR estimates for associations of PA measured by wrist-worn accelerometer with serum amino acids

Outcome	Method	N SNPs	OR	SE	95% Upper	95% Lower	P-value	FDR-corrected P-value	Heterogeneity P-value
Accelerometery – Overall acceleration average									
Serum alanine	IVW	2	1.00	0.01	0.97	1.02	0.99	0.99	0.40
Serum glutamine	IVW	2	1.00	0.02	0.96	1.04	0.95	0.99	0.09
Serum glycine	IVW	2	1.02	0.05	0.92	1.13	0.63	0.84	**4.20e-05**
Serum histidine	IVW	2	1.01	0.03	0.95	1.07	0.65	0.84	0.02
Serum isoleucine	IVW	2	1.02	0.02	0.98	1.05	0.31	0.55	0.13
Serum leucine	IVW	2	1.02	0.01	1.00	1.05	0.06	0.28	0.67
Serum phenylalanine	IVW	2	1.02	0.01	0.99	1.04	0.13	0.29	0.51
Serum tyrosine	IVW	2	1.04	0.03	0.99	1.09	0.13	0.29	0.03
Serum valine	IVW	2	1.05	0.01	1.02	1.07	**<0.001**	**<0.01**	0.26

Table S17. MR estimates for associations of body fat percentage with serum amino acids

Outcome	Method	N SNPs	OR	SE	95% Upper	95% Lower	P-value	FDR-corrected P-value	Heterogeneity P-value
Serum alanine	IVW	11	0.91	0.09	0.73	1.09	0.31	0.42	**5.17e-08**
	Weighted median		1.09	0.06	0.97	1.2	0.17		
	MR-Egger		1.85	0.29	1.28	2.42	0.06		
	Weighted mode		1.11	0.06	1.00	1.22	0.10		
	Simple mode		1.13	0.08	0.98	1.28	0.15		
	MR-PRESSO		1.00	0.07	0.86	1.14	1.00	-	
Serum glutamine	IVW	11	0.92	0.08	0.77	1.07	0.30	0.42	**3.72e-05**
	Weighted median		0.89	0.06	0.78	1.00	**0.04**		
	MR-Egger		0.87	0.32	0.24	1.5	0.68		
	Weighted mode		0.89	0.06	0.77	1.01	0.08		
	Simple mode		0.90	0.08	0.74	1.06	0.25		
	MR-PRESSO		0.90	0.04	0.82	0.97	**0.02**		
Serum glycine	IVW	11	0.95	0.11	0.72	1.17	0.63	0.74	**2.36e-15**
	Weighted median		0.88	0.06	0.76	1.01	**0.05**		
	MR-Egger		0.62	0.45	-0.26	1.50	0.31		
	Weighted mode		0.86	0.06	0.74	0.98	**0.04**		

	Simple mode		0.85	0.08	0.69	1.01	0.07		
	MR-PRESSO		0.87	0.05	0.77	0.96	**0.02**		
Serum histidine	IVW	11	0.96	0.06	0.85	1.07	0.49	0.61	**0.03**
	Weighted median		0.98	0.06	0.87	1.09	0.72		
	MR-Egger		1.23	0.22	0.8	1.66	0.37		
	Weighted mode		0.99	0.07	0.85	1.13	0.87		
	Simple mode		0.96	0.1	0.77	1.15	0.69		
	MR-PRESSO		0.99	0.05	0.89	1.09	0.83		
Serum isoleucine	IVW	11	1.01	0.12	0.76	1.25	0.96	0.96	**5.77e-17**
	Weighted median		1.15	0.07	1.01	1.3	**0.04**		
	MR-Egger		2.75	0.38	2.02	3.49	**0.02**		
	Weighted mode		1.23	0.06	1.11	1.35	**0.01**		
	Simple mode		1.14	0.1	0.93	1.34	0.24		
	MR-PRESSO		1.00	0.09	0.82	1.18	0.98		
Serum leucine	IVW	11	0.98	0.13	0.73	1.24	0.91	0.93	
	Weighted median		1.19	0.06	1.06	1.32	**0.01**		
	MR-Egger		2.60	0.42	1.76	3.43	**0.05**		
	Weighted mode		1.24	0.06	1.12	1.37	**0.01**		
	Simple mode		1.24	0.09	1.06	1.42	0.04		

	MR-PRESSO		0.96	0.09	0.79	1.14	0.70		
Serum phenylalanine	IVW	11	1.12	0.06	1	1.24	0.07	0.15	**1.88e-20**
	Weighted median		1.12	0.06	0.99	1.24	0.08		
	MR-Egger		1.59	0.23	1.15	2.03	0.07		
	Weighted mode		1.17	0.09	0.99	1.35	0.12		
	Simple mode		1.1	0.12	0.87	1.33	0.43		
	MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*		
Serum tyrosine	IVW	11	1.19	0.09	1.01	1.37	**0.05**	0.14	**<0.01**
	Weighted median		1.16	0.08	1.01	1.32	**0.05**		
	MR-Egger		2.63	0.26	2.12	3.14	**<0.01**		
	Weighted mode		1.34	0.11	1.12	1.56	**0.03**		
	Simple mode		1.10	0.18	0.74	1.46	0.61		
	MR-PRESSO		1.14	0.08	0.99	1.30	0.15		
Serum valine	IVW	11	1.02	0.15	0.72	1.31	0.91	0.93	**6.55e-27**
	Weighted median		1.25	0.07	1.11	1.38	**<0.01**		
	MR-Egger		2.96	0.50	1.99	3.94	0.06		
	Weighted mode		1.31	0.06	1.19	1.43	**<0.01**		
	Simple mode		1.28	0.08	1.11	1.44	**0.02**		
	MR-PRESSO		1.15	0.07	1.01	1.29	0.11		

Table S18. Two-sample MR results for genetically predicted serum amino acids and risk of type 2 diabetes

Outcome	Exposure	Method	nSNPs	OR	SE	Lower CI	Upper CI	Pvalue	p.FDR
Type 2 diabetes	Alan	IVW	60	1.50	0.16	1.18	1.81	**0.01**	0.06
		MR Egger		0.45	0.41	-0.36	1.27	0.06	
		Weighted median		0.99	0.07	0.85	1.12	0.83	
		Weighted mode		1.02	0.07	0.88	1.16	0.78	
		Simple mode		1.02	0.11	0.8	1.24	0.86	
		MR-PRESSO		1.13	0.06	1.00	1.27	0.07	
	Glutamine	IVW	65	1.14	0.14	0.87	1.41	0.34	0.73
		MR Egger		1.57	0.31	0.97	2.17	0.14	
		Weighted median		1.04	0.07	0.91	1.16	0.60	
		Simple mode		1.03	0.12	0.78	1.27	0.84	
		Weighted mode		1.00	0.08	0.85	1.16	0.99	
		MR-PRESSO		1.05	0.05	0.95	1.15	0.35	
	Glycine	IVW	52	0.95	0.08	0.80	1.11	0.53	0.94
		MR Egger		1.15	0.12	0.92	1.37	0.23	
		Weighted median		1.00	0.05	0.90	1.09	0.99	
		Simple mode		0.95	0.09	0.77	1.13	0.58	
		Weighted mode		1.00	0.03	0.94	1.07	0.89	
		MR-PRESSO		0.99	0.08	0.92	1.08	0.26	
	Histidine	IVW	29	1.22	0.12	0.99	1.46	0.09	0.33
		Weighted median		0.96	0.09	0.80	1.13	0.66	
		MR Egger		0.97	0.26	0.46	1.49	0.91	
		Simple mode		0.95	0.15	0.66	1.25	0.76	
		Weighted mode		0.99	0.09	0.81	1.17	0.90	
		MR-PRESSO		1.08	0.08	0.92	1.24	0.33	
	Isoleucine	IVW	28	1.23	0.20	0.83	1.63	0.32	0.71
		MR Egger		0.61	0.48	-0.32	1.54	0.31	

		Weighted median		1.20	0.10	1.00	1.40	0.08	
		Simple mode		1.00	0.16	0.69	1.31	0.98	
		Weighted mode		1.32	0.10	1.13	1.5	**<0.01**	
		MR-PRESSO		1.24	0.07	1.07	1.43	**<0.01**	
	Leucine	IVW		1.19	0.13	0.93	1.46	0.18	0.58
		MR Egger	36	0.99	0.29	0.43	1.55	0.97	
		Weighted median		1.35	0.08	1.19	1.52	**<0.001**	
		Simple mode		1.19	0.16	0.88	1.5	0.27	
		Weighted mode		1.33	0.08	1.17	1.5	**0.001**	
		MR-PRESSO		1.29	0.07	1.30	1.47	**<0.001**	
	Phenylalanine	IVW	30	1.13	0.12	0.9	1.36	0.30	0.71
		MR Egger		0.84	0.21	0.43	1.25	0.42	
		Weighted median		1.03	0.06	0.91	1.16	0.60	
		Weighted mode		1.06	0.06	0.94	1.18	0.37	
		Simple mode		1.12	0.13	0.86	1.38	0.39	
		MR-PRESSO		1.05	0.06	0.93	1.17	0.42	
	Tyrosine	IVW	59	0.99	0.09	0.82	1.15	0.87	0.99
		Weighted median		0.83	0.05	0.72	0.93	**<0.001**	
		MR Egger		0.78	0.15	0.49	1.07	0.10	
		Weighted mode		0.85	0.05	0.74	0.96	**<0.01**	
		Simple mode		0.97	0.12	0.73	1.22	0.82	
		MR-PRESSO		0.91	0.06	0.79	1.02	0.14	
	Valine	IVW	50	1.52	0.16	1.21	1.82	**<0.01**	**0.04**
		MR Egger		0.84	0.3	0.25	1.43	0.57	
		Weighted median		1.30	0.07	1.16	1.44	**<0.001**	
		Simple mode		1.02	0.11	0.8	1.25	0.83	
		Weighted mode		1.23	0.06	1.11	1.35	**0.001**	
		MR-PRESSO		1.28	0.06	1.13	1.43	**<0.001**	

Table S19. Two-sample MR results for genetically predicted serum amino acids and lipid levels

Outcome	Exposure	Method	nSNPs	OR	SE	Lower CI	Upper CI	Pval	p.FDR
TG	Alan	IVW	35	1.59	0.14	1.32	1.87	**< 0.001**	**<0.01**
		MR Egger		3.73	0.33	3.07	4.38	**< 0.001**	
		Weighted median		1.08	0.05	0.98	1.17	0.12	
		Simple mode		1.03	0.08	0.87	1.19	0.70	
		Weighted mode		1.07	0.07	0.94	1.21	0.30	
		MR-PRESSO		1.20	0.05	1.08	1.35	**< 0.001**	
TC		IVW		1.27	0.07	1.12	1.41	**0.001**	**<0.01**
		MR Egger		2.06	0.17	1.72	2.39	**0.0001**	
		Weighted median		1.13	0.05	1.04	1.23	**< 0.01**	
		Simple mode		1.14	0.08	0.98	1.30	0.13	
		Weighted mode		1.11	0.06	1.00	1.23	0.08	
		MR-PRESSO		1.09	0.04	1.01	1.19	**0.04**	
LDL		IVW		1.13	0.05	1.03	1.23	**0.01**	**0.03**
		Weighted median		1.15	0.05	1.04	1.25	**0.01**	
		MR Egger		1.39	0.13	1.14	1.63	**0.01**	
		Simple mode		1.08	0.11	0.86	1.29	0.50	
		Weighted mode		1.19	0.08	1.04	1.33	**0.03**	
		MR-PRESSO		1.06	0.04	0.97	1.15	0.19	
HDL		IVW		0.95	0.06	0.84	1.07	0.41	0.48
		MR Egger		1.00	0.15	0.70	1.30	0.99	
		Weighted median		0.95	0.05	0.86	1.04	0.23	
		Simple mode		0.93	0.08	0.78	1.08	0.33	
		Weighted mode		0.92	0.06	0.8	1.03	0.14	
		MR-PRESSO		0.97	0.03	0.90	1.04	0.36	
TG	Glutamine	IVW	42	0.66	0.11	0.45	0.86	**<0.001**	**0.001**
		MR Egger		0.50	0.20	0.10	0.89	**0.001**	
		Weighted median		1.11	0.04	1.03	1.19	**0.01**	

		Simple mode		1.14	0.05	1.03	1.24	**0.02**	
		Weighted mode		1.13	0.04	1.06	1.20	**0.001**	
		MR-PRESSO		0.79	0.07	0.66	0.92	**0.001**	
TC		IVW	43	0.88	0.06	0.76	1.01	**0.05**	0.13
		MR Egger		0.73	0.12	0.49	0.97	**0.01**	
		Weighted median		0.99	0.05	0.9	1.09	0.90	
		Simple mode		1.09	0.07	0.95	1.23	0.25	
		Weighted mode		1.05	0.04	0.96	1.14	0.27	
		MR-PRESSO		0.99	0.04	0.91	1.07	0.89	
LDL		IVW	42	0.96	0.05	0.86	1.06	0.43	0.57
		MR Egger		0.90	0.10	0.70	1.09	0.29	
		Weighted median		0.97	0.05	0.87	1.07	0.60	
		Simple mode		1.17	0.09	0.99	1.36	0.09	
		Weighted mode		0.96	0.06	0.83	1.08	0.48	
		MR-PRESSO		1.01	0.04	0.94	1.09	0.76	
HDL		IVW	43	1.14	0.05	1.03	1.24	**0.01**	**0.05**
		MR Egger		1.08	0.10	0.87	1.28	0.47	
		Weighted median		1.06	0.04	0.99	1.14	0.11	
		Simple mode		1.03	0.06	0.91	1.16	0.63	
		Weighted mode		1.06	0.03	0.99	1.12	0.10	
		MR-PRESSO		1.07	0.03	1.01	1.14	**0.03**	
TG	Glycine	IVW	33	1.01	0.05	0.91	1.10	0.89	0.10
		MR Egger		1.05	0.07	0.92	1.19	0.46	
		Weighted median		1.00	0.03	0.94	1.06	0.96	
		Simple mode		1.01	0.08	0.85	1.17	0.87	
		Weighted mode		0.99	0.03	0.95	1.04	0.83	
		MR-PRESSO		1.02	0.03	0.96	1.07	0.54	
TC		IVW		0.91	0.05	0.81	1.01	0.07	0.25
		MR Egger		0.85	0.08	0.70	1.00	**0.04**	
		Weighted median		1.00	0.03	0.94	1.05	0.90	
		Simple mode		1.03	0.08	0.87	1.18	0.74	

		Weighted mode		1.00	0.02	0.95	1.04	0.89	
		MR-PRESSO		0.98	0.04	0.9	1.05	0.57	
		IVW		0.93	0.05	0.84	1.01	0.09	0.26
		MR Egger		0.88	0.07	0.75	1.01	**0.05**	
		Weighted median		1.00	0.03	0.94	1.06	0.10	
		Simple mode		1.03	0.07	0.9	1.16	0.68	
LDL		Weighted mode		1.00	0.02	0.95	1.05	0.88	
		MR-PRESSO		0.99	0.04	0.92	1.06	0.79	
		IVW		0.91	0.07	0.77	1.04	0.16	0.40
		MR Egger		0.82	0.1	0.62	1.01	**0.05**	
		Weighted median		0.96	0.03	0.91	1.01	0.07	
		Simple mode		0.95	0.06	0.83	1.07	0.37	
HDL		Weighted mode		0.96	0.02	0.92	1.00	0.07	
		MR-PRESSO		0.96	0.03	0.91	1.01	0.15	
	Histidine	IVW	15	0.66	0.33	0.02	1.3	0.20	0.42
		MR Egger		0.67	0.8	-0.89	2.23	0.62	
		Weighted median		1.01	0.05	0.91	1.11	0.87	
		Simple mode		1.21	0.11	0.99	1.43	0.10	
TG		Weighted mode		1.01	0.05	0.91	1.11	0.81	
		MR-PRESSO		0.91	0.09	0.73	1.08	0.31	
		IVW		0.92	0.19	0.55	1.28	0.64	0.80
		MR Egger		0.78	0.45	-0.1	1.66	0.60	
		Weighted median		1.07	0.05	0.96	1.17	0.23	
		Simple mode		1.11	0.12	0.88	1.33	0.40	
TC		Weighted mode		1.07	0.05	0.98	1.17	0.17	
		MR-PRESSO		0.92	0.09	0.74	1.09	0.36	
		IVW		0.96	0.14	0.68	1.23	0.74	0.87
		MR Egger		0.67	0.32	0.04	1.3	0.23	
		Weighted median		0.9	0.06	0.78	1.01	0.07	
LDL		Simple mode		0.98	0.11	0.75	1.2	0.84	
		Weighted mode		0.9	0.06	0.78	1.03	0.13	

		MR-PRESSO		0.9	0.04	0.83	0.97	0.02	
		IVW		1.27	0.07	1.14	1.41	**<0.001**	**<0.01**
		MR Egger		1.75	0.14	1.48	2.01	**0.001**	
		Weighted median		1.39	0.05	1.29	1.5	**1.13E-09**	
		Simple mode		1.14	0.14	0.87	1.41	0.35	
HDL		Weighted mode		1.43	0.05	1.33	1.54	**8.68E-06**	
		MR-PRESSO		1.25	0.05	1.15	1.35	**0.001**	
TG	Isoleucine	IVW	18	2.47	0.26	1.97	2.97	**<0.001**	**<0.01**
		MR Egger		4.02	0.67	2.7	5.33	**0.05**	
		Weighted median		1.26	0.08	1.09	1.42	**<0.01**	
		Simple mode		1.14	0.13	0.89	1.4	0.32	
		Weighted mode		1.10	0.09	0.92	1.27	0.32	
		MR-PRESSO		1.83	0.10	1.51	2.22	**<0.001**	
TC		IVW		1.62	0.11	1.4	1.84	**2.04E-05**	**<0.001**
		MR Egger		2.57	0.28	2.03	3.12	**<0.01**	
		Weighted median		1.27	0.08	1.12	1.42	**0.001**	
		Simple mode		1.4	0.14	1.13	1.67	**0.02**	
		Weighted mode		1.26	0.09	1.1	1.43	**0.01**	
		MR-PRESSO		1.37	0.08	1.18	1.59	**<0.001**	
LDL		IVW		1.32	0.09	1.15	1.49	**0.001**	**<0.01**
		MR Egger		1.63	0.23	1.19	2.08	**0.05**	
		Weighted median		1.18	0.08	1.03	1.33	**0.03**	
		Simple mode		1.08	0.13	0.83	1.34	0.55	
		Weighted mode		1.19	0.09	1.01	1.37	0.07	
		MR-PRESSO		1.24	0.06	1.10	1.39	**<0.01**	
HDL		IVW		0.82	0.08	0.65	0.98	**0.01**	**0.03**
		MR Egger		1.08	0.21	0.68	1.49	0.70	
		Weighted median		0.83	0.06	0.71	0.95	**<0.01**	

		Simple mode		0.82	0.11	0.61	1.02	0.07	
		Weighted mode		0.81	0.07	0.68	0.95	**<0.01**	
		MR-PRESSO		0.83	0.05	0.75	0.92	**<0.01**	
TG	Leucine	IVW	23	8.73	0.51	7.73	9.74	**<0.001**	**0.001**
		MR Egger		1.36	0.08	1.19	1.52	**<0.001**	
		Weighted median		2.71	0.2	2.32	3.09	**4.02E-07**	
		Simple mode		1.09	0.13	0.84	1.35	0.50	
		Weighted mode		1.16	0.13	0.9	1.42	0.26	
		MR-PRESSO		1.95	0.10	**1.58**	2.40	**1.32E-05**	
TC		IVW		3.82	0.33	3.18	4.46	**<0.001**	**0.002**
		MR Egger		1.21	0.08	1.04	1.37	**0.02**	
		Weighted median		1.41	0.14	1.15	1.68	**0.01**	
		Simple mode		1.11	0.14	0.83	1.38	0.48	
		Weighted mode		1.13	0.17	0.8	1.46	0.48	
		MR-PRESSO		1.34	0.07	1.17	1.53	**<0.001**	
LDL		IVW		2.09	0.26	1.57	2.61	**0.01**	**0.02**
		MR Egger		1.31	0.08	1.16	1.47	**<0.001**	
		Weighted median		1.18	0.1	0.98	1.38	0.10	
		Simple mode		1.10	0.17	0.76	1.44	0.59	
		Weighted mode		1.49	0.09	1.31	1.66	**<0.001**	
		MR-PRESSO		1.30	0.07	0.97	1.31	0.12	
HDL		IVW		0.79	0.25	0.3	1.29	0.38	0.50
		MR Egger		0.81	0.06	0.68	0.93	**<0.001**	
		Weighted median		0.70	0.09	0.53	0.87	**3.17E-05**	
		Simple mode		1.08	0.11	0.86	1.30	0.50	
		Weighted mode		0.82	0.07	0.68	0.97	**0.01**	
		MR-PRESSO		0.85	0.07	0.74	0.96	**0.02**	
TG	Phenylalanine	IVW	15	0.98	0.07	0.84	1.11	0.73	0.84

		MR Egger		0.94	0.12	0.71	1.17	0.59	
		Weighted median		0.97	0.04	0.89	1.05	0.47	
		Simple mode		1.03	0.12	0.8	1.26	0.78	
		Weighted mode		0.97	0.04	0.89	1.05	0.51	
		MR-PRESSO		0.98	0.0487	0.88	1.07	**0.04**	
TC		IVW		0.92	0.06	0.81	1.03	0.14	0.45
		MR Egger		0.95	0.1	0.76	1.13	0.58	
		Weighted median		0.91	0.04	0.82	0.99	**0.02**	
		Simple mode		0.93	0.1	0.74	1.12	0.45	
		Weighted mode		0.91	0.04	0.82	1	**0.05**	
		MR-PRESSO		0.94	0.0409	0.86	1.02	0.14	
LDL		IVW		0.92	0.06	0.8	1.03	0.13	0.45
		MR Egger		0.95	0.1	0.76	1.13	0.57	
		Weighted median		0.91	0.04	0.82	0.99	**0.02**	
		Simple mode		0.97	0.1	0.77	1.18	0.80	
		Weighted mode		0.91	0.05	0.82	1.00	**0.05**	
		MR-PRESSO		0.93	0.04	0.85	1.02	0.13	
HDL		IVW		0.99	0.07	0.86	1.13	0.88	0.88
		MR Egger		0.93	0.11	0.71	1.16	0.55	
		Weighted median		0.97	0.04	0.89	1.06	0.54	
		Simple mode		1.26	0.17	0.93	1.59	0.19	
		Weighted mode		0.94	0.04	0.86	1.02	0.16	
		MR-PRESSO		1.17	0.06	1.05	1.29	**0.02**	
TG	Tyrosine	IVW	45	1.12	0.09	0.95	1.3	0.18	0.85
		MR Egger		0.90	0.23	0.45	1.36	0.66	
		Weighted median		0.97	0.05	0.86	1.07	0.55	
		Simple mode		0.96	0.14	0.69	1.23	0.75	
		Weighted mode		0.92	0.13	0.68	1.17	0.52	
		MR-PRESSO		1.07	0.04	0.99	1.15	0.09	
TC		IVW		1.05	0.11	0.84	1.26	0.64	0.91
		MR Egger		0.97	0.28	0.42	1.52	0.92	

		Weighted median		1.04	0.06	0.93	1.15	0.50	
		Simple mode		1.18	0.1	1	1.37	0.08	
		Weighted mode		1.1	0.07	0.97	1.24	0.15	
		MR-PRESSO		1.08	0.03	1.02	1.15	0.02	
		IVW		1.02	0.11	0.81	1.24	0.83	0.92
		MR Egger		0.9	0.29	0.34	1.47	0.72	
		Weighted median		1.06	0.05	0.95	1.16	0.29	
		Simple mode		1.06	0.09	0.88	1.24	0.55	
LDL		Weighted mode		1.09	0.07	0.96	1.22	0.21	
		MR-PRESSO		1.02	0.04	0.93	1.1	0.71	
		IVW		0.99	0.08	0.83	1.15	0.91	0.92
		MR Egger		1.26	0.21	0.85	1.68	0.27	
		Weighted median		0.99	0.05	0.9	1.08	0.78	
		Simple mode		0.92	0.08	0.77	1.07	0.30	
HDL		Weighted mode		0.98	0.05	0.88	1.07	0.64	
		MR-PRESSO		1.05	0.03	0.99	1.1	0.13	
TG	Valine	IVW	28	2.12	0.17	1.8	2.45	**6.08E-06**	**<0.001**
		MR Egger		3.76	0.39	2.98	4.53	**<0.01**	
		Weighted median		1.13	0.08	0.97	1.28	0.13	
		Simple mode		1.09	0.2	0.71	1.48	0.65	
		Weighted mode		1.00	0.07	0.86	1.15	0.96	
		MR-PRESSO		1.70	0.08	1.45	1.98	**9.42E-07**	
TC		IVW		1.45	0.12	1.22	1.69	**<0.01**	**0.01**
		MR Egger		2.33	0.29	1.77	2.89	**<0.01**	
		Weighted median		1.18	0.07	1.05	1.31	**0.02**	
		Simple mode		1.25	0.11	1.04	1.46	**0.04**	
		Weighted mode		1.17	0.08	1.02	1.32	**0.04**	
		MR-PRESSO		1.25	0.05	1.14	1.37	**9.54E-05**	

LDL		IVW		1.28	0.11	1.07	1.49	**0.02**	**0.05**
		MR Egger		1.53	0.27	1.01	2.06	0.12	
		Weighted median		1.15	0.06	1.03	1.28	**0.02**	
		Simple mode		1.06	0.12	0.83	1.29	0.63	
		Weighted mode		1.21	0.08	1.05	1.37	**0.03**	
		MR-PRESSO		1.21	0.04	1.10	1.32	**<0.001**	
HDL		IVW		0.76	0.1	0.57	0.95	**<0.01**	**0.02**
		MR Egger		1.06	0.23	0.6	1.51	0.81	
		Weighted median		0.81	0.06	0.69	0.93	**<0.001**	
		Simple mode		0.89	0.11	0.68	1.1	0.28	
		Weighted mode		0.87	0.08	0.71	1.03	0.09	
		MR-PRESSO		0.82	0.05	0.73	0.91	**<0.01**	

Table S20. Two-sample MR results for genetically predicted serum amino acids and risk of high SBP

Outcome	Exposure	Method	nSNPs	OR	SE	Lower CI	Upper CI	Pval	p.FDR
SBP	Alan	IVW	87	2.50	0.30	1.38	4.55	**<0.01**	**0.03**
		MR Egger		2.48	0.78	0.54	11.52	0.25	
		Weighted median		2.22	0.34	1.15	4.29	**0.02**	
		Simple mode		1.87	0.72	0.45	7.68	0.39	
		Weighted mode		1.96	0.56	0.66	5.87	0.23	
		MR-PRESSO		2.50	0.27	1.97	3.02	**0.001**	
	Glutamine	IVW	111	1.33	0.31	0.72	1.94	0.36	0.73
		MR Egger		1.59	0.67	0.28	2.9	0.49	
		Weighted median		1.23	0.37	0.51	1.96	0.57	
		Simple mode		4.54	1.00	2.58	6.51	0.13	
		Weighted mode		0.48	0.64	-0.78	1.74	0.26	
		MR-PRESSO		1.39	0.28	0.84	1.94	0.24	
	Glycine	IVW	89	0.79	0.29	0.23	1.35	0.42	0.73
		MR Egger		1.3	0.45	0.42	2.17	0.56	
		Weighted median		1.06	0.24	0.58	1.53	0.82	
		Simple mode		0.75	0.52	-0.26	1.76	0.58	
		Weighted mode		0.99	0.21	0.59	1.39	0.96	
		MR-PRESSO		0.82	0.20	0.43	1.21	0.33	
	Histidine	IVW	46	1.11	0.61	-0.09	2.30	0.87	0.90
		MR Egger		8.87	1.30	6.33	11.42	0.10	
		Weighted median		1.95	0.48	1.01	2.88	0.16	
		Simple mode		2.01	0.92	0.22	3.81	0.45	
		Weighted mode		3.05	0.56	1.96	4.14	0.05	
		MR-PRESSO		1.40	0.44	0.54	2.26	0.45	
	Isoleucine	IVW	48	3.98	0.42	1.76	8.99	**<0.001**	**0.02**
		MR Egger		4.20	0.97	0.63	28.03	0.15	
		Weighted median		2.15	0.49	0.82	5.64	0.12	

		Simple mode		0.77	1.11	0.09	6.75	0.82	
		Weighted mode		2.33	0.56	0.77	7.02	0.14	
		MR-PRESSO		3.35	0.40	2.57	4.13	**<0.01**	
	Leucine	IVW	56	4.54	0.59	1.42	14.58	**0.01**	**0.05**
		MR Egger		1.56	1.26	0.13	18.49	0.73	
		Weighted median		1.57	0.44	0.67	3.72	0.30	
		Simple mode		1.30	0.92	0.21	7.94	0.78	
		Weighted mode		1.72	0.46	0.69	4.25	0.25	
		MR-PRESSO		3.22	0.37	2.49	3.94	**<0.01**	
	Phenylalanine	IVW	36	1.26	0.39	0.49	2.04	0.55	0.73
		MR Egger		1.35	0.73	-0.07	2.78	0.68	
		Weighted median		1.79	0.38	1.05	2.54	0.12	
		Simple mode		0.17	1.11	-2.02	2.35	0.12	
		Weighted mode		1.71	0.38	0.97	2.46	0.17	
		MR-PRESSO		1.43	0.30	1.13	2.32	0.25	
	Tyrosine	IVW	85	1.78	0.32	1.16	2.41	0.07	0.52
		MR Egger		1.46	0.53	0.41	2.50	0.48	
		Weighted median		1.13	0.3	0.55	1.71	0.67	
		Simple mode		1.66	0.64	0.4	2.93	0.43	
		Weighted mode		1.44	0.28	0.9	1.98	0.19	
		MR-PRESSO		1.86	0.25	1.36	2.34	**0.01**	
	Valine	IVW	75	3.41	0.46	1.38	8.41	**0.01**	**0.05**
		MR Egger		1.47	0.87	0.26	8.13	0.66	
		Weighted median		1.55	0.35	0.78	3.07	0.21	
		Simple mode		0.74	0.9	0.13	4.38	0.75	
		Weighted mode		1.60	0.35	0.80	3.20	0.19	
		MR-PRESSO		2.33	0.32	1.70	2.96	**<0.01**	

Table S21. Two-sample MR results for genetically predicted serum amino acids and risk of high DBP

Outcome	Exposure	Method	nSNPs	OR	SE	Lower CI	Upper CI	Pval	p.FDR
DBP	Alan	IVW	87	1.92	0.22	1.49	2.35	**<0.01**	**0.04**
		MR Egger		1.69	0.57	0.58	2.80	0.36	
		Weighted median		1.89	0.22	1.46	2.32	**<0.01**	
		Simple mode		2.47	0.56	1.37	3.56	0.11	
		Weighted mode		1.62	0.36	0.93	2.32	0.18	
		MR-PRESSO		1.83	0.17	1.50	2.16	**<0.001**	
	Glutamine	IVW	112	1.36	0.2	0.97	1.75	0.12	0.50
		MR Egger		2.00	0.42	1.17	2.83	0.10	
		Weighted median		1.36	0.21	0.95	1.78	0.15	
		Simple mode		0.92	0.64	-0.33	2.18	0.90	
		Weighted mode		0.97	0.36	0.26	1.68	0.93	
		MR-PRESSO		1.09	0.17	0.76	1.42	0.60	
	Glycine	IVW	89	1.08	0.17	0.74	1.42	0.66	0.90
		MR Egger		1.28	0.27	0.74	1.82	0.37	
		Weighted median		0.97	0.13	0.72	1.22	0.84	
		Simple mode		1.12	0.28	0.58	1.66	0.69	
		Weighted mode		1.07	0.1	0.87	1.27	0.52	
		MR-PRESSO		0.97	0.12	0.73	1.21	0.83	
	Histidine	IVW	49	1.09	0.38	0.35	1.84	0.82	0.93
		MR Egger		3.79	0.8	2.23	5.35	0.10	
		Weighted median		1.63	0.27	1.10	2.16	0.07	
		Simple mode		1.72	0.5	0.74	2.69	0.28	
		Weighted mode		1.93	0.41	1.13	2.72	0.11	
		MR-PRESSO		0.96	0.21	0.55	1.37	0.86	
	Isoleucine	IVW	49	1.98	0.26	1.48	2.49	**<0.01**	**0.05**
		MR Egger		1.36	0.6	0.18	2.53	0.61	
		Weighted median		1.06	0.29	0.50	1.62	0.84	

		Simple mode		1.3	0.65	0.02	2.58	0.69	
		Weighted mode		1.1	0.36	0.40	1.80	0.79	
		MR-PRESSO		1.77	0.21	1.35	2.18	**0.01**	
	Leucine	IVW	56	2.29	0.41	1.49	3.08	**0.04**	0.20
		MR Egger		0.7	0.85	-0.95	2.36	0.68	
		Weighted median		0.98	0.25	0.48	1.48	0.94	
		Simple mode		0.88	0.54	-0.18	1.93	0.81	
		Weighted mode		0.98	0.28	0.43	1.53	0.95	
		MR-PRESSO		1.97	0.24	1.47	2.41	**<0.01**	
	Phenylalanine	IVW	36	1.03	0.22	0.6	1.45	0.91	0.93
		MR Egger		1.2	0.4	0.42	1.98	0.65	
		Weighted median		1.28	0.24	0.81	1.74	0.30	
		Simple mode		1.45	0.58	0.30	2.59	0.53	
		Weighted mode		1.36	0.22	0.93	1.78	0.17	
		MR-PRESSO		1.09	0.18	0.73	1.46	0.62	
	Tyrosine	IVW	86	1.42	0.22	0.98	1.86	0.12	0.50
		MR Egger		1.1	0.37	0.37	1.83	0.81	
		Weighted median		0.9	0.20	0.51	1.29	0.60	
		Simple mode		0.7	0.44	-0.16	1.55	0.41	
		Weighted mode		0.82	0.19	0.45	1.18	0.29	
		MR-PRESSO		1.57	0.17	1.23	1.90	**0.01**	
	Valine	IVW	75	1.84	0.31	1.23	2.46	**0.05**	0.20
		MR Egger		0.89	0.59	-0.27	2.04	0.84	
		Weighted median		0.98	0.21	0.58	1.39	0.93	
		Simple mode		1.21	0.41	0.41	2.00	0.65	
		Weighted mode		1.04	0.18	0.68	1.40	0.84	
		MR-PRESSO		1.41	0.17	1.09	1.75	**0.04**	

Table S22. Two-sample MR results for genetically predicted serum amino acids and risk of coronary artery disease

Outcome	Exposure	Method	nSNPs	OR	SE	Lower CI	Upper CI	Pval	p.FDR
CAD	Alan	IVW	98	1.11	0.05	1.01	1.20	**0.03**	0.25
		MR Egger		1.06	0.06	0.94	1.18	0.36	
		Weighted median		0.84	0.12	0.61	1.06	0.13	
		Simple mode		1.05	0.14	0.78	1.32	0.73	
		Weighted mode		0.98	0.10	0.78	1.18	0.83	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Glutamine	IVW	128	0.95	0.07	0.81	1.08	0.41	0.82
		MR Egger		0.98	0.13	0.73	1.22	0.84	
		Weighted median		1.04	0.07	0.9	1.19	0.55	
		Simple mode		0.82	0.16	0.51	1.13	0.20	
		Weighted mode		0.97	0.07	0.84	1.11	0.70	
		MR-PRESSO		0.97	0.04	0.89	1.05	0.44	
	Glycine	IVW	2	1.09	1.65	0.98	1.20	0.11	0.59
		MR Egger		*NA*	*NA*	*NA*	*NA*	*NA*	
		Weighted median		*NA*	*NA*	*NA*	*NA*	*NA*	
		Simple mode		*NA*	*NA*	*NA*	*NA*	*NA*	
		Weighted mode		*NA*	*NA*	*NA*	*NA*	*NA*	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Histidine	IVW	54	0.97	0.1	0.78	1.15	0.72	0.82
		MR Egger		0.97	0.06	0.84	1.09	0.58	
		Weighted median		1.25	0.16	0.93	1.57	0.18	
		Simple mode		0.95	0.07	0.81	1.09	0.45	
		Weighted mode		1.05	0.09	0.86	1.23	0.62	
		MR-PRESSO		1.08	0.07	0.93	1.23	0.31	
	Isoleucine	IVW	54	0.94	0.07	0.8	1.08	0.41	0.58
		MR Egger		0.69	0.16	0.37	1.00	**0.02**	

		Weighted median		0.93	0.09	0.75	1.12	0.47	
		Simple mode		0.77	0.20	0.38	1.15	0.19	
		Weighted mode		0.89	0.10	0.69	1.09	0.25	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Leucine	IVW	61	0.97	0.09	0.8	1.14	0.71	0.71
		MR Egger		0.83	0.18	0.48	1.17	0.29	
		Weighted median		0.94	0.08	0.78	1.1	0.45	
		Simple mode		0.9	0.15	0.61	1.19	0.49	
		Weighted mode		0.92	0.08	0.75	1.08	0.31	
		MR-PRESSO		0.92	0.06	0.80	1.04	0.16	
	Phenylalanine	IVW	42	1.38	0.20	0.99	1.77	0.11	0.59
		MR Egger		1.16	0.09	0.99	1.33	0.09	
		Weighted median		1.18	0.16	0.88	1.49	0.28	
		Simple mode		1.21	0.11	1.00	1.43	0.08	
		Weighted mode		0.55	1.65	-2.68	3.79	0.72	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Tyrosine	IVW	96	0.94	0.05	0.85	1.04	0.22	0.65
		MR Egger		1.03	0.10	0.83	1.23	0.78	
		Weighted median		1.02	0.05	0.92	1.13	0.67	
		Simple mode		1.03	0.11	0.82	1.24	0.78	
		Weighted mode		1.02	0.06	0.9	1.14	0.74	
		MR-PRESSO		1.07	0.04	0.99	1.16	0.11	
	Valine	IVW	85	1.08	0.08	0.93	1.23	0.32	0.58
		MR Egger		0.83	0.14	0.55	1.11	0.20	
		Weighted median		0.97	0.07	0.83	1.11	0.65	
		Simple mode		0.91	0.14	0.63	1.19	0.51	
		Weighted mode		0.94	0.07	0.8	1.07	0.34	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	

Table S23. Two-sample MR results for genetically predicted serum amino acids and risk of breast cancer

Outcome	Exposure	Method	nSNPs	OR	SE	Lower CI	Upper CI	Pval	p.FDR
Breast cancer	Alan	IVW	20	1.03	0.17	0.7	1.36	0.88	0.96
		MR Egger		0.72	0.4	-0.06	1.51	0.42	
		Weighted median		0.95	0.22	0.52	1.38	0.81	
		Simple mode		1.25	0.37	0.53	1.96	0.55	
		Weighted mode		0.87	0.26	0.35	1.38	0.6	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Glutamine	IVW	21	0.97	0.16	0.66	1.27	0.82	0.96
		MR Egger		1.06	0.31	0.45	1.67	0.85	
		Weighted median		0.94	0.21	0.52	1.36	0.78	
		Simple mode		0.51	0.49	-0.45	1.46	0.18	
		Weighted mode		1.32	0.29	0.76	1.88	0.34	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Glycine	IVW	22	0.98	0.09	0.8	1.15	0.81	0.96
		MR Egger		0.94	0.14	0.67	1.21	0.66	
		Weighted median		1.06	0.10	0.87	1.26	0.55	
		Simple mode		1.43	0.25	0.94	1.92	0.16	
		Weighted mode		1.01	0.09	0.84	1.19	0.89	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Histidine	IVW	7	1.16	0.4	0.37	1.95	0.71	0.96
		MR Egger		2.62	0.82	1.02	4.23	0.29	
		Weighted median		1.19	0.24	0.71	1.67	0.47	
		Simple mode		2.06	0.41	1.25	2.86	0.13	
		Weighted mode		1.26	0.27	0.73	1.78	0.42	
		MR-PRESSO		1.40	0.24	0.93	1.87	0.22	
	Isoleucine	IVW	11	0.74	0.23	0.29	1.19	0.18	0.96
		MR Egger		0.92	0.65	-0.35	2.20	0.91	
		Weighted median		0.70	0.3	0.11	1.29	0.24	

		Simple mode		0.97	0.53	-0.06	2.00	0.96	
		Weighted mode		0.61	0.32	-0.01	1.23	0.15	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Leucine	IVW	15	1.14	0.22	0.7	1.58	0.55	0.96
		MR Egger		0.91	0.59	-0.24	2.06	0.87	
		Weighted median		1.35	0.28	0.8	1.89	0.29	
		Simple mode		1.32	0.51	0.31	2.32	0.6	
		Weighted mode		1.54	0.33	0.9	2.18	0.21	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Phenylalanine	IVW	5	0.86	0.43	0.02	1.70	0.73	0.98
		MR Egger		2.37	1.28	-0.14	4.89	0.55	
		Weighted median		0.94	0.51	-0.06	1.94	0.91	
		Simple mode		1.08	0.64	-0.18	2.33	0.91	
		Weighted mode		1.01	0.60	-0.17	2.18	0.99	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Tyrosine	IVW	22	0.87	0.20	0.47	1.26	0.48	0.96
		MR Egger		1.34	0.68	0.01	2.67	0.67	
		Weighted median		1.23	0.28	0.67	1.79	0.46	
		Simple mode		1.46	0.51	0.47	2.45	0.46	
		Weighted mode		1.43	0.45	0.55	2.31	0.44	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Valine	IVW	16	0.87	0.17	0.53	1.20	0.40	0.98
		MR Egger		0.44	0.41	-0.36	1.24	0.06	
		Weighted median		0.75	0.23	0.29	1.21	0.22	
		Simple mode		0.87	0.4	0.08	1.65	0.72	
		Weighted mode		0.71	0.24	0.24	1.19	0.19	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	

Table S24. Two-sample MR results for genetically predicted serum amino acids and risk of prostate cancer

Outcome	Exposure	Method	nSNPs	OR	SE	Lower CI	Upper CI	Pval	p.FDR
Prostate cancer	Alan	IVW	74	0.99	0.06	0.87	1.1	0.83	0.94
		MR Egger		0.94	0.16	0.63	1.25	0.71	
		Weighted median		0.99	0.07	0.86	1.12	0.87	
		Simple mode		1.04	0.15	0.75	1.33	0.81	
		Weighted mode		0.96	0.13	0.7	1.23	0.77	
		MR-PRESSO		1.03	0.05	0.93	1.13	0.48	
	Glutamine	IVW	97	0.99	0.06	0.87	1.11	0.88	0.94
		MR Egger		0.76	0.13	0.51	1.02	**0.04**	
		Weighted median		0.94	0.06	0.81	1.06	0.28	
		Simple mode		1.13	0.15	0.84	1.42	0.41	
		Weighted mode		0.82	0.07	0.67	0.96	**0.01**	
		MR-PRESSO		0.98	0.05	0.88	1.08	0.63	
	Glycine	IVW	79	1.01	0.03	0.94	1.07	0.85	0.94
		MR Egger		1.06	0.05	0.96	1.16	0.28	
		Weighted median		1.03	0.04	0.95	1.1	0.52	
		Simple mode		0.95	0.09	0.78	1.12	0.57	
		Weighted mode		1.02	0.03	0.95	1.08	0.60	
		MR-PRESSO		1.01	0.03	0.97	1.05	0.72	
	Histidine	IVW	43	0.95	0.08	0.79	1.11	0.53	0.94
		MR Egger		1.02	0.18	0.67	1.38	0.90	
		Weighted median		0.98	0.08	0.81	1.14	0.78	
		Simple mode		1.07	0.19	0.69	1.45	0.72	
		Weighted mode		0.93	0.1	0.74	1.13	0.50	
		MR-PRESSO		0.98	0.07	0.84	1.11	0.80	
	Isoleucine	IVW	41	1.06	0.07	0.92	1.2	0.41	0.94
		MR Egger		1.04	0.16	0.72	1.36	0.81	
		Weighted median		1.05	0.1	0.85	1.25	0.61	

		Simple mode		0.91	0.18	0.55	1.26	0.59	
		Weighted mode		1.04	0.11	0.82	1.26	0.74	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Leucine	IVW	44	0.98	0.07	0.84	1.11	0.73	0.94
		MR Egger		0.99	0.14	0.72	1.26	0.93	
		Weighted median		1.03	0.09	0.85	1.20	0.77	
		Simple mode		0.88	0.18	0.54	1.23	0.49	
		Weighted mode		1.03	0.09	0.86	1.20	0.73	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Phenylalanine	IVW	22	1.04	0.09	0.87	1.2	0.67	0.94
		MR Egger		1.14	0.22	0.71	1.56	0.57	
		Weighted median		1.09	0.12	0.86	1.32	0.44	
		Simple mode		0.95	0.21	0.53	1.37	0.82	
		Weighted mode		1.06	0.14	0.8	1.33	0.65	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Tyrosine	IVW	74	1.00	0.06	0.88	1.12	0.98	0.98
		MR Egger		1.08	0.11	0.87	1.29	0.48	
		Weighted median		1.05	0.06	0.93	1.18	0.39	
		Simple mode		1.03	0.13	0.78	1.28	0.83	
		Weighted mode		1.05	0.06	0.94	1.16	0.39	
		MR-PRESSO		1.01	0.05	0.91	1.10	0.75	
	Valine	IVW	60	1.01	0.06	0.9	1.13	0.84	0.94
		MR Egger		1.08	0.11	0.88	1.29	0.45	
		Weighted median		1.05	0.07	0.91	1.18	0.51	
		Simple mode		1.26	0.15	0.96	1.56	0.14	
		Weighted mode		1.06	0.07	0.93	1.19	0.40	
		MR-PRESSO		1.03	0.05	0.93	1.30	0.57	

Table S25. Two-sample MR results for genetically predicted serum amino acids and risk of Alzheimer's disease

Outcome	Exposure	Method	nSNPs	OR	SE	Lower CI	Upper CI	Pval	p.FDR
AD	Alan	IVW	100	1.2	0.06	1.02	1.38	0.98	1.00
		MR Egger		1.55	0.15	1.14	1.95	0.93	
		Weighted median		1.33	0.09	1.09	1.57	0.58	
		Simple mode		1.46	0.24	1.13	1.78	0.58	
		Weighted mode		1.34	0.24	1.07	1.61	0.63	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Glutamine	IVW	130	0.90	0.11	0.69	1.12	0.37	0.48
		MR Egger		0.88	0.08	0.72	1.03	0.09	
		Weighted median		0.93	0.05	0.83	1.03	0.16	
		Simple mode		0.89	0.16	0.58	1.21	0.48	
		Weighted mode		0.85	0.09	0.67	1.04	0.09	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Glycine	IVW	2	1.17	0.50	0.20	2.15	0.75	0.90
		MR Egger		*NA*	*NA*	*NA*	*NA*	*NA*	
		Weighted median		*NA*	*NA*	*NA*	*NA*	*NA*	
		Simple mode		*NA*	*NA*	*NA*	*NA*	*NA*	
		Weighted mode		*NA*	*NA*	*NA*	*NA*	*NA*	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Histidine	IVW	57	1.05	0.09	0.88	1.22	0.60	0.92
		MR Egger		1.17	0.18	0.81	1.54	0.39	
		Weighted median		1.04	0.13	0.79	1.29	0.75	
		Simple mode		1.09	0.25	0.6	1.59	0.72	
		Weighted mode		1.14	0.18	0.8	1.49	0.45	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Isoleucine	IVW	55	1.23	0.09	0.96	1.51	**0.05**	0.19
		MR Egger		1.18	0.21	0.97	1.4	**0.04**	
		Weighted median		1.24	0.14	0.98	1.51	**0.03**	

		Simple mode		1.19	0.17	0.98	1.4	**0.03**	
		Weighted mode		1.44	0.29	0.88	2.00	0.21	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Leucine	IVW	63	0.88	0.61	0.41	1.34	0.52	0.76
		MR Egger		1.87	1.26	-0.6	4.34	0.62	
		Weighted median		0.83	0.14	0.33	1.33	0.11	
		Simple mode		0.68	0.14	-0.52	1.87	0.14	
		Weighted mode		0.95	0.26	0.77	1.13	0.47	
		MR-PRESSO		0.97	0.09	0.79	1.14	0.74	
	Phenylalanine	IVW	42	1.15	0.08	1.00	1.31	0.06	0.73
		MR Egger		0.97	0.12	0.73	1.21	0.80	
		Weighted median		1.09	0.14	0.81	1.37	0.56	
		Simple mode		0.82	0.32	0.19	1.45	0.53	
		Weighted mode		0.93	0.14	0.67	1.20	0.61	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	
	Tyrosine	IVW	99	0.81	0.23	0.35	1.27	0.37	0.48
		MR Egger		0.95	0.4	0.18	1.73	0.91	
		Weighted median		0.98	0.09	0.8	1.16	0.86	
		Simple mode		0.91	0.2	0.51	1.3	0.62	
		Weighted mode		0.97	0.1	0.78	1.16	0.72	
		MR-PRESSO		0.94	0.06	0.82	1.06	0.29	
	Valine	IVW	86	1.00	0.07	0.88	1.12	0.65	0.76
		MR Egger		0.89	0.12	0.43	1.36	**0.02**	
		Weighted median		1.03	0.11	0.9	1.16	0.10	
		Simple mode		1.01	0.11	0.72	1.31	0.13	
		Weighted mode		1.00	0.21	0.58	1.42	1.00	
		MR-PRESSO		*NA*	*NA*	*NA*	*NA*	*NA*	