UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1565

Molecular Epidemiology of Cardiovascular Disease

MARKUS STENEMO

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2019

ISSN 1651-6206
ISBN 978-91-513-0634-6
urn:nbn:se:uu:diva-381234

Dissertation presented at Uppsala University to be publicly examined in Humanistiska teatern, Engelska parken, Thundbersvägen 3, Uppsala, Tuesday, 4 June 2019 at 09:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Rikard Landberg (Chalmers tekniska högskola).

Abstract

Stenemo, M. 2019. Molecular Epidemiology of Cardiovascular Disease. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1565. 45 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0634-6.

Cardiovascular disease is a major cause of morbidity and mortality, with increasing prevalence worldwide.

Identification of risk markers may enable improved prevention by targeting high-risk individuals, earlier disease diagnosis and treatment, as well as stratification of disease subtypes with different treatment options, thereby minimizing side effects while increasing success rates.

The **overall aim** of this thesis was to investigate associations between proteomic and metabolomic biomarkers, and the development of heart failure and ischemic stroke. **Specific objectives** were to examine potential causal pathways, and the added value in risk prediction of the identified risk markers.

In **Studies I–II**, we performed proximity extension assay based **proteomic profiling** of ≥80 circulating proteins in the Swedish cohorts Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS n=901, median age 70), and the Uppsala Longitudinal Study of Adult Men (ULSAM, n=685, median age 77). In **Study I**, we identified nine proteins involved in apoptosis, inflammation, matrix remodeling, and fibrinolysis associated with incident heart failure, including growth differentiation factor-15 (GDF-15). In **Study II**, we identified several proteins associated with incident ischemic stroke, including GDF-15. Both studies revealed potential to improve disease risk prediction by using proteomic data.

In **Study III**, we performed mass spectrometry-based **metabolomic profiling** in plasma or serum samples from PIVUS, ULSAM, and TwinGene (total n=3,924). The metabolites urobilin and sphingomyelin (30:1) were associated with incident heart failure.

In **Study IV**, we followed up on the results of **Studies I–II**, performing **Mendelian randomization** analyses (a framework for causal analysis using genetic variants) in 1,053,527 individuals, with 88,448 coronary artery disease cases, 70,305 ischemic stroke cases, and 1,420 heart failure cases. This study supports a causal role of genetically elevated GDF-15 levels in heart failure development, but not in coronary artery disease or ischemic stroke.

In conclusion, we identified multiple biomarkers associated with incident heart failure and ischemic stroke, potentially involved in early disease development. We also saw potential to improve disease risk prediction for incident heart failure and ischemic stroke using proteomics data.

Our findings encourage further large-scale proteomic, metabolomic, and genetic studies to give new insights into heart failure and stroke pathogenesis.

Keywords: Biomarkers, ischemic stroke, heart failure, omics, proteomics, metabolomics, epidemiology, risk marker

Markus Stenemo, Department of Medical Sciences, Molecular epidemiology, Husargatan 3, Biomedicinskt Centrum (BMC), Uppsala University, SE-75123 Uppsala, Sweden.

ISSN 1651-6206
ISBN 978-91-513-0634-6
urn:nbn:se:uu:diva-381234 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-381234)

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Stenemo M, Nowak C, Byberg L, Sundström J, Giedraitis V, Lind L, Ingelsson E, Fall T,* Ärnlöv J.* (2018). Circulating proteins as predictors of incident heart failure in the elderly. *Eur J Heart Fail.* 2018 Jan;20(1):55–62.

II Lind L, Siegbahn A, Lindahl B, Stenemo M, Sundström J, Ärnlöv J. (2015). Discovery of new risk markers for ischemic stroke using a novel targeted proteomics chip. *Stroke*. 2015 Dec;46(12):3340–3347.

III Stenemo M, Ganna A, Salihovic S, Nowak C, Sundström J, Giedraitis V, Broeckling C.D., Prenni J.E., Svensson P, Magnusson P, Lind L, Ingelsson E, Ärnlöv J,* Fall T.* The metabolites urobilin and sphingomyelin (30:1) are associated with incident heart failure in the general population. *Submitted manuscript.*

IV Stenemo M, Nowak C, Burgess S, Ärnlöv J, Fall T. The role of growth differentiation factor 15 (GDF-15) in cardiovascular disease – a Mendelian Randomisation study. *Manuscript*.

*Shared last authors

Contents

Abbreviations

ARIC	Atherosclerosis Risk in Communities
AUC	area under the receiver operating characteristic curve
BMI	body mass index
CI	confidence interval
DALY	disability-adjusted life year
ECP	eosinophil cationic protein
ELISA	enzyme-linked immunosorbent assay
FDR	false discovery rate
FS	follistatin
GDF-15	growth differentiation factor 15
GWAS	genome-wide association study
HDL-C	high-density lipoprotein cholesterol
HFrEF	heart failure with reduced ejection fraction
HFpEF	heart failure with preserved ejection fraction
HR	hazard ratio
ICD	International Classification of Diseases
LC	liquid chromatography
LDL-C	low-density lipoprotein cholesterol
MMP-12	matrix metalloproteinase-12
MRI	magnetic resonance imaging
MR	Mendelian randomization
MS	mass spectrometry
NT-proBNP	N-terminal prohormone of brain natriuretic peptide
OPG	osteoprotegerin
OR	odds ratio
PIVUS	Prospective Investigation of the Vasculature in Uppsala Seniors
QTOF	quadrupole time-of-flight
SALT	Screening Across the Lifespan Twin
SD	standard deviation
SE	standard error
SNP	single nucleotide polymorphism
SPON1	spondin-1
ST2	suppression of tumorigenicity 2
TIM-1	T-cell immunoglobulin and mucin domain 1

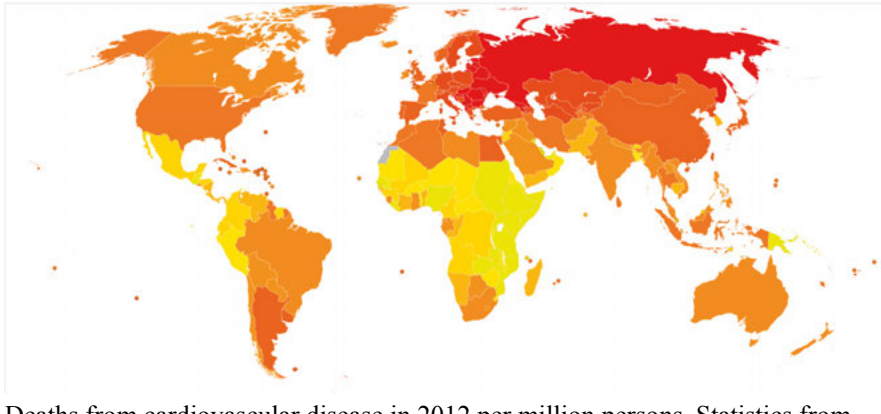
TRAIL-R2	tumor necrosis factor-related apoptosis-inducing ligand receptor 2
ULSAM	Uppsala Longitudinal Study of Adult Men
U-PAR	urokinase plasminogen activator surface receptor
UPLC	ultra performance liquid chromatography

Introduction

Cardiovascular disease refers to a group of diseases that affect the heart, brain, or blood vessels. Cardiovascular disease can broadly be sub-classified into stroke, heart failure, coronary artery disease (which includes myocardial infarction and angina pectoris), peripheral artery disease, and cardiomyopathies, among others. Cardiovascular disease accounted for one third of all deaths in 2017,(1) and the burden is expected to increase due to an aging population, sedentary lifestyle, and poor diet.(2)

Importantly, non-fatal cardiovascular events are a major cause of disability, and the number of disability-adjusted life years (DALYs) lost due to cardiovascular disease was estimated at 366 million in 2017.(3)

Figure 1. World map showing the distribution of cardiovascular disease mortality rates per million persons.



Deaths from cardiovascular disease in 2012 per million persons. Statistics from World Health Organisation, grouped by deciles.(4)

Heart failure

The term ‘heart failure’ refers to when the heart is unable to provide oxygenated blood to the rest of the body, and is usually first noticed when there is high demand for oxygen, e.g., during strenuous exercise. Heart failure is a clinical syndrome, and the underlying pathology varies with different types of damage to the heart, such as myocardial infarction, arrhythmias, valvular disease, endocardial disease or cardiomyopathies, all of which can cause cardiac dysfunction.

Early signs of heart failure include shortness of breath, fatigue, and swollen legs (edema). Since heart failure has many different underlying causes and pathophysiological changes to the heart, there is no single diagnostic test for heart failure. Instead, heart failure diagnosis is based on the medical history of the patient, signs and symptoms at the clinical examination, biomarker assessments, and imaging techniques, such as an echocardiographic examination.

When a patient shows several of the early warning signs, a blood test measuring the circulating biomarker N-terminal pro brain natriuretic peptide (NT-proBNP) is frequently performed. If the levels are above a certain threshold, an ultrasound of the heart (echocardiography) can be performed to better characterize the function and geometry of the myocardium. In addition to being an important tool in diagnosing heart failure, the echocardiographic examination may help to differentiate between different types of heart failure, such as heart failure with preserved or reduced ejection fraction (HFpEF vs. HFrEF, respectively).(5)

Because of the multifactorial etiology and pathology, heart failure research has many challenges. Better understanding of the mechanisms underlying heart failure development and progression is needed in order to improve treatment and prevention.

Epidemiology

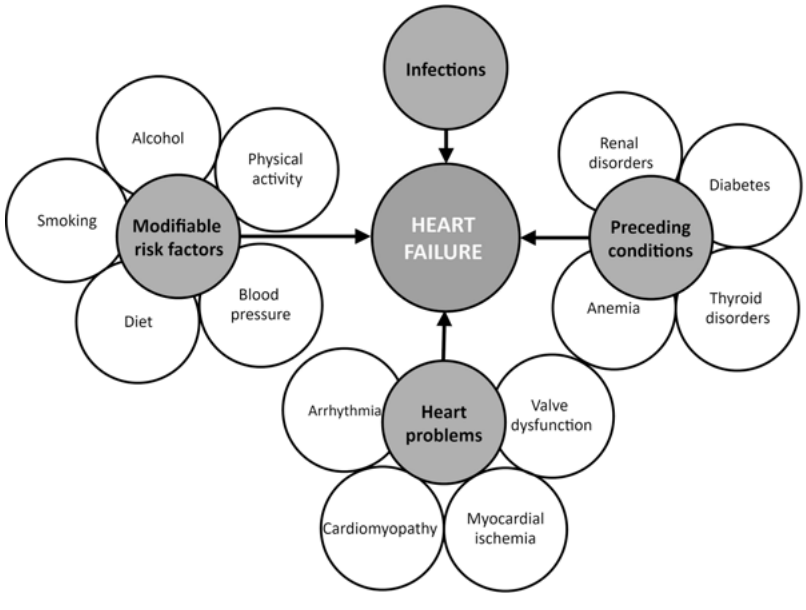
Although heart failure mortality has declined in high-income countries in the last few decades thanks to improved treatment regimes, 50% of individuals diagnosed with heart failure are still estimated to die within five years.(6) The global burden of heart failure is increasing, measured in both DALYs and economic terms, as a result of increasing rates of obesity, smoking, type 2 diabetes, and an aging population.(6) The international annual cost of heart failure has been estimated at 108 billion dollars.(7) The prevalence and incidence of heart failure increase with age; the lifetime risk of developing heart failure has been estimated at over 45%.(8) Heart failure needs to be addressed through earlier detection of high-risk individuals, as well as new therapeutic and preventive strategies.(9)

Etiology

Heart failure is a multifactorial syndrome with several modifiable risk factors involved, including hypertension, dyslipidemia, hyperglycemia, and obesity (**Figure 2**). These have in common that they can be affected by lifestyle factors, such as diet, physical activity, alcohol intake, and smoking. Intervening on the risk factors has the potential to substantially decrease the risk of developing heart failure.

In the Framingham cohort, more than 75% of the heart failure events were preceded by hypertension,(10) while in other cohorts, coronary heart disease has been reported as a more important risk factor.(11) Still, much of the complexity in the development of heart failure is incompletely understood. For instance, Chang et al. showed that approximately half of patients hospitalized for heart failure had a reduced left ventricular ejection fraction, while the rest had a preserved ejection fraction, indicating different underlying mechanisms.(12)

Figure 2. Causes of heart failure.



Ischemic stroke

Stroke occurs when there is poor blood flow to the brain, resulting in hypoxia and brain cell death. The two main types of stroke are ischemic stroke, caused by limited blood flow due to a thrombus or embolus, and hemorrhagic stroke, which is caused by bleeding. This thesis will focus on ischemic stroke. Symptoms of a stroke include neurological deficits, such as inability to move or feel one side of the body, dizziness, speech impairment, and problems understanding speech.

Epidemiology

In 2016, stroke caused more than 5 million deaths and 116 million DALYs worldwide.(13) Important risk factors are shared with heart failure, and include age, hypertension, atrial fibrillation, diabetes, high cholesterol levels, and smoking. Due to this, the global burden is increasing for stroke as well, measured in both DALYs and economic terms.

In the US, 15% of individuals below 65 years of age die within a year of their first stroke, and one third of individuals above age 75 die within one year,(14) highlighting the need to find preventive measurements and earlier detection.

Etiology

Embolic infarction in the brain is preceded by an embolus forming in the circulatory system, which subsequently breaks off from its location. The embolus enters the circulation and if it gets stuck in vessels in the brain it can partially or completely block blood flow, causing an ischemic cascade, which leads to permanent brain damage unless treated within hours. Clinical diagnosis is based on clinical signs of stroke and examinations with computed tomography and magnetic resonance imaging (MRI).

Biomarker discovery using omics data

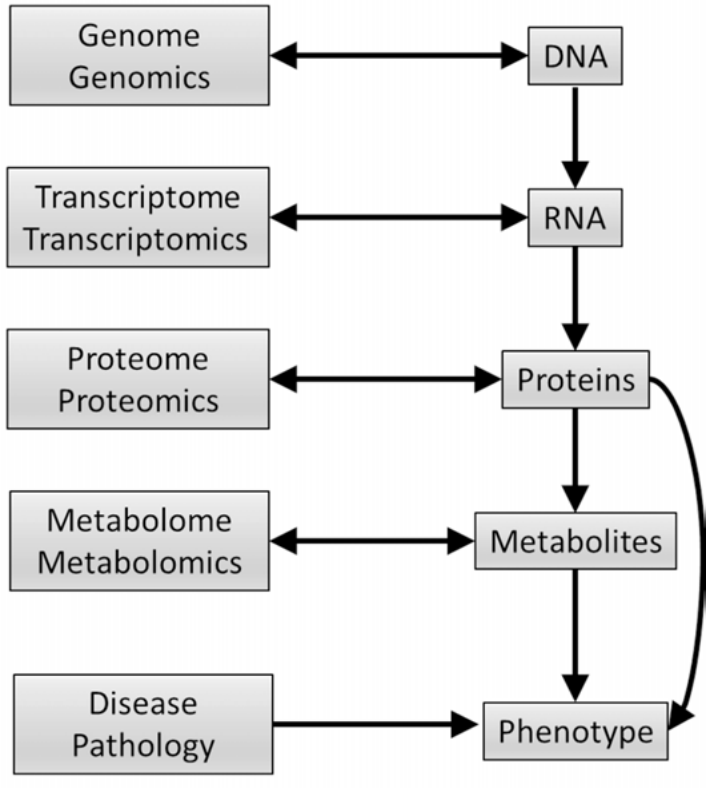
The term "*biomarker*" has been defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention."(15)

The discovery of novel clinically relevant biomarkers could lead to earlier disease diagnosis (facilitating earlier treatment and improved prognosis), improved risk stratification and prediction, and personalized novel treatments (stratification of patients based on disease subgroup, leading to treatments

with fewer side effects and improved prognosis). Finding new biomarkers could suggest novel targets for drug development.

This thesis uses genomic, proteomic, and metabolomic data to evaluate potential biomarkers and test causality. The genome, transcriptome, proteome, and metabolome are an organism's complete set of deoxyribonucleic acid (DNA), ribonucleic acid (RNA), proteins, and metabolites, respectively. The term "*omics*" refers to the large-scale analysis and characterization of biological molecules, and include genomics, transcriptomics, proteomics, and metabolomics (**Figure 3**).

Figure 3. Overview of different omics research fields.



Recent advances in high-throughput technologies have made it possible to perform metabolic and proteomic analyses in large study cohorts, and this has provided novel possibilities for biomarker discovery. However, few studies have investigated the associations between circulating proteins or metabolites and the risk of heart failure or stroke.(16, 17)

Genomics is the study of the genetic makeup in biological systems. Genome-wide association studies (GWAS) look at genetic variants (typically single nucleotide polymorphisms (SNPs) of the entire genome) at a population

level, to identify genetic differences between individuals and evaluate if there is an association between genetic variation and disease.

Transcriptomics is the study of the production of different forms of RNA, their functions, and their regulation.

Proteomics, or proteomic profiling, is the study of proteins in a biological sample. The human plasma proteome consists of all circulating proteins with a molecular mass exceeding the renal filtration threshold (4.5 kDa). Both signaling proteins and proteins leaked from damaged cells are found in the circulation.

Metabolomics, or metabolomic profiling, is the study of low-weight molecules (< 1.5 kDa) in biological systems, derived from either the environment (exogenous metabolites, e.g., diet and medications) or produced by the organism (endogenous metabolites).(18) Depending on the analysis method, detectable molecules include amino acids, carbohydrates, lipids, and nucleotides. The two most common methods for quantifying metabolites are mass spectrometry (MS), which is usually preceded by liquid chromatography (LC) or gas chromatography to separate metabolites by molecular weight; and nuclear magnetic resonance spectroscopy.

Mendelian randomization

One crucial challenge for biomarker research is to differentiate associations between biomarkers and clinical outcomes from causal associations. Observational studies cannot firmly establish whether an association between exposure and outcome is causal (x influences y), due to potential unmeasured or residual confounding (confounder influences x and y), or reverse causality (y influences x). Mendelian randomization (MR) exploits a genetically predisposed exposure to determine the causal effect of the exposure on the outcome and has been proposed as a way of circumventing the limitations in traditional observational studies. (**Figure 4**)(19)

One example of a MR analysis is studying a genetic mutation of the gene *FTO* that leads to increased adiposity, and is associated with heart failure, supporting a causal effect of increased adiposity on heart failure.(20)

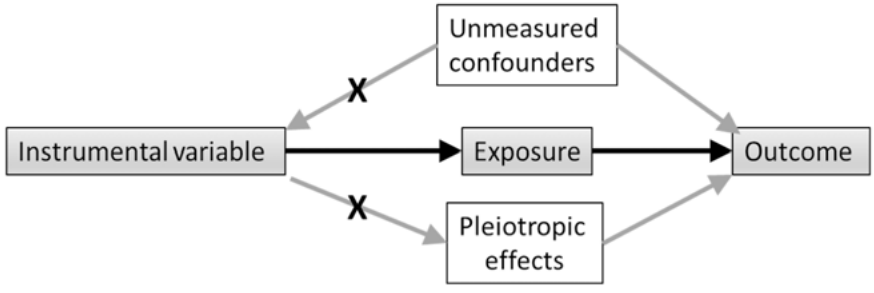


Figure 4. Principles of Mendelian randomization. Genetic variation leads to differences in the exposure (e.g. increased BMI), which in turn leads to the health outcome (e.g. disease). The effect of the instrumental variable on the exposure can be used to estimate the causal effect of the exposure on the outcome.

Mendelian randomization has been described as using naturally occurring randomized trials in which genetic factors are randomly assigned at birth. This makes it possible to study the causal effect of many different exposures on clinical outcomes, particularly in epidemiological settings where clinical trials may be difficult to implement.(21)

The three main principles underlying MR are that the instrumental variable 1) is associated with the exposure, 2) is not affected by confounding, and 3) the effect of the variable on the outcome is mediated through the exposure.(22)

However, there are many assumptions that need to be fulfilled for a MR study to be valid. Important biases that have to be considered include:

- **Population stratification**, which refers to when there are genetic differences between subgroups that share other characteristics.
- **Canalization**, which occurs when there are biological mechanisms that compensate for a genetic effect.
- **Horizontal Pleiotropy**, which is when an instrumental variable affects multiple downstream pathways, some not going through the exposure.

Some of these assumptions and biases cannot be tested, and MR should therefore not be seen as certain evidence of a specific effect on an outcome, but rather as a guide for what should be investigated further.

Aims

Overall aims

The overall aim of this thesis was to discover proteomic and metabolomic risk markers associated with development of heart failure and ischemic stroke. We also wanted to examine potential causal pathways between the biomarkers and the outcomes, and the added value in risk prediction of the identified risk markers.

Specific aims of each study

The aims of Study I were to investigate and validate the associations between 80 cardiovascular disease-associated proteins and heart failure incidence, and to assess whether the proteins could improve the prediction of heart failure beyond established risk factors in ULSAM and PIVUS, using a discovery/replication approach.

The aims of Study II were to investigate and validate the associations between 85 cardiovascular disease-associated proteins and ischemic stroke incidence, and to assess whether the proteins could improve the prediction of ischemic stroke beyond established risk factors in the same two cohorts as in Study I.

The aim of Study III was to investigate the association between 209 metabolites, measured using untargeted ultra performance liquid chromatography coupled with mass spectrometry (UPLC-QTOF-MS/MS), and heart failure incidence in ULSAM, PIVUS, and TwinGene, using a discovery/replication approach.

The aim of Study IV was to investigate the causal relationship of circulating GDF-15 levels on coronary artery disease, ischemic stroke and heart failure, using instrumental variable analysis with a single locus two-sample MR method.

Study populations

This thesis is based on the Swedish cohorts ULSAM, PIVUS, TwinGene, and the British cohort UK Biobank. We also used publicly available data from the international genetic consortia MEGASTROKE and CARDIoGRAMplusC4D. **Table 1** shows an overview of the study populations. The regional ethical review boards at Uppsala University, Karolinska Institutet and the UK Biobank approved the studies and all participants gave written informed consent.

Table 1. Overview of the studies, their contributions, and the samples used.

Cohorts	N	Age	Omics	Outcome	Study
ULSAM	839 1,221	77 70	P, M	HF, IS	I–II III
PIVUS	1,016	70	P, M	HF, IS	I–III
TwinGene	2443	60-70	M	HF	III
UK Biobank	337,488	40-70	G	HF, IS, CAD	IV
MEGASTROKE	521,612	18-100	G	IS	IV
CARDIoGRAMplusC4D	194,427	25-100	G	CAD	IV

CAD; coronary artery disease, HF, heart failure; IS, ischemic stroke; G, genomics; M, metabolomics; P, proteomics.

ULSAM

In the **Uppsala Longitudinal Study of Adult Men (ULSAM)**, all men born in Sweden between 1920 and 1924 and living in Uppsala were invited to participate in a health assessment in 1970.(23) In a third examination cycle between 1991 and 1995, 1,221 of the 1,691 men invited (then aged 70 years) chose to participate, which served as the baseline examination for Study III. In the fourth examination cycle, 839 of the 1,398 men invited (then aged 77 years) chose to participate, serving as the baseline examination for Studies I and II. The men have since been invited to re-examinations at ages 82, 88, and

93 years, and information about each individual has been gathered continuously, including annual updates on mortality and morbidity from national registers.

PIVUS

In the **Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS)**, all 70-year-old residents of Uppsala were invited to participate in a health survey and detailed clinical assessment between 2001 and 2004, which has been described in detail previously.(24) Of 2,025 invited, 1,016 (50.2%) participated in the baseline assessment within one month of their 70th birthday; serving as the baseline examination for Studies I–III. The participants have since been invited to re-examinations at ages 75 and 80 years, with annual updates on mortality and morbidity from national registers, and using medical records for certain outcomes.

TwinGene

The Swedish Twin Registry is a population-based national register of 194,000 Swedish twins born between 1886 and 2008.(25) **TwinGene** is a longitudinal study of 12,591 individuals nested within the Swedish Twin Registry. All twins born before 1958 who participated in the Screening Across the Lifespan Twin (SALT) telephone screening between 1998 and 2002 were contacted again between 2004 and 2008, and 12,591 participated in the new study (http://ki.se/en/meb/twingene-and-genomeeutwin). Metabolomics was performed in a subset of TwinGene using a case-cohort design, where all incident cases of type 2 diabetes (n = 218), coronary artery disease (n = 282), ischemic stroke (n = 186), and dementia (n = 114) prior to 31 December 2010 were included. A sub-cohort (controls) of 1,643 individuals (43% women) stratified by age and sex was also included.(26)

UK Biobank

A total of 502,655 volunteers aged 40–69 years were enrolled in the UK Biobank from 2006 to 2010 (http://www.ukbiobank.ac.uk) at 22 assessment centers across England, Wales, and Scotland. The assessment visit comprised informed consent, a touch-screen questionnaire, a brief verbal interview, physical and functional measures, and collection of biological samples.(27) Of the 488,377 individuals who underwent genotyping, 337,488 individuals had

available genetic data that passed quality control, were unrelated at third degree or closer, reported white British ancestry, and had not withdrawn consent from the study.

CARDIoGRAMplusC4D

For the analyses of coronary artery disease, we included genetic summary statistics from The Coronary Artery Disease Genome-Wide Replication And Meta-Analysis consortium and the Coronary Artery Disease (C4D) Genetics (CARDIoGRAMplusC4D), a two-stage meta-analysis of studies involving 194,427 individuals of European or South Asian ancestry (63,746 cases and 130,681 controls).(28) We included the association of each SNP measured as the per-allele odds ratio and standard error with the outcome of interest.

MEGASTROKE

For the analyses of ischemic stroke, we included genetic summary statistics from MEGASTROKE, a meta-analysis of 521,612 individuals (67,162 cases and 454,450 controls), predominantly of European ancestry.(29) We included the association of each SNP measured as the per-allele odds ratio and standard error with the outcome of interest.

Methods

Outcome definition

Heart failure

In PIVUS and ULSAM, medical records in the National Patient Register (inpatient) for all individuals hospitalized for heart failure (International Classification of Diseases, Ninth Revision (ICD-9): 428; Tenth Revision (ICD-10): I50), or hypertensive heart disease with heart failure (ICD-9: 402; ICD-10: I11) during follow-up were reviewed by physicians blinded to the baseline data.(30) They classified heart failure events as definite, questionable, or miscoded, in accordance with the European Society of Cardiology definitions.(31) We used only cases of heart failure classified as definite in our analyses for definition of the outcome.

In TwinGene, heart failure diagnosis was obtained from the National Patient Register or the Cause of Death Register (ICD-9: 428; ICD-10: I50).

In UK Biobank, heart failure was defined using hospital and death registers, if individuals had heart failure as main diagnosis (ICD-9: 428; ICD-10: I50), or if participants stated that they had heart failure at the verbal interview.

Coronary artery disease

In UK Biobank, coronary artery disease was defined using hospital and death registers, as having a main diagnosis of ischemic heart disease (ICD-9: 410-413; ICD-10: I21-I25), or having undergone coronary interventions (Office of Population Censuses and Surveys Classification of Interventions and Procedures, version 4 (OPCS-4) codes: K40-K46, K49, K50 and K75). Individuals were also classified as having coronary artery disease if they reported angina pectoris or a history of myocardial infarction at the verbal interview. CARDIoGRAMplusC4D used standard criteria for defining cases of coronary artery disease, with some studies including angiography-confirmed stenosis, as well as stable or unstable angina.(28)

Ischemic stroke

Medical records for all individuals with incident ischemic stroke (ICD-10: I63) in PIVUS and ULSAM were collected and validated, using both the Swedish National Patient Register and the Swedish Cause of Death Register.

In UK biobank, ischemic stroke was defined using hospital and death registers, as those individuals with a main diagnosis of ischemic stroke (ICD-9: 434, 436; ICD-10: I63, I64) or reporting ischemic stroke at the verbal interview. The MEGASTROKE consortium defined stroke in accordance with World Health Organization guidelines, i.e., rapidly developing signs of focal or global disturbance of cerebral function lasting more than 24 hours or leading to death with no apparent cause other than that of vascular origin. Strokes were classified as ischemic stroke based on clinical and imaging criteria.(29)

Clinical characteristics

In ULSAM and PIVUS, participants were investigated in the morning after an overnight fast. Venous blood samples were frozen immediately after separation of plasma and stored at -80 °C until analysis. The investigations in PIVUS and ULSAM were performed using standardized methods, including measurements of blood pressure, biochemistry (lipids and glucose), anthropometry, and echocardiography (PIVUS only). Participants in TwinGene were sent blood sampling kits and went to their local health care centers for blood sampling and health check-ups. Participants were instructed to perform sample collection in the morning after an overnight fast, and samples were sent by overnight mail to the Karolinska Biobank, where they were frozen at -80 °C until analysis.

In all Swedish cohorts, information on lifestyle and medication at baseline was collected through questionnaires. Data on myocardial infarction prior to baseline or during follow-up were retrieved from the Swedish hospital discharge register (ICD9: 410; ICD10: I21, I22). Diabetes was defined as having a fasting glucose level > 7 mmol/l, taking antidiabetic medication, or having a HbA1c level > 6.5 mmol/mol.

Proteomic profiling

In **Studies I and II**, Plasma (PIVUS) or serum (ULSAM) samples were assessed with the Proseek Multiplex CVD I^{96x96} proximity extension assay (Olink Bioscience, Uppsala, Sweden) (for details, see (32)). The assay simultaneously measures 92 proteins using two specific antibodies per protein, and includes one negative control and three positive controls (spiked in IL-6, IL-

8, and VEGF-A). The antibodies bind pair-wise to their specific protein, causing a polymerase chain reaction sequence. This is followed by quantitative real-time polymerase chain reaction for quantification.(33) The resulting relative values were $\log_2$-transformed for subsequent analysis, and each protein level was normalized by plate, by setting the mean to zero and standard deviation to one within each plate and for each storage time (correction based on the observed values and predicted values from a spline model). Values below the lower limit of detection were imputed to half the lower limit of detection.

Metabolomic profiling

In **Study III**, metabolomic profiling was performed in ULSAM, PIVUS, and TwinGene, using a Waters Acquity ultra performance liquid chromatography (UPLC) system coupled to a Waters Xevo G2-Quadrupole-Time-Of-Flight Mass Spectrometry (QTOF MS) platform at Colorado State University (Fort Collins, CO, USA). For details on sample handling and data processing using XCMS in R, see (34) and (26), and for the metabolomics pipeline, see (35) and the publicly available code at https://github.com/andgan/metabolomics_pipeline.

Statistical analyses

Studies I–III

Age- and sex-adjusted Cox proportional hazards regression for new-onset heart failure (Studies I and III) and ischemic stroke (Study II) was performed separately for each of the proteins and metabolites. The Cox proportional hazards assumption was assessed by visually inspecting plots of Schoenfeld residuals against time for biomarkers that were found to be significant in the replication step of the analysis.

We used a discovery validation where proteins associated with outcome at a 5% false discovery rate (FDR, estimated using the Benjamini & Hochberg method(36)) or metabolites associated with outcome at a 15% FDR were taken forward to replication in the validation sample, where a nominal significance threshold of p value < 0.05 was used. In Studies I and II, PIVUS was used as the discovery cohort, and ULSAM as the replication cohort. In Study III, ULSAM and PIVUS were used as discovery cohorts and TwinGene as the replication cohort. In TwinGene, we used Cox proportional hazards models where controls were weighted on diabetes, coronary artery disease, ischemic stroke, and dementia, to account for the sampling method.(26) For some analyses, we merged the discovery and validation cohort into one study sample.

We also performed additional multivariable adjustments for the established heart failure risk factors (body mass index (BMI), circulating lipid levels (low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol), triglycerides, lipid medication, glycemic status (fasting glucose, treated with insulin, treated with oral antidiabetics), blood pressure (systolic and diastolic blood pressure, blood pressure medication), smoking status, left ventricular hypertrophy, myocardial infarction prior to/during study (time updated (the covariate is updated to the day an individual is hospitalized with acute myocardial infarction)), atrial fibrillation prior to/during study (time updated), and kidney function (glomerular filtration rate)); or the established stroke risk factors (age, sex, LDL-C, HDL-C, systolic blood pressure, BMI, diabetes mellitus, atrial fibrillation, and smoking). Separate analyses were performed to investigate the influence of NT-proBNP on these associations.

To assess if proteomic or metabolomic data could improve risk prediction beyond the Atherosclerosis Risk in Communities (ARIC) heart failure risk score in Studies I and III,(37) we used Lasso penalized Cox proportional hazards regression to select a parsimonious model that maximized discrimination performance, while minimizing the number of proteins used for prediction. To assess risk prediction improvement, the area under the receiver operating curve (AUC) was used.

Study IV

We performed a single locus two-sample MR study using individual level data from UK Biobank (n = 337,488) and publicly available summary level data for coronary artery disease and ischemic stroke from CARDIoGRAMplusC4D (n= 194,427) and MEGASTROKE (n = 521,612), respectively (total n = 1,053,527). We used three previously reported single nucleotide polymorphisms (SNPs) in the *GDF15* locus on chromosome 19 as a genetic instrument for GDF-15 concentrations, explaining > 20% of the variance.(38, 39) We assessed associations between each of the three SNPs with the outcomes (coronary artery disease, ischemic stroke, heart failure) in UK Biobank, using logistic regression models, adjusting for age, sex, and the first 21 genetic principal components. We performed fixed-effects inverse variance weighted meta-analyses, pooling the estimates for coronary artery disease from the UK Biobank and CARDIoGRAMplusC4D cohorts to investigate the influence of raised GDF-15 on cardiovascular disease risk; and pooling the UK Biobank and MEGASTROKE cohorts to investigate the influence of raised GDF-15 on ischemic stroke risk.

To estimate the causal association between circulating GDF-15 and each outcome based on single SNPs, we used the instrumental variable estimator. This was calculated by dividing the estimated SNP effect on circulating GDF-

15 and estimated SNP effect on each outcome (Equation 1). The standard errors for the instrumental variable estimators were calculated using the delta method (Equation 2).

Equation 1.

$$\beta_{instrumental\ variable\ Estimator} = \frac{\beta_{SNP \rightarrow CVD}}{\beta_{SNP \rightarrow GDF-15}}$$

Equation 2.

$$se_{\substack{instrumental \\ variable}} = abs(\beta_{\substack{instrumental \\ variable}}) \sqrt{(\frac{se_{SNP \rightarrow GDF-15}}{\beta_{SNP \rightarrow GDF-15}})^2 + (\frac{se_{SNP \rightarrow CVD}}{\beta_{SNP \rightarrow CVD}})^2}$$

We also used pooled estimates based on all three SNPs. This was done using inverse variance weighted method for correlated variants to account for their correlations.(40)

Main results

Study I

In PIVUS, 29 proteins were associated with incident heart failure, after adjusting for age and sex at the 5% FDR threshold. Of these, 18 were positively associated with incident heart failure at the nominal significance level in ULSAM.

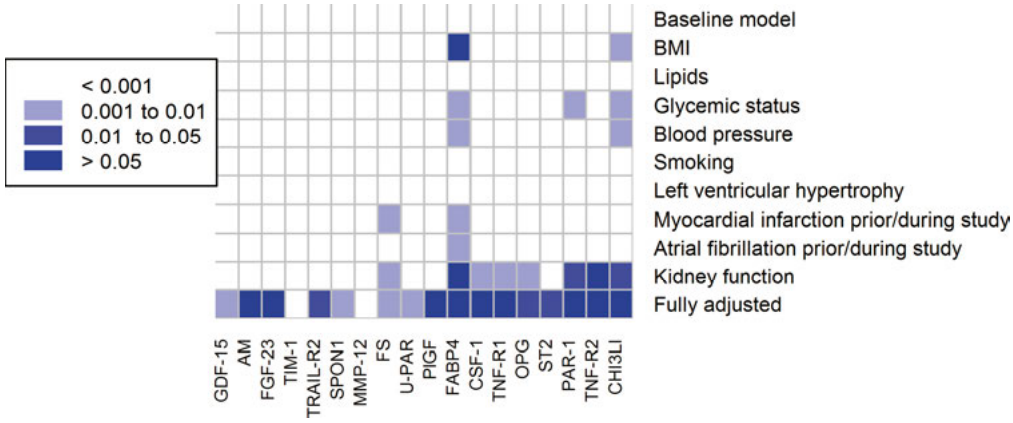
Additional multivariable adjustment models for established heart failure risk factors were performed after merging the two cohorts. Nine of the 18 proteins were positively associated with heart failure incidence in the fully adjusted model (**Figure 5**): GDF-15, T-cell immunoglobulin and mucin domain 1 (TIM-1), tumor necrosis factor-related apoptosis-inducing ligand receptor 2 (TRAIL-R2), spondin-1 (SPON1), matrix metalloproteinase-12 (MMP-12), follistatin (FS), urokinase plasminogen activator surface receptor (U-PAR), osteoprotegerin (OPG), suppression of tumorigenicity 2 (ST2). As seen in **Figure 5**, separate adjustments for heart failure risk factors did not strongly influence associations for most proteins. Kidney function (glomerular filtration rate) was the risk factor that influenced the associations most.

When NT-proBNP was added to the fully adjusted model for each of the nine proteins, MMP-12, OPG, U-PAR, and TIM-1 remained associated with incident heart failure ($P < 0.05$ for each).

Risk prediction

Lasso regression selected two sets containing 24 and 11 proteins as the optimal addition to the ARIC score when either excluding or including NT-proBNP levels in the random training set of 915 individuals. In the validation sample of 457 individuals, the model excluding NT-proBNP showed improvement in risk prediction after including the protein measurements (from AUC 0.751 to 0.852, change in AUC: 0.101, 95% confidence interval 0.030–0.173, p value = 0.006; change in model fit p value < 0.001). Discrimination in the model including NT-proBNP was not improved by added protein markers (from AUC 0.821 to AUC 0.841, change in AUC: 0.020, 95% confidence interval -0.027–0.068, p value = 0.40; change in model fit p value = 0.06).

Figure 5. Heat map of the associations between the 18 proteins that replicated in ULSAM and heart failure, after adjusting for different confounders. All models adjusted for age and sex (baseline model); associations were in a positive direction.



Footnote: Growth differentiation factor 15 (GDF-15), adrenomedullin (AM), fibroblast growth factor 23 (FGF-23), T-cell immunoglobulin and mucin domain 1 (TIM-1), TNF-related apoptosis-inducing ligand receptor 2 (TRAIL-R2), spondin-1 (SPON1), matrix metalloproteinase-12 (MMP-12), follistatin (FS), urokinase plasminogen activator surface receptor (U-PAR), placenta growth factor (PlGF), fatty acid-binding protein 4 (FABP4), macrophage colony-stimulating factor 1 (CSF1), tumor necrosis factor receptor 1 (TNF-R1), osteoprotegerin (OPG), suppression of tumorigenicity 2 (ST2), proteinase-activated receptor 1 (PAR-1), tumor necrosis factor receptor 2 (TNF-R2), chitinase-3-like protein 1 (CHI3LI).

Study II

In PIVUS, 16 proteins were associated with incident ischemic stroke after adjusting for age and sex at the 5% FDR threshold. Of these, three (NT-proBNP, adrenomedullin, and eosinophil cationic protein (ECP)) were associated with incident ischemic stroke at the nominal significance level in ULSAM, after adjustment for established risk factors.

In a meta-analysis of individual level data of the PIVUS and ULSAM cohorts, nine proteins were associated with ischemic stroke incidence, after adjustment for ischemic stroke risk factors at the Bonferroni threshold (**Table 2**).

Table 2. Association between proteins and incident ischemic stroke in the merged PIVUS and ULSAM cohorts. Only the proteins with an age-adjusted p value below the Bonferroni threshold ($p < 0.000588$) are shown. Hazard ratios (HR) and 95% confidence limits (CI) are given adjusted for age (only males in ULSAM), and after adjustment for age, LDL and HDL cholesterol levels, systolic blood pressure, BMI, diabetes, atrial fibrillation and smoking.

	Age-adjusted only		Multiple adjusted	
Protein	**HR (95% CI)**	**p value**	**HR (95% CI)**	**p value**
N-terminal pro-B-type natriuretic peptide (NT-proBNP)	1.49 (1.28, 1.72)	$1.30*10^{-7}$	1.43 (1.22, 1.67)	$8.03*10^{-6}$
Growth differentiation factor 15 (GDF-15)	1.37 (1.17, 1.61)	0.000089	1.33 (1.11, 1.59)	0.0017
Adrenomedullin (AM)	1.40 (1.18, 1.67)	0.00011	1.40 (1.16, 1.69)	0.00037
Interleukin-27 subunit alpha (IL27-A)	1.37 (1.16, 1.61)	0.00017	1.37 (1.14, 1.63)	0.00056
Urokinase plasminogen activator surface receptor (U-PAR)	1.38 (1.17, 1.64)	0.00018	1.41 (1.18, 1.69)	0.00020
Tumor necrosis factor receptor superfamily member 6 (FAS)	1.30 (1.13, 1.49)	0.00020	1.25 (1.08, 1.45)	0.0026
Macrophage colony-stimulating factor 1 (CSF-1)	1.37 (1.15, 1.62)	0.00030	1.30 (1.09, 1.55)	0.0039
Eosinophil cationic protein (ECP)	1.39 (1.16, 1.66)	0.00034	1.55 (1.28, 1.88)	$7.92*10^{-6}$
Matrix metalloproteinase-7 (MMP-7)	1.28 (1.11, 1.46)	0.00044	1.23 (1.06, 1.42)	0.0050

Risk prediction

In meta-analysis of the PIVUS and ULSAM studies, NT-proBNP, adrenomedullin, and ECP combined increased the area under the curve (AUC) from 0.629 (95% CI, 0.582–0.676) in a model adjusting for ischemic stroke risk factors (age, sex, LDL and HDL cholesterol levels, systolic blood pressure, BMI, diabetes mellitus, atrial fibrillation, and smoking) to 0.689 (95% CI, 0.641–0.737; $p = 0.0012$). The addition of NT-proBNP alone increased the AUC to 0.663 (95% CI, 0.616–0.709; $p = 0.030$), while adrenomedullin increased it to 0.657 (95% CI, 0.610–0.704; $p = 0.052$) and ECP to 0.656 (95% CI, 0.610–0.709; $p = 0.073$).

Study III

Three metabolites, urobilin, sphingomyelin (30:1), and ceramide phosphoethanolamine (34:1), were associated with incident heart failure after adjusting for age and sex at the 15% FDR threshold in both PIVUS and ULSAM (**Table 3**). Of these, urobilin and sphingomyelin (30:1) were also associated with incident heart failure at p value < 0.05 in TwinGene (**Table 3**).

Table 3. Association between metabolites and incident heart failure in PIVUS, ULSAM, and TwinGene – age- and sex-adjusted.

Metabolite	ULSAM and PIVUS HR (95% CI)	TwinGene HR (95% CI)
Ceramide phosphoethanolamine (34:1)	0.78 (0.70, 0.87)‡	0.86 (0.70, 1.05)
Sphingomyelin (30:1)	0.80 (0.72, 0.90)‡	0.82 (0.68, 0.99)*
Urobilin	1.45 (1.19, 1.76)‡	1.3 (1.05, 1.6)*

Data are hazard ratios (HRs) with 95% CI adjusted for age and gender, expressed per standard deviation increase of metabolite levels. ‡, p value < 0.001; †, p value < 0.01; *, p value < 0.05.

We pooled the three cohorts and performed multivariable adjustment for established heart failure risk factors. As seen in **Table 4**, adjustment for established heart failure risk factors had a modest impact on these associations.

Table 4. Association between urobilin and sphingomyelin (30:1), and incident heart failure in the PIVUS, ULSAM, and TwinGene cohorts.

Metabolite	PIVUS and ULSAM[a]	TwinGene[a]	Meta-analysis[a]	Meta-analysis, adjusted[b]
Urobilin	1.45 (1.19, 1.76)‡	1.29 (1.03, 1.63)*	1.38 (1.19, 1.60)‡	1.30 (1.10, 1.52)†
Sphingomyelin (30:1)	0.80 (0.72, 0.90) ‡	0.72 (0.58, 0.89)*	0.78 (0.71, 0.87)‡	0.85 (0.75, 0.95)†

Data are hazard ratio (HR) with 95% confidence intervals expressed per standard deviation increase of metabolite levels. *, p value < 0.05; †, p value < 0.01; ‡, p value < 0.001.

[a]Age- and sex-adjusted.

[b]Established heart failure risk factors: age, sex, BMI, LDL-C, HDL-C, triglycerides, lipid medication, diabetes, systolic and diastolic blood pressure, blood pressure medication, kidney function (glomerular filtration rate), smoking status, myocardial infarction prior to or during study (time updated).

Study IV

Coronary artery disease

No evidence of a causal effect on CAD was found (pooled estimate OR 1.00 (95% CI, 0.98, 1.02)), although we identified some heterogeneity in associations of SNP with CAD across the three SNPs.

Ischemic stroke

No evidence of a causal effect on ischemic stroke was found ischemic stroke (OR for IV estimator combined: 0.99 (95% CI, 0.96, 1.02)).

Heart failure

The pooled causal estimate was OR 1.19 (95% CI, 1.07, 1.34) per SD increase of GDF-15 (p value 2.28×10^{-3}) in line with a causal effect of GDF-15 on heart failure. There was no sign of heterogeneity driving this result (p value = 0.84).

Discussion

Principal findings

In **Study I**, we identified several novel associations between circulating proteins involved in apoptosis, inflammation, matrix remodeling, fibrinolysis, and incident heart failure, independent of established risk factors.

In **Study II**, we identified three proteins (NT-proBNP, adrenomedullin, and ECP) as being associated with incident ischemic stroke, independent of established risk factors. We also identified GDF-15 as a stroke risk marker candidate.

In both **Study I** and **Study II**, we observed potential to improve disease risk prediction using proteomics data.

In **Study III**, we performed mass spectrometry-based **metabolomic profiling** in plasma or serum samples from PIVUS, ULSAM, and TwinGene. The metabolites urobilin and sphingomyelin (30:1) were associated with incident heart failure, independent of established risk factors.

In **Study IV**, based on the results from Studies I–II, we performed **Mendelian randomization (MR)** analyses in 1,053,527 individuals. We found evidence supporting a causal role of genetically elevated GDF-15 levels in heart failure development, also our findings indicated that there was no causal link with coronary artery disease or ischemic stroke.

To the best of my knowledge, **Study I** was the first longitudinal association study between high-throughput proteomics and heart failure risk, and **Study III** was the first study to report urobilin and sphingomyelin (30:1) as heart failure risk markers.

Studies I–III demonstrate the possibility to identify new cardiovascular biomarkers that indicate specific subclinical cardiovascular concerns many years before diagnosis.

Biomarkers identification

Previous biomarker research has predominantly used targeted approaches, i.e. evaluated specific biomarkers that have been highlighted as causal factors in earlier experimental research. This approach has the benefit that it is often based on a clear hypothesis regarding the pathophysiological role of the specific biomarker. Since it is primarily based on previous experimental research,

the number of potential biomarker candidates which may add information about underlying mechanisms is limited. In fact, the targeted approach has only resulted in a small number of biomarkers being introduced to clinical practice, i.e., a successful progression from bench-to-bedside is rare.

In the present thesis, I have chosen a different approach for the discovery of novel biomarkers. I have taken advantage of recent developments in high-throughput proteomics and metabolomics, making it possible to perform simultaneous analyses of multiple biomarkers on a larger scale than previously feasible. This approach meant that we could evaluate several potential biomarkers at the same time, without taking any underlying hypotheses into account. One benefit of untargeted analysis is that it is possible to discover previously unknown targets that warrant further investigation. However, there are many limitations to this approach, including the risk of false positive findings due to the multiple testing problem, and a lack of a hypothesis. Therefore, a more conservative statistical analysis plan with external replication would be needed to draw firm conclusions, and experimental studies would be needed to understand the underlying pathology. There is no consensus on how to perform this kind of research, and this thesis could serve as example of how biomarker research can be performed in the future.

The thesis supports the notion that untargeted approaches are a promising approach in biomarker research. Continuous improvements in high-throughput proteomics and metabolomics suggest great potential to discover unknown biomarkers and underlying mechanisms that could be used for risk prediction purposes and as targets of interventions.

Risk prediction

New biomarkers are often evaluated based on their improvement of the prediction model beyond the established risk factors. This is based on the assumption that the established risk factors are optimal and generalizable to all populations, which in complex diseases such as cardiovascular disease, is uncertain and would require further investigation. When used for screening, markers that are not causally related to the outcome may still be beneficial, as they can be valid proxies for hard-to-measure underlying risk factors. There is also ongoing development of machine learning-based methods with potential to improve the methods used.

In **Studies I–II**, we showed that protein blood measurements improved the ability to differentiate between individuals with higher versus lower disease risk, several years before the events. Considering the high cost of each cardiovascular disease event, both economically and in healthy life-years lost, these results warrant further investigations as to whether or not it is possible to target high-risk individuals.

While our results were promising, larger studies are needed to confirm if these results could be translated into clinical utility.

Potential mechanisms

We also mapped the types of pathways that the identified biomarkers have been reported to be involved in, with regard to potential development of heart failure and stroke. These mechanisms include apoptosis, inflammation, immune response, signal transduction/regulation, matrix remodeling, myocardial stretch, and fibrinolysis.

Several biomarkers were found to be involved in the modulation of the nuclear factor kappa B (NF-kB) pathway (related to inflammation, immune response, and apoptosis), and several were involved in cardioprotective pathways, indicating that the associations found between increased circulating levels and increased disease risk might be due to compensatory mechanisms caused by an underlying impairment, rather than a direct detrimental effect.

The association between urobilin and heart failure risk was attenuated when adjusting for baseline NT-proBNP, suggesting that impaired left ventricular contractility may be an important mediating factor leading to heart failure.

However, the underlying mechanisms behind the associations observed in our studies are largely unknown and need to be confirmed in experimental studies.

Growth differentiation factor-15

In **Studies I and II**, we identified growth differentiation factor-15 (GDF-15) as a heart failure and ischemic stroke risk marker. Based on this, we decided to investigate whether or not the relationship between GDF-15 and cardiovascular disease was causal. **Study IV** serves as an example of how high-throughput techniques can be followed up and validated using MR analyses.

In **Study IV**, we found evidence of a causal effect of genetically elevated circulating levels of GDF-15 on the risk of heart failure, using a large-scale single locus two-sample MR study.

GDF-15 is a member of the transforming growth factor beta cytokine superfamily, and is produced in many different parts of the body and by many different cell types.(41) The expression is upregulated in myocardial and vascular cells upon oxidative stress and inflammation,(42) and circulating levels increase after tissue injury, ischemia, and inflammation, through release from macrophages,(43) cardiomyocytes, adipocytes,(44) vascular smooth muscle cells,(45) and endothelial cells.(46)

GDF-15 has been proposed to protect cardiac tissues via inhibition of ventricular hypertrophy, apoptosis,(47) and reduced thrombus formation.(48)

Higher levels of circulating GDF-15 have been associated with many heart diseases and co-morbidities, including atrial fibrillation,(49, 50) cardiomyopathies,(51) heart valve disease,(52) diabetes,(53) kidney disease,(54) and obesity.(55)

Our results also indicated that it is unlikely that circulating GDF-15 is causally involved with coronary artery disease or ischemic stroke. Previous observational studies have reported associations between higher GDF-15 levels and coronary artery disease and ischemic stroke,(32, 56, 57) but our data indicate that GDF-15 is a non-causal risk marker.

Previous community-based observational studies have reported associations between higher circulating GDF-15 levels and left ventricular dysfunction(58) and increased risk of incident heart failure,(59) respectively, and observational studies in heart failure patients have linked higher GDF-15 levels to more severe heart failure.(60-62)

The lack of a causal effect of GDF-15 on coronary artery disease in **Study IV** indicates that the association between genetically elevated circulating levels of GDF-15 and heart failure is predominantly mediated via pathways other than myocardial ischemia. Experimental studies based on expression analyses in myocytes and knockout mouse models suggest that intracellular GDF-15 holds a protective role in the myocardium following myocardial ischemia.(63) Thus, the discrepancy between our MR analysis and previous experimental animal studies might be due to differences in biological function of intracellular and extracellular GDF-15. This is also supported by a study in non-ischemic heart failure patients indicating that the circulating GDF-15 is likely produced in peripheral tissues and not in the heart.(64)

Strengths and limitations

The strengths of **Studies I–III** include the longitudinal study designs with long follow-ups, detailed assessments of participants, use of state-of-the-art omics methods, and a conservative discovery/replication approach in several independent cohorts.

Limitations include generalizability, since most participants were elderly Europeans, resident in a geographically defined part of Sweden, and the ULSAM cohort consisted of only men. We excluded individuals previously hospitalized for heart failure or stroke, but it is possible that some participants had asymptomatic forms of heart failure or stroke at baseline. We did not have data on primary care diagnoses or the incidence of subtypes or severity of heart failure, such as heart failure with preserved or reduced ejection fraction. The estimated hazard ratios should be interpreted with caution, as non-linear relationships may exist. The proteomic and metabolomic techniques used did not provide concentration units, making comparisons with clinically applied cut-offs difficult. Since the circulating biomarkers exhibit pleiotropic effects

on myocardial and vascular function, and in many cases are involved in other functions, much research is needed in order to understand the mechanistic pathways underlying these associations. Our observational studies cannot establish causality, but may indicate future directions for experimental studies into the causal mechanisms leading to cardiovascular disease.

The use of a discovery-replication design in **Studies I–III** limits the number of false positives, with the downside that it also limits power, increasing the risk of false negatives. Therefore, we also performed predefined meta-analyses.

Another concern in all population-based cohort studies is the healthy cohort effect. It is likely that subjects who participated in these cohort studies were healthier than those in the general population who chose not to participate. In the case of ULSAM, individuals who survived and agreed to participate again in the 70- and 77-year investigations used in our studies would also be healthier than the original cohort. This might lead to underestimation of the associations found in our studies, driving the results towards the null hypothesis.

Further, there are no studies showing that reducing the circulating levels of these biomarkers would influence the cardiovascular disease risk.

The **MR** study design in **study IV** has several strengths, making it possible to overcome the two major limitations of observational studies: reverse causality and confounding.(65) Since the genetic instruments explained a large proportion of the exposure, and the large sample size, the study had power to detect even a weak causal association between GDF-15 and each outcome. Other strengths include the use of a strong single locus genetic instrument (so-called *cis*-MR) to minimize bias from genetic effects on other intermediate exposures through signals in *trans*.(66) Moreover, by basing the discovery of the genetic instruments on cohorts independent from those where we estimated the effects of the genetic exposure, we limited the risk of inflated estimates due to Winner's curse.(67)

Limitations in the MR approach include that the number of events of heart failure was quite limited (n = 1,420). Since it is not possible to completely exclude the possibility that the results were affected by violations to some of the assumptions underlying MR, including undetected pleiotropy and remaining bias due to canalization (influence of environment on the observed association), our findings need to be validated.

Conclusions

The **main conclusions** of the four studies in this thesis were that we

1) Identified proteins and metabolites associated with incident heart failure and ischemic stroke, including GDF-15, using a high-throughput omics approach.
2) Saw potential to improve disease risk prediction for incident heart failure and ischemic stroke using proteomics data.
3) Identified novel associations between proteins involved in apoptosis, inflammation, matrix remodeling, fibrinolysis, and risk of developing cardiovascular disease.
4) Found evidence supporting a causal role of genetically elevated GDF-15 levels in heart failure development using a MR approach.

Future perspectives

Cardiovascular disease is a global health problem that needs to be addressed in terms of better understanding of disease mechanisms, drug target identification, and exploration of new therapeutic and preventive opportunities. Many of the underlying genetic and molecular causal pathways leading to cardiovascular disease are poorly understood, and still unknown pathways are likely to be discovered.

The four studies in this thesis contribute to increased understanding of the development of cardiovascular disease in multiple ways, and demonstrate the value of interdisciplinary research involving clinicians, clinical chemists, bioinformaticians, and epidemiologists. The biomarkers that were associated with incident heart failure and ischemic stroke may be involved in early disease development, which motivates further experimental investigations to determine if these biomarkers could be targets for drug development.

While these studies encourage further large-scale proteomic, metabolomic, and genetic studies to give new insights into heart failure and stroke pathogenesis and prediction models, more research is needed before these result could be applied in the clinical setting. This includes replicating in independent cohorts, and performing mechanistic studies in order to elucidate the causality of the observed results.

Rapid technological advances in metabolomics, proteomics, DNA sequencing, computational capacity and statistical methodology, in combination with an increased availability to large clinical datasets are dramatically improving the possibilities of understanding of the complex interplay between genes, proteins, metabolites and their consequences for cardiovascular disease. In this thesis, we performed separate analyses of specific "omics" such as proteomics and metabolomics. As the amount of biological information continues to increase, a system biology approach, where unsupervised machine learning methods such as neural networks might be preferred to identify novel patterns in the complex and multi-layered omics data.

Our findings encourage further large-scale proteomic, metabolomic, and genetic studies to give new insights into heart failure and stroke pathogenesis, and to further investigate the utility of different omic techniques in risk prediction. The fast advances in regard to high-throughput omics, larger datasets, and increased computer power are expected to lead to great improvements in several research areas, and to translate into benefits for global health.

Swedish summary

Kardiovaskulär sjukdom är en av de största orsakerna till global ohälsa och dödlighet. Nya riskmarkörer skulle kunna identifiera personer med extra hög risk för att utveckla sjukdom redan innan sjukdomen brutit ut. Hos dessa högriskindivider skulle en intensifierad förebyggande behandling göra extra stor nytta. Nya riskmarkörer skulle också kunna vägleda vilka behandlingsalternativ som ger bäst effekt och minst biverkningar.

Det **övergripande syftet** med denna avhandling var att undersöka sambanden mellan proteiner, metaboliter, och utvecklingen av hjärtsvikt och hjärninfarkt (ischemisk stroke). Specifika mål var att identifiera nya orsakssamband som leder till kardiovaskulär sjukdom, och att utvärdera mervärdet av att använda biomarkörerna vid riskbedömning i klinisk praxis.

I **studie I och II** analyserade vi ≥80 proteiner i blodet hos studiedeltagare i de svenska studiematerialen **PIVUS** (901 studiedeltagare, medianålder 70 år) och **ULSAM** (685 studiedeltagare, medianålder 77 år).

I **Studie I** identifierade vi nya samband mellan nio proteiner och utvecklingen av hjärtsvikt, varav ett var proteinet Growth differentiation factor-15 (GDF-15). Dessa 9 proteiner är bland annat involverade i celldöd (apoptos) och inflammation, vilket skulle kunna förklara kopplingen med utvecklingen av hjärtsvikt.

I **Studie II** identifierade vi flera proteiner som var associerade med utvecklingen av ischemisk stroke, inklusive GDF-15. Både **studie I och II** indikerade att man kunde förbättra bedömningen av sjukdomsrisk genom att mäta proteiner i blodet.

I **Studie III** mätte vi metaboliter i blod med metoden masspektrometri i de svenska kohorterna **PIVUS, ULSAM** och **TwinGene** (totalt 3,924 studiedeltagare). Metaboliterna urobilin och sphingomyelin(30:1) var associerade med utvecklingen av hjärtsvikt.

I **Studie IV** följde vi upp resultaten i **studie I och II**, och utförde Mendelsk randomisering i flera internationella studiematerial med mer än en miljon studiedeltagare. Mendelsk randomisering är en metod som använder genetiska variationer för att undersöka om en riskfaktor är kausalt kopplad till en sjukdom. Denna studie visar att genetiskt förhöjda nivåer i blodet av proteinet GDF-15 orsakar hjärtsvikt, men inte hjärtinfarkt eller ischemisk stroke.

Sammanfattningsvis identifierade vi flera nya samband mellan riskmarkörer och utvecklingen av hjärtsvikt och ischemisk stroke. Våra fynd talar för

att dessa proteiner och metaboliter kan vara involverade redan tidigt i sjukdomsförloppet. Vi såg också att dessa riskmarkörer har potential att förbättra riskbedömningen för både hjärtsvikt och ischemisk stroke i befolkningen.

Utifrån våra fynd skulle det kunna vara motiverat att utföra ännu större studier där man på ett mer detaljerat sätt kartlägger människans gener, proteiner och metaboliter för att få nya viktiga insikter om det komplexa samspelet som ligger bakom kardiovaskulära sjukdomar som hjärtsvikt och stroke.

Acknowledgements

I am deeply grateful to my main supervisor, Johan Ärnlöv for all support, enthusiasm and encouragement I have received throughout my PhD. My co-supervisor, Tove Fall has help me immensely throughout my PHD, and I am very lucky to have had her as a co-supervisor. I would also like to thank my co-supervisor Erik Ingelsson for always taking time out of your very busy schedule to give feedback and help whenever I have needed it.

I have had many great colleagues that have been with me throughout most of my journey, including Mwenya, Samira, Chris, Stefan, Mona-Lisa and Beatrice; and many more who have been part of the Molecular Epidemiology, cell research, and zebrafish groups during my PHD, including Emily, Charlotta, Andrea, Anastasia, Bene, Dan, Manoj, Marcel, Sari, Jenny, Tiffany and Åsa. Thank you all for your support, feedback and suggestions that have helped me improve my work substantially. I am grateful to Christian for contributing during our weekly meetings with his knowledge and different perspective.

I also have colleagues at EpiHubben who have made my lunch discussions enjoyable, energizing, and great learning experiences: Elise, Frida, Liisa, Susanne, Erika, Jonas, Adam, Anna-Karin, Eva, Carl, Karl, and Kalle.

Everyone that I have been involved with at the Department of Medical Sciences, at Uppsala University, and at the Science for Life infrastructure have help me a lot, and especially thank you Birgitta, Anna, Gabriella and Susanne, for always helping me with all my administrative questions and challenges.

My research would not have been possible without all the volunteers who participated in the studies in this thesis, the healthcare and administrative staff of the cohorts, my collaborators, co-authors, and all the funders who believed in us enough to fund our research.

I also want to thank my great family and friends, who have supported me throughout, and without whom I would not have made it this far. To my mother Gunilla who always asked to make sure I did everything correctly, and my father Dick, who always supported me in my various projects. Thank you both for always being there for me and cheering me on.

To my amazing partner Anna, who has made my life more amazing than it has ever been before.

My utmost thanks to everyone who in different ways made my journey wonderful.

References

1. Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet. 2018;392(10159):1736-88.
2. Ingelsson E. The Grand Challenge of Cardiovascular Epidemiology: Turning the Tide. Frontiers in cardiovascular medicine. 2014;1:2-.
3. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1859-922.
4. Chris55. Cardiovascular diseases world map. https://enwikipediaorg/wiki/File:Cardiovascular_diseases_world_map-Deaths_per_million_persons-WHO2012svg. 2019.
5. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Jr., Colvin MM, et al. 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Circulation. 2017;136(6):e137-e61.
6. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, et al. Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association. Circulation. 2017.
7. Cook C, Cole G, Asaria P, Jabbour R, Francis DP. The annual global economic burden of heart failure. Int J Cardiol. 2014;171(3):368-76.
8. Huffman MD, Berry JD, Ning HY, Dyer AR, Garside DB, Cai X, et al. Lifetime Risk for Heart Failure Among White and Black Americans. Journal of the American College of Cardiology. 2013;61(14):1510-7.
9. McMurray JJ, Petrie MC, Murdoch DR, Davie AP. Clinical epidemiology of heart failure: public and private health burden. Eur Heart J. 1998;19 Suppl P:P9-16.
10. Kannel WB, Belanger AJ. Epidemiology of Heart-Failure. American heart journal. 1991;121(3):951-7.
11. McMurray JJ, Stewart S. Epidemiology, aetiology, and prognosis of heart failure. Heart. 2000;83(5):596-602.
12. Chang PP, Chambless LE, Shahar E, Bertoni AG, Russell SD, Ni H, et al. Incidence and Survival of Hospitalized Acute Decompensated Heart Failure in Four US Communities (from the Atherosclerosis Risk in Communities Study). The American Journal of Cardiology. 2014;113(3):504-10.
13. Johnson CO, Nguyen M, Roth GA, Nichols E, Alam T, Abate D, et al. Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet Neurology. 2019.

14. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart Disease and Stroke Statistics-2019 Update A Report From the American Heart Association. Circulation. 2019;139(10):E56-E528.
15. Group BDW. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69(3):89-95.
16. Zheng Y, Yu B, Alexander D, Manolio TA, Aguilar D, Coresh J, et al. Associations Between Metabolomic Compounds and Incident Heart Failure Among African Americans: The ARIC Study. American Journal of Epidemiology. 2013;178(4):534-42.
17. Yu B, Li AH, Metcalf GA, Muzny DM, Morrison AC, White S, et al. Loss-of-function variants influence the human serum metabolome. Science advances. 2016;2(8):e1600800.
18. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov. 2016;15(7):473-84.
19. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. Nat Rev Drug Discov. 2013;12(8):581-94.
20. Fall T, Hagg S, Magi R, Ploner A, Fischer K, Horikoshi M, et al. The Role of Adiposity in Cardiometabolic Traits: A Mendelian Randomization Analysis. PLoS medicine. 2013;10(6).
21. Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, Leon DA. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. Am J Epidemiol. 2006;163(5):397-403.
22. Burgess S, Timpson NJ, Ebrahim S, Davey Smith G. Mendelian randomization: where are we now and where are we going? Int J Epidemiol. 2015;44(2):379-88.
23. Hedstrand H. A study of middle-aged men with particular reference to risk factors for cardiovascular disease. Ups J Med Sci Suppl. 1975;19:1-61.
24. Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. Arterioscler Thromb Vasc Biol. 2005;25(11):2368-75.
25. Magnusson PKE, Almqvist C, Rahman I, Ganna A, Viktorin A, Walum H, et al. The Swedish Twin Registry: Establishment of a Biobank and Other Recent Developments. Twin Res Hum Genet. 2013;16(1):317-29.
26. Ganna A, Salihovic S, Sundstrom J, Broeckling CD, Hedman AK, Magnusson PK, et al. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. PLoS Genet. 2014;10(12):e1004801.
27. 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 medicine. 2015;12(3).
28. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nature Genetics. 2013;45(1):25-U52.
29. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nature Genetics. 2018.
30. Ingelsson E, Arnlov J, Sundstrom J, Lind L. The validity of a diagnosis of heart failure in a hospital discharge register. Eur J Heart Fail. 2005;7(5):787-91.
31. The Task Force On Heart Failure Of The European Society Of C. Guidelines for the diagnosis of heart failure. European Heart Journal. 1995;16(6):741-51.

32. Lind L, Ärnlöv J, Lindahl B, Siegbahn A, Sundström J, Ingelsson E. Use of a proximity extension assay proteomics chip to discover new biomarkers for human atherosclerosis. Atherosclerosis. 2015;242(1):205-10.
33. Lundberg M, Eriksson A, Tran B, Assarsson E, Fredriksson S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. Nucleic Acids Res. 2011;39(15):e102.
34. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification. Analytical Chemistry. 2006;78(3):779-87.
35. Ganna A, Fall T, Salihovic S, Lee W, Broeckling CD, Kumar J, et al. Large-scale non-targeted metabolomic profiling in three human population-based studies. Metabolomics. 2016;12(1).
36. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
37. Agarwal SK, Chambless LE, Ballantyne CM, Astor B, Bertoni AG, Chang PP, et al. Prediction of incident heart failure in general practice: the Atherosclerosis Risk in Communities (ARIC) Study. Circ Heart Fail. 2012;5(4):422-9.
38. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73-9.
39. Jiang JY, Thalamuthu A, Ho JE, Mahajan A, Ek WE, Brown DA, et al. A Meta-Analysis of Genome-Wide Association Studies of Growth Differentiation Factor-15 Concentration in Blood. Front Genet. 2018;9.
40. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. Stat Med. 2016;35(11):1880-906.
41. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(16):6062-7.
42. Breit SN, Johnen H, Cook AD, Tsai VW, Mohammad MG, Kuffner T, et al. The TGF-beta superfamily cytokine, MIC-1/GDF15: a pleotrophic cytokine with roles in inflammation, cancer and metabolism. Growth Factors. 2011;29(5):187-95.
43. Bermudez B, Lopez S, Pacheco YM, Villar J, Muriana FJG, Hoheisel JD, et al. Influence of postprandial triglyceride-rich lipoproteins on lipid-mediated gene expression in smooth muscle cells of the human coronary artery. Cardiovascular Research. 2008;79(2):294-303.
44. Ding Q, Mracek T, Gonzalez-Muniesa P, Kos K, Wilding J, Trayhurn P, et al. Identification of Macrophage Inhibitory Cytokine-1 in Adipose Tissue and Its Secretion as an Adipokine by Human Adipocytes. Endocrinology. 2009;150(4):1688-96.
45. Schlittenhardt D, Schober A, Strelau J, Bonaterra GA, Schmiedt W, Unsicker K, et al. Involvement of growth differentiation factor-15/macrophage inhibitory cytokine-1 (GDF-15/MIC-1) in oxLDL-induced apoptosis of human macrophages in vitro and in arteriosclerotic lesions. Cell Tissue Res. 2004;318(2):325-33.

46. Ferrari N, Pfeffer U, Dell'Eva R, Ambrosini C, Noonan DM, Albini A. The transforming growth factor-beta family members bone morphogenetic protein-2 and macrophage inhibitory cytokine-1 as mediators of the antiangiogenic activity of N-(4-hydroxyphenyl) retinamide. Clin Cancer Res. 2005;11(12):4610-9.
47. Adela R, Banerjee SK. GDF-15 as a Target and Biomarker for Diabetes and Cardiovascular Diseases: A Translational Prospective. J Diabetes Res. 2015;2015:490842.
48. Rossaint J, Vestweber D, Zarbock A. GDF-15 prevents platelet integrin activation and thrombus formation. J Thromb Haemost. 2013;11(2):335-44.
49. Wallentin L, Hijazi Z, Andersson U, Alexander JH, De Caterina R, Hanna M, et al. Growth differentiation factor 15, a marker of oxidative stress and inflammation, for risk assessment in patients with atrial fibrillation: insights from the Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) trial. Circulation. 2014;130(21):1847-58.
50. Lind L, Sundstrom J, Stenemo M, Hagstrom E, Arnlov J. Discovery of new biomarkers for atrial fibrillation using a custom-made proteomics chip. Heart. 2017;103(5):379-84.
51. Montoro-Garcia S, Hernandez-Romero D, Jover E, Garcia-Honrubia A, Vilchez JA, Casas T, et al. Growth differentiation factor-15, a novel biomarker related with disease severity in patients with hypertrophic cardiomyopathy. Eur J Intern Med. 2012;23(2):169-74.
52. Ljungberg J, Janiec M, Bergdahl IA, Holmgren A, Hultdin J, Johansson B, et al. Proteomic Biomarkers for Incident Aortic Stenosis Requiring Valvular Replacement. Circulation. 2018;138(6):590-9.
53. Bao X, Borne Y, Muhammad IF, Nilsson J, Lind L, Melander O, et al. Growth differentiation factor 15 is positively associated with incidence of diabetes mellitus: the Malmo Diet and Cancer-Cardiovascular Cohort. Diabetologia. 2019;62(1):78-86.
54. Carlsson AC, Ingelsson E, Sundstrom J, Carrero JJ, Gustafsson S, Feldreich T, et al. Use of Proteomics To Investigate Kidney Function Decline over 5 Years. Clin J Am Soc Nephrol. 2017;12(8):1226-35.
55. Vila G, Riedl M, Anderwald C, Resl M, Handisurya A, Clodi M, et al. The relationship between insulin resistance and the cardiovascular biomarker growth differentiation factor-15 in obese patients. Clin Chem. 2011;57(2):309-16.
56. Lind L, Siegbahn A, Lindahl B, Stenemo M, Sundstrom J, Arnlov J. Discovery of New Risk Markers for Ischemic Stroke Using a Novel Targeted Proteomics Chip. Stroke. 2015;46(12):3340-7.
57. Tzikas S, Palapies L, Bakogiannis C, Zeller T, Sinning C, Baldus S, et al. GDF-15 predicts cardiovascular events in acute chest pain patients. PLoS One. 2017;12(8):e0182314.
58. Elmgren A, Larsson A, Lindahl B, Hulthe J, Lind L, Wallentin L, et al. Growth-differentiation factor-15 is an independent marker of cardiovascular dysfunction and disease in the elderly: results from the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) Study. European Heart Journal. 2009;30(19):2346-53.
59. Wang TJ, Wollert KC, Larson MG, Coglianese E, McCabe EL, Cheng S, et al. Prognostic utility of novel biomarkers of cardiovascular stress: the Framingham Heart Study. Circulation. 2012;126(13):1596-604.

60. Anand IS, Kempf T, Rector TS, Tapken H, Allhoff T, Jantzen F, et al. Serial measurement of growth-differentiation factor-15 in heart failure: relation to disease severity and prognosis in the Valsartan Heart Failure Trial. Circulation. 2010;122(14):1387-95.
61. Izumiya Y, Hanatani S, Kimura Y, Takashio S, Yamamoto E, Kusaka H, et al. Growth differentiation factor-15 is a useful prognostic marker in patients with heart failure with preserved ejection fraction. Can J Cardiol. 2014;30(3):338-44.
62. Chan MM, Santhanakrishnan R, Chong JP, Chen Z, Tai BC, Liew OW, et al. Growth differentiation factor 15 in heart failure with preserved vs. reduced ejection fraction. Eur J Heart Fail. 2016;18(1):81-8.
63. Kempf T, Eden M, Strelau J, Naguib M, Willenbockel C, Tongers J, et al. The transforming growth factor-beta superfamily member growth-differentiation factor-15 protects the heart from ischemia/reperfusion injury. Circ Res. 2006;98(3):351-60.
64. Lok SI, Winkens B, Goldschmeding R, van Geffen AJ, Nous FM, van Kuik J, et al. Circulating growth differentiation factor-15 correlates with myocardial fibrosis in patients with non-ischaemic dilated cardiomyopathy and decreases rapidly after left ventricular assist device support. Eur J Heart Fail. 2012;14(11):1249-56.
65. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat Rev Cardiol. 2017;14(10):577-90.
66. Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. Int J Epidemiol. 2016;45(5):1600-16.
67. Taylor AE, Davies NM, Ware JJ, VanderWeele T, Smith GD, Munafo MR. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. Economics and human biology. 2014;13:99-106.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1565

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-381234

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2019