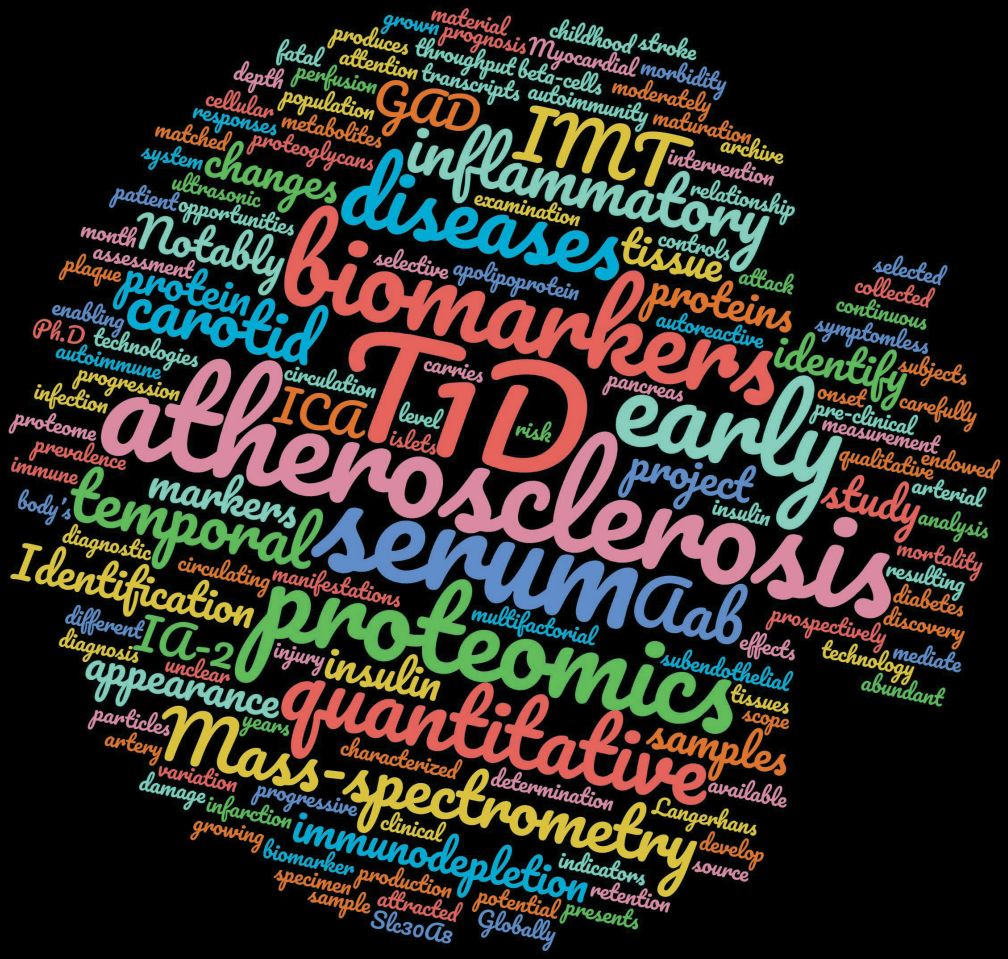
Turun yliopisto
University of Turku



SERUM PROTEOMICS TO DETECT EARLY CHANGES IN TYPE 1 DIABETES AND CAROTID ATHEROSCLEROSIS

Santosh D. Bhosale

TURUN YLIOPISTON JULKAISUJA - ANNALES UNIVERSITATIS TURKUENSIS
Sarja - ser. D osa - tom. 1381 | Medica - Odontologica | Turku 2018

Turun yliopisto
University of Turku

SERUM PROTEOMICS TO DETECT EARLY CHANGES IN TYPE 1 DIABETES AND CAROTID ATHEROSCLEROSIS

Santosh D. Bhosale

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
Sarja - ser. D osa - tom. 1381 | Medica - Odontologica | Turku 2018

University of Turku

Faculty of Medicine
Department of Medical Microbiology and Immunology
Turku Doctoral Programme of Molecular Medicine (TuDMM)
Turku Centre for Biotechnology, University of Turku and
Åbo Akademi University, Turku, Finland

Supervised by

Professor Riitta Lahesmaa, M.D., Ph.D.
Turku Centre for Biotechnology
University of Turku and Åbo Akademi
University Turku, Finland

Dr. Robert Moulder, Ph.D.
Turku Centre for Biotechnology
University of Turku and Åbo Akademi
University Turku, Finland

Reviewed by

Adj. Professor Maciej Lalowski, Ph.D.
Meilahti Clinical Proteomics Core Facility
Helsinki Institue of Life Science (HiLIFE)
Faculty of Medicine
Univerisity of Helsinki, Finland

Adj. Professor Tiina Öhman, Ph.D.
Institue of Biochemistry and
Helsinki Institue of Life Science (HiLIFE)
University of Helsinki, Finland

Opponent

Professor Ole N. Jensen, Ph.D.
Department of Biochemistry and Molecular Biology
University of Southern Denmark
Odense M, Denmark

Cover image by Santosh D. Bhosale (Created using https://www.wordclouds.com/)

ISBN 978-951-29-7386-6 (PRINT)
ISBN 978-951-29-7387-3 (PDF)
ISSN 0355-9483 (PRINT)
ISSN 2343-3213 (PDF)
Painosalama Oy – Turku, Finland 2018

To My Family

ABSTRACT

Santosh D. Bhosale

Serum Proteomics to Detect Early Changes in Type 1 Diabetes and Carotid Atherosclerosis

University of Turku, Faculty of Medicine, Department of Medical Microbiology and Immunology, Turku Doctoral Programme of Molecular Medicine (TuDMM), Turku Centre for Biotechnology, University of Turku and Åbo Akademi University

Turku, Finland, 2018

The detection of early markers is the key issue in predicting the outcome of inflammatory diseases such as type 1 diabetes and atherosclerosis. Whilst biochemical testing approaches have improved prediction of inflammatory diseases, validated biomarkers with better diagnostic specificities are still needed. Currently, majority of the disease-related proteomics studies have focused on their endpoints. The work presented in this thesis includes the first comprehensive proteomics analyses on serum samples collected from two unique Finnish longitudinal cohorts, namely The Diabetes Prediction and Prevention Project (DIPP) and The Cardiovascular Risk in Young Finns Study (YFS), to identify early markers associated with type 1 diabetes and carotid atherosclerosis.

Using mass spectrometry (MS)-based quantitative serum proteomics, profiling was carried out to the study temporal variation in pre-diabetic samples and early markers of plaque formation with the T1D and YFS cohorts, respectively. The analyses revealed consistent differences in the abundance of a number of proteins in subjects having an ongoing asymptomatic changes, several of which are functionally relevant to the disease process. Taken together, the discovered markers are candidates for further validation studies in an independent cohorts and may be used to characterize an increased risk, progression and early onset of these diseases.

Keywords: Serum, Proteomics, iTRAQ, Label-free quantification, T1D, Atherosclerosis, LC-MS/MS, SRM-MS

TIIVISTELMÄ

Santosh D. Bhosale

Tyypin 1 diabeteksen ja ateroskleroosin kehittymiseen liittyvät varhaiset muutokset seerumiproteomissa

Turun yliopisto, Lääketieteellinen tiedekunta, Lääketieteellinen mikrobiologia ja immunologia, Molekyylilääketieteen tohtoriohjelma, Turun biotekniikan keskus, Turun yliopisto ja Åbo akademi

Annales Universitatis Turkuensis, Suomi, 2018

Yksi keskeinen haaste tulehduksellisten sairauksien, kuten tyypin 1 diabeteksen ja ateroskleroosin, ennustamisessa on varhaisten tautimarkkerien löytäminen. Vaikka erilaiset biokemialliset testit ovat jo parantaneet tulehdusperäisten sairauksien ennustamista, uusia tarkempia biomarkkereita tarvitaan edelleen. Tästä huolimatta monissa näiden alojen proteomiikkatöissä on nykyisin keskitytty sairastumishetken tutkimiseen. Tämän väitöskirjatyön aikana olemme tehneet laajamittaiset proteomiikka-analyysit seeruminäytteille, jotka on kerätty osana kahta ainutlaatuista suomalaista seurantatutkimusta: DIPP-tutkimusta (tyypin 1 diabeteksen ennustaminen ja ennaltaehkäisy) ja YFS-tutkimusta (sydän- ja verisuonitautien riski nuorilla suomalaisilla). Näissä tutkimuksissa seerumiproteomiikkaa hyödynnettiin ensimmistä kertaa varhaisten tyypin 1 diabetes- ja ateroskleroosimarkkerien etsimiseen.

Tutkimme tyypin 1 diabeteksen kehittymiseen ja ateroskleroottisten plakkien muodostumiseen liittyviä muutoksia seerumin proteomiprofiileissa massaspektrometriaan perustuvan kvantitatiivisen proteomiikan avulla. Nämä analyysit paljastivat johdonmukaisia eroja lukuisissa proteiineissa myöhemmin sairastuneiden oireettomien henkilöiden ja terveinä pysyneiden kontrollien välillä. Monet näistä proteiineista saattavat myös liittyä olennaisesti tautien kehittymiseen. Tutkimuksissamme löydetyt markkerit tarjoavat lähtökohdan tuleville validointitutkimuksille, ja niitä voitaisiin tulevaisuudessa käyttää yksilön kohonneen sairastumisriskin, taudin etenemisen sekä taudin varhaisen puhkeamisen kartoittamiseen.

Avainsanat: Seerumi, proteomiikka, iTRAQ, Label-free kvantitointi, tyypin 1 diabetes, ateroskleroosi, LC-MS/MS, SRM-MS

TABLE OF CONTENTS

ABBREVIATIONS

1D-PAGE	One dimensional polyacrylamide gel electrophoresis
2D	Two dimensional
2D DIGE	Two-dimensional difference gel electrophoresis
2-DE / 2D PAGE	Two-dimensional polyacrylamide gel electrophoresis
Aab	Autoantibodies
Aab+ve	Autoantibody positive
Aab-ve	Autoantibody negative
ACS	Acute coronary syndrome
ADIPO	Adiponectin
AFAM	Afamin
AMT	Accurate mass and time tag
ag-ab	Antigen-antibody
APEX	Absolute protein expression
APOA1	Apolipoprotein A-I
APOA2	Apolipoprotein A-II
APOA4	Apolipoprotein A-IV
APOB	Apolipoprotein B-100
APOC1	Apolipoprotein C-I
APOC2	Apolipoprotein C-II
APOC3	Apolipoprotein C-III
APOC4	Apolipoprotein C-IV
APOE	Apolipoprotein E
BGH3	Transforming growth factor beta-induced protein ig-H3
C18	Octadecyl silica
CDH13	Cadherin-13
CDK9	Cyclin-dependent kinase 9
CID	Collision-induced dissociation
CNDP1	Beta-ala-his-dipeptidase
CO5	Complement component 5
CO8	Complement component 8
CO9	Complement component 9
CRP	C-reactive protein
DAISY	Diabetes Auto Immunity Study in the Young
DC	Direct current
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DIPP	Type 1 Diabetes Prediction and Prevention Project
EA	Endarterectomy
ECD	Electron capture dissociation

ELISA	Enzyme-linked immunosorbent assay
emPAI	Exponentially modified protein abundance index
ESI	Electrospray ionization
ETD	Electron transfer dissociation
EVs	Enteroviruses
FBLN1C	Fibulin 1 proteoform C
FDR	False discovery rate
FHR-5	Complement factor H-related protein 5
FTICR	Fourier transform ion cyclotron resonance
GAD	Glutamic acid decarboxylase
GSN	Gelsolin
HbA1C	Glycated hemoglobin
HCD	High-energy collisional dissociation
HDL	High-density lipoprotein
HLA	Human leukocyte antigen
IA-2	Protein tyrosine phosphatase
ICA	Islet-cell antibodies
ICAT	Isotope-coded affinity tags
IDDM1	Insulin-dependent diabetes mellitus 1
IGFBP2	Insulin-like growth factor-binding protein 2
IHC	Immunohistochemistry
IL12	Interleukin-12
IL6	Interleukin-6
IMT	Intima-media thickness
iRT	Index retention time
iTRAQ	Isobaric tags for relative and absolute quantification
LC-MS	Liquid chromatography coupled with mass spectrometry
LDL	Low-density lipoprotein
LFQ	Label-free quantification
LIT	Linear ion trap
m/z	Mass to charge ratio
MAC	Membrane attack complex
MALDI	Matrix-assisted laser desorption ionization
MARS	Multiple affinity removal system
MBL2	Mannose binding protein C
MHC	Major histocompatibility complexes
MMP2	72 kDa type IV collagenase
MMTS	Methyl methanethiosulphate
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MuDPIT	Multi-dimensional protein identification technology

NSAF	Normalized spectral abundance factor
OPN	Osteopontin
PAI	Protein abundance index
PFF	Peptide fragment fingerprinting
PFN1	Profilin-1
PMF	Peptide mass fingerprinting
PRM	Parallel reaction monitoring
PTMs	Post-translational modifications
Q	Quadrupole
QQQ	Triple quadrupole
RF	Radio frequency
ROC	Receiver operating characteristics
ROTS	Reproducibility optimized test statistics
RP	Reversed phase
SAX	Strong anion exchange
SCX	Strong cation exchange
SDS	Sodium dodecyl sulfate
SELDI	Surface-enhanced laser desorption ionization
SILAC	Stable isotope labeling with amino acids in cell culture
SMCs	Smooth muscle cells
SRM/MRM	Selected or multiple reaction monitoring
T1D	Type 1 diabetes
TEDDY	The Environmental Determinants of Diabetes in the Young
TMT	Tandem mass tag
TOF	Time of flight
TSP	Top scoring pairs
VLDL	Very-low-density lipoprotein
VTN	Vitronectin
YFS	The Cardiovascular Risk in Young Finn Study

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals I-III.

1. Moulder R, Bhosale SD, Erkkilä T, Laajala E, Salmi J, Nguyen EV, Kallionpää H, Mykkänen J, Vähä-Mäkilä M, Hyöty H, Veijola R, Ilonen J, Simell T, Toppari J, Knip M, Goodlett DR, Lähdesmäki H, Simell O, Lahesmaa R. Serum proteomes distinguish children developing type 1 diabetes in a cohort with HLA-conferred susceptibility. *Diabetes* 2015 Jun;64(6):2265-78.

2. Bhosale SD, Moulder R, Venäläinen MS, Koskinen JS, Pitkänen N, Juonala MT, Kähönen MAP, Lehtimäki TJ, Viikari JSA, Elo LL, Goodlett DR, Lahesmaa R*, Raitakari OT*. Serum Proteomic Profiling to Identify Biomarkers of Premature Carotid Atherosclerosis. *Scientific Reports* 2018 Jun 15;8(1):9209.

3. Bhosale SD, Moulder R, Kouvonen P, Lahesmaa R, Goodlett DR. Mass Spectrometry-Based Serum Proteomics for Biomarker Discovery and Validation. *Methods Mol Biol* 2017;1619:451-466.

1 INTRODUCTION

Inflammation can lead to increased cellular damage in diseases such as type 1 diabetes (T1D) and atherosclerosis (Donath *et al*, 2003; Libby, 2002). These inflammatory responses mediate their effects via circulation leading to infection and tissue injury (Medzhitov, 2008). Identification of changes in tissue and bodily fluid samples, e.g. serum, could provide important insights and/or useful markers of these diseases.

T1D is an autoimmune disease in which the immune system produces autoreactive T cells that can attack the insulin-secreting β cells in the islets of Langerhans of the pancreas, resulting in a progressive loss of insulin production (Van Belle *et al*, 2011a; Sparre *et al*, 2005). The appearance of autoantibodies (Aab) towards islet-cell proteins (ICA), protein tyrosine phosphatase (IA-2), glutamic acid decarboxylase (GAD), insulin and zinc transporter Slc30A8 protein have been identified as manifestations of β cell autoimmunity and increased T1D risk (Zhang *et al*, 2013b). However, the time from the appearance of first Aab to the onset of the clinical disease can vary from 1 month to over 10 years, moreover, not all Aab positive subjects develop T1D. Thus, additional indicators of the early disease process and progression are needed.

Atherosclerosis is a multifactorial disease that is characterized by selective retention of circulating apolipoprotein B particles in the subendothelial space by arterial wall proteoglycans (Williams & Tabas, 1995; Libby *et al*, 2002; Tabas *et al*, 2015). Myocardial infarction and stroke are the fatal end points of the disease. Notably, it has been documented that the atherosclerotic process starts in childhood and may remain symptomless for a long time (McGill H.C. *et al*, 2000). Identification of these early changes could provide new insights into the disease process and presents opportunities for an early intervention. Currently, ultrasonic examination of carotid artery intima-media thickness (IMT) is used as a pre-clinical assessment of atherosclerosis. However, the relationship between carotid plaque and IMT is still unclear, therefore additional biomarkers are required (Tonstad *et al*, 1996; Rundek *et al*, 2015).

Globally, the prevalence of T1D and atherosclerosis have been steadily growing, leading to increased morbidity and mortality. It is, therefore, important to identify biomarkers for the diagnosis and prognosis of these disease states. With current technology, the determination of such markers requires the study of carefully selected patient material and matched controls at a population level. To date, the biological sample of choice has been blood (plasma/serum), which has accordingly attracted much of attention as a source for biomarker discovery. Notably, it is readily available and carries a potential archive of biological information endowed to

it due to its continuous perfusion through the body's tissues (Rifai *et al*, 2006; Sparrow *et al*, 2011). This biological information is in the form of different cells, transcripts, proteins and metabolites, enabling the use of serum as a diagnostic specimen. With the maturation of mass spectrometry-based proteomics technologies, the scope for the qualitative and quantitative analysis of plasma/serum proteins has grown in terms of both throughput and depth (Aebersold & Mann, 2003, 2016; Geyer *et al*, 2017).

In this Ph.D. project, quantitative proteomics approach has been used to identify early serum protein biomarkers associated with two inflammatory diseases, type 1 diabetes, and carotid atherosclerosis. The project involved the proteomics measurement of prospectively collected serum samples and the study of the temporal variation in the moderately abundant serum proteome.

2 REVIEW OF LITERATURE

2.1 Introduction to proteomics:

The word "proteomics" was coined more than 20 years ago in a similar way to genomics, with the original meaning indicating to the study of the total observable proteins in a biological system (Graves & Haystead, 2002; Cox & Mann, 2011; Wilkins *et al*, 1996). In both genomics and proteomics research, significant efforts have been devoted to the building of thorough catalogs of biological information e.g. human genome and proteome sequencing. (Consortium, 2001; Venter, 2001; Wilhelm *et al*, 2014; Kim *et al*, 2014). Today the classical definition of proteomics is no longer restricted to identification of gene products only, but has grown to encompass quantification, modifications, localization, protein-protein interactions, proteoforms, and structural description studies. (Yates *et al*, 2009a; Aebersold & Mann, 2016; Smith & Kelleher, 2013). The revolution in protein identification came with the development of tandem mass spectrometry for sequencing, which has enabled large-scale characterization of proteins (Aebersold & Mann, 2003; Hunt *et al*, 1986). The main lines of investigation in mass spectrometry (MS) based proteomics includes: protein identification, characterization of post-translational modifications (PTMs), protein-protein interaction and quantitative measurements of abundance changes. The information on quantitative protein expression levels (relative or absolute) is prerequisite, especially in clinical proteomics (Aebersold & Mann, 2003; Ong & Mann, 2005). A proteomics experiment can be roughly divided into four stages, 1) Proteomics sample preparation, 2) Separation of proteins and peptides, 3) MS analysis and 4) Data analysis. These are considered as follows.

2.2 Proteomics sample preparation:

Proteins are functional molecules in the physiological system and are amongst the key interacting biomolecules in the complex network within the cell. There is not a single standard recipe for protein extraction and isolation that addresses all organelles and cellular subtypes, thus these require protocol optimization on an individual basis. A number of factors are taken into account for protocol optimization, which include the nature of the sample (physicochemical properties, expression level and location of proteins), experimental goals and analytical strategy (Bodzon-Kulakowska *et al*, 2007). Cell pellets, tissues and biological fluids

(plasma/serum, cerebrospinal fluid, urine and saliva) are commonly used as a protein source. To denature and solubilize proteins a lysis buffer recipe made up off a mixture of detergents and chaotropes (urea and thiourea) are used (Zhang *et al*, 2013c). The denatured proteins are then treated with reducing agents such as dithiothreitol (DTT) followed by alkylation with iodoacetamide (IAA) and finally overnight digestion with a suitable protease, e.g. trypsin (Switzar *et al*, 2013).

2.3 Separation of proteins and peptides:

Cellular lysates and biofluids mostly contain complex mixtures of thousands of proteins and their associated proteoforms in wide-ranging concentrations (Aebersold & Mann, 2016; Smith & Kelleher, 2013). Therefore, in order to examine such level of complexity various forms of sample separation are required. The choice of appropriate separation method is an important step while designing the proteomics experiment. The two major approaches widely used in proteomics are gel-based and chromatography-based methods.

2.3.1 Gel-based methods:

The use of two-dimensional polyacrylamide gel electrophoresis (2D PAGE) has the long-standing history in the characterization of protein mixtures (Klose, 1975; Klose & Kobalz, 1995; O'Farrell, 1975). In the first dimension proteins are separated by their isoelectric point (using an immobilized pH gradient), followed by separation according to their molecular mass in the second dimension. The combination of this orthogonal separation enables the separation of thousands of proteins/proteoforms in a single gel (Braun *et al*, 2007; Monteoliva & Albar, 2004). The separated proteins then can be visualized using Coomassie Brilliant Blue dye or silver staining. To enable the detection of proteins with PTMs, e.g. phosphoproteins and glycoproteins, Pro-Q Diamond and Pro-Q Emerald dyes were developed respectively. The separated proteins can then be identified after in-gel digestion with a suitable protease at the peptide level using matrix-assisted laser desorption ionization time of flight MS (MALDI-TOF-MS) or by liquid chromatography coupled with MS (LC-MS) (Aebersold & Mann, 2003). Despite its advantages for straightforward visualization, 2D-PAGE is limited by laboriousness, lengthy protocols and reproducibility (López, 2007). These limitations have been circumvented by the introduction of fluorescent two-dimensional difference in-gel electrophoresis (2D DIGE). By labeling the compared protein mixtures with distinct fluorescent dyes and separating these on a single 2D gel, this has drastically reduced problems of variation that are encountered with conventional 2 DE gels.

Dedicated software, e.g. DeCyder, is used for protein spot detection, data normalization and relative quantification (Gharbi *et al*, 2002; Unlu *et al*, 1997).

In the context of work presented in this thesis, studies of serum proteomics by gel electrophoresis have not been widely applied due to their inability in detecting low abundant proteins, time-consuming protocols and limited throughput especially when viewed in terms of the large-scale clinical biomarker studies (Anderson & Anderson, 2002a; Rabilloud & Lelong, 2011).

2.3.2 Chromatography-based methods

Chromatography-based methods have most frequently been used to separate mixtures of enzymatically digested proteins. The workflow involves denaturation, reduction, alkylation of proteins and digestion with a trypsin. The resulting peptides are then analyzed with the LC-MS/MS to infer the identities of the proteins present. The scope of proteins identified in complex mixture can be improved by the introduction of pre-fractionation approaches (Aebersold & Mann, 2016; Zhang *et al*, 2013c). Optimally, this should be pursued by use of an orthogonal mode of separation, such that the peptides are separated on the basis of a different mode of interaction. A popular non-gel based method amongst proteomics community has been multi-dimensional protein identification technology (MudPIT). The approach involves separation of the digested peptide by strong cation-exchange (SCX) in the first dimension and RP chromatography in the second dimension, followed by tandem mass spectrometry (MS/MS) analysis (Link *et al*, 1999; Washburn *et al*, 2001). The first dimension can also be anion exchange (SAX) (Holland & Jorgenson, 1995), a mixed bed (Motoyama *et al*, 2007), size exclusion (Wolters *et al*, 2001). MudPIT can be carried out either in manual (offline) or automated (online) modes.

The principle of separation in an ion exchange chromatography is based on the net charge of the peptides. The SCX mode of separation uses a negatively charged ion exchange resin to capture positively charged analytes and vice versa in case of SAX fractionation. The elution of peptides is then achieved with increasing salt concentration or by altering the pH of a mobile phase. (Fournier *et al*, 2007).

RP has been the most common mode of separation of peptides used in shotgun proteomics. With this mode of separation, the stationary phase is composed of a hydrophobic material, for example, C18 (Octadecyl silica) and the mobile phase is usually a mixture of water and an organic solvent such as acetonitrile along with formic acid. The peptides are separated based on the differences of their affinities between the stationary and mobile phase. The elution of the peptides is achieved by increasing the concentration of an organic solvent (Josic & Kovac, 2010).

Alternative methods have included the use of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC). In this mode of separation, the analyte and stationary phase bear the same charge. Although this causes electrostatic repulsion, there is hydrophilic interaction of the analytes with stationary phase due to the organic solvent in the mobile phase (Alpert, 2008). The analytes are then eluted by altering the pH of mobile phase.

2.4 Mass spectrometry in proteomics

Mass spectrometry has progressed in the last two decades and made a significant contribution as an analytical platform in many areas, notably in biological sciences. Proteomics, in particular, has advanced with the emergence of new MS techniques and improved instrumentation (Angel *et al*, 2012).

The mass spectrometer constitutes of three basic components: 1) ion source, 2) mass analyzer and 3) detector.

Briefly, in the ion source, ions are created in a fashion suitable for sampling, followed by their separation in the mass analyzer and finally detection to amplify the signals, which may then be presented as a spectrum. The spectrum is a plot of mass to charge ratios (*m/z*) on *x*-axis versus relative intensities on *y*-axis (El-Aneed *et al*, 2009).

2.4.1 Ionization methods:

The prerequisite is that the analyte must be charged or ionized in the vapor state to enable its analysis by a mass spectrometer. The charged ions are then directed by magnetic and electrical fields of the MS. In proteomics research, soft ionization techniques, i.e. electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI), are commonly used because of their capability for analyzing large molecular weight analytes (Siuzdak, 2004).

2.4.1.1 Electrospray ionization (ESI):

Following the development of the concept of ESI by Malcolm Dole, John Fenn and colleagues pioneered the method to ionize biomolecules for MS (Fenn *et al*, 1989). Briefly, the sample solution or chromatographic eluate is infused through an emitter into the ion source of the MS and a high potential difference (2-4 KV)

is applied to create a spray, which is then directed into the vacuum system. Here the droplets formed in the spray decrease in size (desolvation effect) due to the combined action of heat, vacuum and counter flow of gas. Following the subsequent coulombic explosion of the droplets, ions are transferred into the mass analyzer (Smith *et al*, 1991). Polypeptide ions are usually multiply charged. For tryptic peptides in positive ESI mode the ions are predominately doubly charged. This occurs since trypsin, cleaves the C terminus of arginine and lysine, resulting in peptides that have two sites available for protonation, i.e. N-terminus and a C-terminus (Krusemark *et al*, 2009).

2.4.1.2 Matrix-assisted laser desorption ionization (MALDI):

The basis for the development of MALDI is credited to Franz Gillenkmap and Michael Karas who demonstrated that a sample embedded in a matrix that absorbed radiation from a laser would result in better ionization (Karas & Hillenkamp, 1988). In recognition for the utility of this approach, Koichi Tanaka, who was the first to utilize a laser shot to send large biomolecules to the gas phase, shared half of the 2002 Noble prize with John Fenn, awarded for their separate work on the development of soft ionization methods for mass spectrophotometric analyses of biological macromolecules.

A sample is dissolved in an excess amount of matrix solution. The matrix is an organic compound, e.g. α-cyano-4-hydroxy cinnamic acid, 2, 5-dihydroxy benzoic acid or sinapinic acid, all containing delocalized *pi* bond electrons in the aromatic ring capable of absorbing laser UV energy. The analyte-matrix mixtures are subsequently spotted onto a MALDI plate, followed by drying and co-crystallization. The MALDI plate is then inserted into the mass spectrometer and ablated with a laser, usually nitrogen at λ=337 nm. The matrix absorbs laser energy, which is subsequently transferred to an analyte that then undergoes desorption and ionization. Unlike ESI, it produces mostly singly charged ions and is used for analysis of simple protein mixtures and intact protein analysis (Baldwin, 2005). A number of studies have combined LC with MALDI to characterize complex samples (Wall *et al*, 2002; Zhen *et al*, 2004).

Another variant of MALDI is Surface enhanced laser desorption ionization (SELDI). Here the sample mixture is applied to a modified surface or chip that captures the peptides or proteins and the unbound components are washed away. A matrix solution is then spotted on the sample for co-crystallization followed by analysis with time-of-flight mass spectrometer (Poon, 2007).

2.4.2 Different types of mass analyzers:

The mass analyzer is the heart of a mass spectrometer where ions are separated based on their *m/z* values. Different types of mass analyzers used in proteomics research include quadrupole (Q), linear ion trap (LIT), time of flight (TOF) and Orbitrap. These analyzers vary in terms of their size, mass range, mass accuracy and resolution. Many modern so-called hybrid MS platforms utilize the combination of at least two analyzers in one instrument (Yates *et al*, 2009b). In the work presented in this thesis quadrupole, time of flight, linear ion trap and Orbitrap analyzers were used, and thus will be discussed below.

2.4.2.1 Quadrupole analyzer:

Four parallel rods with a circular cross-section are combined to create the quadrupole. Two of these rods are supplied with the direct current (DC) potential, while the other two with alternating radio frequency (RF) potential. Using an electrical field, ions from the ionization source are directed towards quadrupole. The trajectory of ions is controlled by a combination of constant DC and RF potentials applied to the rods. The positively charged ions will get attracted towards negatively charged rods. However, upon changing the RF potential, ions will experience complex oscillations. With the appropriate values of a given set of parameters at any given time, only ions with a stable trajectory and a narrow *m/z* will survive the path. The other ions with unstable trajectories will eventually hit the rods. The ramping of these parameters results in the transmission of ions with different *m/z* towards the detector. (Dawson, 1986; El-Aneed *et al*, 2009).

2.4.2.2 Time of flight (TOF)

As the name indicates, the principle measurement is the time required for the ions to traverse a known flight path, e.g. 1-2 m in length. Ions of the same charge that enter the analyzer at the same time have the same kinetic energy, however, the time it takes for them to reach the detector is directly proportional to their *m/z* ratios. Unlike in quadrupole where it is necessary to scan a range of frequencies to ensure that ions are sampled, in TOF all ions reach the detector. The resolving power of TOF is improved when used in reflectron mode, where a constant electrostatic field is used to reflect the ion beam towards the detector,

effectively doubling the path length (Gelderman & Simak, 2008; El-Aneed *et al*, 2009). Typically, a TOF instrument can achieve 20-40,000 resolution (Beck *et al*, 2015).

2.4.2.3 Orbitrap mass analyzer:

The Orbitrap consists of an inner (central) and an outer electrode that traps the ions based on the principle of orbital trapping in electrostatic fields (Kingdon, 1923). The ions attain harmonic oscillation as they rotate around the central electrode with a frequency characteristic for their *m/z* values. Using a Fourier transform, an image current of these oscillations is then converted to the frequency spectrum, in a manner similar to the method used in Fourier transform ion cyclotron resonance (FTICR). The Orbitrap was invented and developed by Alexander Makarov (Makarov, 2000). The Orbitrap provides high resolving power (60-280,000), and mass accuracy, and has revolutionized the field of proteomics research (https://en.wikipedia.org/wiki/Orbitrap).

2.4.2.4 Linear ion trap (LIT):

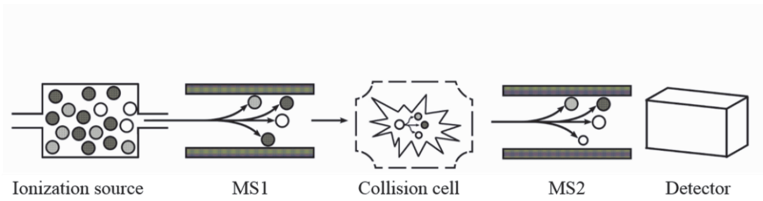
A LIT consists of four hyperbolic rods that are arranged around a central axis. RF voltages are applied radially to the rods, creating a trapping field that confines ions in both *x* and *y*-axes. DC potentials are added to the end electrodes that confine an ion in the *z*-axis. Thus, the combined effects of RF and DC potentials aid in trapping ions in all three dimensions. Linear ion traps or two dimensional (2-D) quadrupole ion traps have become a popular alternative to three-dimensional ion traps due to their higher ion storage capacities (Douglas *et al*, 2005; Schwartz *et al*, 2002).

2.4.3 Peptide fragmentation:

The conventional approach for tandem mass spectrometry (MS/MS) based proteomic analysis has been to use so-called data dependent methods to select the most abundant ions for fragmentation based sequencing (see below). Intact peptide ions are distinguished on the basis of their *m/z* in the first mass analyzer (MS1), followed by selection of the most intense ions (precursor ions) for dissociation in the

collision cell, usually by interaction with a neutral gas (e.g. helium). This fragmentation primarily occurs on polypeptide backbone. **Figure 1 A** represents a schematic of the tandem mass spectrometry mode and **Figure 1 B** depicts the nomenclature used to describe the fragmentation patterns. The most common ions formed as a result of cleavage of a peptide bond are the *b*- and *y*-types. Following cleavage, the charge can be retained on the fragment from the N terminus or C terminus, creating b- and y-type ions, respectively. The most common fragmentation technique used in shotgun proteomics analysis is collision-induced dissociation, (CID) (Cooks, 1995). In the third analyzer (MS2), the *m/z* of the charged fragments are recorded to provide a peptide fingerprint (Graves & Haystead, 2002; Steen & Mann, 2004; Zhang *et al*, 2013c). There are several other methods that can be used to fragment ions, resulting in different fragmentation types and characteristic ions. These methods include high-energy collisional dissociation (HCD) (Olsen *et al*, 2007), and electron capture and transfer dissociation (ECD and ETD)(Zubarev *et al*, 1998; Syka *et al*, 2004).

A



B

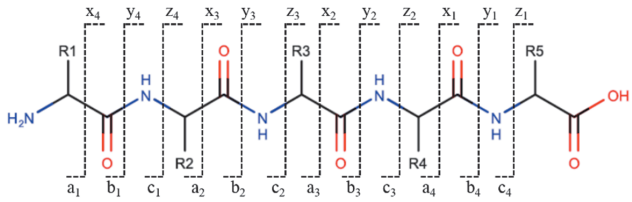


Figure 1: A) Schematic of the tandem mass spectrometry mode: ions are separated based on their *m/z* in MS1, followed by selection and fragmentation in collision cell and separation of fragments in MS2 before reaching to the detector. Adapted and modified from (El-Aneed *et al*, 2009). **B)** Representation of a peptide structure and the nomenclature used to describe CID fragmentation patterns.

2.4.4 MS-based data acquisition strategies:

Currently, the two most common MS-based approaches used for studying proteins are: "top-down" and "bottom-up" proteomics. The Top-Down approach involves analysis of the intact proteins and is not discussed in this thesis (Toby *et al*, 2016). "Bottom-up" proteomics is also referred to as shotgun proteomics (Zhang *et al*, 2013c), whereby peptides from enzymatically digested proteins, are identified and quantified using LC-MS/MS (Aebersold & Mann, 2016). Most "bottom-up" approaches operate in 2D space to identify and quantify peptides, in the first dimension by defining the chromatographic retention using LC, followed by precursor ion's *m/z* detection by MS in the second dimension. The approach can differ on the basis of the mode of selection and fragmentation of the precursors and the detection of fragment ion signals. This leads to the classification of "bottom-up" proteomics into data dependent, targeted and data independent methods as show in **Figure 2**. All of these approaches have been applied at all the different levels of the biomarker discovery pipeline (Sajic *et al*, 2015).

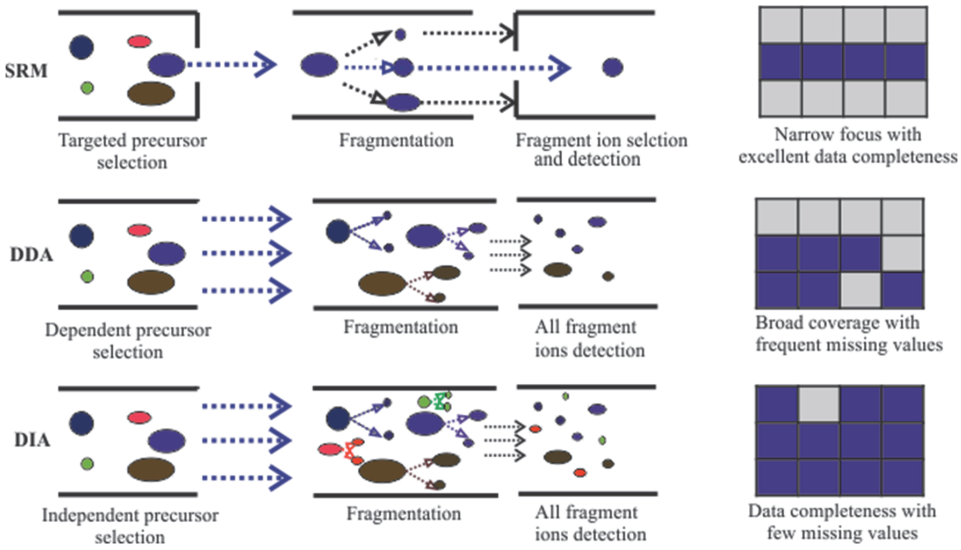


Figure 2: A schematic representation of how peptides are selected, fragmented and analyzed in SRM, DDA and DIA mass spectrometry modes.

2.4.4.1 Data-Dependent Acquisition (DDA)

The DDA mode of data acquisition has been the most widely used in discovery-based proteomics studies. The most abundant peptide precursor ions are selected

from the MS1 scan (survey scan), directed towards the collision cell for fragmentation by CID and peptide fragment ions spectra, i.e. MS2 or MS/MS or tandem spectra are, recorded. The identity of the peptide and its associated protein is then interpolated by comparison with simulated MS2 spectra from an in-silico digested protein sequence database (Domon & Aebersold, 2010a). The selection of 'n' precursor ions (TopN) is based on various factors such as ionization and peptides abundance. The acquisition parameter specifying *dynamic exclusion* helps to prevent the re-selection of same precursor ion in a given time frame, thus improving the scope for peptide identification. As this mode of acquisition relies upon on the fly selection of the most abundant peptides, it results in run to run stochasticity and leads to issues of missing values and questions on the reliability of peptide or protein identification. Typically, shotgun proteomic experiments do not require any prior knowledge of protein contents. (McDonald & Yates, 2002; Sajic *et al*, 2015).

2.4.4.2 Targeted proteomics:

The data acquisition mode traditionally used in targeted proteomics is selected or multiple reaction monitoring (SRM, MRM). Unlike DDA this requires the preselection of the measured proteins/peptides ions. SRM complements the findings of shotgun proteomics with its unique capability in quantifying analytes reliably. It is considered as a gold standard method of proteome quantification for predefined sets of proteins. Since the method is hypothesis-driven, it needs prior information about the target proteins, their proteotypic peptides and transition lists created from the precursor and fragment ion *m/z* values. In SRM, a predefined precursor ion and its corresponding fragment ions are usually monitored on a triple quadrupole (QQQ) MS instrument. Specifically, the first (Q1) and third (Q3) quadrupoles acts as mass filters for the precursor and its selected fragment ions respectively, whereas the second quadrupole functions as a collision cell. To validate a peptide target, a heavy synthetic equivalent of the corresponding peptide(s) including an isotopically labeled, heavy lysine or arginine for tryptic peptides, should be incorporated into the analyzed samples. Precursor and fragment ions selection at two mass filters with narrower *m/z* window imparts high selectivity to an SRM assay (Lange *et al*, 2008; Domon & Aebersold, 2010a; Sajic *et al*, 2015). A scheduled SRM assay includes the monitoring of precursor/fragment ion pairs (transitions) in combination with index retention time (iRT) peptides to establish retention time indices (Escher *et al*, 2012). Typically, many peptides are quantified in a single SRM experiment. Once the SRM chromatogram has been recorded, a data analysis pipeline, such as Skyline can be used to perform the peak detection and peptide quantification (MacLean *et al*, 2010a; Reiter *et al*, 2011).

The advent of newer MS instrumentation has revolutionized accurate mass measurements and high-resolution capabilities, thus enabling the targeted proteomics to be carried out on a quadrupole-Orbitrap configuration instruments. In parallel reaction monitoring (PRM) data acquisition, an Orbitrap replaces the Q3 quadrupole analyzer used in the analogous SRM configuration. The advantages of carrying out targeted proteomics on quadrupole-Orbitrap instruments over SRM include the selectivity provided with the creation of full scan MS/MS for each target and no obligation for selecting the best transitions. (Gallien *et al*, 2012; Peterson *et al*, 2012).

2.4.4.3 Data Independent Acquisition (DIA):

In DIA mode, the MS instrument fragments all precursor ions within a defined *m/z* range to generate multiplex MS2 spectra. The analysis is reiterated until the full *m/z* range is covered. Thus the continuous sequencing enables quantification with less missing values and without being restricted to the predefined selection of the 'n' most intense precursors of interest, like DDA (Chapman *et al*, 2013). However, the resulting MS/MS spectra become very complicated because of fragmentation of all peptides entering into MS in predefined *m/z* window, thus deconvolution of spectra is required. The latter has been reported as the greatest concern for the use of DIA, hence for the broader acceptance of DIA in routine practice, robust data analysis pipelines are required (Gillet *et al*, 2012a; Sajic *et al*, 2015).

2.4.4.4 Data acquisition strategies and their impact:

Ideally the data acquisition strategy selected should be capable of the quantification of as many of proteins as needed across multiple samples with sufficient accuracy (qualitative and quantitative), reproducibility and sensitivity (Domon & Aebersold, 2010b).

DDA is the most popular approach in discovery based proteomics, although suffers from the limitations of reproducibility and missing values. With this mode, precursor ions are monitored, followed by instrument driven dynamic selection of precursors for fragmentation and tandem mass spectra generation. Thus MS2 spectra are sampled discontinuously in the mass and time dimensions. However, due to the stochastic nature of the data acquisition, the reproducibility of both the qualitative and quantitative data influence the results of clinical biomarkers discovery

studies that involves the comparison of many samples (Hu *et al*, 2016). Nevertheless with modern fast scanning instruments, robust chromatographic systems and data analysis pipelines some of these problems can be alleviated.

The major strengths of the SRM method stem from its targeted approach. A list of precursor ions can be isolated and fragmented to measure the intensity of selected fragments. This can be further scheduled across the chromatographic window to enable the sensitive detection of larger number of targets. In this way, SRM analysis can provide consistent and accurate quantification, although is limited in the number of targets that can be determined (Domon & Aebersold, 2010b; Lange *et al*, 2008).

Following early approaches towards the multiplexed acquisition of tandem mass spectra (Gorshkov *et al*, 1998; Masselon *et al*, 2000), the term DIA was first introduced by Venables *et al.* to describe a data acquisition method that was explicitly not data dependent, i.e. not stochastic like DDA (Venable *et al*, 2004). The DIA concept has been developed over the years since its introduction, including the implantation of several difference approaches such as MS^E (Silva *et al*, 2005), PA-cIFIC (Panchaud *et al*, 2009), AIF (Geiger *et al*, 2010) and SWATH (Chapman *et al*, 2013; Gillet *et al*, 2012a). In DIA, the instrument performs untargeted data acquisition like DDA. The DIA process involves comprehensive sampling in MS2 of the precursor space.

The commercial introduction of SWATH (Sequential window acquisition of all theoretical fragment ion spectra), heralded a new era of DIA method combining the strengths of both discovery and targeted approaches i.e. it can provide comprehensive record of the proteins mixture reproducibly with quantitative accuracy. Using wide precursor isolation windows, all precursors falling within the wide precursor isolation windows are selected, fragmented and multiplexed tandem mass spectra are recorded sequentially until the desired *m/z* range is covered (e.g. 400-1200 *m/z* range with 25 Da windows). Thus continuous MS/MS signals are acquired in the time and *m/z* dimensions, resulting in complex mixed MS2 spectra (Gillet *et al*, 2012b). In effect SWATH, provides a near complete digital record of each biological sample, i.e. every peptide will be fragmented and their ion spectrum is acquired which can be compared and re-interpreted on the basis of future knowledge (Guo *et al*, 2015).

As discussed, the tool box for mass spectrometry-based proteomics now provides a range of data acquisition approaches with their individual strengths and weaknesses. For the bulk of the work presented in this thesis a DDA workflow was selected, since, at the time of this study, the approach provided a scalable, reproducible method with a sufficiently matured data analysis pipeline and was suitable

for the available instrumentation. Subsequent validation measurements were made using SRM.

2.4.5 Spectral identifications and database search engines:

The commonly used methodologies for protein identification are 1) peptide mass fingerprinting 2) peptide fragment fingerprinting and 3) *de novo* sequencing.

1) Peptide mass fingerprinting (PMF): Proteolytic digestion of proteins with a specific protease results in generation of a series of peptides with predictable masses. The simultaneous determination of these masses by MS analysis creates a mass fingerprint unique to a particular protein. To identify a protein after MS analysis, proteins from a sequence database are in silico digested to produce theoretical masses, which are then compared with the observed masses to establish a possible identity. The reliability of protein identification depends upon the probability of matching between observed and theoretical masses (Pappin *et al*, 1993; James *et al*, 1993)

2) Peptide fragment fingerprinting (PFF): PFF is currently the main approach in identification of proteins in high throughput manner. Proteins are digested into peptides with a specific protease, e.g trypsin. The digested peptides are analyzed by MS where the peptides are fragmented to produce PFF. Information from both the precursor ion and its fragment ions masses are used in searching the protein sequence database. The identification of proteins from this type of search is based on scoring systems that determine the probability of the best-matched candidate versus the other alternatives (McHugh & Arthur, 2008). SEQUEST (Eng *et al*, 1994), Mascot (Perkins *et al*, 1999) and Andromeda (Cox *et al*, 2011b) have been the most widely used search engines for protein identification.

3) De novo sequencing: This strategy is used when there is insufficient information about the proteome of an organism studied. With this approach, information is deduced from the experimental tandem (MS/MS) spectra of a peptide without the use of a sequence database. Making use of the characteristic patterns of y and b ion series from

tryptic peptides and the mass differences between different combinations of amino acids, the peptides sequence can be derived. Prior to the development of database search engines this was the principle method of MS based peptide identification (Shevchenko *et al*, 2002; Taylor & Johnson, 1997; Ma *et al*, 2003). The interpretation can be conducted in both manual and automated ways. PEAKS (http://www.bioinfor.com/) and Lutefisk (https://omictools.com/lutefisk-tool) are examples of commercial and free software packages used for *de novo* sequencing, respectively.

2.4.6 Quantitative proteomics:

Many first-generation proteomics studies focused on enumerating the proteins present in the biological system studied. With today's technology, it has become possible to provide both quantitative and qualitative data in a relatively high throughput manner. This shift was substantially driven by the advent of MS instrumentation and related sample preparation, chromatographic separation and data processing methods (Aebersold & Mann, 2016; Bantscheff *et al*, 2012).

Quantitative proteomics data is prerequisite to address many biological and clinical questions. Quantitative proteomics can be either absolute or relative. Absolute quantification measures the absolute protein content in a given sample, e.g. the concentration of a protein or number of its copies per cell. Relative quantification measures the relative abundance of a protein between two or more samples (Nikolov *et al*, 2012). The quantification approaches used in mass spectrometry-based proteomics can be divided into two categories, label-based and label-free quantification (Bantscheff *et al*, 2007). The work presented in this thesis involved the use of both label-based and label-free relative quantification, as are discussed in the following sections.

2.4.6.1 Label-based quantification:

This method usually involves chemical or metabolic labeling of peptides and proteins, which introduces a specific mass tag, enabling the different forms to be distinguished by MS and providing a means of quantification (Bantscheff *et al*, 2012; Strassberger *et al*, 2010). Common examples of chemical and metabolic labeling approaches are included in **Table 1**.

Table 1. The commonly used labeling methods in proteomics

Labelling technique
Chemical: • Isotope-coded affinity tags (ICAT) (Gygi *et al*, 1999a) • Isobaric tags for relative and absolute quantification (iTRAQ) (Ross *et al*, 2004b) • Tandem mass tag (TMT) (Thompson *et al*, 2003a)
Metabolic: • Stable isotope labeling with amino acids in cell culture (SILAC) (Ong *et al*, 2002)

Chemical labelling: The reactive group of proteins and peptides are modified using stable isotope labels. The example includes ICAT, iTRAQ and TMT.

The original structure of ICAT reagent chemically consists of three elements (**Figure 3**), 1) A thiol reactive group to bind with cysteine containing amino acids in proteins, 2) a linker (heavy or light) to generate characteristics stable isotope signature and 3) a biotin tag to affinity purify the ICAT labelled peptides using avidin.

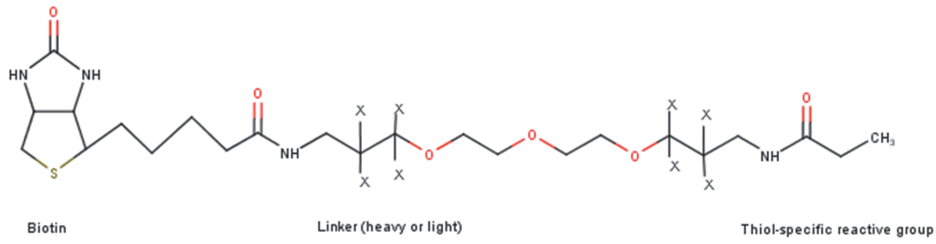


Figure 3. Chemical structure of ICAT reagent. The reagent combines biotin, a linker and a thiol reactive group. X = hydrogen (light) or deuterium (heavy).

For quantitative comparison between two conditions, one of the sample is labelled with light form of the reagent and other sample is labelled with heavy linker. After labelling, the samples are mixed, digested with trypsin followed by avidin affinity purification of ICAT labelled peptides. These peptides are then further analyzed using LC-MS/MS and quantitative information is then extracted from ratios of differentially labelled tags (Gygi *et al*, 1999b). However, the usefulness of ICAT reagents in clinical biomarker discovery project is limited due to its restricted multiplexing capacity and inability to label non-cysteine containing residues in proteins.

In the work presented in this thesis, iTRAQ reagents have been used. iTRAQ is a chemical labeling method which inserts a stable isobaric mass tag at the peptide N-terminus and the epsilon-amino group of lysine residues (Ross *et al*, 2004a). The iTRAQ reagent was commercialized by AB Sciex and is available as 4-plex and 8-plex kits. These can be used to provide quantitative information from four

up to eight different biological conditions. The structure of the iTRAQ reagent consists of three components: a reporter group, a balance group and a peptide reactive group (N-hydroxysuccinimide), as shown in the **Figure 4**. Using different combinations of carbon, nitrogen and oxygen isotopes the distinct isobaric forms of the reagent and components are formed. The reporter ion masses for 4-plex are detected at *m/z* 114-117, with the balance group mass ranging from 28-31 Da. Likewise, the reporter ion masses for the 8-plex reagents are detected at *m/z* 113-119 and 121 with the balance group's mass ranging from 24-31 Da. AB Sciex chose not to create a reagent form where a reporter ion 120 was used because phenylalanine immonium ion is detected at m/z 120.08 (Rauniyar & Yates, 2014). The 4-plex or 8-plex iTRAQ reagents have isobaric masses of 145 or 305 Da, respectively (Beck *et al*, 2012).

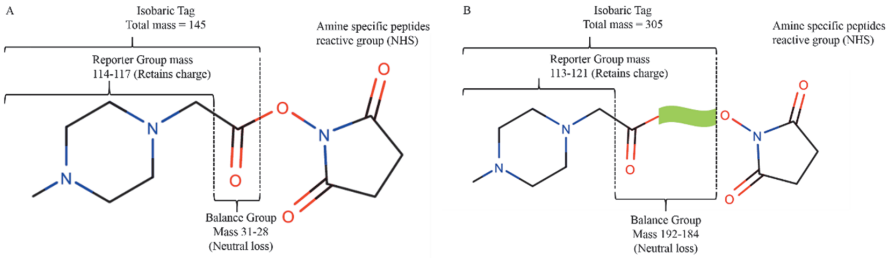


Figure 4. Chemical structure of A) 4-plex iTRAQ reagents and B) 8-plex iTRAQ reagents. The molecule consists of a reporter group (N-methylpiperazine), a balance group (carbonyl) and an amine specific peptides reactive group (NHS).

Peptides from the samples to be studied are first labelled with different forms of the iTRAQ reagents. Thereafter, the samples are pooled. For most comparisons, it is important to reduce the complexity by some form of fractionation (orthogonal). This enables an increased scope for detection of the lower abundant proteins and reduces the influence of chimeric spectra on the quantification. SCX fractionation has been a popular and vendor recommended method. Basic pH reversed-phase fractionation has emerged as a favorable alternative. The fractions are then analyzed with LC-MS or MALDI system. The labeled peptides exhibit same physicochemical properties and during chromatographic separation co-elute as a single precursor in the MS1 scan. Upon fragmentation, MS/MS gives information for the peptide/proteins identification and at the lower *m/z* value the quantitative information can be interpolated from the reporter group ion intensities (Moulder *et al*, 2017a; Rauniyar & Yates, 2014).

The labelling of proteins by iTRAQ had been demonstrated (Wiese *et al*, 2007), although the approach is limited since the labeling of lysine affects the efficiency of trypsin digestion.

Analogous to iTRAQ labelling, tandem mass tags (TMT) are used to label the peptides or proteins for MS based quantification. TMT reagents are available for duplex, sixplex and 10plex comparisons. The concept of TMT tags was first described by Thompson *et al.* prior to the introduction of iTRAQ (Thompson *et al*, 2003b) and later commercialized by Thermo Scientific.

Metabolic labelling: The SILAC technique has been a popular method for the quantitative proteomics analysis of cell cultures due to its accuracy and robustness. It has been applied for example to study cell signaling, PTMs such as phosphorylation, secretomes and protein interactomics. Typically cell populations are grown in culture media containing isotopically distinct amino acids. For binary comparisons, one of the cell population is cultured in growth media containing naturally occurring amino acids, whilst the other is in media containing heavy counterpart of natural amino acids. e.g., arginine and lysine labelled with ^{13}C and ^{15}N. Finally, the two cell populations are combined, processed, analyzed by LC-MS/MS and can be differentiated based on their differences in mass of stable isotope composition. However the technique is unsuitable for analyzing the body fluid samples (Ong *et al*, 2002; Iliuk *et al*, 2009).

In the work presented in this thesis, the iTRAQ labelling strategy was adopted to detect early changes in the serum proteome of T1D developing children because of its robustness and multiplexing capabilities.

2.4.6.2 Label-free quantification (LFQ):

There are two principal approaches to protein Label-free quantification 1) By counting the number of MS/MS events assigned to peptides of a particular protein, named as spectral counting, and 2) based on the measurement of the signal intensity of the mass spectrometric chromatographic peak of a peptide belonging to a specific protein (Bantscheff *et al*, 2007). LFQ approaches are applicable to any sample type, provided that the sample preparation and analytical workflow is reproducible. Moreover, these approaches do not need any labeling strategy, although the data-processing pipeline (e.g. Normalization, retention time alignment, feature detection, feature quantification, feature matching and statistical analysis) is of utmost importance (Mueller *et al*, 2008).

1) Spectral counting based LFQ: The basis for this form of relative protein quantification is that the frequency and number of MS/MS identified for a particular protein can be correlated with its abundance. Thus, a protein with an increased abundance will have a

larger number of unique peptides identified and a higher MS/MS count (Zhu *et al*, 2010; Zhang *et al*, 2013c; Bantscheff *et al*, 2012). The LFQ calculation to estimate the protein abundance is performed by computing a protein abundance index (PAI), which is defined as the number of identified peptides divided by the number of theoretically observed tryptic peptides for a particular protein. Although convenient and fast, the LFQ strategy suffers from a number of drawbacks. The main limitation is the quality of MS/MS spectra, as errors in assigning peptides directly affect the protein quantification. The PAI of high abundant proteins can be calculated with greater accuracy, however, for low abundant proteins, it becomes unreliable. Another important pitfall that is common with other quantification methods is that how to consider spectral counts for peptides shared between proteins ((Strassberger *et al*, 2010; Zhang *et al*, 2013c). To overcome this, a number of methods have been implemented, including the *exponentially modified PAI* (emPAI) (Ishihama *et al*, 2005), *normalized spectral abundance factor* (NSAF) (Zybailov *et al*, 2006) and *absolute protein expression* (APEX) (Braisted *et al*, 2008).

2) Signal intensity based LFQ: In an LC-MS run peptides elute within a particular time window that is recorded by its changing intensity at a particular *m/z* (Zhu *et al*, 2010). The computation of LFQ intensities involves extraction of ion chromatogram for each peptide and integration of their mass spectrometric signals in chromatographic time scale across the LC-MS runs (Strassberger *et al*, 2010). Relative LFQ information is then obtained by comparing the intensity of a peptide/protein from one run with its respective feature from other runs (Bantscheff *et al*, 2007). However, a number of factors must be taken into account. These include the following:
    - The use of high-resolution MS is important to reduce co-eluting peptides interference,
    - The chromatographic separation reproducibility is the key element in controlling the drift in retention time of corresponding peptides across many LC-MS runs,
    - The MS sensitivity should remain constant between successive sample analyses,
    - A sufficient number of MS scans are recorded across the chromatographic peak.

Finding the right balance between the number of MS1 and MS2 scans for signal based quantification, and the protein identification rate is imperative. The issues such as retention time alignment and mass calibration have been addressed by numerous computational methods (Strassberger *et al*, 2010; Zhang *et al*, 2013c).

In the work presented in this thesis the label free quantification approach was selected in as a subsequent method to the iTRAQ labelling approach employed to monitor the early changes in serum proteome of children *en route* to T1D. The change of strategy was due to its ease of implementation, the matured data analysis pipeline, complementary information, scalability and cost effectiveness.

2.4.7 Bioinformatics and data analysis:

Computational methods and bioinformatics have been a central component for generating useful biological information from proteomics experiments. Starting from the first automated peptides sequencing algorithms, it has become increasingly necessary to combine the qualitative and quantitative data from larger and larger data sets (Nesvizhskii, 2010). Both commercial and open source software has been developed with increasing capabilities. For example, ProteinPilot™ was developed by ABSciex for processing iTRAQ data, including the novel Paragon search algorithm (Shilov *et al*, 2007a) and tools for the quantification (used in analysis of iTRAQ data presented in this thesis). The transproteomic pipeline project was initiated as an open source tool to provide a suite of tools for processing data irrespective of the initial data format (i.e. vendor independent) (Kohlbacher *et al*, 2007). The latter was created with the view to use a generic data format and further enable data sharing and archiving. A range of similar and related tools have emerged in recent years, including, for example, Galaxy (Goecks *et al*, 2010), Progenesis (Dakna *et al*, 2009) developed by Nonlinear Dynamics and MaxQuant (Cox & Mann, 2008a). The latter two were used to process the label free data analyzed in the work presented in this thesis, especially MaxQuant has been widely adopted due to its speed and efficacy (freeware). MaxQuant was developed with the stand alone Andromeda search algorithm (Cox *et al*, 2011a) and a downstream data processing package, Perseus (Tyanova *et al*, 2016). Perseus, in particular, enables quick execution of a range of data processing functions including normalization, quality checking, statistical testing and functional annotation.

Downstream of qualitative and quantitative proteomics, additional bioinformatics tools are often used for functional enrichment analysis. In the work presented in this thesis, The Database for Annotation, Visualization and Integrated Discovery (DAVID) (Jr *et al*, 2003).

Furthermore commercial softwares are available for analyzing SRM data, such as Multiquant (AB Sciex) and Pinpoint (Thermo Scientific) (Colangelo *et al*, 2013), as well as open source platforms e.g. Skyline (MacLean *et al*, 2010a). The latter freeware was used for analyzing SRM data presented in this thesis with statistical analysis utilizing the in-built MSstats application (Choi *et al*, 2014).

2.4.8 Quantitative proteomics for plasma/serum biomarker discovery:

In the field of medicine, the clinical chemistry of proteins has achieved significant importance. Currently, protein specific measurements are used to assist in the diagnosis of a number conditions, as indicated in the following examples:

- Prediction of a disease risk, e.g in coronary disease, C-reactive protein (CRP)1 levels increase,
- Diagnosis of acute events, e.g. after myocardial infarction, detection of cardiac troponin
- Detection of disease reoccurrence, e.g. detection of thyroglobulin after removal of the thyroid gland in metastatic thyroid cancer.

These predictive measures have increased the hope for clinical diagnostics using proteins for many disease manifestations, leading to significant attention towards novel biomarkers discovery (Anderson, 2010).

The term *biomarkers* refers to measurable biological molecules that define the healthy and diseased condition. Significant efforts have been devoted towards the use of omics technologies (genomics, transcriptomics, proteomics and metabolomics) for the discovery of new biomarkers. Protein markers, in particular, can be seen as informative and suitable, because of their functionality and accessibility by less invasive methods. (Lyons & Basu, 2012; Sajic *et al*, 2015). Established protein biomarkers are typically measured by an enzyme-linked immunosorbent assay (ELISA) in body fluids, such as blood or urine or cerebrospinal fluid, to provide indications of the subclinical or clinical status of the disease or for monitoring the treatment response. For protein biomarker discovery, MS-based proteomics holds an influential role in biomedical research (Parker *et al*, 2010; Geyer *et al*, 2017).

2.4.8.1 Plasma/serum proteomics and challenges associated with biomarker discovery:

Since the end-point of the biomarker discovery process is typically the development of a classical blood test, blood, plasma or serum are the obvious choices of

biofluids for biomarker discovery (Rifai *et al*, 2006). *Serum, the liquid portion of blood that is devoid of clotting factors. Plasma, the liquid of portion of blood that is pale yellow in color and holds the blood cells in suspension* (Kwasnik *et al*, 2016). *Serum is used as a general term for plasma and serum throughout this thesis, although the appropriate form is specified when referring to cited literature.*

Serum proteomics has attracted a great deal of attention as a potential source of biomarker discovery. Serum is easily accessible and carries a potential archive of biological information that is endowed due to its continuous perfusion through the body's tissues (Issaq *et al*, 2007). Blood is a rich source of proteins, encompassing not only the proteins from plasma-based functionality and blood cells but potentially all other human proteins in miniscule amounts. Thus specific disease may be characterized by serum "signatures" (Schenk *et al*, 2008).

Although serum is routinely used as a specimen in clinical laboratories, it poses significant challenges for proteomics biomarker discovery due to its high dynamic range (spanning over ten orders of magnitude), which can limit the detection of clinically relevant markers (Geyer *et al*, 2017; Anderson & Anderson, 2002b). The inherent wide range of protein abundance for the plasma/serum proteome is represented in the **Figure 5**.

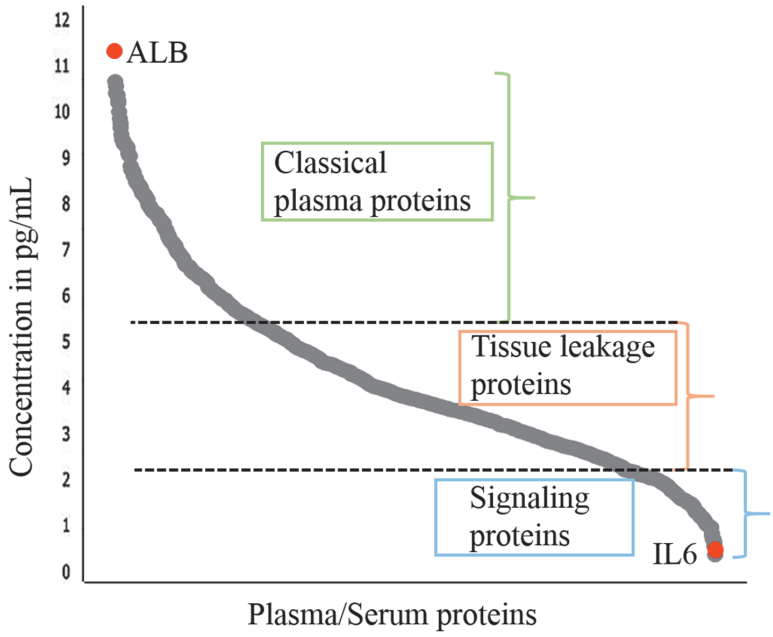


Figure 5: The inherent wide range of protein abundance in the plasma/serum proteome. ALB: Serum albumin and IL6: Interleukin-6.

Figure 5 depicts the log scale abundance of major proteins clinically measured in plasma. These span over >10 orders of magnitude, ranging from albumin at the top, down to, for example, interleukin-6 (IL6) at the bottom. At the upper end of the spectrum are the classical serum proteins, in the middle are the tissue leakage proteins and at the lower extremes signaling proteins and cytokines. Thus, throughout the entire spectrum clinically useful proteins are situated (Anderson & Anderson, 2002b).

The classical plasma protein region is dominated by high abundant proteins including albumin, immunoglobulins, complement proteins, apolipoproteins and peptidases.

Albumin (molecular mass of ~65 kDa), is the most abundant protein and is present at concentrations ranging from 35-50 mg/ml. It acts as a carrier protein in blood serum, and a large number of compounds, such as lipoproteins, cytokines, hormones, peptides and amino acids bind to it (Adkins *et al*, 2002).

Immunoglobulins or antibodies, are highly abundant in serum, function by identifying the antigens and initiating an immune response (Adkins *et al*, 2002).

The complement system's armamentarium comprises of about 35 proteins in serum and tissue fluids. When activated their cascade eradicates bacteria and viruses by interacting with antigen-antibody (Ag-Ab) complexes, cell membranes and other complement protein (Burtis, A *et al*, 2001).

The synthesis of lipids occurs in the liver and intestines and they exist in plasma as a macromolcular complexes called lipoproteins. They are categorized as 1) chylomicrons, 2) very-low-density lipoprotein (VLDL), 3) low-density lipoprotein (LDL), 4) high-density lipoprotein (HDL). The protein components of lipoproteins are called as apolipoproteins. The major forms of these are as follow.

Apolipoprotein A-I (APOA1), Apolipoprotein A-II (APOA2), Apolipoprotein A-IV (APOA4), Apolipoprotein B-100 (APOB), Apolipoprotein C-I (APOC1), Apolipoprotein C-II (APOC2), Apolipoprotein C-III (APOC3), Apolipoprotein C-IV (APOC4), Apolipoprotein E (APOE) and Apo (a).

The apolipoproteins proportion varies for each class of lipoproteins for e.g. APO A1 and APOB are the major components of HDL and LDL, respectively. APOC1, 2, 3 and APOE are present in different proportions on all lipoproteins (Burtis, A *et al*, 2001; Mahley *et al*, 1984).

From the perspective of proteomics analysis, the highly abundant proteins in serum repress the detection of lower abundant and potentially clinically useful proteins, i.e. potential biomarkers. In order to improve the coverage of the lower abundant

proteins a plausible strategy is to remove the highly abundant proteins (Bellei *et al*, 2011). The strategies for removing the most abundant proteins have included such methods as targeted immuno-affinity depletion, combinatorial libraries and precipitation (Tu *et al*, 2010; Dwivedi *et al*, 2010). Improvements in MS-based proteomics technologies have maintained the interest in plasma/serum proteomics (Cox & Mann, 2011; Aebersold & Mann, 2016; Muñoz & Heck, 2014). These advances in combination with the immunodepletion of highly abundant proteins and extensive peptide fractionation methods, has expanded the lists of identified proteins to several thousands (Addona *et al*, 2011; Cao *et al*, 2013) and in some studies even greater than 5000 (Keshishian *et al*, 2015, 2017). Nevertheless, these strategies increase the number of preparative steps and may potentially bias the obtained results (Adkins *et al*, 2002; Geyer *et al*, 2016b). The analyses of serum without depletion of the major components has recently been demonstrated in several large-scale studies, demonstrating the potential utility of this strategy. Liu *et al.* compared the plasma of identical twins (Liu *et al*, 2015), and more recently Geyer and co-workers demonstrated how the analysis of undepleted serum could be conducted with moderately high throughput as means to provide an overview of a person's health state (Geyer *et al*, 2016b, 2016a). Nevertheless, despite the progress made in serum proteomics studies in recent years, many studies are limited to the comparison of a few hundreds of proteins.

2.5 An introduction to type 1 diabetes (T1D):

Type 1 Diabetes (T1D) is an autoimmune disease caused by the complex interplay between genes and environmental factors that precipitate in genetically susceptible individuals. The body's own immune system attacks the insulin-secreting β cells in the islets of Langerhans in the pancreas, resulting in a progressive loss of insulin response. T1D is the most common endocrine and metabolic disorder occuring in children, though somewhat more common in boys than girls (Tuomilehto *et al*, 1998; Ostman *et al*, 2008). The occurrence rate varies amongst countries, with the highest incidence in Finland, where it has peaked at about 50-60 per 100,000 children. The clinical onset of T1D is typically preceded by an asymptomatic period that can last for few months or even years (Purohit & She, 2008; Knip *et al*, 2005; Harjutsalo *et al*, 2013).

2.5.1 Genetic predisposition and the environmental determinants:

The human leukocyte antigen (HLA) region on chromosome 6p21 has been predominantly associated with T1D susceptibility. The region is commonly

referred to as the insulin-dependent diabetes mellitus locus (IDDM1), and is associated with susceptibility for several other autoimmune diseases (Nerup *et al*, 1974; Singal & Blajchman, 1973). Out of all the HLA types, class II has the strongest association with T1D, where haplotypes *DRB1, DQB1* confer susceptibility and *DQA1* confers disease resistance (Erlich *et al*, 2008). Many other genetic factors have been associated with T1D, of which *INS* and *PTPN22* are important (Polychronakos & Li, 2011).

In addition, a number of non-genetic, environmental factors have been suspected to act as triggers, for example, early diet, hygiene and infections may play an important role in the etiology of T1D. Investigated dietary factors have included cow's milk and vitamin D. These are considered as follows.

- Cow's milk: The studies depicting the relationship between consumption of cow's milk and islet autoimmunity of T1D have contradictory results. Wahlberg *et al*. and Virtanen *et al.* have shown that cow's milk consumption has been associated with islet autoimmunity (Wahlberg *et al*, 2006; Virtanen *et al*, 2012) and T1D (Virtanen *et al*, 2000) while others observed not influence at all (Knip *et al*, 2018).
- Vitamin D: This fat-soluble secosteroid actively regulates the immune system and metabolic pathways relevant to diabetes, thus potentially offering a protective role. In addition to the diet, skin also produces vitamin D upon exposure to sunlight (Penckofer *et al*, 2008). The incident rate of T1D varies from spring to autumn and winter seasons, e.g. in autumn and winter many cases are diagnosed, suggesting the importance of vitamin D. However, there are mixed results with some studies suggesting the role of vitamin D in reducing the onset of T1D and others indicating it has no effect (Weets *et al*, 2004; Dong *et al*, 2013; IM *et al*, 2012; Miettinen *et al*, 2012). Overall, there is little evidence to support the use of vitamin D as a diet supplement in order to prevent the onset of islet autoimmunity and subsequent T1D (Zipitis & Akobeng, 2008; Simpson *et al*, 2011; Bizzarri *et al*, 2010; Walter *et al*, 2010).
- Hygiene hypothesis: This postulates that because of improved hygiene and sanitation, the reduced frequency of childhood infections is associated with the increased incidence of autoimmune diseases (Rewers & Ludvigsson, 2016; Knip & Simell, 2011). It is believed that, it is due to the lack of challenges during the maturation of the immune system.
- Infections: A number of studies have drawn attention towards the role of viruses in the progression to T1D. Amongst these, the strongest

evidence has been found for enteroviruses (EVs) in both animal and human studies, in particular, the Coxsackie B virus strain (Filippi & von Herrath, 2008). The first report indicating the role of this virus infection in T1D has shown the presence of elevated neutralizing antibody titers in serum, with comparisons made between healthy controls and recently diagnosed patients (Gamble *et al*, 1969). EVs have been observed to exhibit a tropism to human pancreatic β cells *in vitro* and *in vivo* (Ylipaasto *et al*, 2004; Bennett Jenson *et al*, 1980). In the study of deceased twins that progressed to T1D quite early in life (aged 14 months), the presence of enterovirus RNA was detected (Smith *et al*, 1998). The postmortem biopsies revealed the presence of enteroviruses in their pancreatic islets. Some studies have also reported the increased T-cell response against enterovirus antigens in T1D patients, as well as autoantibody (Aab) positive prediabetic children (Varela-Calvino *et al*, 2002; Juhela *et al*, 2000). Krogvold *et. al.* have identified the presence of EVs in the pancreatic islets of the living recently diagnosed T1D patients (Krogvold *et al*, 2015). The current paradigm is that EV infection first affects the islets of the pancreas, resulting in an inflammatory response. In the susceptible individuals, the EVs undergo replication in the β cells producing viral RNA and proteins, resulting in innate immune response, inflammation and autoimmunity (Hyöty, 2016).

2.5.2 Pancreatic pathology:

The β cells are situated in the islets of Langerhans of the pancreas. They store and release insulin and constitute about 65-80% of the cells in pancreatic islets. Insulin is a hormone required for blood glucose homeostasis. For many years it was thought that the immune system is the culprit for T1D (Van Belle *et al*, 2011b). However, several researchers conceptualized that beta cells are actively involved in their own destruction, and although the exact mechanism is not yet known, this remains a widely accepted hypothesis. The process initiates with the release of autoantigens from beta cells due to cellular damage, metabolic stress or viral infections. As a result, dendritic cells and macrophages (antigen presenting cells), then infiltrate into islets of Langerhans of the pancreas and presents autoantigens to T-helper cells. Naïve CD4+ T cells from blood and pancreatic lymphoid nodes recognize the major histocompatibility complexes (MHC) and beta cell-derived peptides presented by innate (macrophages) and adaptive (dendritic) immune cells in the islets. The interleukin (IL)-12 secreted from macrophages and dendritic cells

activate these CD4+ T cells, in turn activates CD8+ T cells, which then are involved in beta cells destruction. (Yoon & Jun, 2005; Crèvecoeur *et al*, 2015, 2017; Moulder *et al*, 2017b).

2.5.3 The need to identify novel risk factors and prospective sampling:

The pathogenesis of T1D remains unknown. The appearance of a number of autoantigens, including islet-cell autoantibodies (ICA), insulin autoantibodies (IAA), protein tyrosine phosphatase (IA-2), glutamic acid decarboxylase (GAD), insulin and zinc transporter SlC30A8 protein, have been found to indicate the initiation of β cell autoimmunity and an increased T1D risk (Zhang *et al*, 2013b). The detection of high levels of any of these Aab is considered as a sign of the onset of autoimmunity and with the detection of multiple Aab progression to clinical T1D is highly likely. To date, the appearance of Aab and severity of hyperglycemia serve as a measure of the progression to T1D (Kimpimaki & Knip, 2001). The extent of hyperglycemia is usually addressed by estimating glycated hemoglobin (HbA1C), formed as result of the non-enzymatic reaction between glucose and hemoglobin and is used as a biomarker of long-term hyperglycemia. The discovery of HbA1C has lead diabetes research into the chemistry of glycation and its resultant products, i.e. advanced glycation end products and their associated complications in diabetes (The International Expert Committee, 2009; Schalkwijk & Miyata, 2012). However, monitoring of autoantibody status and HbA1C provide early indication of an individual's susceptibility to T1D.

The immune system mediates its function through circulating in the blood. Thus, any activation and/or disruption of the system are reflected in the balance of its circulating components. Blood is used for determining the transcripts and cellular profiles, and serum contains a plethora of proteins and metabolites (Chaussabel *et al*, 2010; Psychogios *et al*, 2011). The signature patterns of serum biomarkers could potentially be used to assist in the T1D prediction. The asymptomatic period from early infancy to the occurrence of T1D associated autoantibodies and finally to clinical T1D provides an opportunity for early disease prediction and diagnosis. Using such a strategy, longitudinal studies have collected samples from children at their birth, through subclinical stages until the onset of clinical T1D. The samples collected in prospective studies have included whole blood samples, blood serum or plasma, and stool samples (Hyöty, 2016; Van Belle *et al*, 2011a). Examples of such studies include the BabyDiab study from Germany (Roll *et al*, 1996), Diabetes Auto Immunity Study in the Young (DAISY) from USA (Rewers *et al*, 1996), the Finnish Diabetes Prediction and Prevention (DIPP) study (Kupila *et al*, 2001), Diabetes Prediction in Skåne (DIPIS) study from Sweden (Lernmark

et al, 2004). A multicenter study, which has involved sample collection from USA, Finland, Germany and Sweden, is The Environmental Determinants of Diabetes in the Young (TEDDY) (Krischer *et al*, 2007).

The Diabetes Prediction and Prevention (DIPP) study was founded in the city of Turku in 1994 to prospectively collect blood samples from children's with a genetic risk for T1D. Newborn children are screened for T1D risk, and blood sample collection is scheduled from at-risk children for every 3 months up until the age of two, with the measurement of serum levels of ICA, IAA, GADA, and IA-2A as represented in **Figure 6**. If the levels of these Aab are above the expected thresholds, sample collection is continued at three months intervals, otherwise every six months for Aab free children older than two years until the age of 15. The children are categorized based on diagnosis and autoantibody status as T1D subjects, autoantibody positive (Aab+ve) or autoantibody negative (Aab-ve). These longitudinal series thus represent the different clinical stages of children and have provided a unique biobank for the study of biochemical changes associated with the onset of T1D. Omics analyses of these samples have included proteomics (Moulder *et al*, 2015), metabolomics (Orešič *et al*, 2008; Oresic *et al*, 2013), transcriptomics (Kallionpää *et al*, 2014; Elo *et al*, 2010) and metagenomics (Giongo *et al*, 2011) analyses.

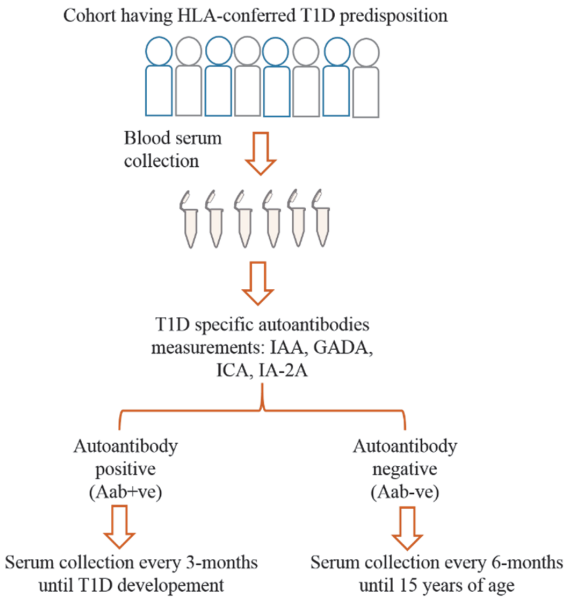


Figure 6. Schematic representation of the study design. Serum samples were collected from the HLA-conferred T1D susceptible children recruited in The Diabetes Prediction and Prevention (DIPP), followed by measurement of T1D specific autoantibodies (Aab). Based on Aab status the children's were classified as Aab+ve and Aab-ve.

2.5.4 Proteomics of T1D:

The sources of human sample materials that have been used in proteomics studies of T1D have included blood, serum, saliva, urine, and pancreatic tissue. The latter is only available from autopsies or from surgery, and thus scarce.

2.5.4.1 The pancreatic islet proteome:

Human pancreatic tissue samples are sometimes available from cadavers or from surgical interventions. In one of the early pancreatic proteomics studies, Hu *et al.* employed 2-DE and MALDI-TOF-MS to analyze the human pancreatic tissue samples and identified 302 proteins. The identified proteins, however, were from both exocrine and endocrine regions of the pancreas (Hu *et al*, 2004). The proteome of human islets was first studied by Ahmed *et al.* using 2-DE and MALDI-TOF/TOF-MS, where they identified 66 proteins specific to islets (Ahmed *et al*, 2005). In their gel-free approach, Metz *et al.* used 2D LC-MS/MS to characterize the proteome of islets obtained from a biobank and identified 3365 proteins (Metz *et al*, 2006). Recently, Burch and co-workers carried out proteomics analysis on human pancreatic tissue obtained from Network for Pancreatic Organ Donors with Diabetes (nPOD). They analyzed the human pancreatic tissue lysate from healthy controls, Aab+ve non-diabetic, T1D and T2D diagnosed samples using a Quadrupole-Orbitrap MS (Q Exactive MS) with label-free quantification approach. They identified a panel of proteins differentially regulated between diabetic and pre-diabetic conditions, that could be used to segregate T1D from T2D (Burch *et al*, 2015). Liu *et al.* characterized the proteome of human pancreatic tissue samples from the nPOD biobank (T1D = 5 and healthy subjects = 5) and identified over 5000 proteins. They used 10-plex TMT based approach and 2D LC-MS/MS with a Q Exactive MS. In their analysis, they observed a unique proteome profile representing exocrine pancreas. The detected differences revealed enrichment of β cell destruction and T1D related pathological pathways, specifically related to cell apoptosis, immune response and viral infections (Liu *et al*, 2016a).

2.5.4.2 Plasma and serum proteome:

Purohit and co-workers compared the serum samples from T1D patients and Aab-ve healthy controls using surface-enhanced laser desorption SELDI-TOF-MS. The samples were obtained from the participants recruited in the prospective assessment in newborns for diabetes autoimmunity study (PANDA), with the mean subject age of 14.3 years. They performed multivariate analysis to identify the putative

T1D biomarkers and to assess the reproducibility of the SELDI technique, and found 146 peptide/protein peaks with a statistical difference between the two conditions. However, due to the poor inter day reproducibility of the technique, poor reproducibility of the multivariate model and protein identification difficulty, they concluded that SELDI was not a suitable option for discovery and validation of biomarkers when small changes in proteomics patterns are expected (Purohit *et al*, 2006). Albrethsen *et al*. also used a SELDI approach to profile 766 serum samples from T1D patients (n=270) collected from 15 countries at 18 different pediatric centers. The samples represented the period after T1D diagnosis, i.e. from 1, 6 and 12 months and the mean age at diagnosis was 9.1 (± 3.7 years). They found that levels of two apolipoproteins (APOC1, APOC3) were increased with time after diagnosis in the T1D patients (Albrethsen *et al*, 2009a).

Metz *et al*. carried out the first major serum proteomics study of T1D patients, analyzing samples from the Diabetes Antibody Standardization Program (DASP) using LC-MS/MS and the accurate mass and time tag (AMT) strategy (Zimmer *et al*, 2006). The AMT approach was developed to make efficient use of the MS1 signal of peptides and record comprehensive chromatographic elution profiles with high mass accuracy instruments, whilst removing the constraints and limitation imposed from recording MS2 signals. A database of confidence peptide identifications is constructed from the LC-MS/MS analysis of related samples and then used to infer peptide identifications from the analysis of the samples by LC-MS (no MS/MS spectra are recorded) by matching the accurate mass and normalized elution time (NET) from the existing database (Pasa-Tolic *et al*, 2004). Utilizing this approach, they identified 5 candidate protein makers, however, due to the small sample size in the discovery phase and the lack of validation experiments, they concluded that the identified panel was not of predictive value for T1D (Metz *et al*, 2008).

In a related follow-up study, Zhang *et al.* used an LC-MS/MS-based approach for 10 sets of pooled controls vs. 10 pooled patient samples from DASP cohort (each pool = 5 individuals) and identified 24 proteins, that were differentially abundant between T1D patients and healthy controls. Further, they performed validation by targeted proteomics of this panel in an independent cohort (100 healthy subjects and 50 T1D patients) followed by blinded validation in additional serum samples (N=10 vs. 10). They found that two proteins, i.e. platelet basic protein and C1 inhibitor classified T1D patients from healthy controls with 100 % specificity and sensitivity. The predictive power of this pair was then further tested using serum samples from 50 age-matched T2D patients, in which C1 inhibitor especially classified the two classes of diabetes. Functional annotation suggested that the dysregulation in innate immune response could be associated with disease status (Zhang *et al*, 2013b). In order to achieve the deeper proteome coverage, their LC-MS/MS

analysis was based on the AMT method, for which they performed immunodepletion of high abundant proteins of pooled serum samples and SCX fractionation.

Zhi *et al.* used a spectral counting approach after 2D LCMS analysis of serum samples from three pooled controls vs. three pooled patient samples (each pool = 10 subjects). Prior to MS analysis, they removed the high abundant serum proteins using a hexapeptide affinity resin and subsequently identified over 2500 serum proteins, of which 21 were differentially expressed between healthy and T1D subjects. The differentially abundant panel were functionally related to autoimmunity, inflammation, metabolic regulation and oxidation. Immunoassays were used to validate these findings in 848 controls and 1139 T1D patients (Zhi *et al*, 2011a). The validation experiments confirmed the higher levels of adiponectin, insulin-like growth factor binding protein 2, C-reactive protein, serum amyloid protein A and lower levels of transforming growth factor beta-induced, myeloperoxidase in the serum of T1D patients. When this result was compared with the results from Zhang *et al.,* the overlap of biomarkers panel was marginal.

In general, the earlier serum proteomics biomarker studies of T1D only compared the disease end points with control groups, i.e. the differences between patients with T1D and healthy controls, with less attention to prediabetic serum analysis. McGuire *et al.* the compared cord blood of children that later on developed T1D (54 cases and 108 controls). They used SELDI to profile the differences and classified T1D developing children, however, the identity of discriminating peak was not established (McGuire *et al*, 2010).

In contrast to these comparisons of diabetics and non-diabetics, we have shown for the first time serum proteomics profile of pre-diabetic serum samples (266 samples from 19 case-control pairs) mapping the changes from early infancy, seroconversion and diagnosis (Moulder *et al*, 2015). The serum samples were obtained from the DIPP project, including pre-diagnosis samples from T1D children (median age at diagnosis 4.1 ±2.9) and healthy controls matched by age, gender, risk group and geographic region. The quantitative profiles of immunodepleted serum samples were measured using LC-MS/MS with the comparison by iTRAQ labeling and a label-free approach. Consistent differences were found in several proteins, even before the appearance of routinely measured autoantibodies in the T1D developing progressors. The proteins APOC-IV and afamin gave a classification of the cases from controls with a high success rate (area under the curve = 0.85). Functionally, the panel of biomarkers suggests that dysregulated immune pathways, complement activation and lipid metabolism may be associated with T1D development.

Recently, Toerne *et al*. studied the serum proteome of 15 children who progressed to T1D within 3.5 years, 15 children who progressed to T1D in 9.5 years or more and 15 Aab negative controls from the German T1D risk cohorts

(BABYDIAB/BABYDIET study). They used label-free LC-MS/MS and found 46 proteins to be significantly different between islets Aab+ve and Aab-ve children. The significant hits were validated using SRM-MS in 140 samples (70 Aab+ve vs. 70 Aab-ve) from the same cohorts. Amongst the distinct peptide signatures, apolipoprotein M and apolipoprotein C-IV enabled discrimination autoantibody positive from autoantibody negative children (von Toerne *et al*, 2017). Notably, the latter protein (APOC-IV), was differentially abundant, and in combination with afamin, able to segregate children developing T1D from healthy controls in our study (Moulder *et al*, 2015). In terms of function, Toerne *et al*. reported that the differences they observed could be associated with lipid metabolism and homeostasis, proposing that lipid metabolism is altered in the early autoimmune process.

In another recent publication, Liu and co-workers reported longitudinal profiles from the plasma proteome of 10 healthy children across 9 timepoints, covering the period from birth to adolescence (Age: from 9 months to 15 years). The plasma samples used in this study were collected as a part of DAISY cohort. The analytical strategy included depletion of the top 14 must abundant serum proteins, TMT labelling followed by basic pH reversed-phase fractionation and LC-MS/MS. They identified 1828 protein groups, and using statistical modeling categorized 1747 of the plasma protein groups into seven major longitudinal expression patterns, of which 970 proteins had age dependant changes in expression (Liu *et al*, 2016c). Several of these age related changes were verified by ELISA for certain proteins (e.g. IGF1, IGFBP2 and IGFBP3). In a subsequent publication they used the same proteomic approach to compare the temporal expression profile of 11 T1D patients over 9 serial time points with 10 matched healthy controls, samples covering the period from birth to onset of autoimmunity and overt T1D, and identified >2000 proteins. After statistical modeling and correction for multiple hypotheses, 13 protein groups were found to be statistically significant between T1D and healthy subjects. Notably, two proteins with a role in oxidative stress, i.e. catalase and superoxide dismutase, had aberrant expression even before seroconversion, and were verified by ELISA in the same samples.

In addition to serum/plasma, saliva (Cabras *et al*, 2010b; Kuehl *et al*, 2015) and urine (Meier *et al*, 2005; Suh *et al*, 2015a) have been used to identify biomarkers of the T1D. The summary of protein biomarkers identified in proteomics analysis of tissues and biofluids in T1D research are presented in **Table 2.** There exist similarities in potential markers between few studies, however, the overlap amongst the biofluids and tissue study was found to be none.

Table 2: Summary of potential biomarkers identified in proteomics analysis of tissues and biofluids in T1D research.

Specimen used	Subject condition	Method used	Potential biomarkers	Reference
Plasma	T1D & ND	2DE-LC-MS/MS, AMT	AZGP1, CLU, SERPINA6, LUM, TF	(Metz *et al*, 2008)
Serum	T1D	SELDI-TOF-MS	APOC1, APOC3	(Albrethsen *et al*, 2009b)
Saliva	T1D & ND	LC-MS/MS	HST1, PRP1, STATH, S100A9	(Cabras *et al*, 2010a)
Serum	T1D & ND	2DE-LC-MS/MS, Spectral counting, ELISA, Luminex	ADIPOQ, IGFBP2, SAA, CRP, MPO, TGFBI	(Zhi *et al*, 2011b)
Serum	T1D & ND	2DE-MALDI-TOF-MS	GDIβ	(Massa *et al*, 2013)
Serum	T1D & ND	LC-MS/MS, AMT, SRM	C3, GSN, PGLYRP2, TTR, PPBP and SERPING1	(Zhang *et al*, 2013a)
Serum	Pre-T1D, Aab+ve & ND	LC-MS/MS, iTRAQ, LFQ	APOC4, APOC2, AFAM, PFN1, MBL2, FHR5, CO9, BGH3, ADIPO, IGFBP2	(Moulder *et al*, 2015)
Pancreatic tissue	T1D, Aab+ve, T2D & ND	LC-MS/MS, LFQ	OLFM4, ENPP1 and RegIIIα	(Burch *et al*, 2015)
Urine	T1D & healthy siblings	2D-LC-MS/MS	LAMP2, ENPEP, NAGA, MAN2B1, CTSC, GNS, OGN	(Suh *et al*, 2015b)
Pancreatic tissue	T1D & ND	LC-MS/MS	CUL4A, HNRNPK, SPARCL1, TSPAN8 & UBE2F	(Liu *et al*, 2016b)
Serum	Pre-T1D, Aab+ve & ND	LC-MS/MS, LFQ, SRM	APOC4, APOM, CFH, HGFAC, CP	(von Toerne *et al*, 2017)
Plasma	Pre-T1D & ND	LC-MS/MS, TMT	CAT, SOD1	(Liu *et al*, 2017)

Pre-T1D: pre-diabetic phase, ND: non-diabetic samples

2.6 An introduction to atherosclerosis:

Atherosclerosis is one of the leading causes of death and disability worldwide (Murray & Lopez, 2013). It is an inflammatory disease of the large and medium-sized arteries, that results in narrowing of the luminal wall or by thrombi occlusion, which leads to ischemia of the heart (coronary heart disease), brain (stroke) or lower extremities (peripheral vascular disease) (Bentzon *et al*, 2014; Libby *et al*, 2002; Ross, 1999). The most common event out of these is the coronary heart disease, including myocardial infarction and stable angina as the fatal end points (Bentzon *et al*, 2014). The atherosclerotic process starts early in childhood and can remain symptomless for a long time, indicating that the beginning of the disease process to the occurrence of clinical events is because of a continuum of complex biological interactions (McGill H.C. *et al*, 2000). Epidemiological studies have uncovered important risk factors associated with atherosclerosis, such as smoking, hypertension, diabetes, obesity and dyslipidemia. These modifiable risk factors can be controlled, and treatment of these reduce the risk of cardiovascular risk. The non-modifiable risk factors include genetic predisposition, aging and gender (Bentzon *et al*, 2014; Vasan, 2006).

2.6.1 The pathophysiology of atherosclerosis:

A normal artery is made up of three layers **(Figure 7)**. The tunica intima is the innermost layer and is composed of a single layer of endothelial cells towards the luminal side. It overlays the subendothelial space, which consists of elastic fibers sheets and extracellular connective tissue matrix (proteoglycans and collagen). The tunica media is the middle layer that separates from the intimal layer by the elastic lamina, composed of smooth muscle cells (SMCs) layers. The outermost layer, the adventitia, is made up of connective tissues with fibroblasts and SMCs interspersed (Lusis, 2000; Packard & Libby, 2007).

The atherosclerotic process is characterized by selective retention of circulating apolipoprotein B particles in sub endothelial space by arterial wall proteoglycans and their subsequent modification (Williams & Tabas, 1998, 1995; Didangelos *et al*, 2012). However, the exact mechanism is not well understood and has been extensively studied in animals e.g. rodents and non-human primates (Lusis, 2000).

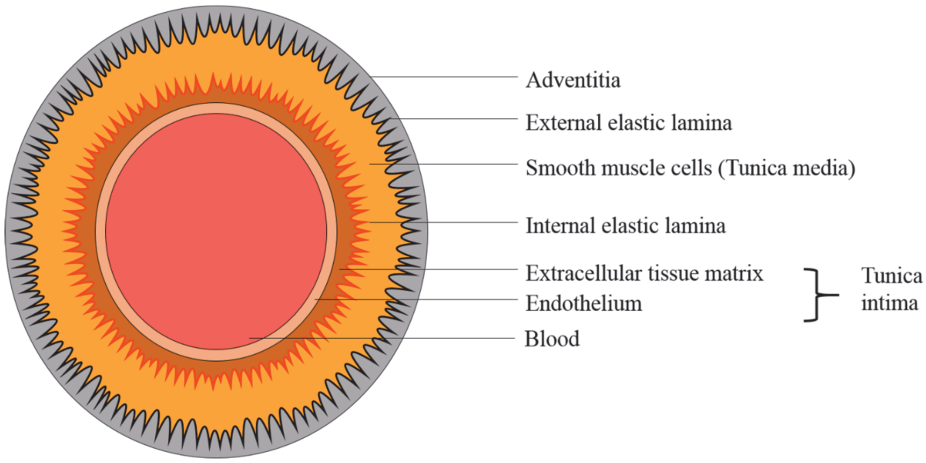


Figure 7. Structural anatomy of a normal artery. This consists of three layers, the innermost is the intima layer, the middle one is the tunica media and the outer layer is the adventitia. The figure was adapted and modified from motifolio.com.

It has been shown that the monolayer of endothelial cells of the intima can be damaged by modifiable risk factors, inducing the expressions of adhesion molecules like vascular adhesion molecule-1 (VCAM-1) and P-selectin, aiding monocytes to adhere to dysfunctional endothelium. The damaged endothelium allows the entry of LDL into the endothelial wall (Kriszbacher *et al*, 2005). The monocytes from the circulation follow this LDL, transmigrate into the inflamed endothelium and proliferate into macrophages, consume the LDL and form foam cells. Further amplification of inflammation occurs through secretion of tumor necrosis growth factor alpha (TNF-α) and interleukin-6. T cells secrete proinflammatory cytokines, such as interferon-γ (IFN-γ), CD40L and CD154, increasing the inflammatory response. The ligation of CD40 ligands induces the expression of metalloproteinases (MMPs), responsible for extracellular matrix degeneration (Packard & Libby, 2007; Blake & Ridker, 2002). Over time, the foam cells die and their lipid-filled contents contribute to the formation of lesions called fatty-streaks. These fatty streaks are thrombogenic (i.e. blood can clot on them). Platelets from the circulation then begin to accumulate over the damaged endothelium and release the platelet-derived growth factor (a regulator of smooth muscle cells growth). Thus, the released platelet-derived growth factor draws smooth muscle cells from the tunica media layer, which in turn migrate into the intimal layer and multiply to secrete collagen, proteoglycans and elastin fibrous cells to form an extracellular matrix wall called as fibrous cap. The fatty streak and fibrous cap together are described as a plaque. With time, the fibrous cap can crack and expose the underlying thrombogenic foam cells to the blood, leading to the formation of a clot in a partially occluded artery. This results in ischemia (reduced blood flow), leading to cell injury and death in the downstream areas relying on

the blood flow. If such an occlusion takes place in the coronary arteries, angina and myocardial infarction can occur. A seriously reduced flow to the internal carotid and middle cerebral artery can lead to stroke and cerebral atrophy respectively (Bentzon *et al*, 2014; Ramsey *et al*, 2010; Businaro *et al*, 2012; Shah, 2007). The buildup of plaque also weakens the structure of the artery and results in aneurysms (abdominal aortic aneurysms) (Michel *et al*, 2011). Occasionally, from the main plaque deposit, some part can break and become an embolism that will drift through the blood stream until it lodged in smaller blood vessels or another artery where the atherosclerotic plaque is building up (Bentzon *et al*, 2014).

2.6.2 The need to identify novel risk factors:

An inflammatory marker, C-reactive protein (CRP), is studied as an atherosclerotic risk marker, however, it is not particularly specific and has only a moderate predictive value (Danesh *et al*, 2004; Ridker, 2004). Furthermore, a number of different inflammatory proteins have been previously studied as potential biomarkers of atherosclerosis, including CD40L (Heeschen *et al*, 2003), interleukin-6 (Ridker, 2004), interleukin-18 (Ridker, 2004), monocyte chemoattractant protein 1 (MCP-1) (De Lemos *et al*, 2003) and myeloperoxidase (Meuwese *et al*, 2007). Moreover, because of the genetic heterogeneity and complexity of the disease etiology, the usefulness of a single protein marker in prediction is limited. Several studies, have therefore investigated the utility of multimarker panels in disease prediction (Kullo & Cooper, 2010). Zethelius and co-workers revealed that a panel of four markers, CRP, cystatin C, troponin I and N-terminal pro-brain natriuretic peptide (NT-proBNP), improved the area of receiver operating characteristics (ROC) curve (C-statistics) analysis for predicting MI and death amongst elderly men during a median follow-up period of 10 years (Sundsten *et al*, 2008). Wang *et al.* assessed a panel of 10 biomarkers in more than 3000 subjects from the Framingham Offspring Study. The participants were followed for the development of cardiovascular diseases with a median follow-up period of 7 years. They found that, out of 10 markers, the B-type natriuretic peptide was associated with the prediction of major cardiovascular events and CRP along with homocysteine predicted mortality (Wang *et al*, 2006). Nevertheless, with this multimarker panel, the risk of prediction improved, though C-statistics incremental was only modest (Kullo & Cooper, 2010)

Whilst the established risk factors have been routinely used in the prognostics of cardiovascular events, it has been reported that 14.4% of events occur in asympto-

matic individuals without any known risk factors (Canto *et al*, 2011). Thus, identification of novel non-established risk factors may help in the better prediction and management of cardiovascular diseases.

2.6.3 Proteomics of atherosclerosis:

Proteomics has made significant progress during the past several years in identifying new biomarkers and understanding cardiovascular disease mechanisms. The two main sources of biological samples that have been used to study the alteration in the proteome of atherosclerosis are tissues and blood (Mokou *et al*, 2017). The major findings of these studies are presented in the following sections.

2.6.3.1 Tissue based proteomics studies:

Tissue samples can be obtained by surgery from arterial, cardiac specimens and the tissue secretome. The latter involves resident cell isolation from surgically obtained tissue, i.e. primary cell cultures. De Klejin and co-workers analyzed the proteome of carotid atherosclerotic plaques from participants of the AtheroExpress biobank, identifying osteopontin (OPN) as a potential plaque biomarker. Subsequently, OPN was validated in femoral plaque samples, indicating its specificity irrespective of plaque localization (De Kleijn *et al*, 2010). Hao *et al.* used electrostatic repulsion hydrophilic interaction chromatography (ERLIC)-MS (Q-Exactive MS) to characterize two pooled human carotid atherosclerotic plaque samples (each pool N=19) and found 4702 proteins. Many low abundant proteins, such as TGF-β, interleukins and other growth factors, were identified. Functional annotation analysis of the differential abundant proteins revealed enrichment of the atherosclerotic signaling pathway, inhibition of MMPs and toll-like receptor signaling pathway (Hao *et al*, 2014). Because of the layered structure of the normal artery, the earlier studies involving proteomics analysis of intact atheromatous tissue of arteries lacked the information from the intimal layer where atherosclerotic progression takes place. To circumvent this, de Cuesta *et al*. used laser capture microdissection to isolate the intimal layer of pre-atherosclerotic and atherosclerotic coronary arteries, then compared these using 2D-DIGE. They found 13 proteins to be differentially regulated, of which three proteins, annexin A4, myosin regulatory light 2 smooth muscle isoform and ferritin light chain, were novel and subsequently validated by immunohistochemistry (de la Cuesta *et al*, 2011). As an alternative to tackling the complexity of whole tissue proteome, Rocchiccioli *et al*. investigated the secretome of carotid endarterectomy specimens of

14 patients. Using label-free LC-MS/MS they found 31 proteins to be differentially regulated, including extracellular and intracellular proteins. Using ELISA assays they further confirmed the higher concentration of Vitamin D-binding protein (GC) and Thrombospondin-1 (THBS1) in plasma samples (Rocchiccioli *et al*, 2013).

Even though the affected tissue may hold the relevant pathological information of the disease process enabling the identification of useful biomarkers, there are a number of potential difficulties associated with the analysis of tissue samples, which include their accessibility and inconvenience for routine diagnosis. Nevertheless, the results of tissue proteomics may allow hypotheses generation for biomarkers in circulation, which could subsequently be tested (Good *et al*, 2007).

2.6.3.2 Plasma and serum biomarkers:

DeGraba *et al*. performed proteomics analysis of serum samples from 38 patients who had undergone endarterectomy (EA) and 40 matched controls in order to investigate the associated serum proteome signature. They carried out serum fractionation using SAX and SELDI chip surfaces followed by MS analysis. These analyses, however, did not classify the asymptomatic from symptomatic EA patients. In the same study, they performed albumin depletion of 20 serum samples, followed by 2D-DIGE analysis and found a decreased abundance of alpha 1 antitrypsin, haptoglobin (HP), GC and increased levels of leucine-rich alpha 2 glycoprotein precursor (LRG) in the subset of symptomatic carotid atherosclerotic patients (DeGraba *et al*, 2011). Lepedda and co-workers purified plasma VLDL, LDL and HDL fractions from carotid atherosclerotic patients undergoing EA and compared this with the equivalent material from matched controls. Using 2-DE, and MALDI-TOF-MS for peptide mass fingerprinting analysis, they identified 23 proteins, of which the increased expression of acute phase SAA was found in all of the EA lipoprotein fractions. Collectively the results depcited the potential role of SAA in inducing an inflammatory response in atherogenesis (Lepedda *et al*, 2013).

Kristensen *et al*. applied a discovery and verification proteomics pipeline to 120 plasma samples obtained from four well-phenotyped patient groups (each group N = 30): 1] subjects with no cardiovascular symptoms and no coronary calcium, 2] subjects with no cardiovascular symptoms but having increased amount of coronary calcium, 3] subjects that had undergone operation due to atherosclerosis and 4] subjects with acute coronary syndrome (ACS). For the discovery phase experiments, they performed immunodepletion (MARS-6), 4-plex iTRAQ labeling, TiO_2 affinity chromatography and hydrophilic interaction liquid chromatography

(HILIC) enrichment, identifying in total 721 proteins. From validations using SRM assay, they verified the differential abundance of vinculin (VCL) and known risk biomarkers of cardiovascular diseases, such as apo(a), serum amyloid protein (SAA), CRP and thrombospondin-4 (TSP-4) in the ACS group. The validations supported the increased expression pattern of apo(a), CRP and SAA from groups 1 to 4. Furthermore, immunoassays provided additional verification of the differences in the levels of apo(a) and CRP (Kristensen *et al*, 2014a).

Recently, a novel targeted proteomics assay based on proximity extension proteomics chip was used to assess the levels of 82 proteins in plasma from the Prospective Study of the Vasculature in Uppsala Seniors (PIVUS). This approach revealed the associations of growth differentiation factor 15 (GDF-15), MMP-12, renin, growth hormone, osteoprotegerin, T-cell immunoglobulin and mucin domain (TIM-1) and tumor necrosis factor ligand superfamily member 14 (TNFSF14) to plaque prevalence independently of each other and conventional cardiovascular risk factors (Lind *et al*, 2015).

Using 2-DE and MALDI-TOF/TOF-MS, Han *et al.* performed serum proteomics of samples from patients with atherosclerotic CAD with their matched controls. The analysis revealed thirty-three differentially expressed proteins between atherosclerotic CAD and matched controls, including increased expression of cyclin-dependent kinase 9 (CDK9). Validation using western and ELISA confirmed the increased expression of CDK9 in serum as well as in peripheral blood mononuclear cells samples. Furthermore, immunohistochemical staining also revealed an increased expression of CDK9 in atherosclerotic plaque tissue sections. Taken together they found elevated levels of CDK9 in serum, monocytes and artery plaque samples indicating, it as a potential biomarker of atherosclerotic CAD (Han *et al*, 2015).

More recently, Lee and co-workers developed a novel approach by integrating plaque imaging, plaque and plasma proteomics to identify biomarkers of atherosclerotic plaque rupture using a human model of angioplasty induced plaque disruption. They identified 491 proteins, of which six proteins were found to be significantly high after plaque disruption. Functional analysis revealed enrichment of liver X receptor (LXR)/retinoic X receptor (RXR) pathway, and specifically found lipopolysaccharide binding protein (LBP) as a biomarker of coronary artery plaques and plaque disruption (Lee *et al*, 2017). A summary of the key biomarkers identified in the proteomics analysis of tissue and biofluids obtained from atherosclerotic patients is represented in **Table 3.**

Table 3. Key protein biomarkers identified in the biofluids and tissue based studies of atherosclerosis patients.

Specimen used	Subject condition	Method used	Potential biomarkers	Reference
Carotid artery tissue	Atherosclerotic patients	2DE-LC-MS/MS	MPO, FGG, FGB	(Hao *et al*, 2014)
Coronary artery plaque	Atherosclerotic & non-atherosclerotic	2D-DIGE-MALDI-TOF-MS, IHC	ANXA4, MYL9, FTL	(de la Cuesta *et al*, 2011)
Secretome	Carotid endarterectomy & controls	LC-MS/MS, LFQ, ELISA	THBS1, GC, VCL	(Rocchiccioli *et al*, 2013)
Serum	Endarterectomy and controls	SELDI-TOF-MS	HP, GC, LRG	(DeGraba *et al*, 2011)
Plasma	No calcification, calcification, stable arterial disease & ACS	2DE-LC-MS/MS, iTRAQ, SRM	VCL, SAA, apo(a), CRP, TSP-4	(Kristensen *et al*, 2014b)
Plasma	Carotid artery atherosclerosis	Olink Proseek® Multiplex CVD I96×96 kit	OR, TIM-1, GDF-15, MMP-12, TNFSF14	(Lind *et al*, 2015)
Serum, coronary artery plaque	Atherosclerotic coronary artery disease & controls	2DE-MALDI-TOF-MS, IHC, Western blotting	CDK9	(Han *et al*, 2015)

Overall, whilst the above studies have targeted serum and tissue from diagnosed patients, there has not been to our knowledge, a published study representing the early phases of plaque presentation. In the past decades there have emerged an increasing numbers of prospective studies that have aimed to follow the individuals at risk from cardiovascular diseases. The Cardiovascular Risk in Young Finn Study (YFS), which started in 1980 was established to study the impact of childhood life style, biological and psychological measures upon cardiovascular risk throughout the individuals life span (Raiko *et al*, 2010). This cohort has been followed up at 3 to 6 years intervals as shown in **Figure 8**. Ultrasonic assessment of carotid arteries to measure the thickness of intima-medial layer were started from the 2001 follow up and continued until the most recent follow-up in 2010-12 (Tonstad *et al*, 1996). During each visit, venous non-fasting blood samples have been collected and separated serum were collected for storage at -70^{0}C.

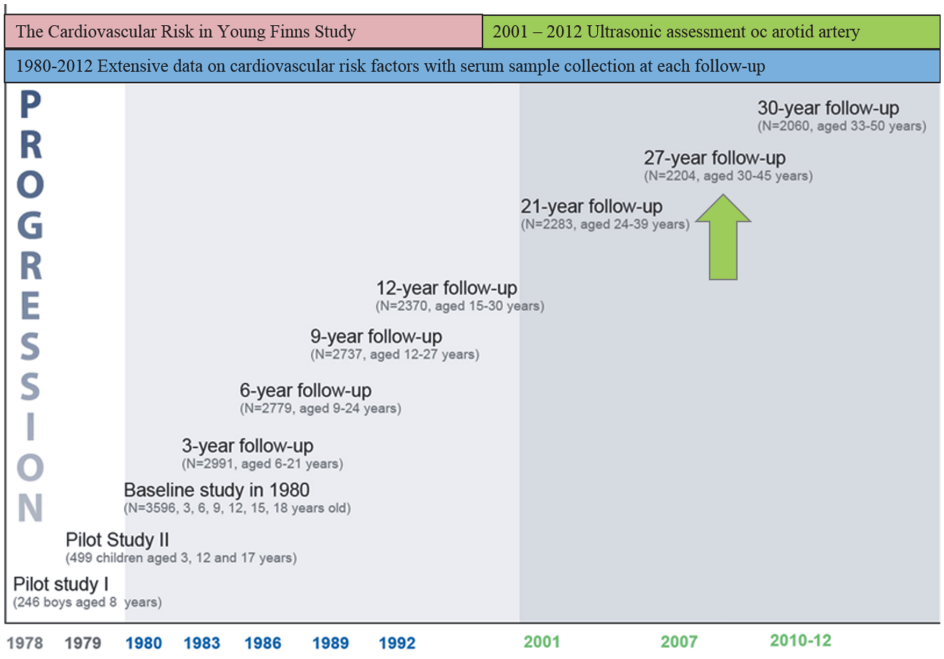


Figure 8. Study design of The Cardiovascular Risk in Young Finns Study. The arrow indicates that the time period from which the subset of serum samples were selected for quantitative MS study to identify biomarkers of premature carotid atherosclerotic phenotype (2007 follow-up). The selected samples were collected from the plaque developing subjects together with their matched controls (age, sex, body mass index and systolic blood pressure). Adapted and modified from (http://youngfinnsstudy.utu.fi/studydesign.html)

3 AIMS OF THE STUDY

The main aims of this thesis were to investigate differences in the serum proteomes and identify potential biomarkers to distinguish healthy and disease developing subjects from prospective cohorts, concerning T1D and carotid atherosclerosis. The work presented in this thesis have addressed two hypotheses. Firstly, whether changes associated with development of T1D can be detected by serum proteomics of longitudinally collected samples from disease developing subjects and carefully matched healthy controls. Secondly whether diagnostically useful serum proteins for carotid atherosclerosis can be detected using cross-sectional analysis of samples obtained from subjects with premature atherosclerotic phenotype and their matched healthy controls. Such findings could provide the basis for identifying potential biomarkers that may be useful for risk prediction and diagnosis. Furthermore, they could be useful in monitoring the disease activity and the effect of pharmaceutical interventions

The specific aims of this thesis were:

1. Identification of serum protein biomarkers at different stages towards the development of T1D using quantitative proteomics (I)

2. Identification of serum protein biomarkers for premature carotid atherosclerosis utilizing label-free quantitative proteomics (II)

3. Development of targeted proteomics pipeline to validate the identified serum protein biomarkers from study I and II using selected reaction monitoring mass spectrometry (III)

4 MATERIALS AND METHODS

4.1 Biobanks/Cohorts:

All the data included in the original publications I and II have been acquired under the permission of the Ethical Committees of the University Hospital of Turku and a written informed consent according to the Declaration of Helsinki was obtained from all study subjects.

4.1.1 The Diabetes Prediction and Prevention Project (DIPP):

The samples used in the study I were collected as a part of the Finnish DIPP study (Kupila *et al*, 2001). In this study, children at risk of T1D have been selected on the basis of HLA genotype and followed longitudinally from birth. At each study visit, sera were separated from venous nonfasted blood samples and stored at -70^{0}C. The levels of Aab specific for T1D (ICA, GADA, IA-2A and IAA) have been measured in the samples using immunofluorescence for ICA and radiobinding assay for the other antibodies (Kulmala *et al*, 1998).

The proteomics analysis were carried out on serum samples from 19 children who progressed to T1D during the DIPP follow-up. The control samples were from children that remained Aab-ve and were matched with T1D samples on the basis of age, place of birth, gender and HLA genotype. Typically seven samples per case and seven samples per control were compared for each subject pair. All together proteomics measurements were performed on 266 serum samples.

4.1.2 The Cardiovascular Risk in Young Finn Study (YFS):

The samples used in study II were collected as a part of the YFS (Raiko *et al*, 2010). The proteomics measurements were performed for serum samples from the 2007 follow-up. These samples were selected on the basis of ultrasound examinations of the carotid artery from participants in whom a distinct plaque was detected, and analyzed together with the equivalent samples from carefully matched controls (matched by age, sex, body mass index and systolic blood pressure; N = 43 vs. 43). The samples were prepared and subsequently analyzed using a label-free quantitative proteomics approach.

4.2 Methods used in this study:

For the studies presented in this thesis a number of preparative steps and methods were used to enable the proteomic characterization, subject classification and data interpretation. These are briefly discussed below. Further information is provided in the original publications.

4.2.1 Immunodepletion of high abundant serum proteins:

To increase the detection of a wider range of serum proteins, the highly abundant serum proteins were immunodepleted using commercially available immunoaffinity columns. The following two columns types were used in the studies.

- ProteomeLab IgY-12, 100 mm x 6.3 mm i.d. (Beckman Coulter):
  - Study I

- Multiple Affinity Removal Human 14, 50 mm x 4.6 mm i.d. (MARS-14; Agilent):
  - Study I and II

The depletion of serum samples was performed by LC using an Ultimate 3000 System and a Hitachi L7100 system for the IgY-12 and MARS-14 columns, respectively. The chromatographic peaks representing the flow through and bound fractions were monitored by UV absorption at λ=280 nm, and collected with an Advantech SF-2100W fraction collector (Advantec Inc., CA, USA). Since the data analysis was based on the use of temporal samples from age and gender-matched controls a single depletion system was used for the compared case and control samples.

4.2.2 Buffer exchange and in-solution digestion:

The immunodepleted serum samples were concentrated, denatured, reduced using tris 2-carboxyethyl phosphine (TCEP) followed by alkylation with methyl methanethiosulphate (MMTS) in dark at room temperature. Finally the samples were digested with trypsin for overnight at 37^0C.

4.2.3 Isobaric tag for relative and absolute quantification (iTRAQ) labeling and strong cation exchange chromatography (SCX):

The digested samples were labeled with iTRAQ reagent (8-plex) according to manufacturer's instructions (Applied Biosystems). The labeled peptides were combined, then vacuum concentrated to reduce the volume to 200-400 µl and subjected to SCX fractionation prior to LC-MS/MS analysis.

4.2.4 LC-MS/MS analysis:

During the course of the studies two LC-MS/MS platforms were used for the analysis of the iTRAQ labeled samples: I) a QSTAR Elite time-of-flight MS (AB-Sciex) coupled to Ultimate 3000 capillary LC, II) an Orbitrap Velos Pro Fourier transform MS (Thermo Fisher Scientific) coupled with an EASY-nLC II. The peptides were first loaded on 2 cm trap column followed by separation on 15 cm long analytical column having an inner diameter of 75 µm. The packing material was 5 µ magic C18-bonded silica material (200 Å). The elution was carried out with a flow rate of 300 nl/min using the binary gradient of acetonitrile and water with 0.2 % formic acid. The tandem mass spectra for iTRAQ labeled samples were carried out using CID and HCD mode for QSTAR Elite and Orbitrap Velos respectively. The label-free data were acquired using only the Orbitrap Velos system in CID mode.

4.2.5 Protein sequence database search:

- Study I:

The ProteinPilot™ software integrated with the Paragon identification algorithm (Shilov *et al*, 2007b) was used for the analysis of iTRAQ data. The spectra were searched against a Human UniProt database (release August 18th, 2011, 20245 entries including 162 common non-human contaminants). The search criteria were specified as 8-plex iTRAQ mode, trypsin digestion and MMTS as a fixed modification of cysteine. The QSTAR Elite MS data were analyzed directly using ProteinPilot™, however the Orbitrap Velos data were converted to mascot generic format using Proteome Discoverer version 1.3 (Thermo Scientific) and parsed to ensure use of the appropriate reported ion fragments (Rissanen *et al*, 2012). A false discovery rate (FDR) of 5% was applied at protein level identification using the PSEP functionality mode of ProteinPilot (Tang *et al*, 2008; Tambor *et al*, 2012).

Proteome Discoverer together with Mascot 2.1 (Matrix Science) was used as the search algorithm to analyze the LFQ data. The search parameters included fixed modification of cysteine residue using MMTS, variable methionine oxidation modification and N/Q demidation, using the aforementioned database.

- Study II:

The MS/MS raw files were searched against UniProt human isoform sequence database (UniProt release, August 2017, entries = 42,210) using Andromeda search engine (Cox *et al*, 2011a) built-in MaxQuant software (Cox & Mann, 2008b). The search parameters allowed up to 2 missed tryptic cleavages, MMTS as a fixed modification, Methionine oxidation and N-acetylation as variable modifications. A FDR of 1% at peptide and protein levels were applied using forward and reverse search of protein sequence database. The functionality "match between runs" was enabled in order to transfer the identifications across the data files (Cox *et al*, 2014).

4.2.6 iTRAQ quantitative analysis (Study I):

The iTRAQ ratios were calculated using ProteinPilot™. For between pair comparisons the reported ion intensities were normalized to a pooled reference, whereas within-pair comparisons were evaluated as the ratio of the case to control intensities.

4.2.7 Label-free quantification (Study I and II):

- The mass spectrometry raw files were imported to Progenesis LC-MS v4.0 (Nonlinear dynamics) (Fischer *et al*, 2012). The software uses feature detection and multiple alignments of the ion intensities, facilitating the comparison of multiple mass spectrometry files. The normalization and intensity based abundance calculation for each protein was performed using Progenesis. The label-free intensity values of each protein were converted to median value across case-control pairs in order to facilitate the comparison with iTRAQ data and similarly protein abundance ratios (case/control) calculated for the paired samples. The statistical analysis were carried out using R (R Development Core Team, 2011) and SPSS (IBM Corporation). The mass spectrometry proteomics raw data and/or peak lists have been deposited to the MassIVE. MSV000078748-65 **(Study I).**

- MaxQuant software was used for LFQ. It performs the peak detection and peptides scoring using a set of algorithms. It does the mass calibration and peptide searching against sequence database to identify proteins, performs the normalization and calculates protein intensity values. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno *et al*, 2016) partner repository with the dataset identifier PXD008278 **(Study II).**

  The SRM verification data are available from the ProteomeXchange Consortium via the PASSEL (Farrah *et al*, 2012) partner repository with dataset identifier PASS01146 **(Study II)**.

4.2.8 DIPP data analysis (Study I):

- Comparison of temporal differences:

A rank product-based method was used to evaluate temporal differences between the children compared from the T1D study. The data were selected and combined to make the following comparisons: 3 to 6 months before seroconversion, 9 to 12 months before seroconversion, 3 to 6 months before diagnosis, 9 to 12 months before diagnosis and 15 to 18 months before diagnosis. This was applied to the $\log_2$ transformed protein relative abundance measurements at selected time intervals, i.e. before and after seroconversion. For these the depletion targets were removed from the data matrix to limit the influence of their variations on the ranking.

In addition to identifying the proteins that were differentially abundant in the study cohort, the longitudinal changes in the protein profiles were evaluated and overviewed in terms of how they correlated relative to seroconversion. For this comparison there were 11 progressors with at least two samples before and after seroconversion, and similarly from the age-matched controls. Spearman's correlation coefficients were calculated for the case/control ratios and the case or control to reference ratios. To estimate the *P*-value the time axis was permutated 10,000 times and an FDR $\leq$5% was considered significant (Benjamini-Hochberg corrected). Using cut offs for the Spearman's correlation coefficients of an absolute value of $\geq$0.4 and a FDR $\leq$5%, the enrichment of GO functions for the proteins satisfying these criteria were compared between the subject classes.

- Changes in the complement proteins:

Frequently the differences were observed between the case and control subjects in the relative abundance of the complement proteins. Hierarchical clustering revealed coordinated changes with this class of proteins. For the comparative analysis of these changes, the Pearson's correlation coefficients were determined for Complement component 5 (CO5) with the other proteins for each subject. The consistency of these correlations between subjects was assessed using rank product analyses of the correlation coefficients. CO5 was selected due to its central role in the formation of the membrane attack complex (MAC).

- Classification Analysis:

The top scoring pairs (TSP) method was applied to identify whether combinations of the quantified proteins could classify the samples and subjects (Xu *et al*, 2005; Geman *et al*, 2004). The analysis was used to determine the difference in protein abundance for the median normalized data at selected and/or averaged time periods.

4.2.9 YFS data analysis (Study II):

- Reproducibility optimized test statistics (ROTS) analysis:

The proteinGroup.txt file, an output of MaxQuant was preprocessed using the Perseus computational pipeline (Tyanova *et al*, 2016). The processed data containing normalized protein abundance values were analyzed using ROTS (Elo *et al*, 2008; Suomi *et al*, 2017).

- Machine learning analysis:

To predict the protein panel with highest discriminative power, a machine learning analysis was performed. The feature selection from the serum proteomics data was based on Lasso regression (Tibshirani, 1996), using the R package glmnet (Friedman *et al*, 2010).

- Targeted proteomics data analysis:

The SRM data were analyzed using Skyline software (MacLean *et al*, 2010b), specifically for the inspection of transitions and peak areas. The MSStats package that can be implements through Skyline was used to perform the comparison between two groups (Choi *et al*, 2014).

4.2.10 Functional annotation analysis (Study I):

The Database for Annotation, Visualization and Integrated Discovery (DAVID) tool (da *et al*, 2009) was used to perform the gene ontology and pathway analysis.

4.3 Development of targeted proteomics pipeline:

A number of differentially abundant serum proteins were detected in the study I (Moulder *et al*, 2015) and II (Bhosale *et al*, 2018). To facilitate the validation of these markers in a wider cohort, SRM mass spectrometry methods to assay these proteins in additional serum samples were developed.

4.3.1 Selection of proteotypic peptides:

About three to five proteotypic peptides (Mallick *et al*, 2007) per protein, representing a panel of 41 and 10 proteins in the study I and II respectively were selected. This selection was based on the acquired discovery proteomics data and aided by the use of Skyline and comparisons with Peptidepicker and the SRM Atlas (Mohammed *et al*, 2014; MacLean *et al*, 2010a; Kusebauch *et al*, 2016).

4.3.2 LC-MS/MS and LC-MRM-MS analysis (Study I, II and III):

Heavy-labelled synthetic equivalents of proteotypic peptides were obtained, pooled and a Q-Exactive MS (Thermo Fischer Scientific) was used to record HCD spectra to create spectral libraries. The pools were spiked with index retention time (iRT) peptides to establish retention time indices for the monitored peptides (Escher *et al*, 2012). Following the database search of the LC-MS/MS data from these pools, Skyline was used to establish suitable transitions for the selected reaction monitoring (SRM) assays. Once the transition targets for the SRM assay were ascertained with the heavy-labeled synthetic analogs, an unscheduled run of serum/plasma samples spiked with iRT peptides was acquired using TSQ triple quadrupole MS (Thermo Fischer Scientific). The unscheduled run was then analyzed using Skyline to establish the actual retention time windows for the target peptides using their iRT values. Finally the scheduled run of serum/plasma was acquired for the samples including the spiked heavy counterparts of the target peptides along with the iRT peptides.

4.3.3 Data analysis (Study I, II and III):

Proteome Discoverer 1.4 was used to analyze the Q-Exactive MS data and targeted proteomics analysis was performed using Skyline. In particular, the MSStats package that is integrated into the Skyline software was used as a graphical interface to carry out the statistical analysis between healthy and diseased conditions.

5 RESULTS

5.1 T1D proteomics (Study I):

5.1.1 Mass spectrometry analysis:

Serum proteomics analysis were carried out on the samples collected as a part of DIPP cohort to identify early changes associated with T1D risk. LC-MS/MS was used to determine the serum proteomics profiles for 266 samples from 19 case-control pairs, covering the prediabetic period and ranging from the age of 3 months to 12 years. The samples were immunodepleted to remove the most abundant proteins. In the analysis of these longitudinal serum proteomes, two different MS-based quantitative strategies were applied to separate sample sets, i.e. iTRAQ, and label-free. Using the iTRAQ labeling based approach (13 case-control pairs), 658 proteins were identified and quantified with ≥2 unique peptides, whilst with the label-free approach 261 proteins were identified and quantified using similar criteria (6 case-control pairs). The overlap between iTRAQ and label-free detections encompassed 248 proteins with two or more unique peptides.

5.1.2 Serum proteome level differences between children who developed T1D and age-matched healthy controls:

Decreased levels of APOC4 and APOC2 were found in the children that progressed to T1D when compared to their matched controls. Similarly lower levels of mannose binding protein C (MBL2) were indicated in the controls. The levels of two complement proteins, i.e. complement factor H-related protein 5 (FHR-5) and complement component 9 (CO9), were increased in abundance in subjects that developed T1D **(Table 2, Study I)**.

The lower levels of APOC4 and APOC2 were already apparent before seroconversion in the T1D developing children. Additionally, on the basis of rank product analysis an increased abundance of profilin-1 (PFN1) was observed in the period of 3-6 months before seroconversion. Similarly, analysis on the period after the appearance of Aab indicated a lower abundance of adiponectin (ADIPO), periostin, sex hormone-binding globulin and a higher abundance of dopamine β-hydroxylase. **(Table 2, Study I)**.

5.1.3 Serum proteome based classification of children progressing to T1D:

The top-scoring pair (TSP) method classified T1D developing subjects with a success rate of 91% **(Figure 9A)**. The segregation was based on the combination of two proteins i.e. APOC4 and afamin (AFAM), whose relative abundances were lower and higher than control subjects respectively **(Figure 9B)**. The evaluation of longitudinal data using TSP method in T1D developing children's revealed segregation of pre and post seroconversion samples with a success rate of ~80%. This classification was based on the relative abundance of apolipoprotein A-IV and insulin-like growth factor-binding protein complex acid labile subunit.

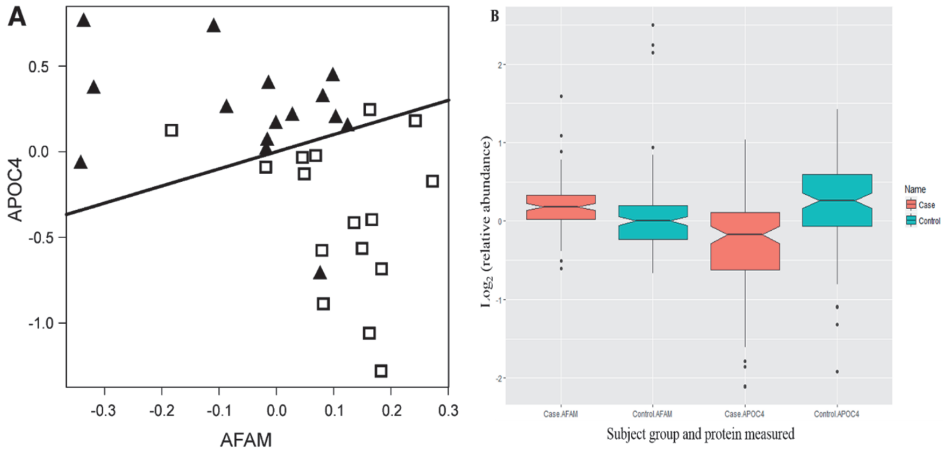


Figure 9: A) Classification of children who developed type 1 diabetes and age-matched control subjects based on the abundance of APOC4 and AFAM. The TSP method was used, yielding a 91% success rate. ▲, control subjects; □, case subjects **(Adapted from Study I)**. **B)** Relative abundance measurements for APOC4 and AFAM for the case and controls subjects.

5.1.4 Functional annotation analysis and hierarchical clustering:

Functional enrichment analysis of the proteins that were positively correlated with age in the children who developed T1D, revealed statistically significant enrichment of inflammation and immune response pathways. However, for a similar analysis of the negatively correlated proteins, there was not any functional enrichment **(Table 3 and 4, Study I)**.

Enrichment analysis was also made on the clusters identified from the case-control paired data after hierarchical clustering. These analyses revealed enrichment of

acute inflammatory response, lipid and cholesterol transport, and innate and humoral immunity pathway in children developing T1D, when compared to their matched controls. These clusters frequently indicated the contrasting behavior of the complement proteins (between cases and controls). As CO5 is a central component in the membrane attack complex (MAC) formation, its correlation with the other quantified proteins was evaluated in the T1D developing children and in their matched controls. This analysis revealed a number of correlations with the other complement proteins (CO6, CO8A, CO8B, CO8G and CO9) with CO5 in addition to a number of contrasting behavior between the subject groups (**Table 5, Study I)**.

5.1.5 Comparison with the published serum proteomics T1D datasets:

To gain an overview of these results in respect to related studies, the findings were compared with two previously published datasets concerning serum proteomics of T1D patients (Zhi *et al*, 2011a; Zhang *et al*, 2013b). The combined list included 38 proteins, 32 of which were detected in our results. The expression patterns determined from the serum proteomes of newly diagnosed T1D patients by Zhi *et al*. included a statistically significant increased abundance of ADIPO, C-reactive protein, insulin-like growth factor-binding protein 2 (IGFBP2) and serum amyloid protein A, and decreased abundances of myeloperoxidase and transforming growth factor beta-induced protein ig-H3 (BGH3) (Zhi *et al*, 2011b). Comparatively, in our data lower abundances of both ADIPO and BGH3 were observed in the children from 18 months prior to diagnosis, whereas both of these and IGFBP2 decreased with age in cases. When compared with the biomarkers evaluated by Zhang *et al.* (Zhang *et al*, 2013b), we observed the increased abundance of β-ala-his dipeptidase (CNDP1) and glutathione peroxidase 3 (GPX3) with age in both T1D developing and control subjects. An increased abundance was also observed for clusterin (CLU) over time in the T1D developing subjects.

5.2 Carotid atherosclerosis proteomics (Study II):

5.2.1 Mass spectrometry analysis:

To investigate the occurrence of moderately abundant serum proteome biomarkers of premature carotid atherosclerosis, samples were selected from participants recruited in YFS cohort. The latter study has evaluated the influence of childhood

lifestyle, biological and psychological measures and their contribution to cardiovascular risk. The health status in the order of two thousand subjects have been longitudinally monitored over the past 30 years and evaluations intervals of 3-5 years with serum collection, biochemical parameters measurement **(Table 1, Study II)** and ultrasound assessment of carotid arteries. Based on the ultrasound examination, serum samples were selected from the subjects who have developed premature atherosclerotic phenotype in carotid arteries (N = 43) and healthy controls (N = 43) matched by age, gender, body mass index and systolic blood pressure. To achieve the broader proteomic coverage, serum samples were immunodepleted to remove the high abundant serum proteins. A label-free quantitative proteomics analysis of 86 serum samples was performed in quadruplicates. Overall, 296 proteins were detected with ≥2 unique plus razor peptides. Out of these 249 proteins with valid values in half of the dataset were considered for statistical analysis.

5.2.2 Serum proteomics changes in premature carotid atherosclerotic patients:

The normalized protein intensity values from label-free quantitative proteomics analysis were used as input for analysis by Reproducibility-Optimized Test Statistics (ROTS) (Elo *et al*, 2009; Suomi *et al*, 2017). The statistical analysis revealed differences in the abundances of seven proteins ($p < 0.05$) in the plaque bearing subjects. The differential abundance of serum proteins in cases and controls using ROTS analysis is shown in **Table 4**. The down-regulated proteins included Fibulin 1 proteoform C (FBLN1C), beta-ala-his-dipeptidase (CNDP1), cadherin-13 (CDH13), gelsolin (GSN) and 72 kDa type IV collagenase (MMP2), whilst apolipoprotein C-III (APOC3) and apolipoprotein E (APOE) were higher in their abundance. However with the correction for multiple hypothesis testing, only FBLN1C remained statistically significant amongst the differentially regulated proteins. These proteins are highlighted in the volcano plot shown in **Figure 10**.

On the basis of the observed differential regulation of APOE and the reported association of *APOE4* with carotid atherosclerosis (Elosua *et al*, 2004; Granér *et al*, 2008), we evaluated the genotype data but did not find any difference in the frequency of distribution of its alleles **(Supplementary Figure S2, Study II)**.

Table 4. Proteins observed to be differentially abundant between cases and their matched controls (N = 43 vs. 43).

UniProt accession numbers	Protein names	No. of unique + razor peptides	Sequence coverage (%)	Log2 fold change	p value	FDR
P23142-4	Fibulin-1 pro-teoform C	6	36.6	-0.27	0.004	0
P02649	Apolipoprotein E	34	80.1	0.22	0.01	NS
P06396	Gelsolin	70	75.8	-0.13	0.03	NS
P02656	Apolipoprotein C-III	8	62.6	0.4	0.03	NS
P08253	72 kDa type IV collagenase	3	7.4	-0.35	0.04	NS
P55290	Cadherin-13	5	9.3	-0.32	0.04	NS
Q96KN2	Beta-Ala-His-di-peptidase	29	57.6	-0.19	0.04	NS

FDR = 0 indicates a value < 0.0001, NS = not significant

Figure 10. A volcano plot depicting the differential levels of serum protein in subjects who developed plaques and its respective controls. Each dot represents a protein.

5.2.3 Machine learning analysis:

To establish whether there was a panel of proteins that could provide an optimal and statistically significant classification of the premature atherosclerotic patients from their matched controls, a machine learning approach was used. A panel of three serum proteins, FBLN1C, APOE and CDH13, was identified to give the best segregation of cases from controls. The area under the receiver operating characteristics curve (AUROC) for FBLN1C alone was 0.67 (95% CI: 0.56-0.79). However, with the inclusion of APOE and CDH13, there was a statistical improvement in AUCROC to 0.79 (95% CI: 0.69-0.88, p = 0.03) **(Figure 11).**

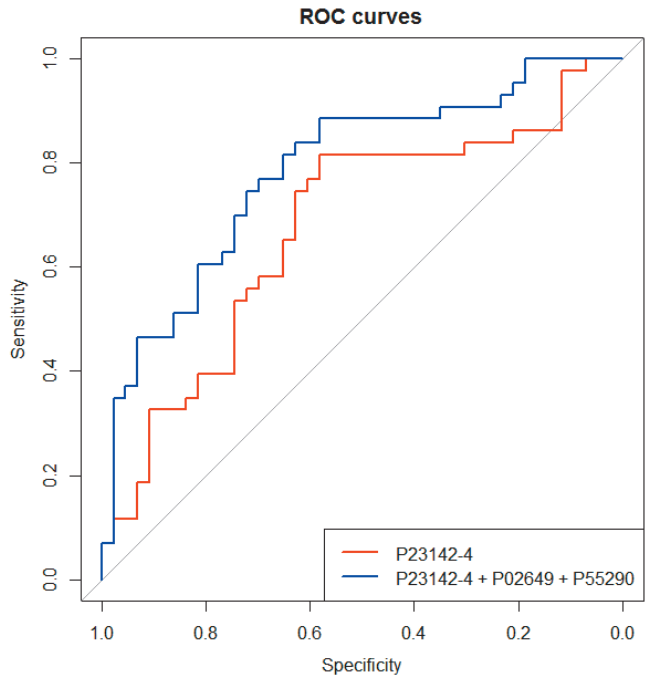


Figure 11. Receiver operator characteristics (ROC) curve analysis for P23142-4 (AUC = 0.67) alone and for panel of P23142-4, P02649 and P55290 (AUC = 0.79).

5.2.4 Targeted proteomics results:

Synthetic peptides were obtained to verify the identity of peptides associated with the differentially abundant proteins. Due to unavailability of similar cohort for validation, the SRM based MS assay was developed and subsequently applied to verify the results with the same samples, although prepared without immunodepletion of the most abundant proteins. For data normalization the assay included two housekeeping proteins (observed to show little variation in between the subjects in

our discovery phase proteomics data). Additionally, APOB, a known cardiovascular risk factor highlighted in the clinical data, and retention time standard peptides were measured **(Table 3, Study II).** This analysis confirmed the decreased abundance of FBLN1C by a ratio of 0.85 (99% CI: 0.73-0.98; FDR < 0.05) in plaque developing subjects compared to controls **(Figure 4; Study II)**. The differences determined for the other targets, however, were not significant.

5.3 Targeted proteomics pipeline results:

A panel of proteins was selected for SRM based validation with the inclusion of key components of the complement and lipid metabolism pathways, as well as several markers of inflammation. Heavy-labelled synthetic equivalents of proteotypic peptides were used to confirm the assigned identities of the peptides detected in discovery experiments. Using the combination of iRT peptides and heavy labeled synthetic peptides scheduled transitions were established for the measurement of 41 and 10 proteins corresponding to study I and study II. To date we have evaluated 200 and 23 heavy-labeled synthetic peptides representing study I and II targets respectively, of which the utility of 90% peptides was confirmed using the SRM approach **(Figure 12 and 13)**.

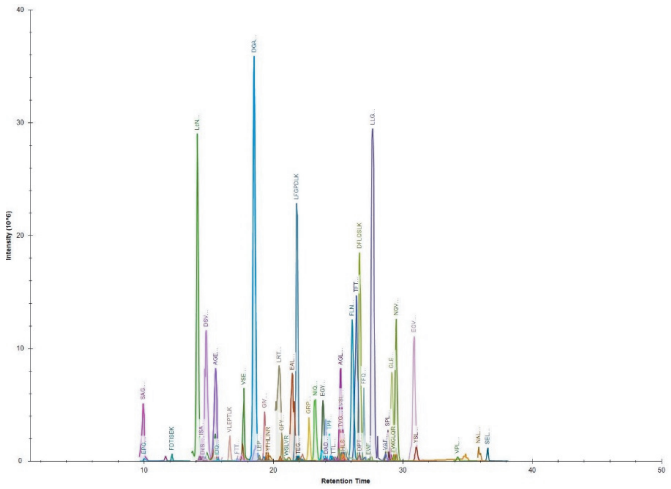


Figure 12. LC-SRM-MS of Study I target proteins. The chromatogram represents 211 peptides corresponding to 41 putative markers and iRT peptides that were detected and scheduled.

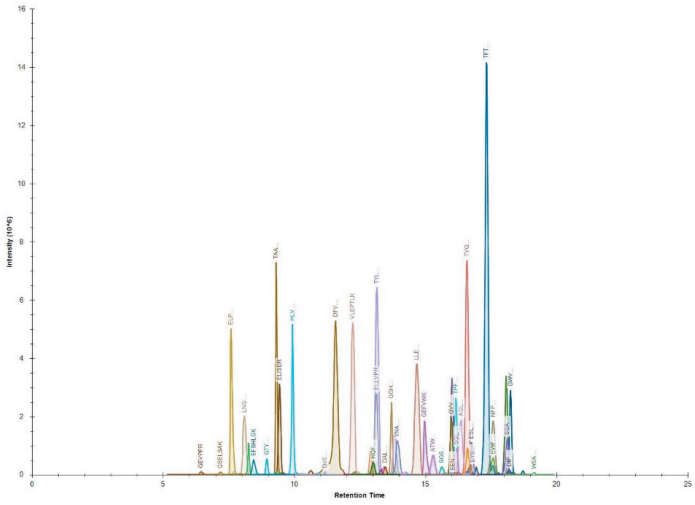


Figure 13. LC-SRM-MS of Study II target proteins. The chromatogram represents 33 peptides corresponding to 10 putative markers and iRT peptides that were detected and scheduled.

6 DISCUSSION

6.1 T1D proteomics (Study I):

Using quantitative proteomics technologies, the results from this study represented the first temporal study of pre-diabetic children to discover serum proteomics patterns. Using samples from the Finnish DIPP study, comparisons were made between children who became Aab+ve and progressed to T1D, with matched healthy controls who remained Aab-ve. In terms of early biomarkers, the lower relative levels of APOC2 and APOC4 were detected even before seroconversion in the children's progressed to T1D. In relation to current hypothesis, Laitinen *et al.* have presented data supporting the role of coxsackievirus B1 in the development of T1D (Laitinen *et al*, 2014), and lower levels of apolipoproteins have been associated with viral infections (Singh *et al*, 1999; Rowell *et al*, 2011). Based on the former, enterovirus data recorded in the DIPP study were evaluated for the subjects considered in our study. Amongst the 19 case-control participants from the present study, measurements had been made for neutralizing antibody data against coxsackievirus B1 in twelve subjects (eight progressors and four controls). The antibodies against the virus were detected in six progressors out of eight, whereas the controls were all antibody negative. Analysis of additional samples with data on viral infections is needed to make further conclusions based on these initial observations.

The decreased serum abundance of ADIPO was observed in children who were Aab+ve. Several studies have indicated the role of ADIPO in the regulation of fat metabolism and in insulin sensitivity in T1D patients (Zhi *et al*, 2011a; Pereira *et al*, 2012). We found an increased abundance of PFN1 prior to seroconversion and functionally it has been related to inflammation and resistance to insulin (Pae & Romeo, 2014).

The differences detected in relation to the correlations of complement proteins were especially notable amongst the MAC components. Usually the components of the MAC circulate independently with a consecutive interaction between them resulting from complement activation where the end-point leads to destruction of bacteria and pathogens (Sodetz & Plumb, 2001). In addition to its protective role, the complement system has been implicated in the pathogenesis of several autoimmune diseases including T1D (Chen *et al*, 2010). Hence, the observation of differential levels of complement proteins in subjects *en route* to T1D may reflect various challenges in mediating immune response (Fink *et al*, 1992).

Whilst the levels of MBL2 were found to be variable within the subjects, in the T1D developing progressors levels were remarkably lower than in their matched controls. MBL2 is an important component of immune response and regulates activation of the lectin complement pathway. The deficiency of MBL2 is linked to a genetic disorder leading to susceptibility to infection (Turner, 2003). However, the relative abundance of MBL2 did not reveal any strong correlation with any of the other quantified proteins.

The analysis using a TSP method classified the children who progressed to T1D based on the relative abundance of APOC4 and AFAM with success rate of 91%. The latter protein plays a role in vitamin E transport and has been linked to secretion of insulin in islet cells (Liu *et al*, 2012). Such associations with APOC4 have not been established. However, von Toerne *et al.* (von Toerne *et al*, 2016) also reported the lower abundance of a peptide from APOC4 and its potential utility in predicting the risk of clinical diabetes in children with multiple autoantibody positivity. Using samples from the BABYDIAB/BABYDIET cohort, they made measurements from 45 islet Aab+ve and negative children for the discovery phase and additional 140 children from the same cohort for validation experiments. Although our study lacks the validation in a large independent cohort, the potential importance of our discovered marker (i.e. APOC4) was confirmed using targeted proteomics methods in a larger population outside of Finland (von Toerne *et al*, 2016). However, another recent study by Liu *et al*. detected APOC4 but did not report any longitudinal behavior (Liu *et al*, 2017). In contrast to these recent publications (von Toerne *et al*, 2016; Liu *et al*, 2017), our study encompassed children's whose average seroconversion age were about 2 years. Such information is missing from the study by Liu *et al*. This is important factor shall be considered in the light of recent reports (Bjelosevic *et al*, 2017) and demonstrating temporal age-associated changes in serum protein levels of young children (Lietzén *et al*, 2018) (see also below).

Overall, our quantitative analyses were mostly limited to the comparison in the order of ~250 proteins. Whilst these mostly represent the moderately abundant serum proteome, the belief of proteomics community is that the key markers for disease process may come from measurement of lower abundant proteins (Gerszten *et al*, 2011). Notably in the study of Liu *et al.* in which they compared plasma samples from children developing T1D and controls, using isotope labelling methodology, and more advanced mass spectrometry instrumentation than was used than in our study, they were able to provide proteomic coverage of more than 2000 proteins (Liu *et al*, 2017). Their application provides a useful overview of the current capabilities of isobaric labelling for plasma proteomics. Also, preceding publication, they used the same approach to determine the temporal changes in the serum proteome of 10 control children's from the DAISY study, between the ages

of 9 months to 15 years. Here they demonstrated consistent quantification for in the order of 1000 proteins and observed age dependent trends in the order of half of the detected proteins. Such temporal data is particularly useful when attempting to distinguish disease related changes from natural development (Liu *et al*, 2016c). However, a notable limitation with their data in relation to study of T1D was the lack of coverage of the early months and years of life. The earliest samples for each subject were from around age of nine months and then from the age of the two years. Data from the DIPP study, in particular, has shown that many children seroconvert to islet cell immunity before the age of 2 years. The recognition of changes during this period is therefore vital and for this reason in the DIPP study, seven temporal samples are often collected during first two years. Furthermore, the average age of seroconversion of the subjects in our study was 2.1 years. Similarly, in studies from our group, proteomics characterization of the serum proteome from control children (not developing T1D) from the DIABIMMUNE study have been made (Lietzén *et al*, 2018). The study included 103 longitudinal serum samples collected from 15 children's between birth to the age of 3 years (including four samples before the age of 2 years and cord blood). These data have provided useful insights into the dynamics of changes of the serum proteome during these early and critical phases.

The evidence of the markers highlighted in our study and subsequent work will required further investigation, which in turn, would benefit from the development of sampling, measurement and validation methods to permit investigation on a wider scale. For our validation studies, a targeted proteomics pipeline has been established **(Study III)** and we are now in the process of validating the putative biomarkers of T1D progression. Importantly the selectivity and sensitivity of the targeted approach enables greater throughput in analysis.

6.2 Carotid atherosclerosis proteomics (Study II):

The proteomics analysis were carried out on the serum samples from the subjects bearing premature carotid atherosclerotic plaques and their matched controls from the YFS cohort. Seven proteins i.e., APOE, APOC3, CNDP1, CDH13, FBLN1C, GSN and MMP2 found to be differentially abundant between cases and controls. Furthermore machine learning analysis predicted a panel of APOE, CDH13 and FBLN1C classified the cases from controls with high discriminative power **(Figure 12)**.

APOE is the surface component of VLDL and chylomicrons. Changes in the biological function of APOE by two common polymorphisms produces three alleles i.e. ε2/ε3/ε4 encoding for *APOE2*, *APOE3* and *APOE4* respectively. In study II,

increased serum levels of APOE were found in the subjects with premature carotid atherosclerotic plaques. This increased expression could be due to the increased levels of VLDL and chylomicrons, which are proatherogenic in nature (Lada & Rudel, 2004; Carmena, 2004). Furthermore, APOE exhibits proinflammatory properties by binding to the plasma lipids and presenting them as antigens to the immune system, thereby inducing an inflammatory response (Van Den Elzen *et al*, 2005). Thus, it could be hypothesized that the increased serum expression of APOE in cases associated with the atherosclerosis is mediated through interaction with lipids and by inflammation.

The levels of CDH13 were found to be decreased in the plaque bearing subjects when compared to their matched controls. T-cadherin is an adiponectin receptor, expressed on endothelial and smooth muscle cells, which regulates circulating and tissue adiponectin levels. Earlier studies have indicated its role in inflammation, vascular wall remodeling and atherosclerosis (Ouchi *et al*, 2000; Kudrjashova *et al*, 2002). Pfaff *et al.* evaluated the association of plasma T-cadherin and risk of atherosclerosis. Using an ELISA assay, they found that the abundance of plasma T-cadherin was negatively correlated with the severity of atherosclerosis (Pfaff *et al*, 2015). This is consistent with the decreased abundance of CDH13 in our study.

Fibulin is a common moderately abundant serum protein that can bind to ECM proteins, e.g. elastin and fibrinogen (Argraves *et al*, 1990). Alteration in the levels of FBLN1 have been linked with many diseases such as atherosclerosis, arterial stiffness, cardiovascular risk and type 2 diabetes (T2D). In the study of Kawata *et al.* reduced plasma levels of FBLN1 were found in patients with acute myocardial infarction and unstable angina (Kawata *et al*, 2001). Based on the latter findings, Argraves and coworkers examined the presence of FBLN1 in human carotid atherosclerosis lesions and found deposition patterns of FBLN1, suggesting its role in the pathophysiology of the disease (Argraves *et al*, 2009). In the plasma of T2D patients both increased and decreased levels of FBLN1 have been reported. With these differences a notable distinction was that one of the group was recently diagnosed (lower levels) and the other with T2D (increased levels). (Laugesen *et al*, 2013; Cangemi *et al*, 2011). Additionally, the lower levels of FBLN1 in recently diagnosed T2D was found to be linked with carotid-femoral stiffness (Laugesen *et al*, 2013). Based on this observation, Paapstel *et al*. examined the association of serum FBLN1 levels and arterial stiffness in atherosclerotic patients and found increased levels in the patients with established atherosclerosis (Paapstel *et al*, 2016). Although, these studies present the contradictory findings about the serum levels of FBLN1, these represent different phases and types or combination of disease and the role of FBLN1 needs to be clearly established in early stages of plaque progression. Moreover, alternative splicing produces four FBLN1 proteoforms (Smith *et al*, 2013; Overgaard *et al*, 2015): A, B, C and D. Amongst these,

FBLN1C has been reported to be the predominant proteoform in plasma (Overgaard *et al*, 2015) and has been identified in the tissue secretome analysis of coronary arteries (de la Cuesta *et al*, 2012).

With the limitation of this observed findings from our study, the speculation is that decrease in abundance of FBLN1C could represent a premature atherosclerotic phenotype, which could over the time become vulnerable to plaque development. Further, this could result from the structural differences in FBLN1C proteoforms potentially contributing to their interaction with ECM proteins. This in turn could subsequently lead to accumulation in arterial intima layers and thus be detected as decreased FBLN1C abundance in serum.

To confirm the differential abundance of serum proteins from the discovery phase, SRM-MS approach was used. However, due to the lack of similar cohort, the targeted proteomics analysis was carried out on the same samples, i.e. verification measurements on undepleted serum samples. The use of undepleted serum removed any potential biases resulting from the depletion approach. The results demonstrated that the initial observation could be conferred with a different method of preparation and analysis. The success with these measurements was enabled by the specificity and sensitivity of SRM-MS method.

These targeted measurements verified the lower abundance of FBLN1C in cases when compared to their matched controls. However the changes in other proteins were not confirmed, possibly reflecting their smaller quantitative differences and individual variability.

Whilst the verifications supported the initial observation of statistically significant difference in the relative abundance of FBLN1C, targeted measurements in an independent cohort are needed to further support the results. Additional structural studies indicating how FBLN1C proteoforms interact with ECM proteins could help in the understanding their contribution in plaque formation.

6.3 Targeted proteomics pipeline (Study III):

A selected reaction monitoring mass spectrometry (SRM-MS) pipeline has been developed for validation experiments for studies I and II. As an alternative to traditionally used Western blotting and ELISA assays, there has been a growing trend toward using SRM-MS for validation of candidate proteins of interest discovered by proteomics. When compared to ELISA and Western blotting, SRM enables the simultaneous analysis of multiple proteins targets in a single analysis. Moreover, the quality of MS data is often better than in standard approaches, in the context

of assay results and performance characteristics (Picotti & Aebersold, 2012). We have already established a targeted MS pipeline for studies I and II as indicated in the above sections, which is also applied to other projects in our group to monitor multiple proteins in a single run. Moreover, in contrast to discovery proteomics where technical replicates are required due to the stochastic nature of the data acquisition, in validation experiments, on account of sensitivity and selectivity, the time can be dedicated to the analysis of biological replicates. Moreover, the chromatographic time required to monitor a panel of a hundred or more peptides is somewhat less than that needed to detect the proteome using discovery phase experiment. Targeted analysis can therefore be performed with a much higher throughput.

The developed workflow is now adopted to validate the discovery phase findings of study I and II.

7 CONCLUSIONS AND FUTURE PROSPECTIVES:

The data presented in this thesis include the first reported serum proteomics profiles of children *en-route* to T1D, covering the period before autoantibody positivity to the development of clinical disease. In particular some of these changes were detected even before the appearance of autoantibodies, which could, with further evaluation, provide potential insights to the pathogenesis of T1D.

Nevertheless the distinguishing signatures detected in T1D developing children need further confirmation in a larger cohort. To facilitate the interpretation of such data, characterization of age related temporal changes in healthy subjects as well as the influence of diet and population with wider genetic variability should be established. Such studies will have a huge impact on the disease driven proteomics biomarker identification. Research towards these goals could proceed with mass spectrometry based technology, as increasingly better methods and instruments are producing more data at a greater speed. However, the serum proteome remains complex and challenging due to the inherent range of protein abundance and subject heterogeneity. Furthermore, although proteomics is usually achieved by analyzing digested proteins, the results are generally wrapped up to represent a single protein. In this manner important information about PTM, splice variants and proteoforms remains hidden. For instance, glycosylation, a common PTM that occurs in approximately 50% of the eukaryotic proteins (Apweiler *et al*, 1999) is also detected in majority of plasma proteins. It will be also important to study low stoichiometry PTMs such as non-enzymatic glycation, which is of particular interest from the view point of long term glycemic index in diabetes (Kulkarni *et al*, 2013). Detection of such aberrant of PTMs in disease states is an important future line of research.

Similarly, using MS-based proteomics of serum samples obtained from the YFS cohort, distinguishing profiles were identified in the subjects bearing premature atherosclerotic plaques when compared to their matched controls. Furthermore, using targeted proteomics, the lower differential abundance of one of the proteins was verified. In addition to the validation experiments, potentially important focus of future study would be to confirm if the structural differences in FBLN1C might influence its association with ECM proteins.

In summary, the serum proteomics analysis of two inflammatory diseases revealed characteristic patterns in the disease developing subjects. Importantly, the changes in these diseases were identified at subclinical stages of the disease revealing the potential of mass spectrometry based serum proteomics in biomarker discovery.

Such early markers could be helpful in prevention and diagnosis of the disease as well as monitoring the effect of pharmaceutical intervention.

The work presented in this thesis has focused on the use of data dependent mass spectrometry (DDA) based analysis of the serum proteomes. In relationship to our own data and much of the published literature (Moulder *et al*, 2017a) many studies of serum have been limited to the consistent comparison of a few hundred proteins. With the increasing performance of targeted methods, such as SRM or PRM, it is becoming possible to achieve confident detection and quantification of panels in the order of hundred or more proteins. Such methods represent a useful hypothesis driven alternative to discovery experiments. The detection of lower abundance proteins still remains a challenge for SRM, although using anti-peptide antibodies as with SISCAPA (Stable Isotope Standards and Capture by Anti-Peptide Antibodies) method have a great potential in mining the low abundant proteome (Anderson *et al*, 2004). Beyond DDA, data independent acquisition analysis of proteomes adds the benefit of recording the fragmentation patterns of almost all of the ionized peptides in the accessible mass range. This in turn may have potential in identifying the hidden proteome in unbiased and reproducible way. These and other current efforts, together with continued developments in MS based technologies, have indicated that the knowledge extracted through serum proteomics could help in predicting the personalized and public proteomic portrait.

Whilst mass spectrometry based proteomics is a useful discovery tool, in the clinic, selected markers would be preferably detected with a simple assay such as ELISA. Due to the inherent complexity of mass spectrometry and the necessary sample preparation work flows, for the use of proteomics in the clinic will likely be limited to specialized centers. In terms of alternatives approaches for protein characterization in biological fluids, antibody arrays have been developed targeting up to over 9000 proteins. These represent an interesting approach, particularly in respect to their simplicity and low sample requirement, although are limited by the specificity of the antibodies. Additionally, use of aptamer based assays, using DNA like amplification strategies for protein detection have shown significant promise in mining the depth of plasma proteome.

By nature, disease driven studies are difficult and require multidisciplinary team of experts with medical, computational and statistical background. Research scientists, clinicians, pharmaceutical industry experts and policymakers should carefully consider all above factors while debating further directions in developing clinical practices and personalized medicine.

ACKNOWLEDGEMENTS

This work was carried out in Turku Centre for Biotechnology and I am really grateful to all the staff and personnel for making the Centre an excellent place to work at.

Kiitos paljon Professor Riitta Lahesmaa for believing in me and giving an opportunity to pursue doctoral training in your group. I want to thanks for your constant encouragement, enthusiasm, freedom and promptness. Many thanks for the challenging projects. I am grateful for your commitment towards my graduation, career development and being a great mentor.

I am very grateful to my second supervisor Dr. Robert Moulder who shared his wide knowledge of proteomics and mass spectrometry, and continuum invaluable motivation and mentoring. Your constructive criticism on daily basis helped in fine tuning my skills and valuable advice to the original papers and the thesis are greatly appreciated. Besides being a fantastic boss, I also want to thank you for the personal discussions and helps.

I would like to acknowledge my thesis committee members. Firstly, Professor David R. Goodlett for his wise wisdom on the progress of my projects. I'm very grateful for your support, trust, motivation and remote availability. I grateful to Professor Harri Lähdesmäki for being another member of my thesis committee.

Kiitos paljon Arttu Heinonen and Pekka Haapaniemi for your excellent technical support. Most things about mass spectrometry I actually learned from you guys. I am grateful to Anne Rokka and Mirva Pääkkönen for their support. I would like to acknowledge Susumu Imanishi for the wonderful scientific discussions. *Kiitos paljon* Dr. Petri Kouvonen for your enthusiasm in teaching SRM, valuable insights, support and remote availability.

I want to thank Docents Maciej Lalowski and Tiina Öhman for reviewing my thesis and giving critical scientific insights.

I would like to acknowledge my opponent, Professor Ole N. Jensen, for accepting the invitation.

I am grateful to have wonderful collaborators that have made this work possible. I would like to thank all the DIPP collaborators and the personnel involved in making this unique cohort available to study. I would like to thank Professor Olli Raitakari for all the scientific insights and his group members especially Juhani Koskinen, Niina Pitkänen for their help during the preparation of YFS manuscript. I would like to express my sincere appreciation to the coauthors of DIPP and YFS

papers. *Kiitos paljon* to the participants of DIPP and YFS cohorts for their continued commitment. *Kiitos paljon* Professor Laura E. Elo for giving computational insights in YFS study. *Kiitos paljon* Dr. Mikko Venäläinen for all the support and availability during data analysis concerning YFS project. I would like to also acknowledge Kartiek, Tomi Suomi and Fatemehsadat Seyednasrollah for their help in R workflow.

Kiitos paljon to the technical and administrative staff, in particular, Anne Lahdenperä, Sirkku Grönroos, Marjo Hakkarainen, Sarita Heinonen, Mårten Hedman and Pasi Viljakainen. Thank you very much for the excellent support ☺ in a timely manner.

Kiitos paljon to Outi Irjala for all the thesis-related submission help and swiftness in the reply.

I would like to thanks all the present and past members of the ATLAS group. Thank you very much Robert Moulder, Niina Lietzén, Karoliina Hirvonen, Subhash Tripathi, Maheswarareddy Emani, Omid Rasool, Ankitha Shetty, Kartiek Kanduri, Verna Salo, Lingjia Kong, Ubaid Ullah, Henna Kallionpää, Mahesh J. Kulkarni, Essi Laajala, Tanja Buchacher, Sameer Jagirdar, Moin Khan, Rafael Sales De Albuquerque, Meraj Khan, Mari Lehti, Syed Andrabi, Marjo Hakkarainen, Sarita Heinonen, Elina Komsi, Viveka Öling, Sari Lehtimäki, Nelly Rahkonen, Senthil Palani, Tapio Lönnberg, Jussi Salmi, Mirkka Heinonen, Elizabeth Nguyen, Obaiah Dirasantha, Minna Kyläniemi, Alexy Sarapulov, Elina Pietilä, Elisa Närvä, Jussi Jalonen, Lotta Oikari, Päivi Junni, Sanna Edelman and Solie Tuomela.

धन्यवाद to Subhash and Mahesh, for the cheerful support in the lab and introducing me to the world of cell culturing.

Kiitos paljon to Niina Lietzén and Karoliina Hirvonen for being wonderful lab mates. Your contribution to scientific discussions, insightful suggestions and support has encouraged me.

I am really grateful to Sachin Kote and Swapnil Bhujbal for their over the years friendship and support.

I feel privileged to have such great friends! धन्यवाद to Subhash, Poonam, Prasanna, Amruta, Ameya, Rama, Mahesh, Bhagyashri, Hasan, Aman, Ankitha, Imran, Rahul, Sachin, Senthil, Naresh, Satish, Raghu, Yashu, Kartiek, Ajit, Swati and Vishal. I have shared many happy moments of life outside research with you guys. Our relaxing weekend evenings, social events, delicious food and trips have put a gilt edge to my life. A special thanks to Devvrat☺.

धन्यवाद to Senthil, Hasan, Ameya, Rahul, Sachin, Athresh, Aravind and Naresh for our fun filled gym sessions and discussion about cricket.

I also want to say *धन्यवाद* to Vipin, Nitin, Arjun, Kalyan, Swathi and Justyna for the lively company.

To my family, Aai (Mother) and Nana (Father)☺, I am thankful for everything that you did and still doing. Without your support, patience and motivation, I would have never reached to this platform. This doctoral hat belongs to both of you. A big thanks to Anil (Brother In-law) and Mau (Sister)☺, Raj, Apurva and Tushar, in my absence you have filled the void space to take care of Aai and Nana. In particular, thanks to Anil for your time, support and everything. I am really grateful to my KATTA friends for a frequent visit to my home and all the support whenever required. I am grateful to my father-in-law and mother-in-law, and their family for the constant encouragement. In particular, *धन्यवाद* to Gaurav (Brother In-law) for all the enjoyable discussion. Thank you very much to my nieces Anvee and Trusha☺ for bringing joy on site and remotely.

Preeti☺, I feel very lucky that we tied up. Many thanks for your love and for all the great times of togetherness. You are my strongest support and toughest critic.

Finally, I am really grateful to the financial support that this work received from National Technology Agency of Finland Distinguish Professor Program The Academy of Finland grants EVO-funding (Hospital District of Southwest Finland and Turku City) Orion Pharma and The Maud Kuistila Memorial Foundation. I appreciate the travel grant support from Turku Centre for System Biology and Turku Doctoral Programme of Molecular Medicine (TuDMM).

Turku, August 2018

Santosh Dilip Bhosale

REFERENCES

Addona TA, Shi X, Keshishian H, Mani DR, Burgess M, Gillette MA, Clauser KR, Shen D, Lewis GD, Farrell LA, Fifer MA, Sabatine MS, Gerszten RE & Carr SA (2011) A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. *Nat. Biotechnol.* **29:** 635–643

Adkins JN, Varnum SM, Auberry KJ, Moore RJ, Angell NH, Smith RD, Springer DL & Pounds JG (2002) Toward a Human Blood Serum Proteome. *Mol. Cell. Proteomics* **1:** 947–955

Aebersold R & Mann M (2003) Mass spectrometry-based proteomics. *Nature* **422:** 198–207

Aebersold R & Mann M (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* **537:** 347–355

Ahmed M, Forsberg J & Bergsten P (2005) Protein profiling of human pancreatic islets by two-dimensional gel electrophoresis and mass spectrometry. *J. Proteome Res.* **4:** 931–940

Albrethsen J, Kaas A, Schonle E, Swift P, Kocova M, Gammeltoft S, Hansen L & Mortensen HB (2009a) Evaluation of a type 1 diabetes serum cohort by SELDI-TOF MS protein profiling. *Proteomics.Clinical Appl.* **3:** 383–393

Albrethsen J, Kaas A, Schönle E, Swift P, Kocova M, Gammeltoft S, Hansen L & Mortensen HB (2009b) Evaluation of a type 1 diabetes serum cohort by SELDI-TOF MS protein profiling. *Proteomics - Clin. Appl.* **3:** 383–393

Alpert AJ (2008) Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal. Chem.* **80:** 62–76

Anderson NL (2010) The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum. *Clin. Chem.* **56:** 177–185

Anderson NL & Anderson NG (2002a) The Human Plasma Proteome. *Mol. Cell. Proteomics* **1:** 845–867

Anderson NL & Anderson NG (2002b) The human plasma proteome: history, character, and diagnostic prospects. *Mol.Cell.Proteomics* **1:** 845–867

Anderson NL, Anderson NG, Haines LR, Hardie DB, Olafson RW & Pearson TW (2004) Mass Spectrometric Quantitation of Peptides and Proteins Using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J. Proteome Res.* **3:** 235–244

Angel TE, Aryal UK, Hengel SM, Baker ES, Kelly RT, Robinson EW & Smith RD (2012) Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem. Soc. Rev.* **41:** 3912

Apweiler R, Hermjakob H & Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta - Gen. Subj.* **1473:** 4–8

Argraves WS, Tanaka A, Smith EP, Twal WO, Argraves KM, Fan D & Haudenschild CC (2009) Fibulin-1 and fibrinogen in human atherosclerotic lesions. *Histochem. Cell Biol.* **132:** 559–565

Argraves WS, Tran H, Burgess WH & Dickerson K (1990) Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain structure. *J. Cell Biol.* **111:** 3155–3164

Baldwin MA (2005) Mass spectrometers for the analysis of biomolecules. *Methods Enzymol.* **402:** 3–48

Bantscheff M, Lemeer S, Savitski MM & Kuster B (2012) Quantitative mass spectrometry in proteomics: Critical review update from 2007 to the present. *Anal. Bioanal. Chem.* **404:** 939–965

Bantscheff M, Schirle M, Sweetman G, Rick J & Kuster B (2007) Quantitative mass spectrometry in proteomics: A critical review. *Anal. Bioanal. Chem.* **389:** 1017–1031

Beck F, Burkhart JM, Geiger J, Zahedi RP & Sickmann A (2012) Robust workflow for iTRAQ-based peptide and protein quantification. *Methods Mol. Biol.* **893:** 101–113

Beck S, Michalski A, Raether O, Lubeck M, Kaspar S, Goedecke N, Baessmann C, Hornburg D, Meier F, Paron I, Kulak N a, Cox J & Mann M (2015) The impact II, a very high resolution quadrupole time-of-flight instrument for deep shotgun proteomics. *Mol. Cell. Proteomics***:** 2014–2029

Van Belle T, Coppieters K & Von Herrath M (2011a) Type 1 Diabetes: Etiology, Immunology, and Therapeutic Strategies. *Physiol Rev* **91:** 79–118

Van Belle TL, Coppieters KT & Von Herrath MG (2011b) Type 1 Diabetes: Etiology, Immunology, and Therapeutic Strategies. *Physiol. Rev.* **91:** 79–118

Bellei E, Bergamini S, Monari E, Fantoni LI, Cuoghi A, Ozben T & Tomasi A (2011) High-abundance proteins depletion for serum proteomic analysis: Concomitant removal of non-targeted proteins. *Amino Acids* **40:** 145–156

Bennett Jenson A, Rosenberg HS & Louis Notkins A (1980) PANCREATIC ISLET-CELL DAMAGE IN CHILDREN WITH FATAL VIRAL INFECTIONS. *Lancet* **316:** 354–358

Bentzon JF, Otsuka F, Virmani R & Falk E (2014) Mechanisms of plaque formation and rupture. *Circ. Res.* **114:** 1852–1866

Bhosale SD, Moulder R, Venäläinen MS, Koskinen JS, Pitkänen N, Juonala MT, Kähönen MAP, Lehtimäki TJ, Viikari JSA, Elo LL, Goodlett DR, Lahesmaa R & Raitakari OT (2018) Serum Proteomic Profiling to Identify Biomarkers of Premature Carotid Atherosclerosis. *Sci. Rep.* **8:** 9209

Bizzarri C, Pitocco D, Napoli N, Di Stasio E, Maggi D, Manfrini S, Suraci C, Cavallo MG, Cappa M, Ghirlanda G, Pozzilli P, Altomare M, Astorri E, Barchetta I, Benevento D, Beretta Anguissola G, Buzzetti R, Capizzi M, Cassone Faldetta MR, Cipolloni L, et al (2010) No protective effect of calcitriol on β-cell function in recent-onset type 1 diabetes: The IMDIAB XIII trial. *Diabetes Care* **33:** 1962–1963

Bjelosevic S, Pascovici D, Ping H, Karlaftis V, Zaw T, Song X, Molloy MP, Monagle P & Ignjatovic V (2017) Quantitative Age-specific Variability of Plasma Proteins in Healthy Neonates, Children and Adults. *Mol. Cell. Proteomics* **16:** 924–935

Blake GJ & Ridker PM (2002) Inflammatory bio-markers and cardiovascular risk prediction. *J Int Med* **252:** 283–94

Bodzon-Kulakowska A, Bierczynska-Krzysik A, Dylag T, Drabik A, Suder P, Noga M, Jarzebinska J & Silberring J (2007) Methods for samples preparation in proteomic research. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **849:** 1–31

Braisted JC, Kuntumalla S, Vogel C, Marcotte EM, Rodrigues AR, Wang R, Huang S-T, Ferlanti ES, Saeed AI, Fleischmann RD, Peterson SN & Pieper R (2008) The APEX Quantitative Proteomics Tool: Generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* **9:** 529

Braun RJ, Kinkl N, Beer M & Ueffing M (2007) Two-dimensional electrophoresis of membrane proteins. *Anal. Bioanal. Chem.* **389:** 1033–1045

Burch TC, Morris MA, Campbell-Thompson M, Pugliese A, Nadler JL & Nyalwidhe JO (2015) Proteomic Analysis of Disease Stratified Human Pancreas Tissue Indicates Unique Signature of Type 1 Diabetes. *PLoS One* **10:** e0135663

Burtis, A C, Edwrd R & Bruns, E D (2001) Fundamentals of Clinical Chemistry :6th Ed

Businaro R, Tagliani A, Buttari B, Profumo E, Ippoliti F, Di Cristofano C, Capoano R, Salvati B & Riganò R (2012) Cellular and molecular players in the atherosclerotic plaque progression. *Ann. N. Y. Acad. Sci.* **1262:** 134–141

Cabras T, Pisano E, Mastinu A, Denotti G, Pusceddu PP, Inzitari R, Fanali C, Nemolato S, Castagnola M & Messana I (2010a) Alterations of the Salivary Secretory Peptidome Profile in Children Affected by Type 1 Diabetes. *Mol. Cell. Proteomics* **9:** 2099–2108

Cabras T, Pisano E, Mastinu A, Denotti G, Pusceddu PP, Inzitari R, Fanali C, Nemolato S, Castagnola M & Messana I (2010b) Alterations of the salivary secretory peptidome profile in children affected by type 1 diabetes. *Mol. Cell. Proteomics* **9:** 2099–2108

Cangemi C, Skov V, Poulsen MK, Funder J, Twal WO, Gall MA, Hjortdal V, Jespersen ML, Kruse TA, Aagard J, Parving HH, Knudsen S, Høilund-Carlsen PF, Rossing P, Henriksen JE, Argraves WS & Rasmussen LM (2011) Fibulin-1 is a marker for arterial extracellular matrix alterations in type 2 diabetes. *Clin. Chem.* **57:** 1556–1565

Canto JG, Kiefe CI, Rogers WJ, Peterson ED, Frederick PD, French WJ, Gibson CM, Pollack C V, Ornato JP, Zalenski RJ, Penney J, Tiefenbrunn AJ & Greenland P (2011) Number of coronary heart disease risk factors and mortality in patients with first myocardial infarction. *JAMA* **306:** 2120–7

Cao Z, Yende S, Kellum JA & Robinson RA (2013) Additions to the Human Plasma Proteome via a Tandem MARS Depletion iTRAQ-Based Workflow. *Int. J. Proteomics* **2013:** 654356

Carmena R (2004) Atherogenic Lipoprotein Particles in Atherosclerosis. *Circulation* **109:** III-2-III-7

Chapman JD, Goodlett DR & Masselon CD (2013) Multiplexed and data-independent tandem mass

spectrometry for global proteome profiling. *Mass Spectrom. Rev.*

Chaussabel D, Pascual V & Banchereau J (2010) Assessing the human immune system through blood transcriptomics. *BMC Biol.* **8:** 84

Chen M, Daha MR & Kallenberg CG (2010) The complement system in systemic autoimmune disease. *J. Autoimmun.* **34:** J276-86

Choi M, Chang C-Y, Clough T, Broudy D, Killeen T, MacLean B & Vitek O (2014) MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics*

Colangelo CM, Chung L, Bruce C & Cheung K-H (2013) Review of software tools for design and analysis of large scale MRM proteomic datasets. *Methods* **61:** 287–298

Consortium IHGS (2001) Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921

Cooks RG (1995) Collision-induced dissociation: Readings and commentary. *J. Mass Spectrom.* **30:** 1215–1221

Cox J, Hein MY, Luber C a & Paron I (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. …* **13:** 2513–2526

Cox J & Mann M (2008a) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech* **26:** 1367–1372

Cox J & Mann M (2008b) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26:** 1367–72

Cox J & Mann M (2011) Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. *Annu. Rev. Biochem.* **80:** 273–299

Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen J V. & Mann M (2011a) Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10:** 1794–1805

Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen J V & Mann M (2011b) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10:** 1794–1805

Crèvecoeur I, Rondas D, Mathieu C & Overbergh L (2015) The beta-cell in type 1 diabetes: What have we learned from proteomic studies? *Proteomics - Clin. Appl.* **9:** 755–766

Crèvecoeur I, Vig S, Mathieu C & Overbergh L (2017) Understanding type 1 diabetes through proteomics. *Expert Rev. Proteomics* **14:** 571–580

da WH, Sherman BT & Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37:** 1–13

Dakna M, He Z, Yu WC, Mischak H & Kolch W (2009) Technical, bioinformatical and statistical aspects of liquid chromatography-mass spectrometry (LC-MS) and capillary electrophoresis-mass spectrometry (CE-MS) based clinical proteomics: A critical assessment. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **877:** 1250–1258

Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GDO, Pepys MB & Gudnason V (2004) C-Reactive Protein and Other Circulating Markers of Inflammation in the Prediction of Coronary Heart Disease. *N. Engl. J. Med.* **350:** 1387–1397

Dawson PH (1986) Quadrupole mass analyzers: Performance, design and some recent applications. *Mass Spectrom. Rev.* **5:** 1–37

DeGraba TJ, Hoehn GT, Nyquist P a., Wang H, Kenney R, Gonzales D a., Kern SJ, Ying S-X, Munson PJ & Suffredini AF (2011) Biomarker Discovery in Serum from Patients with Carotid Atherosclerosis. *Cerebrovasc. Dis. Extra* **1:** 115–129

Didangelos A, Stegemann C & Mayr M (2012) The -omics era: Proteomics and lipidomics in vascular research. *Atherosclerosis* **221:** 12–17

Domon B & Aebersold R (2010a) Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* **28:** 710–721

Domon B & Aebersold R (2010b) Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* **28:** 710–721

Donath MY, Størling J, Maedler K & Mandrup-Poulsen T (2003) Inflammatory mediators and islet beta-cell failure: a link between type 1 and type 2 diabetes. *J. Mol. Med. (Berl).* **81:** 455–70

Dong J-Y, Zhang W-G, Chen J, Zhang Z-L, Han S-F & Qin L-Q (2013) Vitamin D Intake and Risk of Type 1 Diabetes: A Meta-Analysis of Observational Studies. *Nutrients* **5:** 3551–3562

Douglas DJ, Frank AJ & Mao D (2005) Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.* **24:** 1–29

Dwivedi RC, Krokhin O V, Cortens JP & Wilkins JA (2010) Assessment of the reproducibility of random hexapeptide peptide library-based protein normalization. *J. Proteome Res.* **9:** 1144–1149

El-Aneed A, Cohen A & Banoub J (2009) Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Appl. Spectrosc. Rev.* **44:** 210–230

Elo LL, Filén S, Lahesmaa R & Aittokallio T (2008) Reproducibility-optimized test statistic for ranking genes in microarray studies. In *IEEE/ACM Transactions on Computational Biology and Bioinformatics* pp 423–431.

Elo LL, Hiissa J, Tuimala J, Kallio A, Korpelainen E & Aittokallio T (2009) Optimized detection of differential expression in global profiling experiments: case studies in clinical transcriptomic and quantitative proteomic datasets. *Brief. Bioinform.* **10:** 547–555

Elo LL, Mykkänen J, Nikula T, Järvenpää H, Simell S, Aittokallio T, Hyöty H, Ilonen J, Veijola R, Simell T, Knip M, Simell O & Lahesmaa R (2010) Early suppression of immune response pathways characterizes children with prediabetes in genome-wide gene expression profiling. *J. Autoimmun.* **35:** 70–76

Elosua R, Ordovas JM, Cupples LA, Fox CS, Polak JF, Wolf P a, D'Agostino R a & O'Donnell CJ (2004) Association of APOE genotype with carotid atherosclerosis in men and women: the Framingham Heart Study. *J. Lipid Res.* **45:** 1868–75

Van Den Elzen P, Garg S, León L, Brigl M, Leadbetter EA, Gumperz JE, Dascher CC, Cheng TY, Sacks FM, Illarionov PA, Besra GS, Kent SC, Moody DB & Brenner MB (2005) Apolipoprotein-mediated pathways of lipid antigen presentation. *Nature* **437:** 906–910

Eng JK, Mccormack AL & Yates JR (1994) An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *Am. Soc. Mass Spectrom.* **5:** 976–989

Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, Concannon P, Mychaleckyj JC, Todd JA, Bonella P, Fear AL, Lavant E, Louey A & Moonsamy P (2008) HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk analysis of the type 1 diabetes genetics consortium families. *Diabetes* **57:** 1084–1092

Escher C, Reiter L, Maclean B, Ossola R, Herzog F, Chilton J, Maccoss MJ & Rinner O (2012) Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **12:** 1111–1121

Farrah T, Deutsch EW, Kreisberg R, Sun Z, Campbell DS, Mendoza L, Kusebauch U, Brusniak MY, Hüttenhain R, Schiess R, Selevsek N, Aebersold R & Moritz RL (2012) PASSEL: The PeptideAtlas SRMexperiment library. *Proteomics* **12:** 1170–1175

Fenn J, Mann M, Meng C, Wong S & Whitehouse C (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science (80-.).* **246:** 64–71

Filippi CM & von Herrath MG (2008) Viral trigger for type 1 diabetes: pros and cons. *Diabetes* **57:** 2863–2871

Fink TM, Jenne DE & Lichter P (1992) The human vitronectin (complement S-protein) gene maps to the centromeric region of 17q. *Hum. Genet.* **88:** 569–572

Fischer R, Trudgian DC, Wright C, Thomas G, Bradbury LA, Brown MA, Bowness P & Kessler BM (2012) Discovery of candidate serum proteomic and metabolomic biomarkers in ankylosing spondylitis. *Mol. Cell. Proteomics* **11:** M111.013904

Fournier ML, Gilmore JM, Martin-Brown SA & Washburn MP (2007) Multidimensional separations-based shotgun proteomics. *Chem. Rev.* **107:** 3654–3686

Friedman J, Hastie T & Tibshirani R (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33:**

Gallien S, Duriez E, Crone C, Kellmann M, Moehring T & Domon B (2012) Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. *Mol. Cell. Proteomics* **11:** 1709–1723

Gamble DR, Kinsley ML, Fitzgerald MG, Bolton FRCPR & Taylorjj KW (1969) Viral Antibodies in Diabetes Mellitus. *Br. Med. J.* **3:** 627–630

Geiger T, Cox J & Mann M (2010) Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-ion Fragmentation. *Mol. Cell. Proteomics* **9:** 2252–2261

Gelderman MP & Simak J (2008) Functional Proteomics. *Methods* **484:** 79–93

Geman D, d'Avignon C, Naiman DQ & Winslow RL (2004) Classifying gene expression profiles

from pairwise mRNA comparisons. *Stat. Appl. Genet. Mol. Biol.* **3:** Article19

Gerszten RE, Asnani A & Carr SA (2011) Status and prospects for discovery and verification of new biomarkers of cardiovascular disease by proteomics. *Circ. Res.* **109:** 463–474

Geyer PE, Albrechtsen NJW, Tyanova S, Grassl N, Iepsen EW, Lundgren J, Madsbad S, Holst JJ, Torekov SS & Mann M (2016a) Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Mol. Syst. Biol.* **12:** 901

Geyer PE, Holdt LM, Teupser D & Mann M (2017) Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* **13:** 942

Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D & Mann M (2016b) Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* **2:** 185–195

Gharbi S, Gaffney P, Yang A, Zvelebil MJ, Cramer R, Waterfield MD & Timms JF (2002) Evaluation of Two-dimensional Differential Gel Electrophoresis for Proteomic Expression Analysis of a Model Breast Cancer Cell System. *Mol. Cell. Proteomics* **1:** 91–98

Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R & Aebersold R (2012a) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11:** O111.016717

Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R & Aebersold R (2012b) Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **11:** O111.016717

Giongo A, Gano KA, Crabb DB, Mukherjee N, Novelo LL, Casella G, Drew JC, Ilonen J, Knip M, Hyoty H, Veijola R, Simell T, Simell O, Neu J, Wasserfall CH, Schatz D, Atkinson MA & Triplett EW (2011) Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J* **5:** 82–91

Goecks J, Nekrutenko A, Taylor J, Afgan E, Ananda G, Baker D, Blankenberg D, Chakrabarty R, Coraor N, Goecks J, Von Kuster G, Lazarus R, Li K, Taylor J & Vincent K (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11:**

Good D, Thongboonkerd V, Novak J, Bascands J-L, Schanstra J, Coon J, Dominiczak A & Mischak H (2007) Body Fluid Proteomics for Biomarker Discovery: Lessons from the Past Hold the Key to Success in the Future. *J. Proteome Res.*

Gorshkov M V., Tolic P, Paša Tolić L, Udseth HR, Anderson GA, Huang BM, Bruce JE, Prior DC, Hofstadler SA, Tang L, Chen L-Z, Willett JA, Rockwood AL, Sherman MS & Smith RD (1998) Electrospray ionization—Fourier transform ion cyclotron resonance mass spectrometry at 11.5 tesla: Instrument design and initial results. *J. Am. Soc. Mass Spectrom.* **9:** 692–700

Granér M, Kahri J, Varpula M, Salonen RM, Nyyssönen K, Jauhiainen M, Nieminen MS, Syvänne M & Taskinen MR (2008) Apolipoprotein E polymorphism is associated with both carotid and coronary atherosclerosis in patients with coronary artery disease. *Nutr. Metab. Cardiovasc. Dis.* **18:** 271–277

Graves PR & Haystead T a J (2002) Molecular Biologist ' s Guide to Proteomics. *Microbiol. Mol. Biol. Rev. Microbiol. Mol. Biol. Rev.* **66:** 39–63

Guo T, Kouvonen P, Koh CC, Gillet LC, Wolski WE, Rost HL, Rosenberger G, Collins BC, Blum LC, Gillessen S, Joerger M, Jochum W & Aebersold R (2015) Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.* **21:** 407–413

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH & Aebersold R (1999a) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17:** 994–99.

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH & Aebersold R (1999b) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17:** 994–999

Han Y, Zhao S, Gong Y, Hou G, Li X & Li L (2015) Serum cyclin-dependent kinase 9 is a potential biomarker of atherosclerotic inflammation. *Oncotarget; Vol 7, No 2*

Hao P, Ren Y, Pasterkamp G, Moll FL, de Kleijn DP V & Sze SK (2014) Deep proteomic profiling of human carotid atherosclerotic plaques using multidimensional LC-MS/MS. *Proteomics - Clin. Appl.* **8:** 631–635

Harjutsalo V, Sund R, Knip M & Groop P-H (2013) Incidence of Type 1 Diabetes in Finland. *J. Am. Med. Assoc.* **310:** 1–12

Heeschen C, Dimmeler S, Hamm CW, van den Brand MJ, Boersma E, Zeiher AM & Simoons ML (2003) Soluble CD40 ligand in acute coronary syndromes. *N Engl J Med* **348:** 1104–1111

Holland LA & Jorgenson JW (1995) Separation of nanoliter samples of biological amines by a comprehensive two-dimensional microcolumn liquid chromatography system. *Anal. Chem.* **67:** 3275–3283

Hu A, Noble WS & Wolf-Yadlin A (2016) Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research* **5:** 419

Hu L, Evers S, Lu ZH, Shen Y & Chen J (2004) Two-dimensional protein database of human pancreas. *Electrophoresis* **25:** 512–518

Hunt DF, Yates JR, Shabanowitz J, Winston S & Hauer CR (1986) Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci.* **83:** 6233–6237

Hyöty H (2016) Viruses in type 1 diabetes. *Pediatr. Diabetes* **17:** 56–64

Iliuk A, Galan J & Tao WA (2009) Playing tag with quantitative proteomics. *Anal. Bioanal. Chem.* **393:** 503–513

IM S, Joner G, PA J, Eskild A, PA T & LC S (2012) Maternal serum levels of 25-hydroxy-vitamin d during pregnancy and risk of type 1 diabetes in the offspring. *Diabetes* **61:** 175–178

Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J & Mann M (2005) Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol. Cell. Proteomics* **4:** 1265–1272

Issaq HJ, Xiao Z & Veenstra TD (2007) Serum and plasma proteomics. *Chem. Rev.* **107:** 3601–3620

James P, Quadroni M, Carafoli E & Gonnet G (1993) Protein identification by mass profile fingerprinting. *Biochem. Biophys. Res. Commun.* **195:** 58–64

Josic D & Kovac S (2010) Reversed-phase High Performance Liquid Chromatography of proteins. *Curr. Protoc. Protein Sci.* **Chapter 8:** Unit 8.7

Jr GD, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC & Lempicki RA (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4:** P3

Juhela S, Hyoty H, Roivainen M, Harkonen T, Putto-Laurila A, Simell O & Ilonen J (2000) T-cell responses to enterovirus antigens in children with type 1 diabetes. *Diabetes* **49:** 1308–1313

Kallionpää H, Elo LL, Laajala E, Mykkänen J, Ricaño-Ponce I, Vaarma M, Laajala TD, Hyöty H, Ilonen J, Veijola R, Simell T, Wijmenga C, Knip M, Lähdesmäki H, Simell O & Lahesmaa R (2014) Innate immune activity is detected prior to seroconversion in children with HLA-conferred type 1 diabetes susceptibility. *Diabetes* **63:** 2402–2414

Karas M & Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60:** 2299–2301

Kawata K, Tanaka A, Arai M, Argraves WS & Fukutake K (2001) Alteration of plasma fibulin-1 concentrations in ischemic heart diseases. *Jpn. J. Thromb. Hemost.* **12:** 126–132

Keshishian H, Burgess MW, Gillette MA, Mertins P, Clauser KR, Mani DR, Kuhn EW, Farrell LA, Gerszten RE & Carr SA (2015) Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury. *Mol. Cell. Proteomics* **14:** 2375–2393

Keshishian H, Burgess MW, Specht H, Wallace L, Clauser KR, Gillette MA & Carr SA (2017) Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry. *Nat. Protoc.* **12:** 1683–1701

Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LDN, et al (2014) A draft map of the human proteome. *Nature* **509:** 575–581

Kimpimaki T & Knip M (2001) Disease-associated autoantibodies as predictive markers of type 1 diabetes mellitus in siblings of affected children. *J. Pediatr. Endocrinol. Metab.* **14 Suppl 1:** 575–587

Kingdon KH (1923) A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Phys. Rev.* **21:** 408–418

De Kleijn DP V, Moll FL, Hellings WE, Ozsarlak-Sozer G, De Bruin P, Doevendans PA, Vink A, Catanzariti LM, Schoneveld AH, Algra A, Daemen MJ, Biessen EA, De Jager W, Zhang H, De Vries JP, Falk E, Lim SK, Van Der Spek

PJ, Sze SK & Pasterkamp G (2010) Local atherosclerotic plaques are a source of prognostic biomarkers for adverse cardiovascular events. *Arterioscler. Thromb. Vasc. Biol.* **30:** 612–619

Klose J (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26:** 231–243

Klose J & Kobalz U (1995) Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* **16:** 1034–1059

Knip M, Akerblom HK, Al Taji E, Becker D, Bruining J, Castano L, Danne T, de Beaufort C, Dosch HM, Dupre J, Fraser WD, Howard N, Ilonen J, Konrad D, Kordonouri O, Krischer JP, Lawson ML, Ludvigsson J, Madacsy L, Mahon JL, et al (2018) Effect of Hydrolyzed Infant Formula vs Conventional Formula on Risk of Type 1 Diabetes: The TRIGR Randomized Clinical Trial. *JAMA.* **319:** 38–48.

Knip M & Simell O (2011) Environmental triggers of type 1 diabetes. *Cold Spring Harb. Perspect. Biol.* **3:** 1–15

Knip M, Veijola R, Virtanen SM, Hyöty H, Vaarala O & Åkerblom HK (2005) Environmental triggers and determinants of type 1 diabetes. *Diabetes* **54:**

Kohlbacher O, Reinert K, Gröpl C, Lange E, Pfeifer N, Schulz-Trieglaff O & Sturm M (2007) TOPP - The OpenMS proteomics pipeline. In *Bioinformatics*

Krischer J, Rewers M, Barriga K, Baxter J, Deas A, Eisenbarth G, Emery L, Gesualdo P, Hoffman M, Norris J, Waugh K, Weber S, She J-X, Muir A, Schatz D, Hopkins D, Steed L, Silvis K, Shankar M, huang Y-H, et al (2007) The Environmental Determinants of Diabetes in the Young (TEDDY) study: Study design. *Pediatr. Diabetes* **8:** 286–298

Kristensen LP, Larsen MR, Mickley H, Saaby L, Diederichsen AC, Lambrechtsen J, Rasmussen LM & Overgaard M (2014a) Plasma proteome profiling of atherosclerotic disease manifestations reveals elevated levels of the cytoskeletal protein vinculin. *J. Proteomics* **101:** 141–153

Kristensen LP, Larsen MR, Mickley H, Saaby L, Diederichsen AC, Lambrechtsen J, Rasmussen LM & Overgaard M (2014b) Plasma proteome profiling of atherosclerotic disease manifestations reveals elevated levels of the cytoskeletal protein vinculin. *J. Proteomics* **101:** 141–153

Kriszbacher I, Koppán M & Bódis J (2005) Inflammation, atherosclerosis, and coronary artery disease. *N. Engl. J. Med.* **353:** 429-430; author reply 429-430

Krogvold L, Edwin B, Buanes T, Frisk G, Skog O, Anagandula M, Korsgren O, Undlien D, Eike MC, Richardson SJ, Leete P, Morgan NG, Oikarinen S, Oikarinen M, Laiho JE, Hyoty H, Ludvigsson J, Hanssen KF & Dahl-Jorgensen K (2015) Detection of a low-grade enteroviral infection in the islets of langerhans of living patients newly diagnosed with type 1 diabetes. *Diabetes* **64:** 1682–1687

Krusemark CJ, Frey BL, Belshaw PJ & Smith LM (2009) Modifying the Charge State Distribution of Proteins in Electrospray Ionization Mass Spectrometry by Chemical Derivatization. *J. Am. Soc. Mass Spectrom.* **20:** 1617–1625

Kudrjashova E, Bashtrikov P, Bochkov V, Parfyonova Y, Tkachuk V, Antropova J, Iljinskaya O, Tararak E, Erne P, Ivanov D, Philippova M & Resink TJ (2002) Expression of adhesion molecule T-cadherin is increased during neointima formation in experimental restenosis. *Histochem. Cell Biol.* **118:** 281–290

Kuehl MN, Rodriguez H, Burkhardt BR & Alman AC (2015) Tumor Necrosis Factor-alpha, Matrix-Metalloproteinases 8 and 9 Levels in the Saliva Are Associated with Increased Hemoglobin A1c in Type 1 Diabetes Subjects. *PLoS One* **10:** e0125320

Kulkarni MJ, Korwar AM, Mary S, Bhonsle HS & Giri AP (2013) Glycated proteome: From reaction to intervention. *Proteomics - Clin. Appl.* **7:** 155–170

Kullo IJ & Cooper LT (2010) Early identification of cardiovascular risk using genomics and proteomics. *Nat. Rev. Cardiol.* **7:** 309–317

Kulmala P, Savola K, Petersen JS, Vahasalo P, Karjalainen J, Lopponen T, Dyrberg T, Akerblom HK & Knip M (1998) Prediction of insulin-dependent diabetes mellitus in siblings of children with diabetes. A population-based study. The Childhood Diabetes in Finland Study Group. *J. Clin. Invest.* **101:** 327–336

Kupila A, Muona P, Simell T, Arvilommi P, Savolainen H, Hamalainen AM, Korhonen S, Kimpimaki T, Sjoroos M, Ilonen J, Knip M, Simell O & for the Prevention of Type I Diabetes in Finland JDRFC (2001) Feasibility of genetic and immunological prediction of type I diabetes in a population-based birth cohort. *Diabetologia* **44:** 290–297

Kusebauch U, Campbell DS, Deutsch EW, Chu CS, Spicer DA, Brusniak MY, Slagel J, Sun Z, Stevens J, Grimes B, Shteynberg D, Hoopmann MR, Blattmann P, Ratushny A V, Rinner O, Picotti P, Carapito C, Huang CY, Kapousouz M, Lam H, et al (2016) Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell*

Kwasnik A, Tonry C, Ardle AM, Butt AQ, Inzitari R & Pennington SR (2016) Proteomes, their compositions and their sources. In *Advances in Experimental Medicine and Biology* pp 3–21.

de la Cuesta F, Alvarez-Llamas G, Maroto AS, Donado A, Zubiri I, Posada M, Padial LR, Pinto AG, Barderas MG & Vivanco F (2011) A proteomic focus on the alterations occurring at the human atherosclerotic coronary intima. *Mol. Cell. Proteomics* **10:** M110.003517

de la Cuesta F, Barderas MG, Calvo E, Zubiri I, Maroto AS, Darde VM, Martin-Rojas T, Gil-Dones F, Posada-Ayala M, Tejerina T, Lopez JA, Vivanco F & Alvarez-Llamas G (2012) Secretome analysis of atherosclerotic and non-atherosclerotic arteries reveals dynamic extracellular remodeling during pathogenesis. *J. Proteomics* **75:** 2960–2971

Lada AT & Rudel LL (2004) Associations of low density lipoprotein particle composition with atherogenicity. *Curr. Opin. Lipidol.* **15:** 19–24

Laitinen OH, Honkanen H, Pakkanen O, Oikarinen S, Hankaniemi MM, Huhtala H, Ruokoranta T, Lecouturier V, Andre P, Harju R, Virtanen SM, Lehtonen J, Almond JW, Simell T, Simell O, Ilonen J, Veijola R, Knip M & Hyoty H (2014) Coxsackievirus B1 Is Associated With Induction of beta-Cell Autoimmunity That Portends Type 1 Diabetes. *Diabetes* **63:** 446–455

Lange V, Picotti P, Domon B & Aebersold R (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4:**

Laugesen E, Høyem P, Christiansen JS, Knudsen ST, Hansen KW, Argraves WS, Hansen TK, Poulsen PL & Rasmussen LM (2013) Plasma levels of the arterial wall protein fibulin-1 are associated with carotid-femoral pulse wave velocity: A cross-sectional study. *Cardiovasc. Diabetol.* **12:**

Lee R, Fischer R, Charles PD, Adlam D, Valli A, Di Gleria K, Kharbanda RK, Choudhury RP, Antoniades C, Kessler BM & Channon KM (2017) A novel workflow combining plaque imaging, plaque and plasma proteomics identifies biomarkers of human coronary atherosclerotic plaque disruption. *Clin. Proteomics* **14:** 22

De Lemos JA, Morrow DA, Sabatine MS, Murphy SA, Gibson CM, Antman EM, McCabe CH, Cannon CP & Braunwald E (2003) Association between plasma levels of monocyte chemoattractant protein-1 and long-term clinical outcomes in patients with acute coronary syndromes. *Circulation* **107:** 690–695

Lepedda AJ, Nieddu G, Zinellu E, De Muro P, Piredda F, Guarino A, Spirito R, Carta F, Turrini F & Formato M (2013) Proteomic analysis of plasma-purified VLDL, LDL, and HDL fractions from atherosclerotic patients undergoing carotid endarterectomy: Identification of serum amyloid a as a potential marker. *Oxid. Med. Cell. Longev.* **2013:**

Lernmark B, Elding-Larsson H, Hansson G, Lindberg B, Lynch K & Sjöblad S (2004) Parent responses to participation in genetic screening for diabetes risk. *Pediatr. Diabetes* **5:** 174–181

Libby P (2002) Inflammation in atherosclerosis. *Nature* **420:** 868–74

Libby P, Ridker PM & Maseri A (2002) Inflammation and atherosclerosis. *Circulation* **105:** 1135–1143

Lietzén N, Cheng L, Moulder R, Siljander H, Laajala E, Härkönen T, Peet A, Vehtari A, Tillmann V, Knip M, Lähdesmäki H & Lahesmaa R (2018) Characterization and non-parametric modeling of the developing serum proteome during infancy and early childhood. *Sci. Rep.* **8:** 5883

Lind L, ??rnl??v J, Lindahl B, Siegbahn A, Sundstr??m J & Ingelsson E (2015) Use of a proximity extension assay proteomics chip to discover new biomarkers for human atherosclerosis. *Atherosclerosis* **242:** 205–210

Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates 3rd JR & Yates JR (1999) Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17:** 676–682

Liu C-W, Atkinson MA & Zhang Q (2016a) Type 1 diabetes cadaveric human pancreata exhibit a unique exocrine tissue proteomic profile. *Proteomics* **16:** 1432–1446

Liu C-W, Bramer L, Webb-Robertson B-J, Waugh K, Rewers MJ & Zhang Q (2017) Temporal expression profiling of plasma proteins reveals oxidative stress in early stages of Type 1 Diabetes progression. *J. Proteomics*: 0–1

Liu CW, Atkinson MA & Zhang Q (2016b) Type 1 diabetes cadaveric human pancreata exhibit a

unique exocrine tissue proteomic profile. *Proteomics* **16:** 1432–1446

Liu CW, Bramer L, Webb-Robertson BJ, Waugh K, Rewers MJ & Zhang Q (2016c) Temporal profiles of plasma proteome during childhood development. *J. Proteomics*

Liu J, Walp ER & May CL (2012) Elevation of transcription factor Islet-1 levels in vivo increases beta-cell function but not beta-cell mass. *Islets* **4:** 199–206

Liu Y, Buil A, Collins BC, Gillet LC, Blum LC, Cheng LY, Vitek O, Mouritsen J, Lachance G, Spector TD, Dermitzakis ET & Aebersold R (2015) Quantitative variability of 342 plasma proteins in a human twin population. *Mol. Syst. Biol.* **11:** 786

López JL (2007) Two-dimensional electrophoresis in proteome expression analysis. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **849:** 190–202

Lusis AJ (2000) Atherosclerosis. *Nature* **407:** 233–241

Lyons TJ & Basu A (2012) Biomarkers in diabetes: Hemoglobin A1c, vascular and tissue markers. *Transl. Res.* **159:** 303–312

Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A & Lajoie G (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **17:** 2337–2342

MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC & MacCoss MJ (2010a) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26:** 966–968

MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC & MacCoss MJ (2010b) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26:** 966–968

Mahley RW, Innerarity TL, Rall SC & Weisgraber KH (1984) Plasma lipoproteins: apolipoprotein structure and function. *J. Lipid Res.* **25:** 1277–94

Makarov A (2000) Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* **72:** 1156–1162

Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, Ranish J, Raught B, Schmitt R, Werner T, Kuster B & Aebersold R (2007) Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.* **25:** 125–131

Massa O, Alessio M, Russo L, Nardo G, Bonetto V, Bertuzzi F, Paladini A, Iafusco D, Patera P, Federici G, Not T, Tiberti C, Bonfanti R & Barbetti F (2013) Serological Proteome Analysis (SERPA) as a tool for the identification of new candidate autoantigens in type 1 diabetes. *J. Proteomics* **82:** 263–273

Masselon C, Anderson GA, Harkewicz R, Bruce JE, Pasa-Tolic L & Smith RD (2000) Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptide identification from mixtures. *Anal. Chem.* **72:** 1918–1924

McDonald WH & Yates JR (2002) Shotgun proteomics and biomarker discovery. *Dis. Markers* **18:** 99–105

McGill H.C. J, McMahan CA, Herderick EE, Malcom GT, Tracy RE & Jack P (2000) Origin of atherosclerosis in childhood and adolescence. In *American Journal of Clinical Nutrition*

McGuire JN, Eising S, Wagner AM & Pociot F (2010) Screening newborns for candidate biomarkers of type 1 diabetes. *Arch. Physiol. Biochem.* **116:** 227–232

McHugh L & Arthur JW (2008) Computational methods for protein identification from mass spectrometry data. *PLoS Comput. Biol.* **4:**

Medzhitov R (2008) Origin and physiological roles of inflammation. *Nature* **454:** 428–435

Meier M, Kaiser T, Herrmann A, Knueppel S, Hillmann M, Koester P, Danne T, Haller H, Fliser D & Mischak H (2005) Identification of urinary protein pattern in type 1 diabetic adolescents with early diabetic nephropathy by a novel combined proteome analysis. *J. Diabetes Complications* **19:** 223–232

Metz TO, Jacobs JM, Gritsenko MA, Fontes G, Qian W-J, Camp DG, Poitout V & Smith RD (2006) Characterization of the Human Pancreatic Islet Proteome by Two-Dimensional LC/MS/MS. *J. Proteome Res.*

Metz TO, Qian WJ, Jacobs JM, Gritsenko MA, Moore RJ, Polpitiya AD, Monroe ME, Camp DG, Mueller PW & Smith RD (2008) Application of proteomics in the discovery of candidate protein biomarkers in a diabetes autoantibody standardization program sample subset. *J. Proteome Res.* **7:** 698–707

Meuwese MC, Stroes ESG, Hazen SL, van Miert JN, Kuivenhoven JA, Schaub RG, Wareham NJ, Luben R, Kastelein JJP, Khaw KT &

Boekholdt SM (2007) Serum Myeloperoxidase Levels Are Associated With the Future Risk of Coronary Artery Disease in Apparently Healthy Individuals. The EPIC-Norfolk Prospective Population Study. *J. Am. Coll. Cardiol.* **50:** 159–165

Michel JB, Martin-Ventura JL, Egido J, Sakalihasan N, Treska V, Lindholt J, Allaire E, Thorsteinsdottir U, Cockerill G & Swedenborg J (2011) Novel aspects of the pathogenesis of aneurysms of the abdominal aorta in humans. *Cardiovasc. Res.* **90:** 18–27

Miettinen ME, Reinert L, Kinnunen L, Harjutsalo V, Koskela P, Surcel HM, Lamberg-Allardt C & Tuomilehto J (2012) Serum 25-hydroxyvitamin D level during early pregnancy and type 1 diabetes risk in the offspring. *Diabetologia* **55:** 1291–1294

Mohammed Y, Domanski D, Jackson AM, Smith DS, Deelder AM, Palmblad M & Borchers CH (2014) PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. *J. Proteomics* **106:** 151–161

Mokou M, Lygirou V, Vlahou A & Mischak H (2017) Proteomics in cardiovascular disease: recent progress and clinical implication and implementation. *Expert Rev. Proteomics* **14:** 117–136

Monteoliva L & Albar JP (2004) Differential proteomics: an overview of gel and non-gel based approaches. *Brief. Funct. Genomic. Proteomic.* **3:** 220–239

Motoyama A, Xu T, Ruse CI, Wohlschlegel JA & Yates JR (2007) Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. *Anal. Chem.* **79:** 3623–3634

Moulder R, Bhosale SD, Erkkila T, Laajala E, Salmi J, Nguyen E V, Kallionpaa H, Mykkanen J, Vaha-Makila M, Hyoty H, Veijola R, Ilonen J, Simell T, Toppari J, Knip M, Goodlett DR, Lahdesmaki H, Simell O & Lahesmaa R (2015) Serum Proteomes Distinguish Children Developing Type 1 Diabetes in a Cohort with HLA-Conferred Susceptibility. *Diabetes*

Moulder R, Bhosale SD, Goodlett DR & Lahesmaa R (2017a) Analysis of the plasma proteome using iTRAQ and TMT-based Isobaric labeling. *Mass Spectrom. Rev.***:** 1–24

Moulder R, Bhosale SD, Lahesmaa R & Goodlett DR (2017b) The progress and potential of proteomic biomarkers for type 1 diabetes in children. *Expert Rev. Proteomics* **14:** 31–41

Mueller LN, Brusniak MY, Mani DR & Aebersold R (2008) An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J. Proteome Res.* **7:** 51–61

Muñoz J & Heck AJR (2014) From the human genome to the human proteome. *Angew. Chemie - Int. Ed.* **53:** 10864–10866

Murray CJL & Lopez AD (2013) Measuring the Global Burden of Disease. *N. Engl. J. Med.* **369:** 448–457

Nerup J, Platz P, Andersen OO, Christy M, Lyngsoe J, Poulsen JE, Ryder LP, Nielsen LS, Thomsen M & Svejgaard A (1974) HL-A antigens and diabetes mellitus. *Lancet (London, England)* **2:** 864–866

Nesvizhskii AI (2010) A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J. Proteomics* **73:** 2092–2123

Nikolov M, Schmidt C & Urlaub H (2012) Quantitative mass spectrometry-based proteomics: An overview. *Methods Mol. Biol.* **893:** 85–100

O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250:** 4007–21

Olsen J V, Macek B, Lange O, Makarov A, Horning S & Mann M (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **4:** 709–712

Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A & Mann M (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **1:** 376–386

Ong S-E & Mann M (2005) Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1:** 252–62

Oresic M, Gopalacharyulu P, Mykkanen J, Lietzen N, Makinen M, Nygren H, Simell S, Simell V, Hyoty H, Veijola R, Ilonen J, Sysi-Aho M, Knip M, Hyotylainen T & Simell O (2013) Cord serum lipidome in prediction of islet autoimmunity and type 1 diabetes. *Diabetes* **62:** 3268–3274

Orešič M, Simell S, Sysi-Aho M, Näntö-Salonen K, Seppänen-Laakso T, Parikka V, Katajamaa M, Hekkala A, Mattila I, Keskinen P, Yetukuri L, Reinikainen A, Lähde J, Suortti T, Hakalax J, Simell T, Hyöty H, Veijola R, Ilonen J, Lahesmaa R, et al (2008) Dysregulation of lipid and amino acid metabolism precedes islet

autoimmunity in children who later progress to type 1 diabetes. *J. Exp. Med.* **205:** 2975–2984

Ostman J, Lonnberg G, Arnqvist HJ, Blohme G, Bolinder J, Schnell a. E, Eriksson JW, Gudbjornsdottir S, Sundkvist G & Nystrom L (2008) Gender differences and temporal variation in the incidence of type 1 diabetes: Results of 8012 cases in the nationwide Diabetes Incidence Study in Sweden 1983-2002. *J. Intern. Med.* **263:** 386–394

Ouchi N, Kihara S, Arita Y, Okamoto Y, Maeda K, Kuriyama H, Hotta K, Nishida M, Takahashi M, Muraguchi M, Ohmoto Y, Nakamura T, Yamashita S, Funahashi T & Matsuzawa Y (2000) Adiponectin, an adipocyte-derived plasma protein, inhibits endothelial NF-kappaB signaling through a cAMP-dependent pathway. *Circulation* **102:** 1296–301

Overgaard M, Cangemi C, Jensen ML, Argraves WS & Rasmussen LM (2015) Total and isoform-specific quantitative assessment of circulating fibulin-1 using selected reaction monitoring MS and time-resolved immunofluorometry. *Proteomics - Clin. Appl.* **9:** 767–775

Paapstel K, Zilmer M, Eha J, Tootsi K, Piir A & Kals J (2016) Association Between Fibulin-1 and Aortic Augmentation Index in Male Patients with Peripheral Arterial Disease. *Eur. J. Vasc. Endovasc. Surg.* **51:** 76–82

Packard RRS & Libby P (2007) Inflammation in Atherosclerosis: From Vascular Biology to Biomarker Discovery and Risk Prediction. *Clin. Chem.* **54:** 24–38

Pae M & Romeo GR (2014) The multifaceted role of profilin-1 in adipose tissue inflammation and glucose homeostasis. *Adipocyte* **3:** 69–74

Panchaud A, Scherl A, Shaffer SA, von Haller PD, Kulasekara HD, Miller SI & Goodlett DR (2009) Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. *Anal. Chem.* **81:** 6481–6488

Pappin DJC, Hojrup P & Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3:** 327–332

Parker CE, Pearson TW, Anderson NL & Borchers CH (2010) Mass-spectrometry-based clinical proteomics--a review and prospective. *Analyst* **135:** 1830–8

Pasa-Tolic L, Masselon C, Barry RC, Shen Y & Smith RD (2004) Proteomic analyses using an accurate mass and time tag strategy. *Biotechniques* **37:** 621–4, 626–33, 636 passim

Penckofer S, Kouba J, Wallis DE & Emanuele MA (2008) Vitamin D and diabetes: Let the sunshine in. *Diabetes Educ.* **34:** 939–954

Pereira RI, Snell-Bergeon JK, Erickson C, Schauer IE, Bergman BC, Rewers M & Maahs DM (2012) Adiponectin dysregulation and insulin resistance in type 1 diabetes. *J. Clin. Endocrinol. Metab.* **97:** E642-7

Perkins DN, Pappin DJ, Creasy DM & Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20:** 3551–67

Peterson AC, Russell JD, Bailey DJ, Westphall MS & Coon JJ (2012) Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics* **11:** 1475–1488

Pfaff D, Schoenenberger AW, Dasen B, Erne P, Resink TJ & Philippova M (2015) Plasma T-cadherin negatively associates with coronary lesion severity and acute coronary syndrome. *Eur. Hear. J. Acute Cardiovasc. Care* **4:** 410–418

Picotti P & Aebersold R (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9:** 555–566

Polychronakos C & Li Q (2011) Understanding type 1 diabetes through genetics: Advances and prospects. *Nat. Rev. Genet.* **12:** 781–792

Poon TCW (2007) Opportunities and limitations of SELDI-TOF-MS in biomedical research: practical advices. *Expert Rev. Proteomics* **4:** 51–65

Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, Young N, Xia J, Knox C, Dong E, Huang P, Hollander Z, Pedersen TL, Smith SR, Bamforth F, Greiner R, et al (2011) The human serum metabolome. *PLoS One* **6:** e16957

Purohit S, Podolsky R, Schatz D, Muir A, Hopkins D, Huang YH & She JX (2006) Assessing the utility of SELDI-TOF and model averaging for serum proteomic biomarker discovery. *Proteomics* **6:** 6405–6415

Purohit S & She J-X (2008) Biomarkers for type 1 diabetes. *Int. J. Clin. Exp. Med.* **1:** 98–116

R Development Core Team R (2011) R: A Language and Environment for Statistical Computing Available at: http://www.r-project.org

Rabilloud T & Lelong C (2011) Two-dimensional gel electrophoresis in proteomics: A tutorial. *J. Proteomics* **74:** 1829–1841

Raiko JR, Viikari JS, Ilmanen A, Hutri-Kahonen N, Taittonen L, Jokinen E, Pietikainen M, Jula A, Loo BM, Marniemi J, Lehtimaki T, Kahonen M, Ronnemaa T, Raitakari OT & Juonala M (2010) Follow-ups of the Cardiovascular Risk in Young Finns Study in 2001 and 2007: levels and 6-year changes in risk factors. *J. Intern. Med.* **267:** 370–384

Ramsey SA, Gold ES & Aderem A (2010) A systems biology approach to understanding atherosclerosis. *EMBO Mol. Med.* **2:** 79–89

Rauniyar N & Yates JR (2014) Isobaric Labeling-Based Relative Quanti fi cation in Shotgun Proteomics. *J. Proteome Res.* **13:** 5293–5309

Reiter L, Rinner O, Picotti P, Huttenhain R, Beck M, Brusniak MY, Hengartner MO & Aebersold R (2011) mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* **8:** 430–435

Rewers M, Bugawan TL, Norris JM, Blair A, Beaty B, Hoffman M, McDuffie RS, Hamman RF, Klingensmith G, Eisenbarth GS & Erlich HA (1996) Newborn screening for HLA markers associated with IDDM: Diabetes autoimmunity study in the young (DAISY). *Diabetologia* **39:** 807–812

Rewers M & Ludvigsson J (2016) Environmental risk factors for type 1 diabetes. *Lancet* **387:** 2340–2348

Ridker PM (2004) Established and Emerging Plasma Biomarkers in the Prediction of First Atherothrombotic Events. *Circulation* **109:** IV-6-IV-19

Rifai N, Gillette M a & Carr S a (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **24:** 971–983

Rissanen J, Moulder R, Lahesmaa R & Nevalainen OS (2012) Pre-processing of Orbitrap higher energy collisional dissociation tandem mass spectra to reduce erroneous iTRAQ ratios. *Rapid Commun. Mass Spectrom.* **26:** 2099–2104

Rocchiccioli S, Pelosi G, Rosini S, Marconi M, Viglione F, Citti L, Ferrari M, Trivella MG & Cecchettini A (2013) Secreted proteins from carotid endarterectomy: an untargeted approach to disclose molecular clues of plaque progression. *J. Transl. Med.* **11:** 260

Roll U, Christie MR, Fuchtenbusch M, Payton MA, Hawkes CJ & Ziegler AG (1996) Perinatal autoimmunity in offspring of diabetic parents: The German multicenter BABY-DIAB study: Detection of humoral immune responses to islet antigens in early childhood. *Diabetes* **45:** 967–973

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A & Pappin DJ (2004a) Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **3:** 1154–1169

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A & Pappin DJ (2004b) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol.Cell.Proteomics* **3:** 1154–1169

Ross R (1999) Inflammation or Atherogenesis. *N. Engl. J. Med.* **340:** 115–126

Rowell J, Thompson AJ, Guyton JR, Lao XQ, McHutchison JG, McCarthy JJ & Patel K (2011) Serum apolipoprotein C-III is independently associated with chronic hepatitis C infection and advanced fibrosis. *Hepatol. Int.*

Rundek T, Gardener H, Della-Morte D, Dong C, Cabral D, Tiozzo E, Roberts E, Crisby M, Cheung K, Demmer R, Elkind MSV, Sacco RL & Desvarieux M (2015) The relationship between carotid intima-media thickness and carotid plaque in the Northern Manhattan Study. *Atherosclerosis* **241:** 364–370

Sajic T, Liu Y & Aebersold R (2015) Using data-independent, high-resolution mass spectrometry in protein biomarker research: Perspectives and clinical applications. *Proteomics - Clin. Appl.* **9:** 307–321

Schalkwijk CG & Miyata T (2012) Early- and advanced non-enzymatic glycation in diabetic vascular complications: The search for therapeutics. *Amino Acids* **42:** 1193–1204

Schenk S, Schoenhals GJ, de Souza G & Mann M (2008) A high confidence, manually validated human blood plasma protein reference set. *BMC Med. Genomics* **1:** 41

Schwartz JC, Senko MW & Syka JEP (2002) A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **13:** 659–669

Shah PK (2007) Molecular mechanisms of plaque instability. *Curr. Opin. Lipidol.* **18:** 492–499

Shevchenko A, Chernushevic I, Shevchenko A, Wilm M & Mann M (2002) 'De novo' sequencing of peptides recovered from in-gel digested proteins by nanoelectrospray tandem mass spectrometry. *Mol. Biotechnol.* **20:** 107–18

Shilov I V., Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, Hunter CL, Nuwaysir LM & Schaeffer DA (2007a) The Paragon Algorithm, a Next Generation Search Engine That Uses Sequence Temperature Values and Feature Probabilities to Identify Peptides from Tandem Mass Spectra. *Mol. Cell. Proteomics* **6:** 1638–1655

Shilov I V, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, Hunter CL, Nuwaysir LM & Schaeffer DA (2007b) The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics* **6:** 1638–1655

Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ, McKenna T, Nold MJ, Richardson K, Young P & Geromanos S (2005) Quantitative proteomic analysis by accurate mass retention time pairs. *Anal. Chem.* **77:** 2187–2200

Simpson M, Brady H, Yin X, Seifert J, Barriga K, Hoffman M, Bugawan T, Barón AE, Sokol RJ, Eisenbarth G, Erlich H, Rewers M & Norris JM (2011) No association of vitamin D intake or 25-hydroxyvitamin D levels in childhood with risk of islet autoimmunity and type 1 diabetes: The Diabetes Autoimmunity Study in the Young (DAISY). *Diabetologia* **54:** 2779–2788

Singal DP & Blajchman MA (1973) Histocompatibility (HL-A) Antigens, Lymphocytotoxic Antibodies and Tissue Antibodies in Patients with Diabetes Mellitus. *Diabetes* **22:** 429–432

Singh IP, Chopra AK, Coppenhaver DH, Ananatharamaiah GM & Baron S (1999) Lipoproteins account for part of the broad non-specific antiviral activity of human serum. In *Antiviral Research*

Siuzdak G (2004) An introduction to mass spectrometry ionization: An excerpt from The Expanding Role of Mass Spectrometry in Biotechnology, 2nd ed.; MCC Press: San Diego, 2005. *J. Assoc. Lab. Autom.* **9:** 50–63

Smith CP, Clements CB, Riding MH, Collins P, Bottazzo GF & Taylor KW (1998) Simultaneous onset of type 1 diabetes mellitus in identical infant twins with enterovirus infection. *Diabet. Med.* **15:** 515–517

Smith LM & Kelleher NL (2013) Proteoform: A single term describing protein complexity. *Nat. Methods* **10:** 186–187

Smith LM, Kelleher NL & for Top Down Proteomics C (2013) Proteoform: a single term describing protein complexity. *Nat. Methods* **10:** 186–187

Smith RD, Loo JA, Loo RRO, Busman M & Udseth HR (1991) Principles and practice of electrospray ionization—mass spectrometry for large polypeptides and proteins. *Mass Spectrom. Rev.* **10:** 359–452

Sodetz JM & Plumb ME (2001) Complement: Terminal Pathway. In John Wiley & Sons, Ltd

Sparre T, Larsen MR, Heding PE, Karlsen AE, Jensen ON & Pociot F (2005) Unraveling the Pathogenesis of Type 1 Diabetes with Proteomics: Present And Future Directions. *Mol Cell Proteomics* **4:** 441–457

Sparrow RL, Greening DW & Simpson RJ (2011) Serum/Plasma Proteomics. **728:** 259–265

Steen H & Mann M (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5:** 699–711

Strassberger V, Fugmann T, Neri D & Roesli C (2010) Chemical proteomic and bioinformatic strategies for the identification and quantification of vascular antigens in cancer. *J. Proteomics* **73:** 1954–1973

Suh MJ, Tovchigrechko A, Thovarai V, Rolfe MA, Torralba MG, Wang J, Adkins JN, Webb-Robertson BJ, Osborne W, Cogen FR, Kaplowitz PB, Metz TO, Nelson KE, Madupu R & Pieper R (2015a) Quantitative Differences in the Urinary Proteome of Siblings Discordant for Type 1 Diabetes Include Lysosomal Enzymes. *J. Proteome Res.* **14:** 3123–3135

Suh MJ, Tovchigrechko A, Thovarai V, Rolfe MA, Torralba MG, Wang J, Adkins JN, Webb-Robertson BJM, Osborne W, Cogen FR, Kaplowitz PB, Metz TO, Nelson KE, Madupu R & Pieper R (2015b) Quantitative Differences in the Urinary Proteome of Siblings Discordant for Type 1 Diabetes Include Lysosomal Enzymes. *J. Proteome Res.* **14:** 3123–3135

Sundsten T, Zethelius B, Berne C & Bergsten P (2008) Plasma proteome changes in subjects with Type 2 diabetes mellitus with a low or high early insulin response. *Clin. Sci. (Lond).* **114:** 499–507

Suomi T, Seyednasrollah F, Jaakkola MK, Faux T & Elo LL (2017) ROTS: An R package for reproducibility-optimized statistical testing. *PLoS Comput. Biol.* **13:**

Switzar L, Giera M & Niessen WMA (2013) Protein digestion: An overview of the available techniques and recent developments. *J. Proteome Res.*

Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J & Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **101:** 9528–33

Tabas I, García-Cardeña G & Owens GK (2015) Recent insights into the cellular biology of atherosclerosis. *J. Cell Biol.* **209:** 13–22

Tambor V, Hunter CL, Seymour SL, Kacerovsky M, Stulik J & Lenco J (2012) CysTRAQ — A combination of iTRAQ and enrichment of cysteinyl peptides for uncovering and quantifying hidden proteomes. *J. Proteomics* **75:** 857–867

Tang WH, Shilov I V & Seymour SL (2008) Nonlinear fitting method for determining local false discovery rates from decoy database searches. *J. Proteome Res.* **7:** 3661–3667

Taylor JA & Johnson RS (1997) Sequence database searches via de Novo peptide sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **11:** 1067–1075

The International Expert Committee (2009) International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes. *Diabetes Care* **32:** 1327–1334

Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T & Hamon C (2003a) Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **75:** 1895–1904

Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T & Hamon C (2003b) Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75:** 1895–1904

Tibshirani R (1996) Regression Selection and Shrinkage via the Lasso. *J. R. Stat. Soc. B* **58:** 267–288

Toby TK, Fornelli L & Kelleher NL (2016) Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **9:** 499–519

von Toerne C, Laimighofer M, Achenbach P, Beyerlein A, de las Heras Gala T, Krumsiek J, Theis FJ, Ziegler AG & Hauck SM (2017) Peptide serum markers in islet autoantibody-positive children. *Diabetologia* **60:** 287–295

von Toerne C, Laimighofer M, Achenbach P, Beyerlein A, de Las Heras Gala T, Krumsiek J, Theis FJ, Ziegler AG & Hauck SM (2016) Peptide serum markers in islet autoantibody-positive children. *Diabetologia*

Tonstad S, Joakimsen O, Stensland-Bugge E, Leren TP, Ose L, Russell D & Bønaa KH (1996) Risk factors related to carotid intima-media thickness and plaque in children with familial hypercholesterolemia and control subjects. *Arterioscler. Thromb. Vasc. Biol.* **16:** 984–91

Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJ & Liebler DC (2010) Depletion of abundant plasma proteins and limitations of plasma proteomics. *J. Proteome Res.* **9:** 4982–4991

Tuomilehto J, Nejentsev S, Koskinen S, Sjöroos M, Reijonen H, Schwartz EI, Kovalchuk L, Sochnev A, Adojaan B, Podar T, Knip M, Simell O, Koskenvuo M, Åkerblom HK & Ilonen J (1998) Distribution of insulin-dependent diabetes mellitus (IDDM) -related HLA alleles correlates with the difference in IDDM incidence in four populations of the Eastern Baltic region. *Tissue Antigens* **52:** 473–477

Turner MW (2003) The role of mannose-binding lectin in health and disease. *Mol. Immunol.* **40:** 423–429

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M & Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13:** 731–40

Unlu M, Morgan ME & Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18:** 2071–2077

Varela-Calvino R, Ellis R, Sgarbi G, Dayan CM & Peakman M (2002) Characterization of the T-cell response to Coxsackievirus B4: Evidence that effector memory cells predominate in patients with type 1 diabetes. *Diabetes* **51:** 1745–1753

Vasan RS (2006) Biomarkers of cardiovascular disease: Molecular basis and practical considerations. *Circulation* **113:** 2335–2362

Venable JD, Dong M-Q, Wohlschlegel J, Dillin A & Yates JR (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1:** 39–45

Venter JC (2001) The Sequence of the Human Genome. *Science (80-.).* **291:** 1304–1351

Virtanen SM, Laara E, Hypponen E, Reijonen H, Rasanen L, Aro A, Knip M, Ilonen J & Akerblom HK (2000) Cow's milk consumption, HLA-DQB1 genotype, and type 1 diabetes: a nested case-control study of siblings of children with diabetes. Childhood diabetes in Finland study group. *Diabetes* **49:** 912–917

Virtanen SM, Nevalainen J, Kronberg-Kippilä C, Ahonen S, Tapanainen H, Uusitalo L, Takkinen HM, Niinistö S, Ovaskainen ML, Kenward MG, Veijola R, Ilonen J, Simell O & Knip M (2012) Food consumption and advanced β cell autoimmunity in young children with HLA-conferred susceptibility to type 1 diabetes: A nested case-control design. *Am. J. Clin. Nutr.* **95:** 471–478

Vizcaíno JA, Csordas A, Del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu QW, Wang R & Hermjakob H (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44:** D447–D456

Wahlberg J, Vaarala O & Ludvigsson J (2006) Dietary risk factors for the emergence of type 1 diabetes-related autoantibodies in 21/2 year-old Swedish children. *Br. J. Nutr.* **95:** 603–608

Wall DB, Berger SJ, Finch JW, Cohen SA, Richardson K, Chapman R, Drabble D, Brown J & Gostick D (2002) Continuous sample deposition from reversed-phase liquid chromatography to tracks on a matrix-assisted laser desorption/ionization precoated target for the analysis of protein digests. *Electrophoresis* **23:** 3193–3204

Walter M, Kaupper T, Adler K, Foersch J, Bonifacio E & Ziegler A-G (2010) No Effect of the 1alph,25-dihydroxyvitamin D3 on beta-cell residual function and insulin requirement in adults with new-onset type 1 diabetes. *Diabetes Care* **33:** 1443–1448

Wang TJ, Gona P, Larson MG, Tofler GH, Levy D, Newton-Cheh C, Jacques PF, Rifai N, Selhub J, Robins SJ, Benjamin EJ, D'Agostino RB & Vasan RS (2006) Multiple biomarkers for the prediction of first major cardiovascular events and death. *N. Engl. J. Med.* **355:** 2631–2639

Washburn MP, Wolters D & Yates 3rd JR (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19:** 242–247

Weets I, Kaufman L, Van Der Auwera B, Crenier L, Rooman RPA, De Block C, Casteels K, Weber E, Coeckelberghs M, Laron Z, Pipeleers DG & Gorus FK (2004) Seasonality in clinical onset of Type 1 diabetes in Belgian patients above the age of 10 is restricted to HLA-DQ2/DQ8-negative males, which explains the male to female excess in incidence. *Diabetologia* **47:** 614–621

Wiese S, Reidegeld KA, Meyer HE & Warscheid B (2007) Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics* **7:** 340–350

Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese J-H, Bantscheff M, Gerstmair A, et al (2014) Mass-spectrometry-based draft of the human proteome. *Nature* **509:** 582–7

Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez J-C, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL & Hochstrasser DF (1996) From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis. *Nat. Biotechnol.* **14:** 61–65

Williams KJ & Tabas I (1995) The response-to-retention hypothesis of early atherogenesis. *Arterioscler. Thromb. Vasc. Biol.* **15:** 551–61

Williams KJ & Tabas I (1998) The response-to-retention hypothesis of atherogenesis reinforced. *Curr. Opin. Lipidol.* **9:** 471–474

Wolters DA, Washburn MP & Yates JR (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73:** 5683–5690

Xu L, Tan AC, Naiman DQ, Geman D & Winslow RL (2005) Robust prostate cancer marker genes emerge from direct integration of inter-study microarray data. *Bioinformatics* **21:** 3905–3911

Yates JR, Ruse CI & Nakorchevsky A (2009a) Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **11:** 49–79

Yates JR, Ruse CI & Nakorchevsky A (2009b) Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu. Rev. Biomed. Eng.* **11:** 49–79

Ylipaasto P, Klingel K, Lindberg AM, Otonkoski T, Kandolf R, Hovi T & Roivainen M (2004) Enterovirus infection in human pancreatic islet cells, islet tropism in vivo and receptor involvement in cultured islet beta cells. *Diabetologia* **47:** 225–239

Yoon J-W & Jun H-S (2005) Autoimmune destruction of pancreatic beta cells. *Am. J. Ther.* **12:** 580–591

Zhang Q, Fillmore TL, Schepmoes AA, Clauss TR, Gritsenko MA, Mueller PW, Rewers M, Atkinson MA, Smith RD & Metz TO (2013a) Serum proteomics reveals systemic dysregulation of innate immunity in type 1 diabetes. *J. Exp. Med.* **210:** 191–203

Zhang Q, Fillmore TL, Schepmoes AA, Clauss TRW, Gritsenko MA, Mueller PW, Rewers M, Atkinson MA, Smith RD & Metz TO (2013b) Serum proteomics reveals systemic dysregulation of innate immunity in type 1 diabetes. *J. Exp. Med.* **210:** 191–203

Zhang Y, Fonslow BR, Shan B, Baek MC & Yates JR (2013c) Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* **113:** 2343–2394

Zhen Y, Xu N, Richardson B, Becklin R, Savage JR, Blake K & Peltier JM (2004) Development of an LC-MALDI method for the analysis of protein complexes. *J. Am. Soc. Mass Spectrom.* **15:** 803–822

Zhi W, Sharma A, Purohit S, Miller E, Bode B, Anderson SW, Reed JC, Steed RD, Steed L, Hopkins D & She JX (2011a) Discovery and validation of serum protein changes in type 1 diabetes patients using high throughput two dimensional liquid chromatography-mass spectrometry and immunoassays. *Mol. Cell. Proteomics* **10:** M111.012203

Zhi W, Sharma A, Purohit S, Miller E, Bode B, Anderson SW, Reed JC, Steed RD, Steed L, Hopkins D & She JX (2011b) Discovery and validation of serum protein changes in type 1 diabetes patients using high throughput two dimensional liquid chromatography-mass spectrometry and immunoassays. *Mol. Cell. Proteomics* **10:** M111.012203

Zhu W, Smith JW & Huang CM (2010) Mass spectrometry-based label-free quantitative proteomics. *J. Biomed. Biotechnol.* **2010:**

Zimmer JSD, Monroe ME, Qian WJ & Smith RD (2006) Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom. Rev.* **25:** 450–482

Zipitis CS & Akobeng AK (2008) Vitamin D supplementation in early childhood and risk of type 1 diabetes: a systematic review and meta-analysis. *Arch. Dis. Child.* **93:** 512–517

Zubarev RA, Kelleher NL & McLafferty FW (1998) Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **120:** 3265–3266

Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L & Washburn MP (2006) Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J Proteome Res* **5:** 2339–2347

Annales Universitatis Turkuensis

Turun yliopisto
University of Turku

Painosalama Oy, Turku, Finland 2018

ISBN 978-951-29-7386-6 (PRINT)
ISBN 978-951-29-7387-3 (PDF)
ISSN 0355-9483 (PRINT) | ISSN 2343-3213 (PDF)