Technische Universität München

Fakultät für Medizin

Proteomic phenotyping of patient-derived peripheral blood mononuclear cells and serum in recent-onset type 1 diabetes

Marlen Franziska Lepper

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzender: Prof. Dr. Dr. Andreas Pichlmair

Prüfende/-r der Dissertation:

1. Prof. Dr. Anette-Gabriele Ziegler

2. Prof. Dr. Dr. Fabian Theis

Die Dissertation wurde am 29.11.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 04.07.2018 angenommen.

HelmholtzZentrum münchen
German Research Center for Environmental Health

TUM
TECHNISCHE
UNIVERSITÄT
MÜNCHEN

Technische Universität München, Fakultät für Medizin

in Kooperation mit dem Helmholtz Zentrum München,

Abteilungen für Proteinanalytik und Diabetesforschung

Proteomic phenotyping of patient-derived peripheral blood mononuclear cells and serum in recent-onset type 1 diabetes

Marlen Franziska Lepper

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften

Betreuer: Prof. Dr. Anette-Gabriele Ziegler und Dr. Stefanie Hauck

April 2014 – Dezember 2017

Table of Contents

Table of Contents

Summary

Type 1 diabetes (T1D) is an autoimmune disease which arises from the selective destruction of insulin-producing pancreatic ß-cells leading to hyperglycemia and insulin deficiency. The underlying pathogenesis is attributed to a complex interplay of genetic predisposition, environmental factors and aberrant T cell responses. However, the precise etiological agent triggering islet autoimmunity and T1D development remains incompletely understood. In the present thesis, liquid chromatography tandem mass spectrometry (LC-MS/MS) based discovery proteomics was explored to portray the dynamic proteomic changes underlying T1D onset in peripheral CD4+ T cells, CD4-depleted cells and serum. The objective was to identify disease-associated proteomic signatures which lead to a better understanding of T1D pathogenesis, and which ideally exhibit biomarker potential. The first stage of this proteomic thesis is committed to the evaluation of sample preparation strategies applicable to biobanked clinical peripheral blood mononuclear cell (PBMC) samples limited in quantity and viability. As a result, complex surfaceome profiling strategies were ruled out in favor of simple and straightforward whole-proteome digest approaches which achieved high sensitivity when applied to as few as 250,000 cells CD4+ T cells. Specifically, the in-StageTip (iST) method was chosen for experiments conducted on the OrbitrapXL platform due to high reproducibility and compatibility with in-situ peptide fractionation to enhance proteome coverage. In the next stage, proteomic profiles of patient- and control-derived CD4+ T cells and the complementary CD4-depleted cell fraction were generated using a combination of data-dependent and data-independent acquisition (DDA and DIA) LC-MS/MS. In detail, two sets of biobanked PBMC samples collected from 53 pediatric patients with newly diagnosed T1D and 61 age- and sex-matched control children were profiled successively. The introduction of the DIA workflow on the next-generation MS platform QExactive HF greatly enhanced analytical depth compared to conventional shotgun proteomics resulting in >8000 identified proteins in CD4+ T cells and CD4-depleted cells. Using both DDA and DIA LC-MS/MS strategies, a consistent inflammatory signature was identified in patients with T1D across both investigated cellular compartments. Strikingly, this phenotype was primarily attributable to (circulating) inflammatory mediators of the myeloid cell lineage, especially neutrophils, and the complement system, thus highlighting the dominant role that innate immunity plays during the onset of T1D. Key signature proteins were myeloperoxidase (MPO), proteinase 3 (PRTN3) and ribonuclease 3 (RNASE3). Specifically, these findings support the hypothesis that altered neutrophil or myeloid cell activity is implicated in the pathogenic processes accompanying T1D onset. Additionally, proteomic profiles

Summary

revealed mildly reduced protein levels of T cell-specific proteins in both cell fractions at T1D onset indicating an immune-dampened phenotype. Reduced expression levels were observed for key T cell surface markers such as CD3 (δ/ε), CD247 and CD2 as well as T cell receptor signaling-associated tyrosine-protein kinases LCK and ZAP70, and transcription factor NFATC2. Accordingly, effector molecules of cytotoxic lymphocytes including granzymes A and H (GZMA/H) and granulysin (GNLY) were broadly downregulated at T1D onset. These findings implicate a numeric and/or functional impairment of T cells in general and cytotoxic lymphocytes such as natural killer (NK) cells and cytotoxic T lymphocytes (CTLs) in particular. The observed signature in the periphery could likely stem from tissue infiltration or a condition of immune exhaustion characterized by attenuation of effector activity and transitioning into T cell memory phenotype. In this context, a particularly interesting biomarker candidate is GZMA which was recently attributed a role in maintaining immune tolerance in T1D. In the last stage of the thesis, serum samples of 50 patients with newly diagnosed T1D and 50 healthy controls were profiled using DIA LC-MS/MS. The obtained proteomic profiles revealed higher expression of several complement factors, protease inhibitors, apolipoproteins and hemostasis-associated proteins in T1D sera. However, serum profiling did not show evidence for higher abundance levels of myeloid-derived circulatory inflammatory proteins in T1D sera. Importantly, serum profiling confirmed biomarker candidates proposed in current literature and introduced novel candidates such as hepatocyte growth factor-like protein (MST1), a critical regulator of ß-cell dysfunction and death. In summary, the generated proteomic landscape of PBMCs and serum provides evidence for distinct alterations in innate and adaptive immune function in recent-onset T1D. Thus, the present work represents a valuable repository of proteomic signatures potentially implicated in T1D pathogenesis providing a sound basis for future investigation of cellular pathology or biomarker discovery.

Zusammenfassung

Bei der Autoimmunerkrankung Typ-1-Diabetes (T1D) führt die selektive Zerstörung der insulin-produzierenden β-Zellen in der Bauchspeicheldrüse zur Entstehung von Hyperglykämie und Insulinmangel. Die Pathogenese der Erkrankung wird auf das komplexe Zusammenspiel von genetischer Prädisposition, Umweltfaktoren und abnormaler adaptiver Immunität zurückgeführt. Die exakten ätiologischen Faktoren, die zum Ausbruch von Inselautoimmunität und zur Entstehung von T1D führen, werden derzeit jedoch nur unvollständig verstanden. In der vorliegenden Arbeit wurde die Flüssigchromatographie-gekoppelte proteinbasierte Massenspektrometrie (LC-MS/MS) als Werkzeug zur Ergründung der dynamischen proteomischen Veränderungen, die dem Beginn der T1D Erkrankung zugrunde liegen, angewendet. Im Speziellen wurden Proteinprofile von peripheren CD4 positiven T-Zellen und CD4 negativen Zellen sowie Blutserum erstellt. Hierbei lag das Ziel darin, krankheits-assoziierte Proteinsignaturen aufzudecken, die zu einem besseren Verständnis der Pathogenese von T1D führen und im Idealfall Biomarkerpotential besitzen. Der erste Teil dieser Arbeit beschäftigt sich mit der Evaluation verschiedener proteomischer Probenvorbereitungsmethoden, die auf klinische Blutzellproben aus der Biobank angewendet werden sollen. Die Schwierigkeit lag darin eine Methode zu identifizieren, die trotz limitierter Ausgangszellmenge mit eingeschränkter Viabilität hochsensitive Ergebnisse erzielen konnte. Als Ergebnis dieser Methodenfindung wurde die komplexe Aufarbeitung und Abbildung des Zelloberflächenproteoms zu Gunsten von einfacheren Methoden zur Abbildung des Gesamtproteoms verworfen. Letztere erzielten hohe Sensitivität in Bezug auf die Proteinidentifikationsrate, obwohl lediglich 250,000 CD4 positive T-Zellen verarbeitet wurden. Im Detail wurde die in-Stage Tip (iST) Methode für Experimente auf der Orbitrap XL Plattform ausgewählt, da sie sich durch hohe Reproduzierbarkeit und die direkte Integration der Peptidfraktionierung zur verbesserten Abdeckung des Proteoms ausgezeichnete. Im nächsten Schritt wurden Proteinprofile von CD4 positiven T-Zellen und den komplementären CD4 negativen Zellen, die aus Biobankproben von Patienten mit neu diagnostiziertem T1D und gesunden Kontrollen gewonnen wurden, erstellt. Hierzu wurde eine Kombination aus daten-abhängiger und daten-unabhängiger Akquisition in der Massenspektrometrie verwendet (DDA und DIA). Im Detail wurden nacheinander zwei Probensätze an peripheren mononuklearen Blutzellproben (PBMC) aus der Biobank aufgearbeitet, die insgesamt 53 pädiatrische Patienten mit T1D sowie 61 alters- und geschlechts-abgestimmte Kontrollkinder umfassten. Hierbei zeigte sich, dass die Einführung der DIA Methode auf der Q Exactive HF Plattform, einem Massenspektrometer der nächsten Generation, im Vergleich zur konventionellen

Zusammenfassung

shotgun Methode zu außerordentlichen Verbesserungen in der analytische Tiefe führte. Im Detail wurden mit der DIA Strategie sowohl in CD4 positiven T-Zellen als auch in CD4 negativen Zellen mehr als 8000 unterschiedliche Proteine identifiziert. Zusammenfassend wurde basierend auf der Anwendung der proteinbasierten Massenspektrometrie in der DDA und DIA Methode eine konsistente, inflammatorische Proteinsignatur in Patienten mit T1D entdeckt, die sich über beide untersuchte Zellkompartimente erstreckte. Dieser Phänotyp war vorwiegend auf die Präsenz von (zirkulierenden) inflammatorischen Mediatoren der myeloiden Zelllinie, insbesondere neutrophile Granulozyten, zurückzuführen. Darüber hinaus zeigte sich eine verstärkte Aktivierung des Komplementsystems, was zusätzlich hervorhebt, dass während der Diagnosestellung von T1D eine Veränderung in der Antwort des angeborenen Immunsystems vorliegt. Schlüsselproteine dieser Signatur waren Myeloperoxidase (MPO), Proteinase 3 (PRTN3) und Ribonuklease 3 (RNASE3). Im Speziellen stützen diese Ergebnisse die Hypothese, dass eine abweichende Aktivität oder Funktion von Neutrophilen und/oder myeloiden Blutzellen in die pathogenen Prozesse, die der Entstehung von T1D zugrunde liegen, involviert ist. Des Weiteren belegten die zellulären Proteinprofile, dass T-Zell-spezifische Proteine in beiden Zellfraktionen während der Diagnosestellung von T1D geringfügig niedriger abundant waren, was auf einen gedämpften Immunphänotyp hinweist. Reduzierte Expressionsniveaus wurden beispielsweise für die wichtigen T-Zelloberflächenmarker CD3 (δ/ε), CD247 und CD2, die T-Zellrezeptor-assoziierten Tyrosinkinasen LCK und ZAP70, sowie den Transkriptionsfaktor NFATC2 festgestellt. Im Einklang mit diesem Befund zeigte sich, dass auch die Effektormoleküle von zytotoxischen Lymphozyten einschließlich der Granzyme A und H (GZMA/H) sowie Granulysin (GNLY) bei Ausbruch der Erkrankung deutlich verringerte Proteinmengen aufwiesen. Dieses Ergebnis impliziert eine numerische und/oder funktionelle Störung von T Zellen im Allgemeinen und zytotoxischen Lymphozyten wie natürlichen Killerzellen (NK) und zytotoxischen T Lymphozyten (CTL) im Speziellen. Diese beobachtete Proteinsignatur ist möglicherweise die Folge der Abwanderung dieser Zellen von der Peripherie ins Gewebe oder ein Zustand der Immunerschöpfung, der sich durch die Abschwächung von Effektoraktivität und den Übergang zum T-Gedächtniszellenphänotyp auszeichnet. In diesem Kontext ist insbesondere GZMA ein interessanter Biomarkerkandidat, da diesem erst kürzlich eine wichtige Rolle in der Aufrechterhaltung der Immuntoleranz im T1D-Mausmodell zugeschrieben wurde. Im letzten Abschnitt dieser Arbeit wurden Proteinprofile von Blutsera aus 50 Patienten mit neu diagnostiziertem T1D und 50 gesunden Kontrollen mittels DIA LC-MS/MS erstellt. Die proteomische Charakterisierung von Blutserum zeigte eine Erhöhung der Expressionswerte von verschiedenen Komplementfaktoren,

Proteaseinhibitoren, Apolipoproteinen und Hämostaseproteinen bei neu diagnostiziertem T1D. Die Serumuntersuchung lieferte allerdings keine Hinweise auf eine erhöhte Abundanz von myeloid-stämmigen, zirkulierenden Entzündungsproteinen in Blutsera von Patienten mit T1D. Die Erstellung von proteomischen Blutseraprofilen bestätigte in bedeutender Weise einige in der gegenwärtigen Literatur bekannten T1D-assoziierten Biomarkerkandidaten, und stellte darüber hinaus einige neue, potentielle Markerproteine vor. Eine übergeordnete Rolle spielt hierbei die Identifizierung des Hepatozyten Wachstumsfaktor ähnlichen Proteins (MST1), das als kritischer Regulator der ß-Zelldysfunktion und ß-Zell-Apoptose bekannt wurde. Zusammenfassend liefert der proteomische Fingerabdruck von peripheren Blutzellen und Blutserum in neu diagnostiziertem T1D deutliche Hinweise auf eine veränderte Funktion von angeborener und adaptiver Immunität. Die vorliegende Arbeit stellt ein wertvolles Repositorium an Proteomsignaturen dar, das als Basis für zukünftige Untersuchungen zur zellulären Pathologie oder zur Biomarkerentwicklung im Bereich der T1D Forschung dienen könnte.

1 Introduction

1.1 Type 1 Diabetes

Autoimmune disorders arise from the aberrant recognition of self-epitopes by the immune system, leading to inflammation and tissue damage. Autoimmune type 1 diabetes (T1D) is a disorder of chronic hyperglycemia characterized by the selective destruction of insulin-producing pancreatic β-cells, culminating in lifelong insulin deficiency. The pathogenesis of T1D has been attributed to a complex interplay between genetic factors, environmental agents and impaired adaptive immunity. It is commonly accepted that the process underlying ß-cell destruction is T cell-mediated, evidenced by multiple studies in both humans and the non-obese diabetic (NOD) mouse model (Bluestone et al., 2010). However, despite considerable research, the precise event that triggers disease initiation and progression is currently unknown.

The first hallmark of T1D is the onset of islet autoimmunity, which is indicated by the presence of multiple islet-specific autoantibodies in the circulation; this checkpoint marks the beginning of the preclinical phase of T1D (Ziegler et al., 2013). The sustained loss of pancreatic β-cells usually progresses subclinically, eventually manifesting as overt hyperglycemia and metabolic imbalance when approximately two-thirds of the islets become functionally impaired, i.e. devoid of insulin-producing cells (Foulis and Stewart, 1984; Willcox et al., 2009). A recently proposed disease staging classification by Insel et al. (2015) distinguishes three distinct stages of T1D development (Fig. 1). Accordingly, presymptomatic T1D with normoglycemia is initiated in stage 1 with the seroconversion to multiple (≥2) circulating islet autoantibodies marking the onset of ß-cell autoimmunity. In this stage, the 5-year, 10-year and lifetime risk of developing symptomatic disease for genetically at-risk children account for approximately 44%, 70% and 100%, respectively (Ziegler et al., 2013). The presymptomatic phase progresses in stage 2 with ongoing ß-cell autoimmunity; the loss of functional ß-cell mass has now resulted in the development of dysglycemia. Progressive ß-cell destruction eventually leads to the manifestation of clinical disease in stage 3 characterized by the onset of typical symptoms of diabetes such as polyuria, polydipsia, polyphagia, weight loss, fatigue and diabetic ketoacidosis. Remarkably, the rate of progression from onset of ß-cell autoimmunity to development of overt metabolic disease is highly variable in children and adults, ranging from months to decades (Atkinson and Eisenbarth, 2001; Atkinson et al., 2014; Ziegler et al., 2013). Additionally, preservation of residual ß-cell function after diagnosis, indicated by C-

Introduction

peptide production, is heterogeneous in patients and can in some cases still be detected after several decades of long-standing T1D (Keenan et al., 2010a).
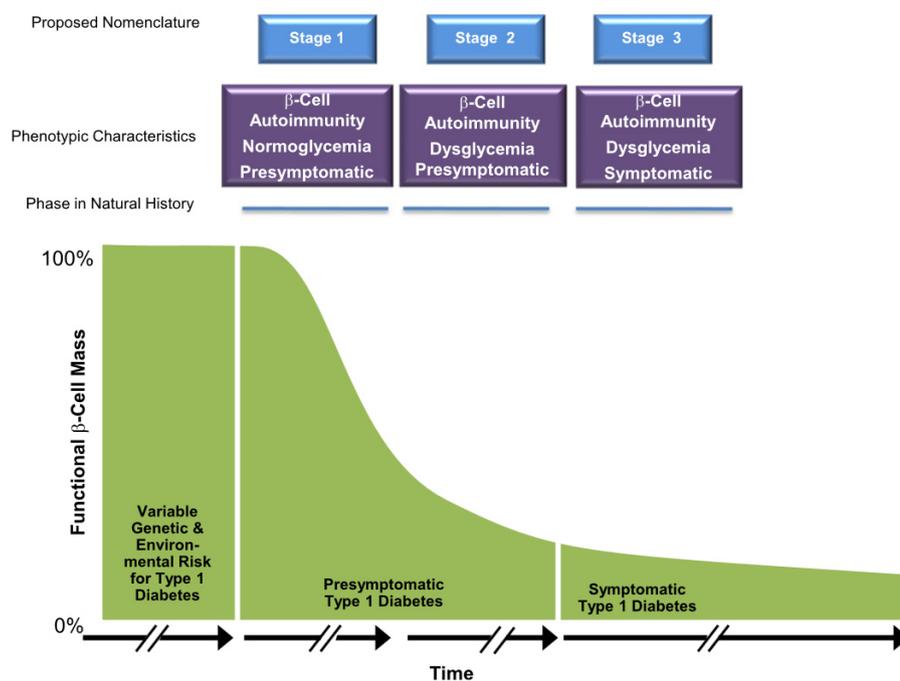
Figure 1 The natural history of type 1 diabetes with its corresponding staging classification.
Genetic and environmental risk factors can variably affect T1D development. The disease model proposed by Insel et al. (2015) describes T1D as a continuous disease progressing sequentially at variable rates through three distinct stages. This model distinguishes presymptomatic T1D (stages 1 and 2) from symptomatic T1D (stage 3). Presymptomatic T1D with normoglycemia is initiated by the onset of ß-cell autoimmunity indicated by multiple circulating autoantibodies (stage 1). The presymptomatic stage 2 is defined as persistent ß-cell autoimmunity with the development of glucose intolerance (dysglycemia). Stage 3 represents the onset of overt metabolic disease with clinical symptoms. Model and graphic source: Insel et al. (2015).

Approximately 300 million people worldwide are affected with diabetes mellitus (DM) and T1D accounts for 10-15% of all cases (Pociot and Lernmark, 2016; Sherwin and Jastreboff, 2012). T1D is one of the most common chronic diseases of early childhood; yet the initiation of symptomatic disease is not restricted to childhood and can be diagnosed at any age (Gale, 2005; Leslie, 2010). Globally, the incidence of T1D has been increasing for several decades (Dabelea, 2009). In Europe, the relative annual increase has been reported to be on average 3.9% with the disease being most common in Finland and Sardinia (Patterson et al., 2009). Notably, the reported increase in incidence does not apply equally to all patient groups but predominantly affects young children under the age of 5 (Harjutsalo et al., 2008). This increase in early diabetes incidence is closely linked to a faster progression from ß-cell autoimmunity to diabetes onset in a recent study investigating genetically at-risk children in Germany (Ziegler et al., 2011). While T1D is common in Europe and North America, the disease is not very

2

prevalent in large parts of Asia or South America (Atkinson et al., 2014; Maahs et al., 2010) (Fig. 2). Surprisingly, even in countries which are geographically close profound differences in incidence have been reported, e.g. T1D incidence in Estonia is beneath one-third of the incidence in Finland (Podar et al., 2001). These vast differences in geographical and population-based incidence suggest genetic factors and environmental agents in triggering or driving the pathogenic process underlying ß-cell destruction. The influence of an environmental component is further supported by studies reporting that T1D incidence is influenced by seasonal changes and birth month (Kahn et al., 2009; Moltchanova et al., 2009).
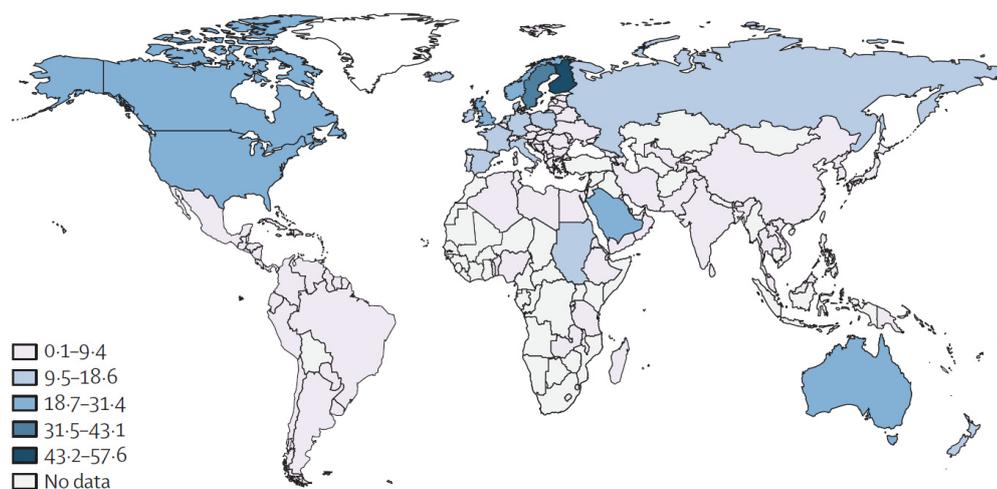
Figure 2 Incidence of type 1 diabetes worldwide.
Depiction of the estimated global incidence of type 1 diabetes for the year 2011 based on data from Whiting et al. (2011). Graphic source: Atkinson et al. (2014).

1.1.1 Pathophysiology

1.1.1.1 Genetics

T1D risk assessment can be performed by patient genotyping. Genetic predisposition for T1D mainly arises from human leukocyte antigen (HLA) coding genes and to a much lesser extent from more than 50 other susceptibility loci (Barrett et al., 2009; Concannon et al., 2009; Cooper et al., 2008; Onengut-Gumuscu et al., 2015). HLA cell surface glycoproteins can be divided into two distinct classes known as major histocompatibility complex (MHC)-I and MHC-II (Fig. 3). While all nucleated cells in the human body express MHC-I molecules on their surface, MHC-II expression is restricted to professional antigen-presenting cells (APCs) such as macrophages, dendritic cells or B cells. Both MHC-I and MHC-II play a key role in the immune response being responsible for peptide antigen presentation to T cells: CD8+ cytotoxic T lymphocytes (CTLs) recognize endogenous antigens in the context of MHC-I, whereas CD4+ T helper cells

recognize exogenous antigen in the context of MHC-II (Fig. 3). Hence, the HLA variants carried by an individual influence peptide binding and signal transduction after T cell receptor (TCR) engagement, hereby strongly influencing thymic selection and peripheral activation of T cells (Pugliese, 2017).
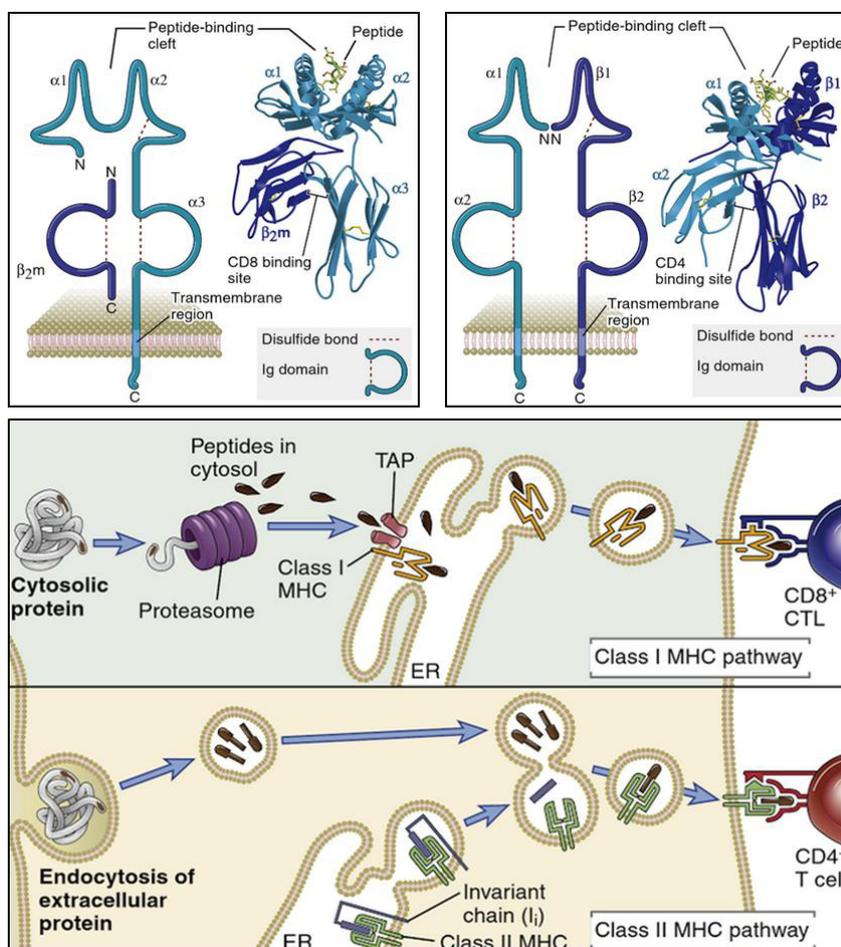
Figure 3 Structure of MHC class I (top left) and II (top right) molecules and corresponding pathways of antigen presentation to T cells.
The top left panel depicts the MHC class I molecule consisting of a polymorphic α-chain which is non-covalently associated with a non-polymorphic ß2-microglobulin. The peptide-binding cleft accommodates peptides of approximately 8-11 residues in length. The MHC class II molecule shown in the top right panel is composed of a polymorphic α- and β-chain which are non-covalently attached. The peptide-binding cleft is for peptides of 10-30 residues or longer in length. The bottom panel illustrates the two distinct pathways of antigen processing and presentation with class I and II MHC molecules. In the endogenous pathway, intracellular proteins in the cytosol are processed by the proteasome, transported into the endoplasmic reticulum by TAP and loaded onto MHC-I molecules for presentation to CD8+ T cells. In the exogenous pathway, extracellular protein antigens are endocytosed, enzymatically processed within the late endosome or lysosome and then loaded onto MHC-II molecules for presentation to CD4+ T cells. Ig, immunoglobulin; TAP, transporter associated with antigen processing; ER, endoplasmic reticulum. Model and graphic source: Abbas et al. (2014).

Notably, the MHC region which encodes the HLA on chromosome 6p21 confers about 50% of the genetic risk for T1D (Clayton, 2009; Noble et al., 1996; Noble et al., 2010b).

In detail, the disease is closely linked to polymorphisms in genes encoding the MHC-II molecule with greatest susceptibility being conferred by specific HLA DR/DQ alleles, e.g. DRB1*0301-DQB1*0201 (DR3/DQ2) or DRB1*0401-DQB1*0302 (DR4/DQ8) (Erlich et al., 2008). Indeed, 80-90% of patients with T1D carry one of the latter high-risk haplotypes, whereas 30-50% of patients have both haplotypes − this heterozygous genotype confers greatest predisposition (Erlich et al., 2008). In contrast, the most protective haplotypes are DRB1*1501 and DQA1*0102-DQB1*0602 providing disease resistance (Erlich et al., 2008; Greenbaum et al., 2000). Presentation of peptide antigens in the context of MHC-II is partially determined by the amino acid sequence of MHC-II α- and ß-chains. Hence, the linkage between genetic polymorphisms in the MHC-II region and T1D susceptibility can reflect a changed ability to present relevant autoantigens (Todd et al., 1988). For example, the HLA-DQ heteromer comprising the DQ2 α-chain and the DQ8 ß-chain is capable of efficiently presenting autoantigen epitopes to T cells (van Lummel et al., 2012). Independent of the risk conferred by HLA DR/DQ alleles, MHC-I genetic loci (HLA-A, -B and -C) also contribute to T1D susceptibility, however to a much lesser extent (Noble et al., 2010a). For instance, the HLA-A*02 allele, occurring with a frequency of >60% in patients with T1D, enhances T1D susceptibility in individuals already carrying the high-risk DR3/4-DQ8 haplotype (Fennessy et al., 1994; Robles et al., 2002; van Belle et al., 2011).

The remaining genetic risk for T1D is conferred by approximately 50 non-HLA loci which were identified by genome-wide association studies (GWAS) and which are predominantly linked to impaired immune regulation (Barrett et al., 2009; Pociot et al., 2010; Todd et al., 2007). Unlike the HLA-locus which confers high T1D susceptibility, the individual non-HLA loci alone exert only a modest effect on disease risk. The most prominent non-HLA T1D susceptibility genes include *INS* (insulin), *PTPN22* (protein tyrosine phosphatase, non-receptor type), *CTLA4* (coding for cytotoxic T lymphocyte–associated protein 4) and *IL2RA* (interleukin 2 receptor α-chain). Susceptibility derived by the *INS* locus on chromosome 11 is based on a variable number of tandem repeat (VNTR) polymorphisms in its promotor region. Here, class I alleles with shorter repeats predispose to T1D, whereas class III alleles with longer repeats provide protection (Bennett et al., 1997; Kockum et al., 1996). The other three genes *PTPN22*, *CTLA4* and *IL2RA* are implicated in TCR signaling, limitation of T cell responses and regulatory T cell function, respectively (Garg et al., 2012; Malek and Castro, 2010; Pugliese, 2017). Remarkably, recent transcriptomic studies prove that numerous non-HLA risk loci are expressed in pancreatic ß-cells suggesting their direct involvement in pathogenic mechanisms at the ß-cell level by modulating processes such as viral infection, islet

inflammation and apoptosis (Bergholdt et al., 2012; Eizirik et al., 2012; Størling and Pociot, 2017). For instance, the T1D susceptibility gene *CTSH* (cathepsin H) protects the ß-cell against cytokine-induced apoptosis (Floyel et al., 2014). Another important non-HLA locus is *IFIH1* (interferon-induced helicase) which functions in virus sensing by mediating the induction of the interferon response (von Herrath, 2009; Winkler et al., 2011). Contrary to the HLA-locus which seems to have limited influence on the rate of progression from ß-cell autoimmunity to onset of symptomatic disease, distinct non-HLA susceptibility loci have been reported to have notable effects on disease progression (Achenbach et al., 2013; Lipponen et al., 2010). Consequently, combined information on single nucleotide polymorphisms (SNPs) and risk allele scores in non-HLA loci have improved risk stratification for the development of ß-cell autoimmunity and progression to symptomatic disease (Winkler et al., 2014; Winkler et al., 2012).

In summary, genetic susceptibility to T1D seems to affect the immune system level and the pancreatic ß-cell itself. On the one hand, the immune system is likely predisposed to an imbalance between autoimmune activation and the maintenance of self-tolerance. On the other hand, genetic susceptibility potentially influences ß-cell responses to environmental insults and the immune system, thereby modulating ß-cell function and survival.

1.1.1.2 Environment

Insights from studies investigating the distribution of HLA genotypes over time among newly diagnosed patients with T1D suggest decreasing frequencies of high-risk HLA haplotypes (Gillespie et al., 2004; Hermann et al., 2003; Vehik et al., 2008). This observation implies an increasing influence of environmental agents in triggering the onset of symptomatic disease in genetically less susceptible individuals. The impact of environmental determinants is further supported by the fact that approximately 90% of newly diagnosed children develop the disease sporadically without family history (Lebastchi and Herold, 2012a) and the observation that monozygotic twins possess a rather low pair-wise concordance for T1D (Barnett et al., 1981; Kaprio et al., 1992). Hence, there is a growing body of evidence connecting a plethora of environmental exposures with the occurrence of islet autoimmunity and/or the onset of symptomatic T1D including viruses, bacteria, dietary factors as well as anthropometric and psychosocial factors (Regnéll and Lernmark, 2013; Rewers and Ludvigsson, 2016).

The association between viral infections and T1D etiology has existed throughout the past century (Gamble et al., 1969; Yoon, 1990). The common assumption is that viruses

may contribute to ß-cell death either directly via cytolytic effects or by triggering autoimmune responses which eventually result in ß-cell destruction (Knip and Simell, 2012). Here, an often discussed pathway which pinpoints viruses as drivers of the autoimmune response is molecular mimicry, where structural homology between viral and ß-cell antigens leads to harmful cross-reactive immunity (Coppieters et al., 2012b). Enteroviruses (e.g. Coxsackievirus B) are certainly among the most intensively investigated virus species in the context of T1D. In consequence, numerous studies in humans have provided proof for a close tie between enterovirus infections and T1D pathology (Stene and Rewers, 2012). Evidence has been provided for the tropism of enteroviruses to pancreatic islets (Roivainen et al., 2000; Tracy et al., 2011; Yoon et al., 1979), their potential to induce islet autoantibodies (Hyoty et al., 1995) and their influence on disease progression (Hyoty and Taylor, 2002; Stene et al., 2010; Yeung et al., 2011). Specifically, the Finnish Diabetes Prediction and Prevention (DIPP) study has observed a close temporal association between the first occurrence of islet autoantibodies and enterovirus infection (Lonnrot et al., 2000; Oikarinen et al., 2011). In the same study a seasonal variation in the initiation of islet autoimmunity was observed which fitted well with the annual enterovirus infection pattern being highest in the fall and winter (Kimpimaki et al., 2001b). Notably, enterovirus infections during pregnancy have been linked to islet autoimmunity in the mother and the offspring (Dahlquist et al., 1995; Resic Lindehammer et al., 2012). Furthermore, the enteroviral capsid protein (VP1) has been detected more frequently in the islets of newly diagnosed patients with T1D than in healthy controls ─ consistent with a persistent enteroviral infection (Krogvold et al., 2015; Richardson et al., 2009). However, two prospective studies in the US and Germany failed to confirm any link between enterovirus infection and islet autoimmunity (Fuchtenbusch et al., 2001; Graves et al., 2003). Whether enteroviruses function as inducers of ß-cell autoimmunity, accelerators of disease progression or general stress factors therefore remains an item of ongoing debate. Beyond enteroviruses, T1D has also been associated with a large repertoire of other virus candidates including rotavirus, cytomegalovirus, mumps, rubella and retroviruses (reviewed in Knip and Hyöty (2008)). For instance, molecular mimicry has been proposed for rotavirus protein VP7 and T cell epitopes in the islet autoantigens insulinoma antigen-2 and glutamic acid decarboxylase 65 (Honeyman et al., 2010).

Contrary to viruses, the involvement of the human microbiome in T1D etiology is largely underinvestigated. The intestinal microbiome is closely intertwined with an individual's diet which has a considerable effect on its composition (De Filippo et al., 2010; Turnbaugh et al., 2009). There is a line of evidence suggesting that both the composition

of the microflora as well as their metabolic products function in systemic immunomodulation affecting the balance between inflammatory and tolerogenic responses (Arpaia et al., 2013; Clarke et al., 2010; Maslowski et al., 2009; Wen et al., 2008). An association between the gastrointestinal system and T1D has been postulated in several recent studies reporting reduced microbial diversity in individuals with islet autoimmunity or patients with T1D as compared to healthy controls (de Goffau et al., 2013; Giongo et al., 2011; Kostic et al., 2015; Murri et al., 2013). Specifically, healthy controls benefited from a higher abundance of lactate- and butyrate-producing species likely to be responsible for maintaining intestinal homeostasis (Brown et al., 2011). Additionally, interesting evidence for the implication of the gut microbiome in T1D etiology has come from the NOD mouse. Wen et al. (2008) reported that abrogation of bacterial sensing in pathogen-free NOD mice (MyD88-negative) protected them from development of spontaneous diabetes. They further showed that this effect was dependent on commensal microbes as germ-free MyD88-negative mice developed diabetes, but had decreased incidence when colonized with a defined microflora representing that commonly found in humans. Hence, Wen and colleagues concluded that normal intestine microbiota induce local tolerance through a MyD88-independent mechanism, thereby alleviating disease progression. This finding is closely linked to the "hygiene hypothesis" proposing that enhanced hygiene standards and fewer childhood infections result in reduced exposure to microbes thereby diminishing protective immunity and increasing susceptibility to autoimmunity (Bach and Chatenoud, 2012).

Besides viruses and the microbiome, dietary factors represent an integral piece of environmental influence and have been intensively investigated in the context of T1D. Among the numerous factors which have been evaluated so far are breast feeding, cow milk, fruit and berry juices, gluten, vitamin D and omega-3 fatty acids (Beyerlein et al., 2014; Lamb et al., 2015; Norris et al., 2007; Raab et al., 2014; Sorkio et al., 2010; Virtanen et al., 2012). However, exploration of the connections between these variables and T1D development has often yielded contradictory results. For example, the association between breastfeeding and T1D development remains a controversial issue; while some prospective studies have found no causality (Virtanen et al., 2006; Ziegler et al., 2003), other studies indicate that short breastfeeding is a risk factor for the emergence of ß-cell autoimmunity (Holmberg et al., 2007; Kimpimaki et al., 2001a). In turn, another focus of investigation has been the early introduction of cow's milk proteins through infant formulas which is generally thought to mark the end of exclusive breastfeeding. Here, bovine insulin has been proposed as a driving exogenous antigen in the disease process potentially leading to cross-reactivity with human insulin (Vaarala

et al., 1998). However, once again, gathered evidence is contradictory with some data indicating early exposure to cow's milk as a predisposing factor (Holmberg et al., 2007; Virtanen et al., 1993), while others found no connection (Norris et al., 1996). Another important nutritional factor implicated in T1D susceptibility is vitamin D. A large-scale prospective study in Finland including >10000 subjects followed up for 30 years reported a reduced risk for T1D development if supplemented with high-dose vitamin D in the first year of life (Hypponen et al., 2001). In 2008, a meta-analysis was conducted which summarized the results from five observational studies confirming reduced risk for T1D development in infants who were supplemented with vitamin D (Zipitis and Akobeng, 2008). Additionally, Raab et al. (2014) reported that vitamin D levels were lower in both children with islet autoimmunity and symptomatic T1D compared to seronegative, healthy controls.

Unraveling the effects of environmental exposures on the development of T1D has proven to be difficult; however, it is very likely that environmental determinants and changes in lifestyle other than increased genetic disease susceptibility are causative of the increasing T1D incidence rates. A better understanding of (i) the interplay between genes and the environment and (ii) the combination or sequence of potentially causal environmental exposures is necessary to gain new insights into the disease process.

1.1.1.3 Islet autoantibodies

While the exact mechanisms triggering ß-cell autoimmunity remain elusive, the diagnosis of islet autoimmunity can be robustly performed with biochemical assays. This primary serological evidence for the autoimmune attack involves the detection of circulating autoantibodies to the four major humoral autoantigens insulin (IAA), glutamic acid decarboxylase 65 (GADA), insulinoma antigen-2 (IA-2A) and zinc transporter-8 (Znt8A) (Baekkeskov et al., 1990; Palmer et al., 1983; Payton et al., 1995; Wenzlau et al., 2007). A common feature of all four autoantigens is their involvement in the regulated secretory pathway of the pancreatic ß-cell with insulin being the prototypical cargo of the secretory granule, constituting up to 50% of total cell protein (Eizirik et al., 2009). IA-2 and Znt8 are located in the granule membrane where they function in granule mobility and exocytosis, and channeling of zinc necessary for insulin hexamer crystallization, respectively (Chimienti et al., 2005; Trajkovski et al., 2008). In contrast, GAD65 — the major enzyme in the synthesis of γ-aminobutyric acid — is a peripheral membrane protein stored in synaptic-like microvesicles within the ß-cell (Wenzlau and Hutton, 2013). Seroconversion to islet autoantibodies usually occurs throughout childhood being most frequent from nine months to two years of age; in fact, the greater part of individuals diagnosed with

Introduction

T1D before puberty had developed autoantibodies under the age of 3 years (Knip et al., 2016; Ziegler and Bonifacio, 2012). Remarkably, approximately 90% of newly diagnosed individuals are seropositive for two or more islet autoantibodies (Simmons and Michels, 2015). Hence, islet autoantibodies remain the strongest predictive marker for T1D development — over 80% of children who seroconvert to multiple islet autoantibodies progress to symptomatic T1D within 15 years (Ziegler et al., 2013). In contrast, only a small proportion of children seropositive for only one autoantibody develop T1D (Ziegler et al., 2013) (see also Fig. 4).
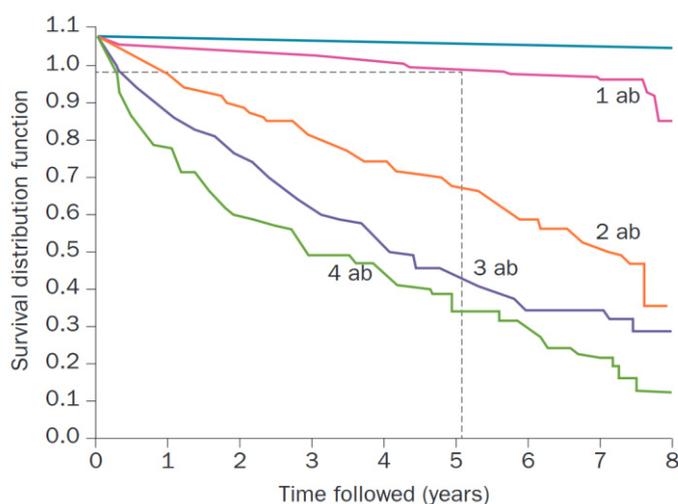
Figure 4 The number of islet autoantibodies influences T1D progression.
The number of islet autoantibodies present in an individual determines the rate of progression to onset of symptomatic T1D. While only a few individuals with a single autoantibody progress to clinical onset within 5 years after seroconversion, 60% of individuals who seroconverted to four islet autoantibodies develop T1D within 5 years. ab, islet autoantibody. Graphic source: Lernmark and Larsson (2013).

Commonly the first sign of islet autoimmunity typically involves IAA and/or GADA and is often followed by prompt spreading to multiple islet autoantibodies (Chmiel et al., 2015). Among children, IAA is usually the first autoantibody to emerge and consistently high levels of IAA are associated with 100% likelihood of progression to symptomatic T1D (Parikka et al., 2012; Steck et al., 2011). Accordingly, a recent study found that the loss of IAA reactivity was associated with delayed progression to T1D in multiple islet autoantibody-positive children (Endesfelder et al., 2016). Altogether, several factors seem to influence the rate of progression from islet autoantibody positivity to symptomatic T1D onset, including the age of seroconversion, IAA levels and the number of circulating autoantibodies (Steck et al., 2011).

Islet autoantibodies are indispensable surrogate markers for monitoring an ongoing ß-cell destruction process and potential screening benefits include identifying individuals at

10

risk, lowering the incidence of diabetic ketoacidosis, and evaluating the efficacy of therapies in clinical intervention trials (Simmons and Michels, 2015). Nonetheless, islet autoantibodies are generally thought to be not directly involved in ß-cell destruction, thus they do not provide considerable knowledge about the mechanisms driving the autoimmune attack. Much of the current understanding of immunopathology rather derives from the investigation of pancreatic specimens and peripheral blood lymphocytes.

1.1.1.4 Islet inflammation

The common perception that the selective destruction of ß-cells in T1D results from an autoimmune-driven — in particular T cell-mediated process — largely stems from the observation of a chronic inflammatory infiltrate affecting pancreatic islets at onset of symptomatic disease (In't Veld, 2011) (Fig. 5). This characteristic inflammatory infiltration was first observed in a pancreatic section of a child which had died from diabetic ketoacidosis over a century ago; the respective phenomenon was subsequently termed "insulitis" in the 1940s (Schmidt, 1902; von Meyenburg, 1940). However, in practice the investigation of human insulitis is hampered by the inaccessibility of the pancreas in living individuals, thus most studies on pancreatic pathology have involved the retrospective analysis of post-mortem pancreata obtained from autopsies or organ donors (Atkinson et al., 2014). As a result of the scarce availability, the analysis of islet inflammation using pathologic specimens has included fewer than 200 recent-onset T1D cases worldwide in the past century (In't Veld, 2011; Morgan et al., 2014). Current knowledge includes that insulitis can be found in the islet periphery (peri-insulitis) which is the predominant form compared to intra-insulitis which can be found within the parenchyma (Krogvold et al., 2016).
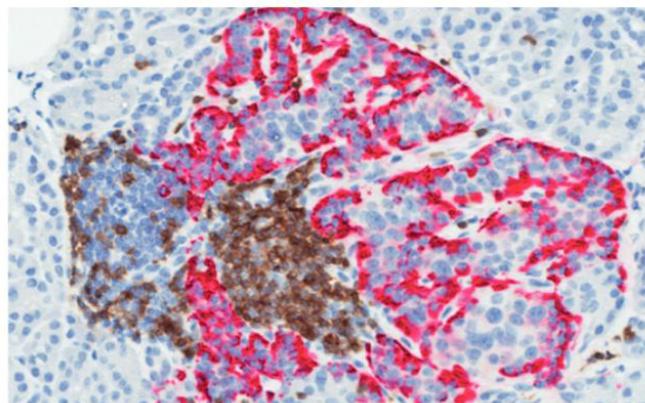
Figure 5 Inflammatory infiltrate in a pancreatic islet of a patient with recent-onset type 1 diabetes.
Immunohistochemistry of a post-mortem pancreatic islet section derived from a patient with newly diagnosed T1D showed evidence for intra-islet presence of infiltrating CD3+ T cells (brown).

Introduction

Glucagon-producing α-cells were stained in pink. Graphic source: Atkinson et al. (2014) directing image courtesy to M. Campbell-Thompson, University of Gainsville, USA.

Studies investigating pancreatic specimen obtained from recent-onset T1D individuals have revealed that approximately 70% of islets exhibited complete insulin loss, with nearly 20% of insulin-containing islets displaying signs of insulitis, whereas only 1% of insulin-deficient islets were inflamed; furthermore, many pancreatic sections proved the presence of non-inflamed insulin-containing islets (Foulis et al., 1986; Gepts, 1965; Gregg et al., 2012; Keenan et al., 2010b). These observations support the notion that insulitis occurs only in a relatively low proportion of islets; possibly attributable to the dissipation of inflammation once ß-cells have been destroyed (Morgan et al., 2014). However, more recently Campbell-Thompson et al. (2016) showed that pancreatic organ donors with long-standing T1D can still display signs of insulitis, highlighting that islet autoimmunity may persist for years after diagnosis. Studies aimed at characterizing the cellular composition of the inflammatory infiltrate revealed a predominantly lymphocytic population with CD8+ T cells being the most prevalent cell subset in the lesion, followed by macrophages (CD68+), CD4+ T cells, B cells (CD20+) and plasma cells (CD128+) (Bottazzo et al., 1985; Willcox et al., 2009). Interestingly, Willcox et al. (2009) observed that the numbers of CD8+ T cells increased when few residual insulin-positive ß-cells remained within islets, these however disappeared when insulin-positivity was lost. Additionally, Willcox and colleagues found that the population of CD20+ cells closely matched the dynamics of CD8+ T cells, while levels of macrophages were more constant. With regards to Natural killer (NK) cells, observations are not consistent with some studies detecting them in the infiltrate, while others did not (Dotta et al., 2007; Willcox et al., 2009). Additionally, the surrounding exocrine pancreas has been reported to be enriched in lymphocytes and neutrophils (Rodriguez-Calvo et al., 2014; Valle et al., 2013). The predominant presence of CD8+ T cells in insulitis may in part be explained by hyperexpression of HLA-I molecules on endocrine cells in insulin-containing islets which is another hallmark of T1D pancreatic pathology (Foulis et al., 1987; Richardson et al., 2016). This hyperexpression may be triggered by viruses or the presence of inflammatory cytokines such as interferon γ (IFN-γ) and tumor necrosis factor α (TNF-α) (Campbell et al., 1988; Krogvold et al., 2015). Interestingly, Coppieters et al. (2012a) provided evidence that autoantigen-specific CD8+ T cells occupied insulitis lesions, a finding which directly links T cell autoreactivity to T1D pathology. The authors further demonstrated that the antigen specificity of the infiltrating CD8+ T cells increased in diversity concomitant with longer disease duration implying post-diagnosis epitope spreading. In conclusion, CD8+ T cells are generally perceived as prime candidates to

mediate ß-cell cytotoxicity in an environment favoring antigen presentation by MHC-I hyperexpression.

In summary, insights gained from studies investigating pancreatic pathology largely pinpoint mononuclear immune cells, and in particular the adaptive arm of the immune system, as mediators of ß-cell destruction.

1.1.1.5 The immune system

Both genetic susceptibility and pancreatic pathology closely link T cells as drivers of autoimmunity. Indeed, several studies in humans and in mice provide compelling evidence that a failure in central and/or peripheral tolerance results in the activation and expansion of autoreactive T cells (Bluestone et al., 2010; Roep and Peakman, 2011). As previously stated, the primary genetic susceptibility for T1D is derived from the HLA class II genes which function in activating CD4+ T cells — which in turn license CD8+ T cell activation (Bevan, 2004) — highlighting that both cell populations are relevant for T1D pathogenesis. For instance, the NOD mouse model suggests that T1D development depends on both CD4+ and CD8+ T cell subsets as only transfer of both splenic CD4+ and CD8+ T cells but not either subset alone resulted in onset of T1D in immunocompromised syngeneic recipient mice (Phillips et al., 2009). The pivotal role of CD4+ T cells in the pathogenesis of the NOD mouse was already illustrated nearly three decades ago, when Shizuru et al. (1988) reported that immunotherapy targeting T helper lymphocytes can prevent the spontaneous onset of diabetes. The involvement of cellular immunity was also demonstrated in studies involving human subjects, for instance, Sibley et al. (1985) investigated the recurrence of T1D in recipients of segmental pancreas grafts obtained from HLA-identical donors and proved a clear role for T cells, especially CD8+ T cells, and monocytes in the recurrent ß-cell destruction. Sibley and colleagues could further show that ß-cell survival in the transplants could only be achieved if the recipient was immunosuppressed, indicating that selective ß-cell destruction was a consequence of a cell-mediated immune response.

The first prerequisite for T1D development in both humans and mice is the induction of an immune response against ß-cell antigens which secondly needs to exhibit strong pro-inflammatory characteristics in order to be destructive (Wallberg and Cooke, 2013). In parallel, the regulatory control of the elicited autoimmune response needs to be compromised resulting in a chronic destruction of ß-cells. Figure 6 illustrates a simplified model of immunopathogenesis in T1D involving components of both the innate and adaptive immune system (Roncarolo and Battaglia, 2007). The first insult which triggers

release of ß-cell antigens could either be virus-induced or physiological (Kassem et al., 2000; Turley et al., 2003). According to the model proposed by Roncarolo and Battaglia (2007) tissue-resident APCs such as dendritic cells subsequently capture, process and present (modified) ß-cell antigens in the context of MHC-I to CD8+ T cells which may directly lyse ß-cells through the perforin-granzyme pathway or induce apoptosis by secretion of cytotoxic cytokines such as IFN-γ. Immature DCs then migrate to the draining pancreatic lymph nodes, where ß-cell antigens are presented to CD4+ T cells; this lymph node priming step is thought to be crucial leading to the expansion of low frequency circulating autoreactive T cells. Clonal expansion of autoreactive CD4+ T cells results in the expression of adhesion molecules such as ICAM1 and LFA1 as well as chemokine receptors (CCR4, CCR5 and CXCR3) which allow homing of effector cells to pancreatic islets by tracing chemokines induced by the early CD8+ T cell inflammatory response. The model by Roncarolo and Battaglia (2007) continues by stating that CD4+ T cells, once in the pancreas, recruit and activate other inflammatory cells such as macrophages, eventually leading to insulitis. The recruited macrophages secrete pro-inflammatory interleukin-1 (IL-1) and TNF which may induce either pro-apoptotic signaling and/or the expression of CD95 (FAS) in ß-cells which allows killing through CD95L (FasL) expressing effector T cells. Additionally, IFN-γ secreted by effector T cells potentiates the capacity of cytotoxic cytokine production by macrophages — this T cell-macrophage crosstalk exacerbates the pro-inflammatory immune cascade (Lehuen et al., 2010). At several checkpoints throughout the model, regulatory T cells such as FOXP3+CD24+CD25+ cells (TReg) and T regulatory type 1 cells (T_R1) could prevent the activation of autoreactive T cells either directly or indirectly through limiting their access to APCs (Tang et al., 2006).
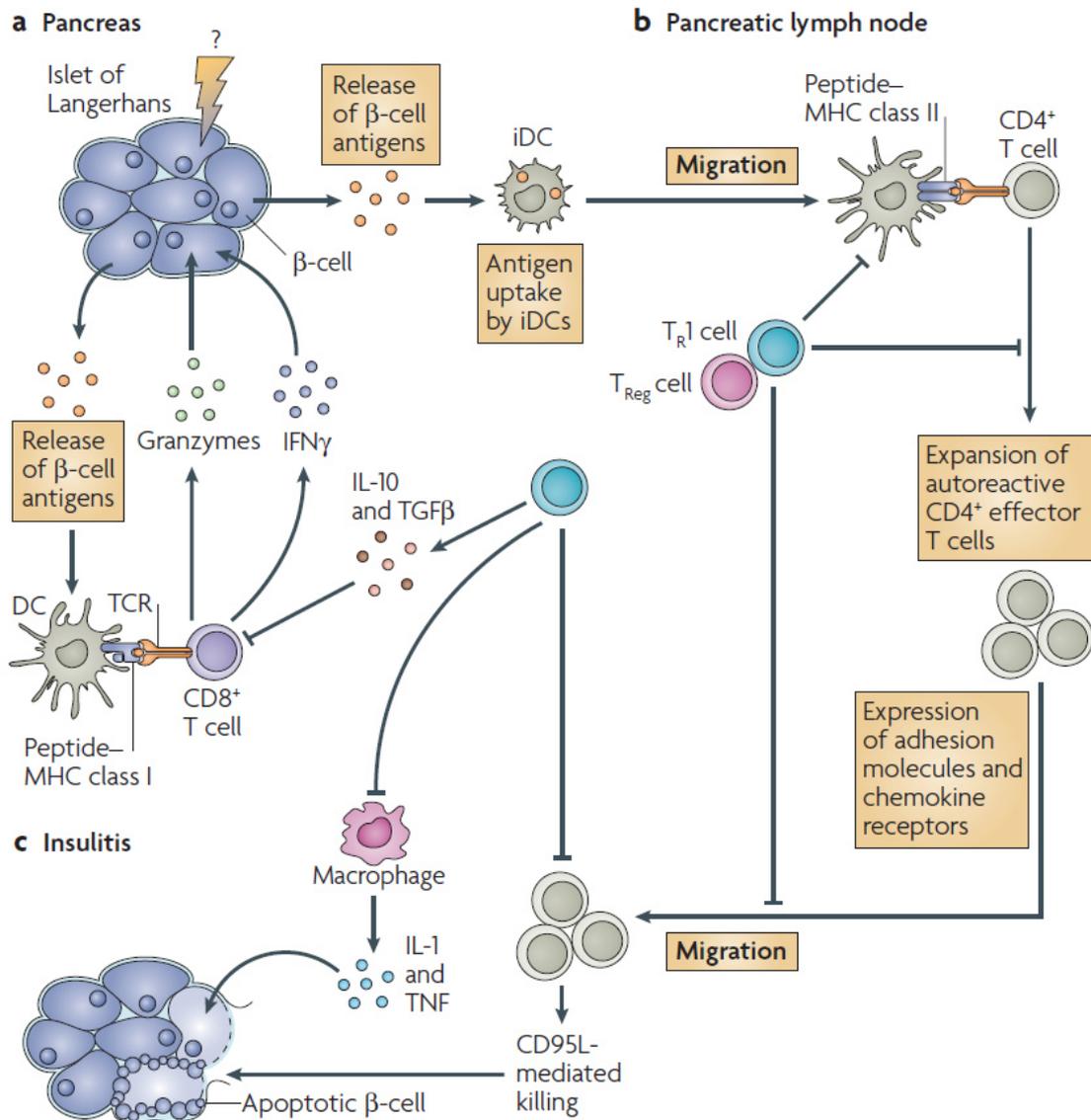
Figure 6 A simplified model of immunopathogenesis in type 1 diabetes.
The release of (modified) ß-cell antigens which is triggered by still undefined pathologic factors is the first step in the pathogenesis of T1D. Subsequently, professional antigen-presenting cells (APCs) such as dendritic cells phagocytose ß-cell proteins and present peptide antigen fragments in the context of MHC-I to islet-specific CD8+ T cells which can damage ß-cells through the perforin-granzyme pathway or by secretion of cytotoxic cytokines (IFN-γ). Autoreactive CD4+ T cells are activated in the draining lymph node, expand and home to pancreatic islets as effector cells where they recruit and activate other inflammatory cells such as macrophages which promote ß-cell destruction through production of cytotoxic cytokines (IL-1 and TNF) or reactive oxygen species. Additionally, effector T cells contribute to ß-cell killing through Fas/FasL (CD95/CD95L) interactions. Activation and expansion of autoreactive T cells can be blocked by regulatory T cells such as TRegs and T_R1 cells at several checkpoints, however in T1D immune tolerance is broken. iDC, immature dendritic cell; IL, interleukin; TNF, tumor necrosis factor; TGF-ß, transforming growth factor β. Graphic and model source: Roncarolo and Battaglia (2007).

The prerequisite for T cells recognizing self-antigens is a failure in thymic deletion. Escape of negative selection by low-affinity T helper cells may result in the response to islet autoantigens or to neoantigens — generated in the target tissue but absent in the thymus — both presented in the context of MHC-II in the periphery (Marrack and Kappler,

Introduction

2012). Autoreactive CD4+ T cells are integral to T1D pathogenesis as they participate in several events of cellular crosstalk; besides recruiting macrophages (Calderon et al., 2006) they may also activate islet-antigen specific B cells which in turn differentiate into antibody-producing B cells (Wallberg and Cooke, 2013). Activated B cells can function as APCs thereby potentially increasing autoimmune ß-cell responses. Moreover, in order to elicit an effective CD8+ T cell response, DCs need to be licensed with the cross-presentation of exogenous antigens in the context of MHC-I (Joffre et al., 2012). Acquisition of cross-presentation capacity relies on activation of DCs which usually occurs through interaction with activated CD4+ T cells (Smith et al., 2004). Thus, CD4+ T cell crosstalk and cytokine production are crucial mediators for priming the inflammatory response directed at ß-cells. The immune cell crosstalk of pro-inflammatory CD4+ T cells (Th1) in the context of T1D pathogenesis is summarized in Fig. 7.
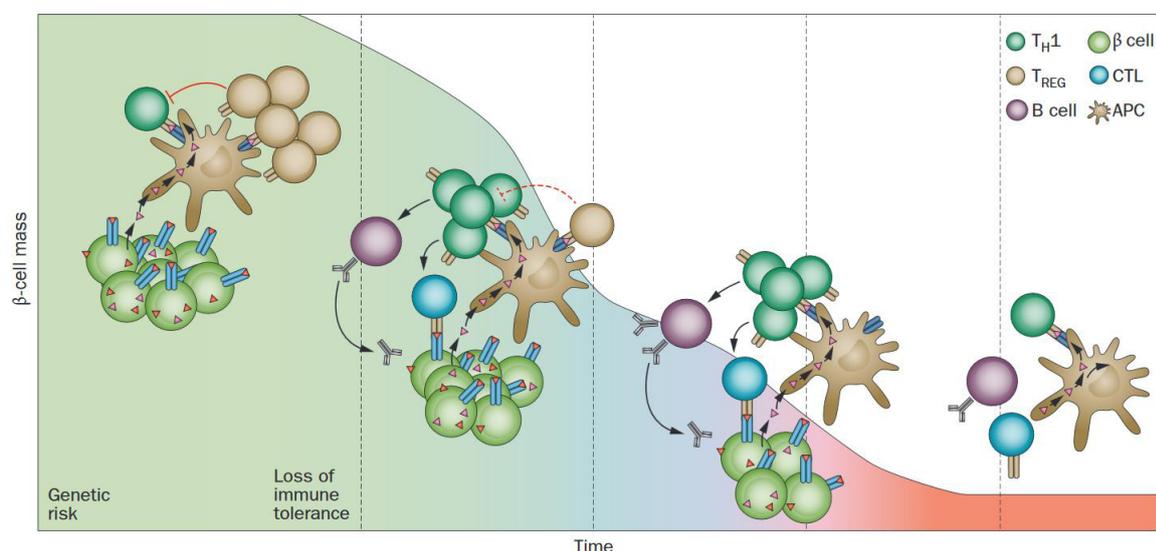
Figure 7 Crosstalk of the T helper cell population in the pathogenesis of T1D.
Preceding the onset of ß-cell autoimmunity, an immune balance persists in which activation of autoreactive T cells is suppressed by regulatory T cells (TRegs). By yet unidentified factors, this immune balance is challenged resulting in compromise of immune tolerance. Professional antigen-presenting cells (APCs) phagocytose ß-cell proteins and present peptide antigen fragments in the context of MHC-II to pro-inflammatory CD4+ T helper cells (Th1) which in turn initiate a cascade of immune responses. CD4+ T cells are capable of activating antibody-producing B cells and islet-antigen specific CD8+ cytotoxic T lymphocytes (CTLs) which may directly lyse ß-cells presenting islet peptides in the context of MHC-I. This model depicts ß-cell loss as a chronic progressive process; however it is currently unknown whether it may also exhibit a relapsing/remitting pattern. Graphic and model source: Roep and Tree (2014).

In the healthy state, autoreactive immune responses are kept in check by regulatory mechanisms commonly referred to as immunological tolerance. As previously stated, regulatory T cells (TRegs and T_R1 cells) are primary controllers of immunological tolerance (see also Fig. 6). TRegs may elicit their protective function through production of immunosuppressive cytokines (TGF-ß, IL-10), cell-cell contact-dependent

mechanisms (CTLA-4 expression), and by depriving pathogenic T cells of IL-2 — mediated by the expression of high affinity IL-2 receptor α-chain (CD25) (Busse et al., 2010; Sojka et al., 2008). Sarween et al. (2004) further demonstrated that TRegs can inhibit the expression of CXCR3 on T helper cells thereby preventing pancreatic tissue infiltration. Additionally, T_R1 cells produce high levels of IL-10 which has been shown to elicit an immunoregulatory role in T1D (Roncarolo et al., 2003). Interestingly, studies aimed at investigating TReg function in the context of T1D found that patients with T1D exhibited impaired TReg function, and resistance of effector T cells to TReg-mediated suppression (Buckner, 2010; Ferraro et al., 2011; Lindley et al., 2005; Schneider et al., 2008). In conclusion, several lines of evidence closely link impaired TReg function with T1D development highlighting that successful restoration of immunological tolerance is a promising therapeutic option for T1D (Hull et al., 2017).

1.1.2 Therapeutic strategies

To date, no effective therapeutic strategies exist which successfully prevent or reverse T1D development. Once symptomatic onset of T1D has been diagnosed, patients have to commit to lifelong insulin administration in the form of daily injections or an insulin pump to maintain good glycemic control (Simmons and Michels, 2015). Heightened awareness of the importance of diabetes management has decreased the risk for long-term complications such as myocardial infarction, hyperglycemic crisis, stroke, renal failure, and lower-extremity amputation since the 1990s (Gregg et al., 2014). Nonetheless, the ability to predict T1D with the measurement of islet autoantibodies has led to numerous clinical trials aimed at preventing, delaying or reversing the onset of T1D (Simmons et al., 2016). Intervention trials with the primary goal of maintaining or restoring residual ß-cell function have mainly focused on patients with newly-diagnosed T1D and included (i) the depletion of adaptive immune cells (Herold et al., 2013; Keymeulen et al., 2010; Pescovitz et al., 2014), (ii) blocking inflammatory cytokines (Mastrandrea et al., 2009; Moran et al., 2013; Sumpter et al., 2011), (iii) autoantigen vaccination (Buzzetti et al., 2011; Ludvigsson et al., 2014; Ludvigsson et al., 2012; Skyler et al., 2005) and (iv) enhancing TReg function (Long et al., 2012). Although some intervention trials achieved beneficial effects in preclinical studies, none of these approaches succeeded in restoring long-term independence of exogenous insulin administration (Herrath et al., 2013). In consequence, a general perception has evolved questioning the sufficiency of single-agent immunomodulation or immunosuppression in restoring residual ß-cell function — suggesting that application of combined therapeutic approaches could pave the way to success. In contrast to single-agent immunotherapies, novel combination therapies may target several different aspects of T1D pathology at

once, e.g. the modulation of innate and adaptive immunity, the induction of immunological tolerance and the stimulation of ß-cell regeneration (Pozzilli et al., 2015; Skyler, 2015). However, design and implementation of the right combinations is challenging and to date the few clinical combination trials which have been enrolled have mostly failed to show benefit (Gottlieb et al., 2010; Griffin et al., 2014; Skyler, 2015). Hence, a greater understanding of the precise immunological events underlying ß-cell destruction is necessary to develop tailored therapeutic strategies. The realization that T1D pathogenesis is a complex disease − harboring a great extent of heterogeneity among individuals − has generated a demand for better biomarkers to complement measurement of the known islet autoantibodies (Heinonen et al., 2015).

1.1.3 "Omics" strategies in biomarker discovery

The World Health Organization (WHO) has defined the term biomarker as "any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease" (Bonifacio, 2015; Strimbu and Tavel, 2010). With regards to T1D, there is an unmet need for the identification of highly sensitive and specific biomarkers which enable (i) more accurate detection of individuals in the presymptomatic stage, (ii) the prediction of progression rates to symptomatic onset of T1D and (iii) facilitation of the understanding of pathologic mechanisms. In other words, the primary goal is to identify biomarkers capable of differentiating disease processes, hereby guiding the development of novel therapeutic strategies and serving as quantifiable measures for monitoring clinical outcome such as the preservation of ß-cell function (Lebastchi and Herold, 2012b). In the long-term, the precise definition of immune signatures of disease heterogeneity holds the potential to ultimately implement the provision of personalized medicine comparable to the standards achieved in cancer therapy (Roep and Peakman, 2010; Roep and Tree, 2014). Currently, the standard biomarkers for diabetes preferred by regulatory agencies such as the Food and Drug Administration (FDA) comprise C-peptide levels, insulin usage and glycated hemoglobin (HbA1c) (Rekers et al., 2015). Although these variables provide clear information about clinical outcome, they convey little insight into the effectiveness of a given immunotherapy since they do not reflect changes in the underlying immunopathology (Rekers et al., 2015). Remarkably, the FDA recommends measurement of immune biomarkers for trials applying immune-modifying drugs in T1D, however fewer than half have included any immune endpoints as of 2015 (Rekers et al., 2015).

Biomarkers which have previously been outlined as predictors of T1D are either directly or indirectly associated with pancreatic islets (Bonifacio, 2015). These include the strong

genetic association founded in the HLA locus and to a lesser extent in the *INS* gene which determine islet autoimmunity, and secondly islet autoantibodies themselves, the only truly established markers for detection of ß-cell autoimmunity (Bonifacio, 2015). Yet, humoral activation does not necessarily correlate with activation of cellular immune responses in T1D (Hummel et al., 1996). Hence, increasing efforts have been made in the past years to complement the prevailing serologic and genomic knowledge with holistic insights into the transcriptomic, metabolomics, epigenetic and proteomic level on a systemwide scale (Heinonen et al., 2015).

The obvious measure of disease activity would likely be in the target tissue itself, however with the inaccessibility of pancreatic islets in living individuals, biomarker discovery studies have largely focused on the investigation of peripheral blood. The easy accessibility of peripheral blood for minimal-invasive sequential sampling and its presumed interaction with the affected tissue through dynamic exchange of cells, proteins (cytokines/chemokines) and other soluble factors (e.g. lipids and microRNAs) makes it a practical surrogate for biopsy material (Cabrera et al., 2016a). Centrifuged peripheral blood is composed of 55% plasma, 45% erythrocytes and 1% buffy coat. The buffy coat contains blood platelets and white blood cells comprising approximately 50-70% neutrophils, 20-30% lymphocytes, 2-8% monocytes, 2-4% eosinophils and less than 1% basophils (Hansen and Netter, 2014). Additionally, peripheral blood mononuclear cells (PBMCs) consisting of lymphocytes (T cells, B cells and NK cells), monocytes and dendritic cells can readily be isolated using Ficoll-based density gradient centrifugation. In consequence, applying "omics" to blood-derived samples such as PBMCs or plasma/serum holds great potential as a discovery strategy integrating the uncovering of pathologic mechanisms on the cellular level with the discovery of circulating biomarker candidates.

1.1.3.1 Transcriptomics

With regards to blood-based profiling, directly measuring the transcriptome of whole blood, PBMCs or their subpopulations in cases and controls is certainly among the most extensively applied techniques. For instance, application of this direct transcriptomic strategy had a significant clinical impact in oncology where it succeeded in diagnosing and classifying acute lymphoblastic leukemia subtypes with sufficient accuracy (Cabrera et al., 2016a; Roberts and Mullighan, 2015). In the context of T1D, transcriptomic profiling has been conducted on whole blood (Elo et al., 2010; Kallionpaa et al., 2014; Reynier et al., 2010), PBMCs (Evangelista et al., 2014; Ferreira et al., 2014; Jin et al., 2013; Kaizer et al., 2007; Rassi et al., 2006; Stechova et al., 2012) and immunocyte

subsets (Beyan et al., 2010; Orban et al., 2007; Padmos et al., 2008). Notably, Kaizer et al. (2007) found that PBMCs of newly diagnosed patients exhibited a transient innate inflammatory transcriptional profile including elevated levels of *IL1B* transcripts. Accordingly, Stechova et al. (2012) investigated PBMC transcriptional profiles in subjects with recent-onset T1D, their healthy seronegative first-degree relatives and unrelated healthy controls and identified IL-1 signaling to be the most significantly altered response pathway. Reynier et al. (2010) explored gene expression profiles of whole blood samples and reported a type 1 IFN-regulated signature associated with recent-onset T1D and autoantibody positivity in first-degree relatives supporting the hypothesis of a viral etiology. Notably, transcriptomic analyses in two longitudinal pediatric cohorts confirmed a role for a transient IFN-driven signature in children prior to seroconversion (Ferreira et al., 2014; Kallionpaa et al., 2014). Ferreira and colleagues further demonstrated that the identified IFN signature was temporally associated with a recent history of upper respiratory tract infections. Altogether, studies aimed at exploring the transcriptomic landscape of peripheral blood cells in different stages of T1D indicate an elevated state of innate inflammation.

A specialized set of transcripts are small, non-coding microRNA molecules (miRNAs) which regulate protein expression at the post-transcriptional level. With regards to T1D, Nielsen et al. (2012) explored circulating miRNA profiles in children with newly diagnosed T1D and identified a signature of twelve upregulated miRNAs implicated in apoptosis and ß-cell function. Specifically, Nielsen and colleagues identified miR-25 to be negatively associated with residual ß-cell function and positively associated with glycemic control. Recently, Garcia-Contreras et al. (2017) investigated miRNA expression profiles in plasma-derived exosomes and confirmed the upregulation of miR-25 in patients with long-standing T1D. Another miRNA candidate which has been suggested for monitoring residual ß-cell function is miR-375, an islet-cell specific miRNA (Erener et al., 2013; Poy et al., 2004). Marchand et al. (2016) demonstrated that miR-375 was lower in children with recent-onset T1D compared to age-matched healthy controls. Accordingly, Samandari et al. (2017) found levels of hsa-miR-375 to be negatively associated with residual ß-cell function 6 months post-diagnosis. Samandari and colleagues further propose hsa-miR-197-3p as a novel biomarker being the strongest predictor of residual ß-cell function one year post-diagnosis in their study cohort.

1.1.3.2 Epigenetics

Beyond microRNAs, regulation of gene expression is determined by epigenetic modifications. The most intensively investigated epigenetic marks are DNA methylation, and histone post-translational modifications (Petronis, 2010). Importantly, environmental factors can induce alterations of gene expression via epigenetic mechanisms — a possible explanation for the incomplete concordance of T1D susceptibility in monozygotic twins (Wang et al., 2017). For example, Stefan et al. (2014) explored DNA methylation patterns in lymphocyte cell lines of monozygotic twin pairs who were either discordant or concordant for T1D. Hereby, Stefan and colleagues could show that differential DNA methylation between affected and unaffected twins occurred in genes functionally annotated for immune and defense response pathways. Integration of these methylation patterns with GWAS data mapped several known T1D susceptibility genes including *HLA*, *INS*, *IL2RB* and *CD226* suggesting that alterations in DNA methylation patterns maybe implicated in T1D pathogenesis (Stefan et al., 2014). A recent study by Paul et al. (2016), investigating DNA methylation levels in immune effector cells to assess discordance of T1D in 52 monozygotic twin pairs, confirmed a role for epigenetic changes in T1D pathology. Specifically, the authors discovered cell type-specific enrichment of differentially variable CpG positions in twins with T1D; these positions were enriched at gene regulatory elements and linked to pathways associated to immune cell metabolism and the cell cycle, especially mTOR signaling (Paul et al., 2016). Another biomarker discovery approach exploiting DNA methylation status is the detection of circulating demethylated *INS* DNA as an indicator of ß-cell death (Zhang et al., 2017a). Interestingly, patients with newly diagnosed T1D exhibited increased levels of circulating ß-cell derived demethylated *INS* DNA compared to healthy controls (Akirav et al., 2011). The association of elevated levels of demethylated *INS* DNA was subsequently investigated and confirmed in additional T1D cohorts including the pre-onset and new-onset period highlighting its potential as a novel biomarker (Cabrera et al., 2016a; Fisher et al., 2015; Herold et al., 2015).

1.1.3.3 Metabolomics

Another "omics" technology which has been explored in T1D biomarker discovery is metabolomics, defined as the detection and quantification of the spectrum of small-molecule metabolites in a biological sample (Fiehn, 2002; Overgaard et al., 2016). Metabolites are perceived as the ultimate response of biological systems to environmental or genetic changes — representing the end products of cellular regulation (Fiehn, 2002). With regards to T1D, metabolomics studies have revealed (i) lower perinatal phospholipid levels in children developing T1D in early childhood, (ii) decreased

methionine levels at seroconversion in young children, and (iii) elevated triglycerides at seroconversion (La Torre et al., 2013; Oresic et al., 2013; Oresic et al., 2008; Overgaard et al., 2016; Pflueger et al., 2011). For instance, Oresic et al. (2008) investigated children enrolled in the DIPP study progressing to T1D and found decreased levels of phosphatidylcholines in cord blood as well as reduced levels of antioxidant ether phospholipids throughout the follow-up in serum. Oresic and colleagues could further demonstrate that these children exhibited increased serum levels of pro-inflammatory lysophosphatidylcholines several months before seroconversion. Interestingly, plasmalogens, which are the most abundant subclass of ether phospholipids in normal serum, can protect cells from oxidative stress (Oresic et al., 2008; Zoeller et al., 2002). Furthermore, Lenzen et al. (1996) have reported reduced intrinsic expression levels of antioxidant enzymes in murine pancreatic ß-cells in comparison to other tissues. In conclusion, the association between low phospholipid levels and elevated risk for T1D development could be explained by increased susceptibility of pancreatic ß-cells to oxidative damage (Oresic et al., 2008).

1.1.3.4 Proteomics

The last prominent "omics" technology which has been applied for biomarker discovery in T1D is proteomics which defines the large-scale analysis of the entire protein complement present in an experimental system in a given state at a given time. The principal dogma of molecular biology describes a unidirectional flow of information from DNA to RNA to proteins (Altelaar et al., 2013; Crick, 1970). With the discovery of epigenetic marks, alternative splicing, non-coding RNAs, protein-protein interaction networks and post-translational modifications (PTMs) this concept has been challenged providing key regulatory mechanisms which demonstrate that the information flow from genotype to phenotype is not uniquely directed by the genome (Altelaar et al., 2013). Accordingly, in comparison to the genome which is relatively static, the proteome is highly dynamic and complex being subject to alterations by environmental changes which result in considerable variations in protein expression between distinct organ tissues and cell types (Lottspeich, 1999). In consequence, the global analysis of proteins — the key functional entities of a cell — provides arguably the most relevant data set characterizing a biological system (Altelaar et al., 2013; Cox and Mann, 2007). Remarkably, many studies have provided evidence that expression changes at the level of transcripts do not necessarily correlate with the proteome effector level (Chen et al., 2002; Ghazalpour et al., 2011; Tian et al., 2004). In brief, the proteome of a biological system is much more complex and dynamic than the genome or the transcriptome, yet elucidation of its composition is indispensable for uncovering the functional level.

With regards to T1D, proteomic profiling studies have explored multiple biological fluids such as serum or plasma, urine and saliva as well as pancreatic tissue (reviewed in Moulder et al. (2017) and Crevecoeur et al. (2015)). Serum and plasma are exceptionally complex biofluids exhibiting a dynamic range of more than 10 orders of magnitude potentially harboring thousands of proteins (Anderson and Anderson, 2002; Nanjappa et al., 2014). For instance, Metz et al. (2008) were among the first to explore the serum proteome in patients with T1D; while their pilot study identified five biomarker candidates, they subsequently extended their profiling to a larger study cohort and highlighted 24 serum proteins closely linked to innate immunity and complement activation to be associated with disease status (Zhang et al., 2013a). Specifically, peptides from platelet basic protein (PPBP) and plasma protease C1 inhibitor (SERPING1) achieved high sensitivity and specificity for classification of samples (Zhang et al., 2013a). Zhi et al. (2011) investigated the T1D-associated systemic changes to the serum proteome applying two-dimensional liquid chromatography tandem mass spectrometry (2D LC-MS/MS) and found significant differences in 21 serum proteins implicated in inflammation, oxidation, metabolic regulation, and autoimmunity. Zhi and coworkers subsequently validated six candidate biomarkers using an immunoassay in a large-scale study population of 848 healthy controls and 1139 patients with T1D. This validation confirmed that patients with T1D exhibited significantly higher levels of inflammatory mediators such as adiponectin (ADIPOQ), C-reactive protein (CRP), serum amyloid A (SAA) and insulin-like growth factor binding protein 2 (IGFBP2), but lower levels of myeloperoxidase (MPO) and TGF-ß. Studies assessing longitudinal serum proteomic profiles which also address the comparison of prediabetic serum have been less frequent (Moulder et al., 2015; von Toerne et al., 2017). Moulder et al. (2015) investigated longitudinal samples from the DIPP study to profile temporal changes in the serum proteome from early infancy to prediabetic transition and onset of symptomatic T1D in genetically at-risk children. Moulder and colleagues identified dynamic changes in the serum protein profiles — some of which were even consistently observed before seroconversion to islet autoantibodies. The authors observed decreased serum protein levels of apolipoprotein family members APOC4 and APOC2 in prospective samples of children progressing to T1D. Remarkably, even before seroconversion lower levels of both proteins were apparent in children who developed T1D later in life compared to healthy controls (Moulder et al., 2015). Additionally, Moulder et al. (2015) measured lower levels of ADIPOQ in the period after detection of seroconversion implying distinct serum protein dynamics over time considering that Zhi et al. (2011) found increased levels of the very same protein in patients with manifested T1D disease. Most recently,

Introduction

von Toerne et al. (2017) compared children with islet autoantibody positivity to healthy controls and revealed higher levels of apolipoprotein M (APOM) and lower levels of APOC4 in seropositive children. Specifically, von Toerne and colleagues provided compelling evidence that the combined quantitative information from one specific peptide of each of these apolipoproteins could be exploited to discriminate seropositive and seronegative children. Moreover, the authors found that three peptides, derived from ceruloplasmin (CP), complement factor H (CFH) and hepatocyte growth factor activator (HGFAC), and age at sampling were predictive covariates for progression time to T1D (von Toerne et al., 2017). Interestingly, Albrethsen et al. (2009) showed that members of the apolipoprotein family (APOC1, APOC3) increased with time in serum of patients with T1D collected one, six and twelve months after diagnosis. Altogether, these results indicate that complex dynamics of apolipoprotein expression seem to underlie islet autoimmunity and T1D development. Recently, increasing progress has further been made in the field of direct proteomic analyses of the target tissue itself using human pancreatic specimens from the Network for Organ Donors with Diabetes (nPOD) (Burch et al., 2015; Liu et al., 2016). Burch et al. (2015) explored pancreatic tissue lysates obtained from autoantibody positive subjects, patients with either T1D or T2D, and healthy controls and identified protein signatures uniquely regulated in the prediabetic and diabetic states. For instance, among the proteins which were significantly higher expressed in T1D pancreata as compared to the other three experimental groups were Olfactomedin 4 (OLFM4), ectonucleotide pyrophosphatase (ENPP1) and regenerating islet-derived protein III α (REG3A). Profiling human pancreatic specimen in distinct stages of T1D disease generates a valuable foundation for studies aimed at identification of pancreas-specific proteins in peripheral blood which hold a great potential in serving as biomarkers for early detection of T1D (Cabrera et al., 2016a).

Taken together, the application of systemwide omics approaches to T1D is still in its infancy when compared to autoantibody assays, yet exploration of blood-based signatures harbors great potential in identifying novel biomarkers for T1D prediction and preclinical disease staging, likewise shedding light on underlying disease mechanisms (Cabrera et al., 2016a). Unsurprisingly, many pieces of "omics information" are still lacking to gain a holistic understanding of T1D pathology — Moulder et al. (2017) recently pointed out that detailed proteomic analyses of peripheral blood cell populations are currently missing although these are expected to provide functional differences between healthy and T1D status. Thus, to investigate changes at the effector cell level, a proteomics profile acquired using quantitative LC-MS/MS can be a powerful tool for obtaining a comprehensive portray of cellular pathology.

1.2 Quantitative Mass Spectrometry

Mass spectrometry (MS) is the current method of choice for investigation of complex protein samples by virtue of its speed, wide dynamic signal range, quantitative capability, and ability to couple with chromatographic separation methods (Perry et al., 2008). By definition, MS is an analytical technique which investigates ionized analytes in the gas phase by measuring their mass-to-charge (m/z) ratio and detecting the number of ions at each m/z value (Aebersold and Mann, 2003). If applied in tandem mode, this technology is not only capable of conveying information about the mass and quantity of an analyte, but may also unravel the analyte's identity, i.e. the primary structure. The challenge of proteome analysis is founded in the exceptional proteomic complexity of a given biological sample in comparison to its corresponding genome (Reinders and Sickmann, 2009). The paramount example is the human plasma proteome with its extraordinary dynamic range of more than 10 orders of magnitude in concentration separating albumin from the rarest protein species (Anderson and Anderson, 2002). Hence, the object of investigation in MS-based proteomics is almost exclusively a complex protein sample which may comprise more than 10,000 different proteins in the case of complete mammalian proteomes (Cox and Mann, 2011). Contrary to nucleic acid based techniques, amplification of proteins prior to analysis is not possible so that analytical difficulty arises from the accurate detection of lower abundant proteins in the presence of a large background of higher abundant proteins (Reinders and Sickmann, 2009). In the past two decades, MS-based proteomic technology has evolved from gel-based methodology to data-dependent shotgun proteomics to data-independent and targeted methods (Moulder et al., 2017). A recognizable milestone in proteomic technology was achieved in 2014 with the mapping of the human proteome at ~20,000 proteins resolution (Kim et al., 2014; Wilhelm et al., 2014).

1.2.1 The proteomics workflow

With regards to the nature of the analyte, proteomic workflows distinguish between top-down and bottom-up methodologies. The top-down approach refers to the characterization of intact proteins which are introduced into the gas phase and subsequently fragmented in the mass spectrometer which results in the acquisition of molecular masses of both the protein and its fragment ions (Chait, 2006). Although top-down proteomics has the great potential to achieve a complete understanding of the primary protein structure including all its post-translational modifications, the method is hampered by limitations in protein fractionation, protein ionization and the generation of sufficient gas phase fragmentation of especially large intact proteins (Chait, 2006; Zhang et al., 2013b). Bottom-up proteomics on the other hand describes the analysis of

peptides obtained from commonly a tryptic digest of a protein or protein mixture. When applied to a protein mixture, this bottom-up technology is referred to as shotgun proteomics, in analogy to shotgun genomic sequencing (Yates, 1998; Zhang et al., 2013b). The crucial advantage of bottom-up over top-down approaches is that peptides in contrast to proteins can be much more easily fractionated, ionized and fragmented (Zhang et al., 2013b). In consequence, bottom-up workflows have been universally adopted and proven successful in the application to a wide range of biological systems and therefore represent the state-of-the-art. In practice, the shotgun proteomics workflow includes three main experimental steps including protein extraction and proteolytic digest (sample preparation), peptide separation and ionization, and ultimately peptide identification and quantification. The underlying principle is that first the masses of the intact tryptic peptides (peptide precursors) are determined and secondly these peptide ions are fragmented in the gas phase to produce sequence information (Chait, 2006). The obtained peptide mass and sequence information then allows inference of parent protein identity. An overview of a canonical shotgun proteomics workflow applying electrospray ionization (ESI) is depicted in Fig. 8.
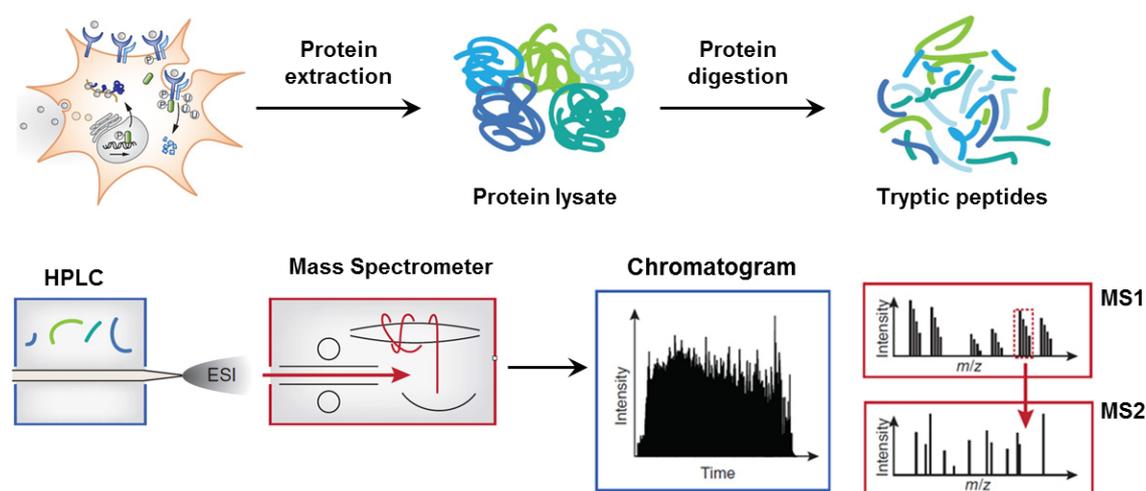
Figure 8 Overview of a shotgun proteomics workflow applying LC-ESI-MS/MS.
The shotgun proteomics workflow encompasses several distinct steps and begins with the protein extraction from tissues, body fluids, cells or subcellular compartments. In the next step, the obtained protein lysate is subjected to proteolytic digest, usually tryptic, to obtain a complex peptide mixture. Peptides are subsequently separated by high performance liquid chromatography (HPLC) and ionized by electrospray ionization (ESI). In the last step, peptide masses (MS1) and peptide fragment masses (MS2) are acquired in the mass spectrometer. At several stages in the workflow additional steps aimed at reducing complexity at the protein or peptide level may be implemented such as subcellular fractionation, multi- or unidimensional protein separation, and peptide pre-fractionation. Model and graphic source: adapted from Meissner and Mann (2014).

1.2.1.1 Sample preparation

Sample preparation — the primary step in a proteomics workflow — is an area of active research which is characterized by its high application dependency (Meissner and Mann, 2014). First, optimal protein extraction from a biological sample needs to be achieved which in case of primary cells or cell lines can be easily facilitated by boiling them in detergent or a chaotropic agent in order to solubilize and denature the proteins. Application of detergent-based protocols may result in better recovery of membrane proteins compared to protocols relying solely on chaotropic salts or organic solvents, yet most detergents are not compatible with the downstream LC-MS/MS application (Meissner and Mann, 2014). An essential step in MS sample preparation also involves the chemical protection of cysteines by reduction and subsequent irreversible alkylation to abrogate disulfide bond formation. Importantly, sample preparation strategies are often aimed at reducing the immense complexity of a biological sample since mass spectrometers are limited in the number of ion signals they are able to resolve in a specific timeframe (Reinders and Sickmann, 2009). Most commonly the initial step in protein pre-fractionation is subcellular fractionation which is closely intertwined with the biological question being addressed enabling specific enrichment of a cellular compartment of interest such as the nucleus, mitochondria or the surface compartment. For instance, isolation of the cell surface compartment (surfaceome) can be performed by biotinylation of sialic acid-containing glycans on living cells using periodate oxidation and aniline-catalyzed oxime ligation (Glyco-PAL) and subsequent enrichment of biotinylated cell surface glycoproteins with streptavidin beads (Zeng et al., 2009). To further reduce sample complexity on the protein level, multidimensional protein separation techniques such as 2D gel electrophoresis or free-flow electrophoresis may be employed. Subsequently, the core step in sample preparation for bottom-up experiments is the proteolytic digest of the protein lysate using endoproteases. Here, most commonly trypsin is used which specifically hydrolyzes peptide bonds on the carboxyl terminal side of the basic arginine (R) and lysine (K) amino acid residues resulting in peptides with an average size of 800-2000 Da which are highly amenable for LC-MS/MS analysis (von Hagen, 2011). After proteolytic digest, sample complexity may further be reduced on the peptide level employing peptide pre-fractionation before the actual LC separation such as strong anion or strong cation exchange chromatography (SAX/SCX), hydrophilic liquid interaction chromatography (HILIC) or high-pH reversed phase (hpH-RP) fractionation. Here, a specific variant is the multidimensional protein identification technology (MudPIT) consisting of a triphasic chromatography column combining a RP-precolumn with a SCX-precolumn, followed by the main RP-separation

column which is directly coupled to the mass spectrometer (Washburn et al., 2001). Depending on the biological question, sample preparation may also include the enrichment of specific subpopulations of peptides, for example PTM-containing peptides. A crucial drawback of the outlined sample preparation strategies (e.g. subcellular fractionation, peptide pre-fractionation or PTM enrichment) is their dependency on sufficient amounts of high quality input sample. The spectrum of applicable sample preparation strategies greatly diminishes when working with samples limited in quality and quantity such as biobanked tissues, cells or biofluids. Here, the operator needs to balance the need for reduction of sample complexity with limiting unnecessary steps potentially leading to sample loss, while at the same time maximizing downstream sample quality. In these cases, several protocols which greatly simplify MS sample preparation by spatially combining multiple integral steps can be exploited. One example is filter-aided sample preparation (FASP) in which protein lysates are retained on a spin filter device by size exclusion, subsequently enabling extensive sample purification by filter washing prior to on-filter tryptic digest (Manza et al., 2005; Wisniewski et al., 2009). The resulting tryptic peptides can be captured by centrifugation, do not require further desalting steps and can be readily loaded onto an HPLC system. More recently, Kulak et al. (2014) described a novel method called in-StageTip (iST) which incorporates all necessary steps of sample preparation including cell lysis, protein denaturation, reduction and alkylation, proteolytic digest and peptide-clean up in a single-pot reaction. These self-built StageTips can be equipped with the chromatographic material of choice, thereby further integrating peptide-based pre-fractionation.

1.2.1.2 Liquid chromatography tandem mass spectrometry (LC-MS/MS)

After successful sample preparation the obtained (tryptic) peptide mixture is separated depending on hydrophobicity using an HPLC system equipped with a C18 packed stationary phase (RP mode). The two main techniques for volatizing and ionizing peptides are ESI and matrix-assisted laser desorption/ionization (MALDI) (Fenn et al., 1989; Karas and Hillenkamp, 1988). While MALDI sublimates and ionizes the peptides out of crystalline matrix using laser pulses, ESI is based on ionizing the peptides out of a solution and can thus be readily coupled to liquid-based separation tools (e.g. HPLC) (Aebersold and Mann, 2003). LC-ESI-MS/MS is the preferred method for analyzing complex samples and both mass spectrometers used for analytics in this thesis relied on just this technology (LTQ Orbitrap XL and Q Exactive HF, Fig. 10 and 11). In ESI, the analyte solution eluting from the chromatography column is pumped through a needle at the inlet of the mass spectrometer to which a high voltage is applied. The liquid then forms a Taylor cone with an excess of charge on its surface resulting from the electric

field gradient between the ESI needle and the counter electrode (Reinders and Sickmann, 2009). Subsequently, small droplets with an excess of charge will detach from the tip of the Taylor cone when Coulombic repulsion of the surface charge equals the surface tension of the solution (Rayleigh limit) (Taflin et al., 1989). These highly charged droplets evaporate as they move toward the entrance of the mass spectrometer eventually producing single molecular ions which can be analyzed for their m/z ratio (Reinders and Sickmann, 2009). There are two major explanations as to why the ESI process ultimately leads to free, charged analyte molecules: the first theory assumes that solvent evaporation results in increased charge density which leads to the formation of smaller and smaller droplets that eventually consist of single ions (Coulomb fission or explosion) (Dole et al., 1968). The second theory suggests that increased charge which results from solvent evaporation leads to the release of ions from the droplet surfaces (ion evaporation) (Iribarne and Thomson, 1976).

After introduction of peptide ions into the gas phase, conventional MS/MS data acquisition occurs in two distinct stages. First, a full scan of all peptide precursors present in the mass spectrometer in a given time window is performed to obtain an MS1 spectrum. For recording of the MS/MS spectra (also called MS2), peptide ions are selected for fragmentation by collision-induced dissociation (CID) based on either their intensity (data-dependent acquisition, DDA) or a predefined mass range (data-independent acquisition, DIA) (Reinders and Sickmann, 2009; Venable et al., 2004). Fragmentation is triggered by interaction with collision gas, usually nitrogen or argon, and primarily occurs along the peptide backbone giving rise to distinct sequence ions (Fig. 9) (Hunt et al., 1981; Hunt et al., 1986). Peptide identity, i.e. the amino acid sequence, is subsequently inferred with the help of software tools by matching the MS/MS spectra obtained from the peptide fragmentation with theoretical MS/MS spectra computed by *in silico* digestion of a protein database (common for DDA) or by assigning the measured MS/MS spectra to experimental MS/MS spectra contained in a peptide spectral library (common for DIA). The former is referred to as peptide-to-spectrum (PSM) matching; the latter is called spectrum-to-spectrum matching.
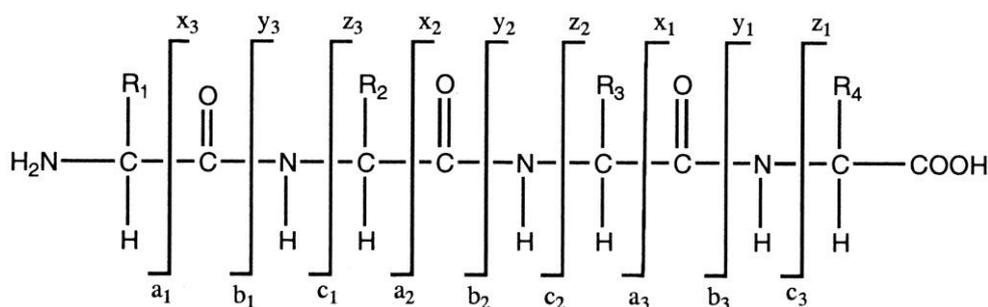
Figure 9 Nomenclature of peptide fragment ions.
The nomenclature for peptide fragment ions was first proposed by Roepstorff and Fohlman (1984). Peptide fragment ions are primarily distinguished by the position of the charge which may be retained on the N-terminus (a, b and c sequence ions) or the C-terminus (x, y and z sequence ions). Peptide sequence ions are further differentiated based on the exact location of the bond cleavage event occurring either before (a, x), in (b, y) or after (c, z) the peptide bond. The subscript numbers indicates the number of amino acid residues contained in the fragment ion. Graphic source: Graves and Haystead (2002).

In order to acquire both MS1 and MS2 spectra, mass spectrometers optimally need to be equipped with two synchronized mass analyzers which function in separating ions based on their m/z ratios. In the field of proteomics, there are four main types of mass analyzers including time-of-flight (TOF), quadrupole, ion trap and the Fourier transform ion cyclotron, each of which exhibits distinct characteristics regarding their key parameters sensitivity, resolution and mass accuracy (Aebersold and Mann, 2003). For the MS analyses conducted in this thesis two hybrid instruments were used; initially LTQ Orbitrap XL machinery (Fig. 10) and at later stages also the next-generation Q Exactive HF technology (Fig. 11). The LTQ Orbitrap XL contains a linear ion trap interfaced with an orbital trap (Orbitrap) mass analyzer, whereas the Q Exactive HF instrument combines dual quadrupole-Orbitrap technology. The core technology in both machines is the Orbitrap which is characterized by its high resolving power and excellent mass accuracy (Perry et al., 2008). The Orbitrap analyzer consists of an outer barrel-like electrode and a central spindle-like electrode which operate by radially trapping ions in an electrostatic field (Hu et al., 2005; Makarov, 2000). Ions are separated by different frequencies of oscillation around the spindle electrode and in-between the outer electrodes which relate to their m/z ratio, corresponding mass spectra are extracted by fast Fourier transformation of the frequency signal (Hu et al., 2005). In the LTQ Orbitrap XL instrument high mass accuracy measurements of precursor peptide ions (MS1) are recorded in the Orbitrap, while precursor selection, fragmentation and MS2 spectra recording occur in the linear ion trap (see also Fig. 10). The linear ion trap operates by confining ions radially and axially by using a 2D radio frequency field and by applying stopping potentials to end electrodes, respectively (Douglas et al., 2005).
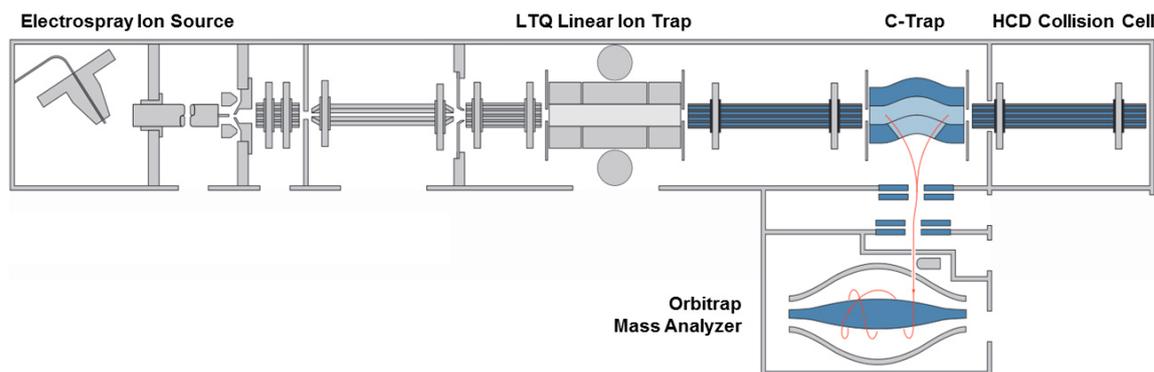
Figure 10 Schematic illustration of Thermo Fisher LTQ Orbitrap XL™ technology.
The Orbitrap XL mass spectrometer combines a linear ion trap with an Orbitrap mass analyzer. Peptides entering the MS from the HPLC system are ionized by electrospray ionization (ESI), collected in the linear ion trap, followed by ejection into the C-shaped storage trap (C-Trap). From the C-Trap ions are transferred into the orbital trap (Orbitrap) where they are captured by rapidly increasing the electric field. Signals from the outer electrodes of the Orbitrap are amplified and transformed into a frequency spectrum by fast Fourier transformation which is converted into a mass spectrum (MS1). Precursor ions can be selected in the linear ion trap and subjected to fragmentation by either collision-induced dissociation (CID) or higher energy collisional dissociation (HCD) in a separate collision cell. MS/MS spectra are recorded in the linear ion trap. Graphic and model source: www.planetorbitrap.com. LTQ, linear trap quadrupole.

The Q Exactive HF instrument on the other hand was launched almost a decade after first commercial introduction of Orbitrap technology. The Q Exactive HF machine is notable for an ultra-high field Orbitrap mass analyzer characterized by much faster scan speed and higher resolution than older generation instruments; the device is further equipped with an advanced quadrupole technology enabling improved precursor selection and transmission (Scheltema et al., 2014). Generally, quadrupole mass filters separate ions by applying time-varying electric fields between four metal rods permitting stable trajectories only for ions of desired m/z (Aebersold and Mann, 2003). Besides conventional DDA workflows, the Q Exactive HF also opens the door to targeted measurements and all-ion fragmentation workflows.
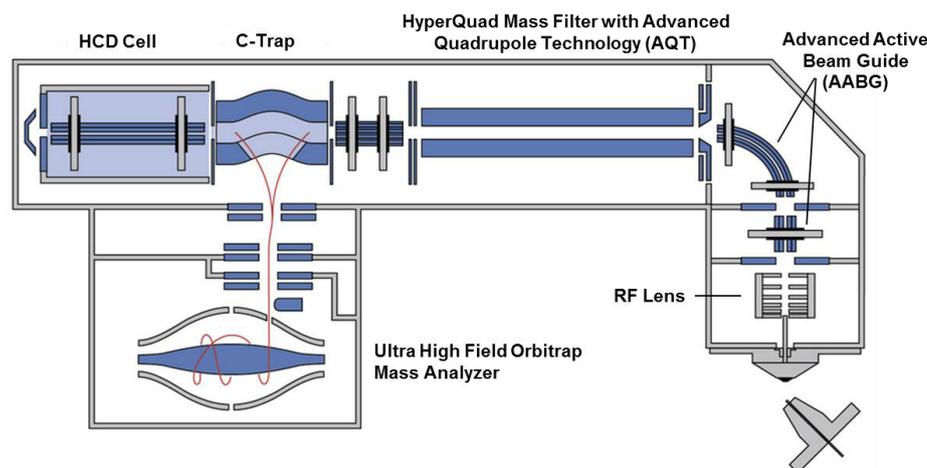
Figure 11 Schematic illustration of Thermo Fisher Q Exactive HF™ technology.
HPLC separated peptides enter the Q Exactive HF mass spectrometer by electrospray ionization (ESI). The Q Exactive HF setup comprises a mass selection pre-filter for removal of undesirable ions, a segmented quadrupole mass filter, a C-Trap, a higher energy collisional dissociation (HCD) cell, and ultimately an ultra-high field (HF) Orbitrap mass analyzer. For acquisition of full scan MS1 spectra, precursor ions pass through the quadrupole into the C-Trap where they are stabilized, followed by ejection into the Orbitrap for detection. For recording of MS2 spectra, quadrupole filtered precursor ions are fragmented by collision-induced dissociation in the HCD cell and resulting fragment ions are collected in the C-Trap and detected in the Orbitrap. Graphic source: www.planetorbitrap.com.

1.2.1.3 Data acquisition strategies

In traditional shotgun proteomics, the mass spectrometer is operated in DDA mode in which the system continuously acquires series of full mass range survey scans (MS1) followed by subordinated scans of fragment ion spectra (MS2) for selected precursor ions detectable in the directly preceding survey scan (Fig. 12) (Domon and Aebersold, 2010). The process directing peptide precursor selection is fully automated and solely based on signal intensity which means that the mass spectrometer randomly samples peptides for fragmentation being biased to pick those with the strongest signal (Doerr, 2015). Already fragmented peptide masses are temporarily stored in an exclusion list to avoid their repeated sequencing in consecutive scans, a principle called dynamic exclusion. To date, most instruments can perform a DDA cycle with one survey scan (MS1) and ten subordinated MS2 scans in two seconds (Hu et al., 2016). Operating the instrument in DDA mode typically yields thousands of protein identifications, yet the stochastic and irreproducible nature of precursor selection diminishes reproducibility and prevents the measurement of lower abundant peptides (Gillet et al., 2012; Liu et al., 2004). This is in particular relevant for very complex samples in which many peptides co-elute and thus occur together in a survey scan; here DDA would fragment and record only the most abundant ones (usually a Top 10 method), omitting the rest, a phenomenon which has become known as "undersampling" or "missing value problem"

(Michalski et al., 2011). Limited reproducibility of the DDA mode can be mainly attributed to the fact that even in repeated measurements of the same sample a different population of precursors will be selected for fragmentation (Tabb et al., 2010). Additionally, the exact time point of MS2 acquisition within the visibility window of a specific precursor is variable resulting in differences in MS2 spectra intensities (Lehmann, 2010). As a result, the quantitative comparison between MS2 spectra in independent LC-MS/MS runs is difficult. Moreover, owing to the dynamic exclusion strategy each peptide species is deliberately sampled only once or twice which further prevents precise MS2-based quantification which would require multiple measurements per peptide (Hu et al., 2016). Despite its drawbacks, the DDA strategy is flexible, enables a breadth of detection, and is rather simple in its setup and post-analysis; these key factors have made it the preferred method across a large scientific community (Hu et al., 2016).
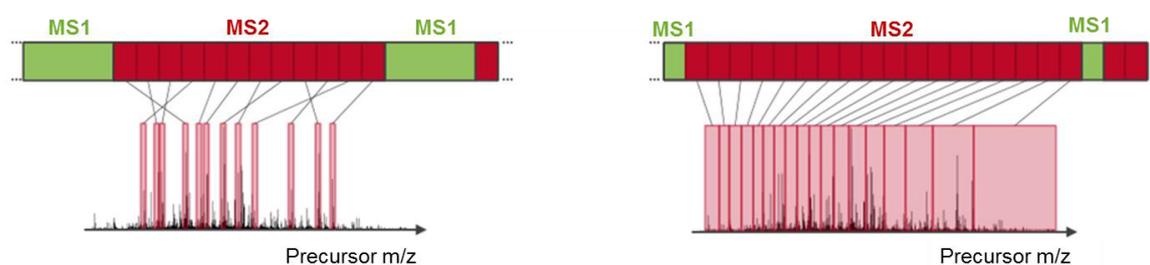
Figure 12 Data-dependent versus data-independent acquisition.
(Left) In data-dependent acquisition (DDA) LC-MS/MS the instrument first records a full scan of all ion masses (m/z) introduced in the mass spectrometer at a given time point, referred to as survey scan (MS1). The machine then sequentially isolates the Top *N* most intensive peptide precursors, fragments them and records the corresponding fragment ion spectra (MS2). This process, also called a cycle, is repeated over and over again applying the dynamic exclusion principle which puts the mass of an already fragmented precursor temporarily into an exclusion list. (Right) In data-independent acquisition (DIA) LC-MS/MS the instrument operates in an unbiased mode, neither requiring knowledge nor detection of the precursor ion to be fragmented. Here, a survey scan is followed by recordings of fragment ion spectra for all precursors located in a predefined mass range, also termed isolation window. As a result, precursor ions spanning the entire mass range will be selected for fragmentation and resulting fragment ion spectra will be recorded. MS1, survey scan of the entire m/z range; MS2, fragment ion spectra. Graphic source: reprinted with permission from Biognosys AG, www.biognosys.com.

To circumvent the limitations of the DDA mode, strategies which do not require prior detection of the peptide precursor to be selected for fragmentation have gained increasing attention. These workflows operate in the DIA mode, also referred to as all-ion fragmentation, in which all peptides within a predefined m/z window are subjected to fragmentation (Venable et al., 2004). In other words, the machine sequentially fragments all precursor ions introduced to the mass spectrometer by scanning the entire mass range step-by-step in predetermined asymmetrical isolation windows of distinct sizes in-

between MS1 survey scans (Fig. 12). As a result, the DIA procedure guarantees that all peptide precursors present within the selected mass range will be sampled, theoretically allowing identification of all sufficiently abundant peptides if the corresponding spectra can be properly interpreted (Bilbao et al., 2015; Hu et al., 2016). One drawback of the DIA mode is that the link between the fragment ions and the originating parent ion is lost and that the recorded MS2 spectra are highly complex containing fragment ions from mixtures of precursor peptides, thus complicating downstream data analysis (Gillet et al., 2012). Fortunately, recent advances in bioinformatics software development have helped to overcome these post-analysis difficulties (Bruderer et al., 2015; Bruderer et al., 2016; Navarro et al., 2016). Another limitation of the DIA workflow is its limited precision in measuring very low abundant peptides which is most possibly attributed to signal dwarfing by more abundant co-eluting peptides (Hu et al., 2016).

With regards to post-acquisition data analysis, as previously outlined, in the DDA mode MS2 spectra are matched *in silico* to peptide sequences in a database on the basis of the observed and expected fragment ions (Steen and Mann, 2004). There are several distinct algorithms which may be implemented to query sequence databases including Mascot, Sequest, Byonic, MS Amanda or Andromeda (Bern et al., 2012; Cox et al., 2011; Dorfer et al., 2014; Eng et al., 1994; Perkins et al., 1999). In the next step, the assigned peptide sequences are statistically validated using for instance decoy search strategies in which MS2 spectra are competitively matched against random databases to estimate the false discovery rate (FDR) (Altelaar et al., 2013; Elias and Gygi, 2007; Kall et al., 2007a; Nesvizhskii et al., 2007). In the last step, protein inference is achieved by assigning the validated peptide sequence to proteins, either uniquely to a single protein or shared to multiple proteins (Zhang et al., 2013b). For the DIA workflow, peptide identification requires the comparison of the obtained DIA spectra with sets of annotated PSMs from previous DDA experiments contained in spectral libraries which reflect accurate, empirically determined fragmentation patterns for each peptide in the library (Hu et al., 2016). However, if the goal is to detect novel proteins or very low abundant proteins, the application of spectral library search has its limitations since spectrum-to-spectrum matching may only occur if the spectral library was populated with high quality experimental DDA spectra. Fortunately, the availability of high quality DDA spectra as well as ready-to-use comprehensive spectral libraries via public repositories has expanded over the last years (Rosenberger et al., 2014). A specific variant for DIA data analysis was recently introduced by Tsou et al. (2015) describing the open-source Umpire computational workflow. The DIA-Umpire detects precursor and fragment chromatographic features and assembles them into pseudo-MS2 spectra which are

subsequently matched by conventional database searching and protein inference tools enabling untargeted, spectral library-free analysis of DIA data (Tsou et al., 2015; Tsou et al., 2016).

1.2.1.4 Quantification strategies

Discovery proteomics is almost exclusively aimed at comparing protein expression in a biological system under two or more experimental conditions. Therefore, the bottom-up proteomics workflow does not only require a comprehensive qualitative analysis but also the implementation of an elaborate strategy for quantitative comparison. The implementation of quantitative strategies in bottom-up proteomics has been an active field of research over the past 15 years which has yielded a myriad of different solutions to the problem (Fig. 13). Generally, it can be distinguished between relative and absolute quantification, the latter aiming to precisely determine the amount of the substance in question, e.g. ng/ml of a biomarker or the copy number of a protein per cell (Ong and Mann, 2005). Absolute quantification relies on stable isotope-labeled peptides such as AQUA or QconCAT which are spiked into the sample in known concentrations as internal standard, i.e. the sequence of the peptide(s) of interest needs to be known beforehand (Gerber et al., 2003; Pratt et al., 2006). While absolute quantification is for example important for biomarker measurement, it is not applicable in discovery proteomics where thousands of proteins are monitored. Here, relative quantification is the method of choice determining the relative change in protein amount between two states usually indicated by fold change of protein or peptide abundance (Ong and Mann, 2005). In relative quantification, two primary concepts can be differentiated which rely either on the introduction of stable isotope labels during different experimental stages or the label-free approach which exploits advanced software solutions (Fig. 13). In isotopic labeling approaches, the proteomes which are to be compared are labeled with different isotopes (e.g. ^{13}C, ^{15}N and ^{2}H) and combined prior to analysis by LC-MS/MS resulting in the measurement of peptides with slightly different masses which can be differentiated based on their m/z values (MS1). In a specific variant of isotopic labeling different isobaric tags are incorporated into the proteomes of interest resulting in the measurement of identical m/z values on the precursor level (MS1); however these precursors give rise to distinct reporter ions in the fragment spectra (MS2) indicative of relative abundance. When working with primary cells or cell cultures, isotopic labels may be introduced metabolically by using heavy or light isotope-containing amino acids in growth media for incorporation into all proteins (SILAC) (Ong et al., 2002). Chemical labeling on the other hand may introduce isotopic labels either on the protein or peptide level. Two prominent strategies for label introduction at the protein level are ICAT

(isotype coded affinity tag) exploiting the reactivity of cysteine residues and ICPL (isotope coded protein label), a chemical tag attached to lysine residues (Gygi et al., 1999; Schmidt et al., 2005). Alternatively, chemical labeling can also be initiated at the peptide level applying isobaric labeling such as TMT (tandem mass tag) or ITRAQ (isobaric tags for relative and absolute quantitation) (Ross et al., 2004; Thompson et al., 2003). Generally, the earlier isotope labeling is introduced into the proteomics workflow, the more accurate will be the quantitative results due to fewer upstream steps which may introduce experimental variability (Bantscheff et al., 2007). A major advantage of isotopic labeling is the mixing of samples which may reduce experimental process variation if introduced early and further immensely saves MS machine time. Moreover, isotopic labeling strategies exhibit a higher quantitative precision and accuracy compared to label-free approaches. On the other hand, drawbacks of labeling are the rather high costs and the strong limitations concerning multiplexing capacity. Currently, using commercially available isobaric labels no more than 10 different experimental conditions can be compared (TMT 10-plex) precluding proteomic analyses of large study cohorts without sample pooling.
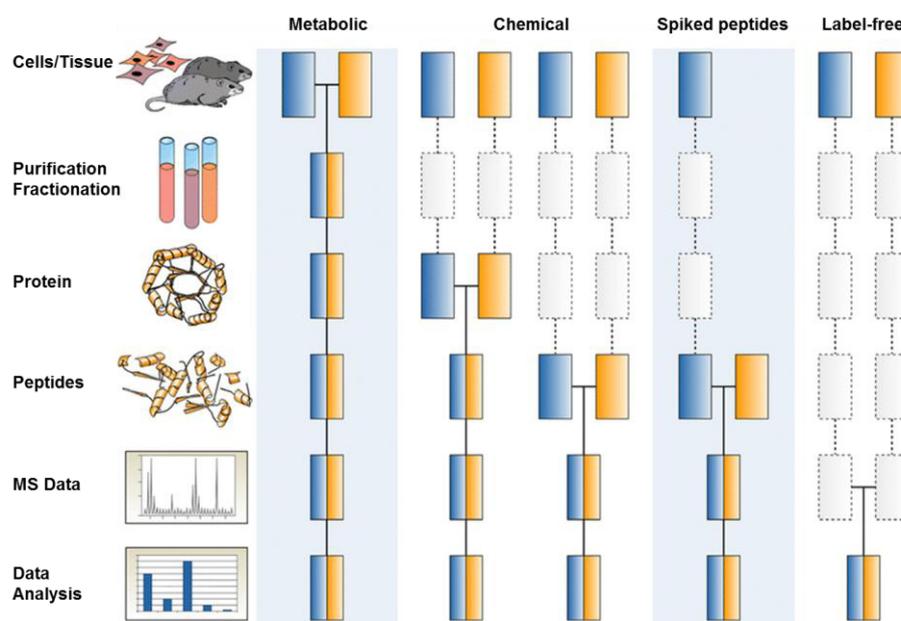
Figure 13 Quantitative strategies in the bottom-up proteomics workflow.
The chart illustrates the spectrum of labeling strategies in a typical bottom-up proteomics workflow aimed at quantitatively comparing two experimental conditions (blue and yellow). Quantification strategies range from metabolic or chemical labeling to spike-in standards and label-free approaches. The horizontal lines define at which time point of the workflow samples are combined, while the dashed lines indicate stages which may contribute to quantification errors due to possible experimental variation during parallel processing. Model and graphic source: adapted from Bantscheff et al. (2007) and Ong and Mann (2005).

The alternative to isotopic labeling is label-free quantification, which does not include any form of sample mixing, rather relying on comparison of individually recorded spectra using advanced software solutions. If the spectra are recorded in DDA mode, there are two major approaches for label-free quantification based either on comparison of extracted ion chromatograms (XIC) of the survey scan (MS1) or spectral counting (Old et al., 2005). The latter is based on the empirical observation that the number of MS2 spectra recorded for a specific peptide (i.e. PSMs) is correlated to its abundance in the sample (Liu et al., 2004). Hence, relative quantification can be achieved by comparing the number of such spectra between different sample runs. In MS1-based quantification, first the intensity signal of the peptide eluting from the chromatographic column is plotted over time and secondly the area under the peak is extracted (i.e. the XIC) which is linearly related to its amount and can be compared across experimental conditions (Ong and Mann, 2005). Thus, integrating ion intensities of the same peptide precursor measured across multiple experimental conditions can be exploited for determining relative abundance. Relative changes in protein abundance may then be inferred from the collective quantitative differences of all assigned peptides or a limited subset of assigned peptides (e.g. Top N method or proteotypic peptides). The MS1-based strategy requires highly sophisticated software which typically first carries out mass calibration, noise and data reduction, feature detection and generation of peptide elution profiles (Bantscheff et al., 2012). The software then performs the alignment of peptide features across sample runs based on retention time and m/z values as well as cross-run intensity normalization (Bantscheff et al., 2012). The MS1-based strategy was applied for all DDA experiments in this thesis employing the commercial Progenesis QI software platform, other freely available tools include Maxquant or OpenMS (Cox and Mann, 2008; Kohlbacher et al., 2007). Among the key advantages of using MS1-based quantification in DDA studies is its cost- and work-effectiveness since no labels need to be introduced. Secondly, an important asset is founded in its applicability to large sample sets; the number of experimental conditions to be compared is not restricted by a predefined number of different isotope labels. Another important aspect is the alleviation of the previously described "missing value problem" by matching peptide features across sample runs, hereby transferring peptide identifications in-between samples. As previously outlined, the dynamic exclusion strategy minimizes repeated fragmentation events for the same peptide precursor, thus the MS1-based strategy generates much more accurate results than spectral counting. Important prerequisites for MS1-based quantification are highly reproducible sample preparation and robust HPLC performance characterized by high retention time stability to achieve high alignment quality. In summary, label-free quantification is the method of choice when working with large

sample sets. This is valid not only for experiments conducted in the DDA mode but also to DIA quantification. Although the DIA workflow was introduced more than a decade ago, interest has only been rekindled in the recent years, so that the development and optimization of software tools is currently an active field of research (Navarro et al., 2016; Venable et al., 2004). Software solutions which have implemented label-free quantification in the DIA workflow include Spectronaut, OpenSWATH, SWATH2.0, Skyline and the DIA-Umpire (Bruderer et al., 2015; MacLean et al., 2010; Rost et al., 2014a; Tsou et al., 2015). In this thesis the Spectronaut workflow was applied for DIA experiments, here the operator can choose between XIC-based label-free quantification of precursor (MS1) or fragment ion intensities (MS2) comparing either the area under the peak or peak height across sample runs.

1.3 Aim of this study

T1D is an organ-specific autoimmune disorder often diagnosed in early childhood which despite decades of research continues to be incurable. T1D pathology is characterized by its heterogeneity involving a complex interplay between genetic susceptibility, environmental influences and aberrant immune responses. Yet, the precise mechanisms underlying initiation of islet autoimmunity and onset of clinical disease remain elusive. To date, the only available biomarker indicative of ongoing ß-cell autoimmunity are serum-borne islet autoantibodies. However, islet autoantibodies are not believed to play a direct pathologic role and thus convey little insights about immunopathology. Consequently, there is an unmet need for the discovery of novel immune biomarkers which reflect the underlying disease processes. The application of systemwide omics technologies to gain a holistic understanding of T1D disease pathology is still in its infancy. For instance, the proteomic fingerprint of peripheral immune cells in T1D is missing due to a lack of MS-based studies.

This study is aimed at proteomic profiling of biobanked PBMC specimen in recent-onset T1D using quantitative label-free LC-MS/MS to provide novel insights into cellular pathology. The overall study aims were to

- develop a sensitive proteomic method to profile low numbers of PBMCs,
- increase sensitivity by exploring different mass spectrometric acquisition methods,
- establish protein expression signatures in peripheral CD4+ T cells and CD4-depleted cells from patients with recent-onset T1D,
- compare PBMC proteome dynamics with serum signatures in recent-onset T1D,
- identify a (circulating) protein marker indicative of disease development and/or ß-cell destruction.

Primarily, proteins represent the functionally most diverse class of biomolecules and determination of their relative abundance holds the promise to establish disease-specific signatures at the effector level. Secondly, biobanked PBMCs are the exploratory tool of choice being (i) immediately available for investigation and (ii) exhibiting the necessary prerequisites for monitoring of a potential prospective biomarker. Specifically, PBMCs are highly accessible by minimal-invasive venipuncture and thus represent an inexpensive and widely-available surrogate for biopsy material. PBMC specimen harbor a wide array of different immune cells implicated in T1D disease processes — among them circulating autoreactive diabetogenic T cells (Monti et al., 2007). A specific focus of

this investigation will be the CD4+ T cell population — a key player in T1D pathogenesis — which upon activation follows distinct differentiation fates profoundly affecting downstream effector function by cytokine production, immune cell crosstalk and migration (Haskins and Cooke, 2011).

Figure 14 The experimental strategy employed in this thesis is divided into three stages. First, a sample preparation method was established which needed to be applicable to small cell numbers expected in biobank samples. Secondly, the established protocol was transferred to a first set of patient-derived PBMC samples. Lastly, identified proteomic signatures were validated in a different study population with improved LC-MS/MS methods (DIA), additional sample types (serum) and independent methods (immunoblotting).

The experimental strategy of this thesis with the ultimate goal of portraying the dynamic proteomic changes underlying T1D onset in peripheral CD4+ T cells and CD4-depleted cells can be subdivided into three distinct stages (Fig. 14):

1. Development of a sample preparation method applicable to biobanked samples

Different sample preparation methods were compared including the surfaceome profiling approach Glyco-PAL and whole-proteome profiling strategies. These experiments were conducted with limited numbers of freshly isolated CD4+ T cells and PBMCs from healthy donors to simulate the expected low cell count in biobank samples. A suitable standard operating procedure (SOP) was chosen based on sample requirements and LC-MS/MS performance on Orbitrap XL machinery indicated by high sensitivity and high reproducibility. Purity of the CD4+ T cell isolation process was evaluated, too. During this stage no next-generation mass spectrometer such as the Q Exactive HF was available.

2. Proteomic profiling of CD4+ T cells and CD4-depleted cells in sample set I

Processing of the first biobanked PBMC sample set derived from 23 patients with newly diagnosed T1D and 31 age- and sex-matched healthy control children was performed. To this end, each PBMC sample was separated into the CD4+ T cell subset and the CD4-depleted cell fraction. Both cell fractions were prepared with the chosen SOP (iST method), profiled by label-free quantitative LC-MS/MS and analyzed for differential protein signatures. During this stage two next-generation mass spectrometers (Q Exactive HF) were purchased which tremendously expanded the proteomic possibilities of this thesis.

3. Validation of identified protein signatures using next-generation LC-MS/MS

The next aim was to validate the observed protein signatures from sample set I in a second study population of recent-onset T1D (n=30 pairs case-control). Here, the aforementioned purchase of Q Exactive HF technology changed the scope of the technical possibilities. The second patient-derived PBMC sample set was therefore profiled using DDA and DIA LC-MS/MS on the Q Exactive HF to (i) thoroughly compare both methods and (ii) to extract differential protein expression in a situation of in-depth proteome coverage. Additionally, a set of 100 serum samples was profiled using the novel DIA LC-MS/MS approach to confirm characteristics of the observed cellular signature and to discover novel circulatory signatures. Lastly, several protein candidates associated with T1D onset were validated in pooled CD4+ T cell and CD4-depleted cell lysates using immunoblotting.

2 Material and Methods

2.1 Material

2.1.1 Chemicals

Chemical	Manufacturer
Acetonitrile (ACN), HPLC grade	Sigma Aldrich, Steinheim, Germany
Aminooxy-biotin	Biotium, Fremont, USA
Ammonium bicarbonate (ABC)	Sigma Aldrich, Steinheim, Germany
Ammonium formate (NH$_4$HCO$_2$)	Sigma Aldrich, Steinheim, Germany
Ammonium hydroxide (NH$_4$OH)	Sigma Aldrich, Steinheim, Germany
Aniline	Sigma Aldrich, Steinheim, Germany
ß-Mercaptoethanol	Sigma Aldrich, Steinheim, Germany
Blotting grade blocker	Bio-Rad, Munich, Germany
Bromphenol blue	Merck Millipore, Darmstadt, Germany
Bovine serum albumin (BSA), Fraction V	Biomol, Hamburg, Germany
Chloracetamide	Sigma Aldrich, Steinheim, Germany
Complete protease inhibitor	Roche Diagnostics, Mannheim, Germany
Diemethyl sulfoxide (DMSO)	Sigma Aldrich, Steinheim, Germany
Dithiotreitol (DTT)	Merck Millipore, Darmstadt, Germany
Ethanol	Merck Millipore, Darmstadt, Germany
Ethylenediaminetetraacetic acid (EDTA)	Merck Millipore, Darmstadt, Germany
Formic acid	Sigma Aldrich, Steinheim, Germany
Glycerol	Merck Millipore, Darmstadt, Germany
Guanidine hydrochloride (GdmCl)	Sigma Aldrich, Steinheim, Germany
Hydrochloric acid (HCl)	Merck Millipore, Darmstadt, Germany
H$_2$O, HPLC grade	Merck Millipore, Darmstadt, Germany
Iodacetamid (IAA)	Merck Millipore, Darmstadt, Germany
Isopropanol	Carl Roth, Karlsuhe, Germany
Methanol	Merck Millipore, Darmstadt, Germany
NP-40	Roche Diagnostics, Mannheim, Germany
Paraformaldehyde	Sigma Aldrich, Steinheim, Germany
Sodium azide (NaN$_3$)	Sigma Aldrich, Steinheim, Germany
Sodium carbonate (Na$_2$CO$_3$)	Sigma Aldrich, Steinheim, Germany
Sodium dodecyl sulfate (SDS)	Sigma Aldrich, Steinheim, Germany
Sodium chloride (NaCl)	Sigma Aldrich, Steinheim, Germany
Sodium hydroxide (NaOH)	Applichem, Darmstadt, Germany
Sodium periodate (NaIO$_4$)	Merck Millipore, Darmstadt, Germany

Chemical	Manufacturer
Strep Tactin Supeflow	IBA, Göttingen, Germany
Trifluoroacetic acid (TFA)	Applied Biosystems, Foster City, USA
Tris(2-carboxyethyl)phosphine (TCEP)	Sigma Aldrich, Steinheim, Germany
Tris(hydroxymethyl)aminomethane (Tris)	Carl Roth, Karlsruhe, Germany
Tween 20	Serva, Heidelberg, Germany
Urea	Sigma Aldrich, Steinheim, Germany

2.1.2 Consumables

Consumable	Manufacturer
Cell counting chamber	NanoEnTek, Seoul, South Korea
Cell strainer, 70 µm	Corning, New York, USA
Empore™ SDB-RPS disks	3M Bioanalytical Technologies, St. Paul, USA
Falcon conical tubes	Corning, New York, USA
Mini-Protean TGX™ precast gels 4-15%	Bio-Rad, Munich, Germany
MS/LS columns	Miltenyi Biotec, Bergisch Gladbach, Germany
Parafilm, "M"	Bemis, Neenah, USA
Pipette tips	Eppendorf, Hamburg, Germany
Pre-separation filters, 30 µm	Miltenyi Biotec, Bergisch Gladbach, Germany
Protein LoBind Tubes	Eppendorf, Hamburg, Germany
PVDF membrane, Hybond-P	Amersham, Little Chalford, UK
Serological pipettes	Greiner, Frickenhausen, Germany
Twin.tec PCR Plates 96 LoBind, semi-skirted	Eppendorf, Hamburg, Germany
Vacutainer (Na-Heparin)	Becton Dickinson, Franklin Lakes, USA
30 kDa centrifugal filter device, Vivacon 500	Sartorius, Göttingen, Germany
96 well plates	Corning, New York, USA

2.1.3 Buffers, cell culture media and reagents

Buffer/Medium/Reagent	Manufacturer
G7 buffer	New England Biolabs, Ipswich, USA
Hank's Balanced Salt Solution (HBSS)	Thermo Fisher Scientific, Waltham, USA
Lymphoprep™	Stemcell Technologies, Vancouver, Canada
MACS buffer	Miltenyi Biotec, Bergisch Gladbach, Germany
Phosphate-buffered saline (PBS) no calcium, no magnesium	Thermo Fisher Scientific, Waltham, USA
RPMI Medium 1640 GlutaMAX™	Thermo Fisher Scientific, Waltham, USA

Buffer/Medium/Reagent	Manufacturer
Tris/Glycine/SDS (TGS)	Bio-Rad, Munich, Germany
Trypan blue solution	Sigma Aldrich, Steinheim, Germany
X-Vivo™ 15 medium	Lonza, Basel, Switzerland

2.1.4 Enzymes

Enzyme	Manufacturer
Lysyl Endopeptidase (Lys-C), MS grade	Wako, Osaka, Japan
Peptide-N-Glycosidase F (PNGase F)	New England Biolabs, Ipswich, USA
Trypsin, MS grade	Promega, Madison, USA

2.1.5 Kits and Standards

Kit/Standard	Manufacturer
Amersham ECL Select Western Blotting Detection Reagent	GE Healthcare, Little Chalfont, UK
Bovine serum albumin (BSA) standard	New England Biolabs, Ipswich, USA
CD4 MicroBeads, human	Miltenyi Biotec, Bergisch Gladbach, Germany
CD4+ T cell isolation kit II, human	Miltenyi Biotec, Bergisch Gladbach, Germany
HRM Calibration Kit	Biognosys, Schlieren, Switzerland
PageRuler™ Prestained Protein Ladder	Thermo Fisher Scientific, Waltham, USA
Pierce™ BCA Protein Assay Kit	Thermo Fisher Scientific, Waltham, USA
Pierce™ SDS-PAGE Sample Prep Kit	Thermo Fisher Scientific, Waltham, USA
PreOmics iST Kit	PreOmics, Planegg, Germany

2.1.6 Laboratory and Analytical equipment

Instrument	Manufacturer
Acquity UPLC M-Class HSS T3 Column	Waters, Milford, USA
BD LSRFortessa™	Becton Dickinson, Franklin Lakes, USA
Centrifuge Mikro 200R	Hettich, Tuttlingen, Germany
Centrifuge PerfectSpin 24R	Peqlab, Erlangen, Germany
Centrifuge 5424	Eppendorf, Hamburg, Germany
Digital Developer, Fusion Fx7	Vilber Lourmat, Eberhardzell, Germany
Incubator Heracell 150i	Thermo Fisher Scientific, Waltham, USA
Laminar flow hood, Herasafe KS	Thermo Fisher Scientific, Waltham, USA
Magnetic stirrer	IKA Labortechnik, Staufen, Germany

Instrument	Manufacturer
Mass spectrometer Orbitrap XL	Thermo Fisher Scientific, Waltham, USA
Mass spectrometer Q Exactive HF	Thermo Fisher Scientific, Waltham, USA
Microcentrifuge	Starlab, Hamburg, Germany
Microscope, Stemi DV4	Carl Zeiss Microscopy, Munich, Germany
Mini-Protean Tetra Electrophoresis Cell	Bio-Rad, Munich, Germany
Mini Trans-Blot Cell	Bio-Rad, Munich, Germany
Multifuge X3R	Thermo Fisher Scientific, Waltham, USA
Nano trap column	LC Packings, Sunnyvale, USA
PepMap100 C18 HPLC column	LC Packings, Sunnyvale, USA
pH Meter FE20 FiveEasy	Mettler Toledo, Columbus, USA
Pipette 1 ml, 200 µl, 20 µl, 10 µl	Gilson, Middleton, USA
Power supply, Power Pac 300	Bio-Rad, Munich, Germany
Rotator SB3	Cole-Parmer, Vernon Hills, USA
RSLC Ulitmate 3000	Dionex, Idstein, Germany
Spectral photometer EMax plus	Molecular Devices, Sunnyvale, USA
SpeedVac, Concentrator plus	Eppendorf, Hamburg, Germany
Thermomixer comfort	Eppendorf, Hamburg, Germany
Top Spin rotator, Intelli-Mixer	LTF Labortechnik, Wasserburg, Germany
Tube roller, Stuart	Cole-Parmer, Vernon Hills, USA
Ultrasonic bath, Bandelin Sonorex	Bandelin electronic, Berlin, Germany
Ultrasonic bath, T310-H	Elma Schmidbauer, Singen, Germany
Vortex, Genie 2	Scientific Industries, Bohemia, USA
Water Bath, Type 1008	GFL, Burgwedel, Germany

2.1.7 Antibodies

Antibody	Clonality/ Host species	Dilution	Manufacturer
CD4-APC (M-T466)	mouse	1:12.5	Miltenyi Biotec, Bergisch Gladbach, Germany
Endoplasmic Reticulum Aminopeptidase 2 (ERAP2)	monoclonal, mouse	1:500	R&D Systems, Minneapolis, USA
Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH)	monoclonal, mouse	1:10000	Merck Millipore, Darmstadt, Germany
Granzyme A (GZMA)	polyclonal, rabbit	1:500	Cell Signaling technology, Danvers, USA
Myeloperoxidase (MPO)	monoclonal, rabbit	1:500	Cell Signaling technology, Danvers, USA

Antibody	Clonality/ Host species	Dilution	Manufacturer
Nuclear Factor Of Activated T Cells 2 (NFATC2)	monoclonal, rabbit	1:1000	Cell Signaling Technology, Danvers, USA
Perforin 1 (PRF1)	monoclonal, mouse	1:500	Abcam, Cambridge, UK
Peroxidase AffiniPure F(ab')$_2$ Fragment Goat Anti-Mouse IgG (H+L)	goat	1:10000	Jackson ImmunoResearch, West Grove, USA
Peroxidase AffiniPure F(ab')$_2$ Fragment Goat Anti-Rabbit IgG (H+L)	goat	1:10000	Jackson ImmunoResearch, West Grove, USA
Proteinase 3 (PRTN3)	monoclonal, rabbit	1:500	Abcam, Cambridge, UK
Ribonuclease 3 (RNASE3)	monoclonal, rabbit	1:250	Abcam, Cambridge, UK
Ribosomal Protein L7 (RPL7)	polyclonal, rabbit	1:1000	LifeSpan BioSciences, Seattle, USA
Zeta Chain Of T Cell Receptor Associated Protein Kinase 70 (ZAP70)	monoclonal, rabbit	1:1000	Cell Signaling technology, Danvers, USA

2.1.8 Clinical Samples

The Ficoll-isolated PBMC samples and the serum samples used in this study were obtained from the biobank of the Institute of Diabetes Research at Helmholtz Zentrum München. Patients with T1D were enrolled in the Diabetes Mellitus Incidence cohort study (DiMelli), which provides a registry of children and adolescents under the age of 20 who were recently diagnosed with diabetes mellitus in Bavaria, Germany (Thumer et al., 2010; Warncke et al., 2013). The diagnosis of diabetes was established in accordance with criteria developed by American Diabetes Association (ADA) and the WHO (2003). The DiMelli registry has been used to characterize diabetes phenotypes based on immunological, metabolic, and genetic markers. Healthy, autoantibody-negative control samples were obtained from the TEENDIAB study, an observational cohort study in children with a familial risk of T1D; this study was designed to investigate genetic and environmental factors that affect the development of islet autoimmunity and T1D in pubescent children (Weber et al., 2015; Ziegler et al., 2012). The DiMelli registry was approved by the medical ethics committee of Bavaria, Germany (Bayerische Landesärztekammer, No. 08043). The TEENDIAB cohort was conducted in accordance with the Declaration of Helsinki, and all procedures involving human subjects were approved by the ethics committee of the Technical University Munich (No. 2149/08) and Medizinische Hochschule Hannover (No. 5644). Written informed consent to participate was obtained from each patient and/or parent. All PBMC samples from patients with T1D

were age- and sex-matched to healthy control samples. An overview of the clinical PBMC sample sets analyzed in this thesis is provided in Tab. 1.

Table 1 Characteristics of the T1D and control groups in PBMC sample sets I and II.
IQR = interquartile range.

	PBMC sample set I		PBMC sample set II	
	T1D	control	T1D	control
n	23	31	30	30
female	14	15	15	15
median age in years (IQR)	7.4 (6.6-7.7)	7.8 (7.2-8.3)	7.6 (6.5-8.9)	7.9 (6.9-8.8)
median T1D disease duration [days]	9	-	9	-
median sample storage period [days]	735	603	880	813

The serum sample set analyzed in this thesis was a selection of 50 pediatric patients with T1D (DiMelli) and 50 healthy controls (TEENDIAB). The sexes were equally distributed among T1D group and control groups; samples were however not matched for age. The characteristics of the serum sample set can be found in Tab. 2.

Table 2 Serum sample set characteristics.
IQR = interquartile range.

	T1D group	Control group
n	50	50
female	23	24
median age (IQR)	10.4 (5.6-12.1)	13.6 (10.8-14.8)
median T1D disease duration (days)	9	-

An overview of the processing workflows employed for all clinical samples can be found in Tab. 3.

Table 3 Workflow characteristics for all processed clinical samples.
GdmCl, guanidinium chloride; SAM = significance analysis of microarrays; BH = Benjamini Hochberg; ANCOVA = analysis of covariance; DDA = data-dependent acquisition; DIA = data-independent acquisition.

	PBMC sample set I		PBMC sample set II		Serum set
sample type	CD4+ T cells	CD4- cells	CD4+ T cells	CD4- cells	serum
lysis buffer	Urea	GdmCl	GdmCl	GdmCl	-

Material and Methods

	PBMC sample set I		PBMC sample set II	Serum set
sample preparation method	in-StageTip (iST)		Filter-aided sample preparation (FASP)	PreOmics
mass spectrometer	Orbitrap XL	Q Exactive HF	Q Exactive HF	
LC-MS/MS method	DDA 170 min	DDA 130 min	DDA 130 min DIA 130 min	DIA 70 min
Data analysis software	Progenesis QI		Progenesis QI (DDA) Spectronaut (DIA)	Spectronaut
identification settings	1% peptide FDR match between runs		1% peptide FDR match between runs (DDA) q-value sparse (DIA)	1% peptide FDR q percentile 25%
quantification settings	all unique peptides		all unique peptides (DDA) Top 5 unique peptides (DIA)	all unique peptides
statistics	unpaired T-Test		paired one-sample T-Test	unpaired ANCOVA (covariate age)
p-value adjustment	FDR adjustment with SAM (Storey)		FDR adjustment (Storey)	FDR adjustment (BH)

2.1.9 Software

Software	Version	Manufacturer/Developer
Adobe Illustrator CS3	13.0.0	Adobe Systems, Mountain View, USA
Byonic	2.0	Protein Metrics, San Carlos, USA
EndNote X7	9325	Thomson Reuters, New York, USA
Image J	1.50i	National Institutes of Health (NIH), USA https://imagej.nih.gov/ij/
FACS Diva	8.0.1	Becton Dickinson, Franklin Lakes, USA
FlowJo	7.6	Tree Star, Ashland, USA
Generanker	2016	Genomatix Software, Munich, Germany
Genomatix Pathway System (GePS)	2016	Genomatix Software, Munich, Germany
GraphPad Prism	6.07	GraphPad Software, La Jolla, USA
Mascot	2.5.1	Matrix Science, London, UK
Perseus	1.5.6.0	Max Planck Institute of Biochemistry, Munich, Germany http://www.perseus-framework.org
Phobius	-	Stockholm Bioinformatics Center, Sweden http://phobius.sbc.su.se/
Progenesis QI	2.0	Nonlinear Dynamics, Newcastle, UK
Proteome Discoverer	2.1	Thermo Fisher Scientific, Waltham, USA
R Studio	0.99.902	R Studio, Boston, USA

Software	Version	Manufacturer/Developer
Signficance Analysis of Microarrays (SAM)	-	https://github.com/MikeJSeo/SAM
Spectronaut	10	Biognosys, Schlieren, Switzerland
String	10.5	String Consortium https://string-db.org/
XL STAT 2015	4.01.20555	Addinsoft, Paris, France

2.2 Methods

2.2.1 Isolation of peripheral blood mononuclear cells from whole blood

Isolation of PBMCs from whole blood was carried out by density gradient centrifugation using Lymphoprep™ (#07851, Stemcell Technologies). Lymphoprep™ has a density of 1.077 g/mL; granulocytes and erythrocytes have a higher density than mononuclear cells resulting in sedimentation through the Lymphoprep™ layer during centrifugation. Approximately 50 ml whole blood was obtained from healthy donors by venipuncture into Na-Heparin containers (#367876, BD Vacutainer). Containers were inverted 10 times and let cool down to room temperature (RT) for 30 min. Whole blood was diluted 1:1 with RPMI medium (#61870-010, Thermo Scientific) and 15 ml diluted whole blood was layered over 10 ml Lymphoprep™. Next, density gradient centrifugation was carried out for 40 min at 400xg and RT with deactivated breaks. After centrifugation erythrocytes, granulocytes and debris formed a sediment at the bottom of the tube. The layer of Lymphoprep™ could be found above this sediment, it was separated from the blood plasma layer by the interphase ring which contained the PBMCs. If required, samples of blood plasma were aliquoted and frozen at -80 °C. The remaining plasma supernatant was discarded and the PBMC interphase ring was transferred into 30 ml prewarmed RPMI medium. Cells were centrifuged for 15 min at 300xg and RT, the supernatant was discarded and cells were resuspended in 30 ml prewarmed RPMI medium. The centrifugation step was repeated and cells were resuspended in 10 ml RPMI. Cell counting was performed manually with a C-Chip counting chamber (Neubauer Improved, #DHC-N01, NanoEnTek). In brief, 6 µl cells were diluted 1:1 with trypan blue solution (#T8154, Sigma Aldrich) and 10 µl of the cell dilution were loaded onto the chamber. Cell count was determined by manually counting all viable cells in 4 large squares of the chamber. If the cell count in a large square exceeded 80-100, the cell suspension was further diluted. The viable cell count per ml was calculated by multiplying the mean large square viable cell count with the chamber factor (= 10^4) and the employed dilution factor.

2.2.2 Thawing of biobanked PBMC samples

Biobanked PBMC samples were thawed in a water bath at 37 °C and immediately transferred to 10 ml of prewarmed RPMI medium without supplement. Cells were washed in RPMI by centrifugation for 5 min at 400xg and RT, resuspended in 10 ml RPMI, and filtered through a 70-µm cell strainer (#352350, Corning). After two additional washing steps in RPMI, PBMCs were subjected to CD4+ T cell enrichment. Cell counting was performed optionally as described before.

2.2.3 Enrichment of CD4+ T cells *via* magnetic activated cell sorting

Isolation of CD4+ T cells from PBMCs using magnetic activated cell sorting (MACS) can be achieved by either positive or negative sorting. In positive sorting, magnetic beads coated with anti-CD4 antibodies are used to specifically bind and enrich CD4+ T cells. In negative mode, CD4+ T cells are left untouched while all CD4- cells present in a PBMC preparation are depleted by using a cocktail of biotin-conjugated antibodies against CD8, CD14, CD16, CD19, CD36, CD56, CD123, TCR γ/δ and Glycophorin, and anti-biotin MicroBeads.

2.2.3.1 Isolation of CD4+ T cells from freshly isolated PBMC

Enrichment of CD4+ T cells from freshly isolated PBMCs of healthy blood donors was performed with both positive and negative sorting for evaluation of CD4+ T cell preparation purity. If the PBMC count was $\leq 10^7$ total cells, MS columns (#130-042-201, Miltenyi Biotec) were used with an OctoMACS separator device; if the PBMC count exceeded 10^7 total cells LS columns (#130-042-401, Miltenyi Biotec) were used with a MidiMACS separator device. All columns were equipped with 30-µm pre-separation filters (#130-041-407, Miltenyi Biotec) to prevent cell clumps from entering the column. Positive sorting was performed with CD4 MicroBeads (#130-045-101, Miltenyi Biotec) and negative sorting with CD4+ T cell isolation kit II (#130-091-155, Miltenyi Biotec). Both protocols were performed in accordance with the manufacturer's instructions with slight modification. For positive sorting, pelleted PBMCs were resuspended in 80 µl MACS buffer (#130-091-222, Miltenyi Biotec) per 10^7 cells containing PBS (pH 7.2) and 2 mM EDTA supplemented with 0.5% (w/v) BSA. After the addition of 20 µl CD4 MicroBeads, the cells were mixed and incubated for 15 min at 4 °C. The cells were then washed in 1 ml MACS buffer per 10^7 cells, centrifuged for 10 min at 300xg at 4 °C, and resuspended in 500 µl MACS buffer without BSA. In all subsequent steps, BSA was omitted in order to minimize interference with mass spectrometry. After the filters were equilibrated with 500 µl MACS buffer (3 ml for LS columns), the cells were loaded onto the columns and washed three times with 500 µl MACS buffer (3 ml for LS columns). To

elute the CD4+ T cell fraction, the column was removed from the separator and flushed with 1 ml MACS buffer (5 ml for LS columns).

For negative sorting, pelleted PBMCs were resuspended in 40 µl MACS buffer (#130-091-222, Miltenyi Biotec) per 10^7 cells containing PBS (pH 7.2) and 2 mM EDTA supplemented with 0.5% (w/v) BSA. After the addition of 10 µl biotin-antibody cocktail (per 10^7 cells), the cells were mixed and incubated for 10 min at 4 °C. Subsequently 30 µl MACS buffer and 20 µl anti-biotin MicroBeads were added, the cells were mixed and incubated for an additional 15 min at 4 °C. Cells were then washed with 10-20x the labeling volume of MACS buffer, centrifuged for 10 min at 300xg at 4 °C and resuspended in 500 µl MACS buffer without BSA. As in positive sorting, BSA was omitted in all subsequent steps in order to minimize interference with mass spectrometry. After the filters were equilibrated with 500 µl MACS buffer (3 ml for LS columns), the cells were loaded onto the columns and washed three times with 500 µl MACS buffer (3 ml for LS columns). The entire effluent containing the untouched CD4+ T cells was collected.

For both protocols, aliquots of CD4+ T cells were taken for quality control flow cytometry (2.2.4). Cells were then either subjected to Glyco-PAL proteomics (2.2.6.1) or pelleted (8 min at 400xg and 4 °C). Pelleted cells were snap-frozen on dry ice and stored at -80 °C before further processing.

2.2.3.2 Isolation of CD4+ T cells from biobanked PBMC samples

CD4+ T cells were enriched from biobanked PBMC samples from the DiMelli and TEENDIAB studies after thawing (see 2.2.2). CD4+ T cells were isolated using MACS positive sorting in accordance with the manufacturer's instructions (CD4 MicroBeads, Miltenyi Biotec, #130-045-101) as described above (2.2.3.1.). In all cases, magnetic separation was performed using MS columns (Miltenyi Biotec, #130-042-201) equipped with 30-µm pre-separation filters (Miltenyi Biotec, #130-041-407) on an OctoMACS separator device. In addition to enrichment of CD4+ T cells, the washing effluent containing the CD4-depleted cells was also collected for each sample. The cells were optionally counted, pelleted by centrifugation for 8 min at 400xg and 4 °C, snap-frozen on dry ice, and stored at -80 °C until further processing.

2.2.4 Flow cytometry

Purity of CD4+ T cell preparations from healthy blood donors obtained by both positive and negative sorting was assessed using flow cytometry. Due to limited patient material, quality control flow cytometry was not performed with CD4+ T cell preparations from biobanked samples. In brief, isolated CD4+ T cells were centrifuged for 10 min at 300xg

and RT and resuspended in 100 µl FACS buffer (PBS pH 7.4 supplemented with 5% BSA) containing CD4-APC antibody (1:12.5, #130-091-232, Miltenyi Biotec). Cells were mixed and incubated for 20 min at 4 °C in the dark. Cells were washed with 500 µl FACS buffer (300xg, 5 min, and 4 °C), resuspended in 100 µl 1.5% formalin and fixed for 30 min at 4 °C. Cells were washed again and resuspended in up to 500 µl FACS buffer. Data were acquired using BD LSRFortessa cell analyzer. Data analysis was performed with FlowJo (Version 7.6).

2.2.5 Cell lysis and BCA assay

Frozen cell pellets derived from healthy blood donors were lysed by resuspension in up to 100 µl lysis buffer. The lysis buffer used during SOP development was Urea buffer containing 8M Urea, 100 mM Tris (pH 8.5) and 1x complete protease inhibitor (#11697498001, Roche). All CD4+ T cell samples derived from the clinical sample set I were lysed in 25 µl Urea buffer. During later stages of this thesis, the lysis buffer was changed to GdmCl buffer containing 6M guanidinium chloride, 100 mM Tris (pH 8.5) and 1x complete protease inhibitor. CD4-depleted cell pellets from clinical sample set I and all cell pellets derived from sample set II were lysed in GdmCl buffer (25-100 µl). After lysis in the respective lysis buffer samples were shortly boiled (5 min, 95 °C) and sonicated for 5 min (model T310-H, Elma Transsonic or Sonorex, Bandelin electronic). Protein concentration was measured using Pierce™ BCA Protein Assay Kit (#23225, Thermo Scientific) according to the manufacturer's instructions. In brief, BSA standard (#B900S, NEB) was serially diluted to concentrations ranging from 0.125-8 µg/µl. Then, 2 µl of BSA standard or 2 µl sample were pipetted in triplicates into a 96-well microtiter plate (#353072, Corning). Kit Reagent A and B were mixed in a proportion of 50:1 and 200 µl of the solution was added to each well. The plate was incubated for 30 min at 37 °C and net absorbance was measured at 562 nm using a spectral photometer (EMax Plus, Molecular Devices).

2.2.6 Sample preparation for mass spectrometry

2.2.6.1 Glycocapture proteomics

The glycocapture proteomics protocol was adapted for primary human CD4+ T cells from Zeng et al. (2009) and Graessel et al. (2015). The protocol is based on biotinylation of sialic acid-containing glycans on living cells by periodate oxidation and aniline-catalyzed oxime ligation (Glyco-PAL, Fig. 15).
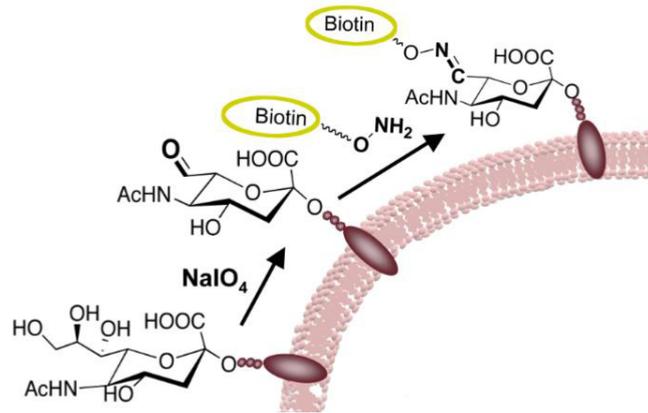
Figure 15 Principle of cell surface protein labeling with PAL.
Aldehydes on cell surface sialic acids are introduced by periodate oxidation (NaIO$_4$) and subsequently labeled with aminooxy-biotin. Graphic source: Zeng et al. (2009)

In brief, replicates of 1 million viable CD4+ T cells were subjected to the Glyco-PAL protocol (replicates differed in the respective buffer used for labeling). Tested labeling buffers included PBS with 2 mM EDTA; Hanks buffer (#14025-050, Thermo Scientific) adapted to pH 6.7 and X-Vivo medium (#BE04-418F, Lonza). All following steps were performed on ice. Cells were washed once in the respective labeling buffer (300xg, 3 min) and biotinylated by addition of 1 mM NaIO$_4$, 500 µM aminooxy-biotin (#90113, Biotium) and 10 mM aniline (# 242284, Sigma Aldrich) in a one-pot reaction of 1 ml labeling buffer (in test 2 the total volume was reduced to 250 µl). Cells were incubated on a rotating wheel (Rotator SB3, Cole-Parmer) for 30 min at 4 °C in the dark (in test 2 adapted to 10 min). The biotinylation reaction was then quenched by addition of glycerol to a final concentration of 1 mM and further 5 min incubation (4 °C, rotating in the dark). Cells were washed once with ice-cold tris-buffered saline (TBS, 20 mM Tris-HCl, 150 mM NaCl), lysed in 200 µl lysis buffer (1% NP40, 10 mm NaCl, 10 mm Tris pH 7.6, 1 × complete protease inhibitor) and frozen at -20 °C. For capturing the surfaceome, the lysed cells were thawed on ice and vortexed. Crude lysates were cleared by centrifugation (6000xg, 10 min 4 °C), the pellet was discarded and the supernatant (soluble fraction) was diluted 1:5 with TBS. Affinity purification of the biotin-labeled surfaceome was performed with 25 µl Streptavidin beads (Strep Tactin Supeflow, #2-1206-10, IBA) per sample. Beads were prewashed thrice with washing buffer (TBS, 0.2% NP40) and subsequently incubated with the diluted cell lysates for 2 hours at 4 °C on a Top Spin rotator (Intelli Mixer, LTF Labortechnik). To obtain bead-protein complexes, samples were pelleted at 2000xg for 1-2 min. The following washing and incubation procedure was carried out in a volume of 200 µl per step with centrifugation steps in between at 2000×g for 2 min. First, beads were washed with washing buffer, followed by a washing step with 0.2% SDS in TBS. Bead-bound proteins were then reduced for 30

min with 100 mM DTT in TBS at 300 rpm and RT. Beads were subsequently washed with UC buffer (6M Urea, 100 mM Tris pH 8.5) and proteins were alkylated for 30 min in UC buffer containing 50 mM IAA at RT. Beads were then washed 4 times, once each with: UC buffer, 5 M NaCl,100 mM Na_2CO_3 pH 11.5 and 50 mM pH 8.5. Bead-bound proteins were digested in a volume of 40 µl 50 mM Tris pH 8.5 containing 1 µg trypsin (#V511A, Promega). The digestion step was performed under mild shaking in a Thermomixer (comfort, Eppendorf) at 37 °C o/n. Beads were centrifuged and the supernatant containing the tryptic peptides was transferred into a new tube. Beads were resuspended in 40 µl 50 mM Tris pH 8.5 and centrifuged again; the supernatant was pooled with the first tryptic fraction. The beads were washed with G7 buffer (#B3704S, NEB) and glycopeptides were subsequently eluted by incubation with PNGase F. The digestion step was performed at 37 °C in a volume of 20 µl G7 buffer containing 500 U PNGase F (#P0705L, NEB) under mild shaking for 6 h. Beads were then centrifuged and the supernatant containing the glycopeptides was transferred into a new tube. Beads were again resuspended in 20 µl G7 buffer and centrifuged; the supernatant was pooled with the first PNGase F fraction. Tryptic and PNGase F fractions were acidified with 3 µl 5% TFA and stored at -20 °C until MS analysis was carried out.

2.2.6.2 iST sample preparation

The iST sample preparation method was adapted from Kulak et al. (2014). This adaption was the result of the SOP development for DDA LC-MS/MS on Orbitrap XL machinery. All cellular samples (CD4+ and CD4-depleted) derived from clinical sample set I were processed for LC-MS/MS using the iST method. In brief, approximately 10 µg total protein lysate was reduced with 10 mM TCEP and alkylated with 40 mM CAA. Protein lysates were diluted 1:10 in dilution buffer containing 10% (v/v) ACN, 25 mM Tris (pH 8.5), 1 µg trypsin (#V511A, Promega), and 1 µg Lys-C (#121-05063, Wako), then digested overnight at 37 °C. After digestion, the peptides were acidified in a final concentration of 1% TFA. Prior to loading the peptides on the StageTip, SDB-RPS (#2241, 3M Bioanalytical Technologies, 14-gauge plug SDB-RPS) was activated by the sequential application of 50 µl acetone, isopropanol, methanol, and water, followed by centrifugation for 1 min at 1000xg. Acidified peptides were transferred to the StageTip and loaded onto the activated SDB-RPS material by centrifugation for 1-2 min at 1500xg. This loading procedure was repeated three times, followed by three washing steps with 100 µl 0.2% TFA. Peptides were eluted in three fractions by the sequential application of 60 µl of elution buffers 1-3 and centrifugation for 1-2 min at up to 1500xg (Buffer 1: 100 mM NH_4HCO_2, 40% ACN, and 0.5% formic acid; Buffer 2: 150 mM NH_4HCO_2, 60% ACN, 0.5% formic acid, and 50 mM NaOH; Buffer 3: 5% NH_4OH, 80% ACN). Optionally a 4th

fraction was generated by application of 60 µl 5 % NH_4OH, 80% ACN/isopropanol (v/v, 1:1) and 2 mM NaOH. Eluates were collected and evaporated using a SpeedVac centrifuge (Concentrator plus, Eppendorf). The peptides were resuspended in 50 µl loading buffer (2% ACN and 0.5% TFA), sonicated briefly, and analyzed using LC-MS/MS.

2.2.6.3 Filter-aided sample preparation

The FASP method was used in a modified version adapted from Wisniewski et al. (2009). The FASP protocol was used for all cellular samples (CD4+ and CD4-depleted) from clinical sample set II due to its high compatibility with spectral library generation and DIA LC-MS/MS. If not stated otherwise, 10 µg total protein lysate were subjected to FASP. In brief, protein lysates were adjusted to a volume of 200 µl with ABC buffer (50 mM ammonium bicarbonate) and reduced by addition of 1 µl of 1 M DTT with incubation for 30 min at 60 °C. The samples were then diluted in urea buffer containing 8 M urea and 100 mM Tris (pH 8.5) to a final concentration of 4 M urea. To alkylate the cysteine residues, 10 µl of freshly prepared 300 mM IAA solution was added, and the samples were incubated in the dark for 30 min at RT. After the unreacted IAA was quenched with 2 µl of 1 M DTT, the samples were centrifuged through a 30 kDa cut-off filter (# VN01H23ETO, Sartorius) and washed three times with urea buffer and twice with ABC buffer by centrifugation for 15 min at 14,000xg. The proteins were digested with 1 µg Lys-C in 40 µl ABC buffer for 2 hours at RT followed by 1 µg trypsin in 10 µl ABC buffer for 16 hours at 37 °C. If less than 10 µg protein lysate were used as starting material, the amount of enzymes was adapted, i.e. 5 µg CD4+ T cell lysate were digested with 0.5 µg Lys-C and trypsin each. The peptides were collected by centrifugation for 10 min at 14,000xg, and the filters were washed with 20 µl ABC buffer containing 5% ACN. Finally, the peptides were acidified with 2 µl 100% TFA prior to mass spectrometric analysis.

2.2.6.4 PreOmics

The serum sample set consisting of 50 pediatric patients with T1D (DiMelli) and 50 healthy controls (TEENDIAB) was processed for LC-MS/MS with the PreOmics iST Kit (#00027, PreOmics) according to the manufacturer's instructions. The PreOmics iST Kit is a commercial adaption of the iST method described in Kulak et al. (2014) and consisted of a collection of buffers ("lyse", "resuspend", "stop", "wash1", "wash2" and "elute"), lyophilized enzyme ("digest") and several custom-made microtiter plates ("device", "collection" and "adapter"). In brief, 2 µl of serum were mixed with 50 µl lyse buffer and boiled for 10 min at 95 °C under constant shaking. Samples were let cool down to RT and transferred to the device plate containing the peptide purification

columns. The lyophilized enzyme was reconstituted by adding 210 µl resuspend buffer to a digest tube, vortexed and incubated for 10 min at 500 rpm and RT. For tryptic digest, 45 µl of resuspended enzyme were added to each sample and the device plate was incubated on a block heater (comfort, Eppendorf) for 2 hours at 300 rpm and 37 °C. For this purpose the block heater was equipped with a custom-made heating shaker adapter (#00004, PreOmics). Next, the device plate was placed into a collection plate for collection of effluent during the centrifugation steps. Tryptic digest was ended by adding 100 µl stop buffer to each sample; the plate was subsequently centrifuged for 3 min at 3800xg. The columns in the device plate were washed in two consecutive washing steps by loading 200 µl wash1 and wash2 buffer to the plate. Each time the plate was centrifuged for 3 min at 3800xg and effluent was disposed of. Next, the device plate was stacked up with the adapter plate and an elution plate (Twin.tec LoBind PCR plate, #0030129504, Eppendorf) and 100 µl elute buffer were loaded to the columns. The plates were centrifuged for 3 min at 2800xg and the elution step was repeated with another volume of 100 µl elute buffer. The eluted peptides (200 µl) were completely dried using a SpeedVac (Concentrator plus, Eppendorf) and frozen at -80 °C until further processing. For LC-MS/MS, samples were resuspended in 100 µl loading buffer (2% ACN and 0.5% TFA) and sonicated for 5 min (Sonorex, Bandelin electronic). For PreOmics processing and subsequent LC-MS/MS, the sample order was randomized and additionally a pooled sample was measured 5 times throughout the sequence for evaluation of technical reproducibility (1st, 25th, 50th, 75th and 100th). This quality control sample was generated by pooling 10 µl of 10 randomly picked samples (5 each from T1D and control group).

2.2.7 Mass spectrometry

2.2.7.1 Data-dependent acquisition

LC-MS/MS in DDA mode was performed on Orbitrap XL and Q Exactive HF mass spectrometers (both Thermo Scientific). Orbitrap XL machinery was used for all LC-MS/MS experiments during SOP development and for LC-MS/MS of CD4+ T cell samples from clinical sample set I. Q Exactive HF machinery was used for LC-MS/MS analysis of CD4-depleted cell samples from clinical sample set I and all analyses conducted on clinical sample set II (see Tab. 3).

For Orbitrap XL, DDA LC-MS/MS was performed as previously described (Molin et al., 2015). Approximately 1 µg of each sample was loaded automatically onto the HPLC system. A nano trap column was used (300 µm inner diameter × 5 mm, packed with Acclaim PepMap100 C18, 5 µm, 100 Å; LC Packings) before separation using RP

chromatography (PepMap, 25 cm, 75 µm ID, 2µm/100 Å pore size, LC Packings) operated on a rapid separation LC (RSLC) system (Ultimate 3000, Dionex). After 5 min, peptides were eluted using the following gradient of increasing ACN concentration in 0.1% formic acid over a period of 145 minutes: 135 minutes of 6-31% ACN, followed by 5 minutes at 31-71% ACN, and then 5 minutes at a fixed concentration of 71% ACN. Between each gradient, the ACN (in 0.1% formic acid) concentration was set back to starting conditions for 20 minutes. A high-resolution (60,000 full-width at half-maximum) MS spectrum was acquired with the Orbitrap using a mass range from 300 to 1500 m/z. From this high-resolution MS pre-scan, the ten most abundant peptide ions were selected for fragmentation in the linear ion trap if they were at least doubly charged, with a dynamic exclusion of 60 seconds.

The Q Exactive HF was coupled to an RSLC system (Ultimate 3000, Dionex). For DDA LC-MS/MS approximately 1 µg of sample was automatically loaded on the HPLC system, which was equipped with a nano trap column (300 µm inner diameter × 5 mm, packed with Acclaim PepMap100 C18, 5 µm, 100 Å; LC Packings). After 5 min, the peptides were eluted from the trap column and separated using RP chromatography (Acquity UPLC M-Class HSS T3 Column, 1.8 µm, 75 µm x 250 mm; Waters) using a gradient of increasing ACN concentration over a period of 125 minutes: 80 minutes of 7-27% ACN, followed by a two short gradients of 27–41% ACN (15 min) and 41–85% ACN (5 min). After 5 min at a fixed concentration of 85% ACN, the gradient was set back to 3% ACN over a period of two minutes and allowed to equilibrate for 18 min. All ACN solutions contained 0.1% formic acid. A high-resolution (60,000 full-width at half-maximum) MS spectrum was acquired in the Orbitrap ranging from 300 to 1500 m/z with automatic gain control target set to $3x10^6$ and a maximum injection time of 50 ms. From this high-resolution MS pre-scan, the ten most abundant peptide ions were selected for fragmentation if they exceeded an intensity of at least 2xe4 counts and if they were at least doubly charged. MS/MS spectra were also recorded in the Orbitrap at a resolution of 15,000 with a maximum injection time of 50 ms. Dynamic exclusion was set to 30 sec.

2.2.7.2 Data-independent acquisition

For LC-MS/MS in DIA mode, the same RSLC/Q Exactive HF system as described above was used. In brief, all samples were spiked with 1 injection unit of the HRM Calibration Kit (#Ki-3003, Biognosys) for retention time indexing. Again, approximately 1 µg of sample was automatically loaded onto the nano trap equipped HPLC system. Tryptic digests of cell lysates were analyzed using 130 min runs; digests of serum samples were analyzed in shorter 70 min runs. After 5 min, the peptides were eluted from the trap

column and separated using reversed-phase chromatography using a gradient of 7–27% ACN over a period of 90 min and 35 min, respectively. The main gradient was followed by a two short gradients of 27–41% ACN (15 min and 10 minutes, respectively) and 41–85% ACN (5 min). After 5 min at 85% ACN, the gradient was set back to 3% ACN over a period of two minutes and allowed to equilibrate for 8 min. All ACN solutions contained 0.1% formic acid. The DIA method for 130 min runs consisted of a full MS scan at 120,000 resolution ranging from 300 to 1650 m/z with automatic gain control target set to 3×10^6 and a maximum injection time of 120 ms. Subsequently, 37 DIA windows with a variable width spanning from 300 to 1650 m/z were acquired at a resolution of 30,000 (Tab. 4). Normalized collision energy was set to 28, and the spectra were recorded in profile type.

Table 4 DIA method used for ion fragmentation on the Q Exactive HF with 130 min runs.
Start, End and Isolation Window size is given in m/z.

Window Number	Start	End	Window size
1	300	348	48
2	347	373	26
3	372	393	21
4	392	411	19
5	410	429	19
6	428	446	18
7	445	463	18
8	462	480	18
9	479	496	17
10	495	512	17
11	511	529	18
12	528	545	17
13	544	561	17
14	560	577	17
15	576	594	18
16	593	611	18
17	610	627	17
18	626	643	17
19	642	661	19
20	660	678	18
21	677	696	19
22	695	715	20
23	714	734	20
24	733	754	21
25	753	775	22
26	774	796	22
27	795	819	24
28	818	843	25
29	842	868	26
30	867	896	29
31	895	928	33
32	927	963	36
33	962	1003	41

Window Number	Start	End	Window size
34	1002	1051	49
35	1050	1113	63
36	1112	1214	102
37	1213	1650	437

The DIA method for 70 min runs consisted of a full MS scan at 120,000 resolution ranging from 300 to 1500 m/z with automatic gain control target set to 3×10^6 and a maximum injection time of 120 ms. Subsequently, 16 DIA windows with a variable width spanning from 300 to 1500 m/z were acquired at a resolution of 30,000 (Tab. 5). Normalized collision energy was set to 28, and the spectra were recorded in profile type.

Table 5 DIA method used for ion fragmentation on the Q Exactive HF with 70 min runs.
Start, End and Isolation Window size is given in m/z.

Window Number	Start	End	Window size
1	300	364	64
2	363	405	42
3	404	445	41
4	444	485	41
5	484	519	35
6	518	558	40
7	557	596	39
8	595	632	37
9	631	669	38
10	668	712	44
11	711	759	48
12	758	808	50
13	807	867	60
14	866	948	82
15	947	1063	116
16	1062	1500	438

2.2.8 Data analysis

2.2.8.1 Progenesis QI for label-free quantification of DDA LC-MS/MS data

Spectra acquired from the samples in DDA experiments were analyzed using Progenesis QI software for proteomics (Version 2.0, Nonlinear Dynamics, Waters) for label-free quantification as previously described (Grosche et al., 2016). If Q Exactive HF raw files were analyzed the Progenesis QI import filer was set to 2. The profile data of the MS scans were transformed into peak lists with respective m/z values, intensities, abundances, and m/z width. MS/MS spectra were transformed similarly and then stored in peak lists comprised of m/z and abundance. Using one sample as a reference, the retention times of the other samples were aligned using automated alignment to a

maximal overlay of 2D features. Features with one or more than seven charges were masked at this point and excluded from further analyses. After alignment and feature exclusion, the samples were allocated to their respective experimental group (e.g. T1D and control group), and the raw abundances of all features were normalized to correct for technical variations. All MS/MS spectra were exported from the Progenesis QI software as Mascot generic files (mgf) and used for peptide identification. The respective identification settings for the individual experiments can be found in Tab. 6. If peptide identification was performed with Mascot (Version 2.5.1, Matrix Science), a percolator decoy database search was always implemented. For combination of search engines Proteome Discoverer platform (Version 2.1, Thermo Scientific) with integrated Mascot and Byonic (Version 2.0, Protein Metrics) search nodes was used. The search parameters always included 10 ppm peptide mass tolerance and 0.6 Da fragment mass tolerance (0.02 Da for Q Exactive HF raw files). Carbamidomethylation was set as a fixed modification, and if not stated otherwise methionine oxidation and deamidation of asparagine and glutamine were set as variable modifications. One missed cleavage was allowed as default setting for Mascot searches and two missed cleavages for Byonic searches. Database searches were set to a FDR of 1% on the peptide level when the concatenated mgf files were searched. Search engine identifications were (optionally) combined in a multi-consensus result file maintaining 1% peptide FDR and exported in pepXML format for re-import into Progenesis QI.

Table 6 Peptide identification settings for database searches with DDA LC-MS/MS data.
Default settings (*) are defined as methionine oxidation and deamidation of asparagine and glutamine in variable modification settings, and one and two missed cleavages allowed for Mascot and Byonic searches, respectively. K, lysine; R, arginine; M, methionine; FDR, false discovery rate.

Experiment	Platform	Search engine	Settings	Database
SOP development	Mascot server	Mascot	default settings* 1% peptide FDR ion score 15 cut-off	Ensembl human Release 75
CD4+ T cells sample set I	Proteome Discoverer	Mascot and Byonic	2 missed cleavages allowed variable modifications: Oxidation (M) Carbamylation (N-term., K, R) 1% peptide FDR no ion score cut-off	Swissprot human spiked Release 2016.02 www.uniprot.org
CD4- cells sample set I and CD4+/CD4- (T) cells sample set II	Proteome Discoverer	Mascot and Byonic	default settings* 1% peptide FDR no ion score cut-off	Swissprot human spiked Release 2016.02 www.uniprot.org

For quantification, all unique peptides from an identified protein group were included, and total cumulative normalized abundance was calculated by summing the abundances of all unique peptides allocated to that respective protein. All fold changes reported by Progenesis QI in the DDA experiments refer to the protein level.

2.2.8.2 Spectral library generation

To generate custom spectral libraries for peptide identification in DIA LC-MS/MS experiments, select DDA LC-MS/MS data (acquired on Q Exactive HF) spiked with the HRM Calibration Kit (# Ki-3003, Biognosys) were analyzed using Proteome Discoverer (Version 2.1, Thermo Scientific) with embedded Byonic (Version 2.0, Protein Metrics) search engine node. Identifications were filtered to satisfy 1% protein and peptide FDR and search results were stored as Proteome Discoverer magellan storage files (msf). The Proteome Discoverer result file (pdResult) necessary for spectral library generation was created by combinining multiple individual search results (msf) in a multi-consensus result file (pdResult) maintaining the 1% protein and peptide FDR threshold. The peptide spectral library was then generated in Spectronaut (Version 10, Biognosys) with default settings by importing the respective combined result file. Spectronaut was equipped with the Swissprot human database (Release 2016.02, 20165 sequences, www.uniprot.org) with a few spiked proteins (e.g. Biognosys indexed retention time (iRT) peptide sequences). Details on the two spectral libraries used for analysis of DIA LC-MS/MS runs can be found in Tab. 7.

Table 7 Spectral libraries used for analysis of DIA LC-MS/MS data.
HeLa, human cervical cancer cell line; ECM, extracellular matrix. The values listed in the size column correspond to values reported by Spectronaut.

Spectral Library	Size [Proteins/precursors]	DDA Input [sample type]	Used for DIA LC-MS/MS
Meta human library	11403/432673	CD4+ T cells/CD4- cells PBMCs, neutrophils, bone marrow diverse cell line nucleic extracts HeLa, fibroblasts, ECM retina, retinal pigment epithelia astrocytes lung tissue adipose tissue skin biopsies duodenal juice, sweat	CD4+ T cells/CD4- cells Clinical sample set II
Plasma and serum library	715/27057	plasma and serum	Serum set

2.2.8.3 Spectronaut for label-free quantification of DIA LC-MS/MS data

The DIA LC-MS/MS data was analyzed using Spectronaut (Version 10, Biognosys) with slightly adapted default settings for the spectral library search. Spectronaut HTRMS

converter was used to convert the raw files. In brief, the data and XIC extraction settings were set to dynamic with a correction factor of 1. Automatic calibration mode was chosen with precision iRT enabled for applying the non-linear iRT calibration strategy. Peptide identification was filtered to satisfy a FDR of 1%. The q-value sparse mode for peptide precursor signal matching between runs was enabled for analysis of all cellular samples (CD4+ T cells/CD4-depleted cells). For serum set samples, the q-value percentile mode with a threshold of 25% was employed, i.e. only peptide precursor signals exhibiting q<0.01 in at least 25% of all samples were considered for quantification. In both cases, calculation of protein quantities was based only on proteotypic (unique) peptides of a respective protein. Protein grouping was defined by protein group ID and peptide grouping by stripped sequence. Summed precursor quantities (based on MS2 area quantity) were used for protein group quantification. For cellular samples, protein quantity calculation was based on the sum of the top 5 stripped sequences which themselves were calculated based on the top 3 precursors. For serum set samples, protein quantification was based on the sum of all available proteotypic precursors.

2.2.8.4 Statistics

For all quantitative data obtained in DDA or DIA LC-MS/MS experiments, statistical significance regarding differential protein abundance was evaluated by group comparisons (T1D versus control groups). For data obtained in DDA experiments with Progenesis QI software, an unpaired Student's t-test was performed on the log transformed normalized protein abundances. Significance analysis of microarrays (SAM) was used to correct for multiple hypothesis testing (Tusher et al., 2001); a statistical technique developed for identifying significant genes in a set of microarrays likewise applicable to protein expression data. All one-peptide-based protein identifications were excluded from SAM analysis and thus not considered for candidate protein evaluation. SAM was used in the R package available via the SAM project Github homepage (https://github.com/MikeJSeo/SAM). SAM uses the FDR and q-value method published by Storey (2002). SAM was performed on log2 transformed normalized protein abundances by selecting two class unpaired T-statistic as test setting. The significance filter was set to meet a median FDR of 5%, thus all proteins below the threshold of q<0.05 were considered significantly different between the T1D and control groups.

For quantitative DIA LC-MS/MS data obtained for all cellular samples (i.e. CD4+ T cells/CD4-depleted cells of clinical sample set II), the tools built into Spectronaut were used for statistical significance testing. Here, differential protein expression between the T1D and control groups was tested using a one-sample two-tailed Student's t-test on

log2 transformed peptide ratios. Accordingly, samples were paired based on sex- and age-matching. Again, multiple hypothesis testing was corrected for using the Storey method. All proteins with q<0.05 were considered significantly different; again one-peptide-based protein identifications were excluded from further analyses such as pathway enrichment.

Quantitative DIA LC-MS/MS data obtained for the serum sample set could not be analyzed with Spectronaut built-in statistics tools (based on sample pairing) due to the fact that T1D and control groups were not matched for age (see Tab. 2). Here, differential protein expression testing was performed with analysis of covariance (ANCOVA) using the Excel (Microsoft) XLStat add-in included in the OMICS module (Addinsoft). ANCOVA was performed with default settings correcting for age as covariate; the resulting p-values were corrected for multiple hypothesis testing using the Benjamini Hochberg method (Benjamini and Hochberg, 1995). Again all proteins with q<0.05 were considered significantly different between T1D and control groups and one-peptide-based protein identifications were excluded from all further analyses.

Correlation analyses for select signature proteins measured in sample set I and II were computed for all patients with T1D using GraphPad Prism (Version 6, GraphPad Software, La Jolla, USA). Correlations calculated for protein abundances in sample set II were based on DIA data. All protein abundance values were log2 transformed prior to analysis. The clinical parameters glycated hemoglobin (HbA1C) and fasting C-peptide levels as well as patient age were selected for correlation assessment. HbA1c measurements were available in 18 (set I) and 30 (set II) T1D cases; C-peptide measurements in 16 (set I) and 23 (set II) T1D cases, respectively. Normal distribution of HbA1c, C-peptide and Age was assessed with D'Agostino and Pearson omnibus test and correlations were computed using the Pearson method with default settings. All assessed parameters followed a normal distribution with the exception of age in sample set II; here correlation was assessed using the Spearman method with default settings. All obtained p-values were corrected for using the Bonferroni method.

2.2.8.5 Phobius Analysis

To determine the subcellular localization of the identified proteins, FASTA sequence files were submitted to Phobius, a combined transmembrane topology and signal peptide predictor (Kall et al., 2004, 2007b). The web-based Phobius tool is available via http://phobius.sbc.su.se/. Proteins were then classified into three categories depending on the presence of an exclusive secretory signal peptide (secretory proteins), at least

one transmembrane domain (TMD) with or without a secretory signal peptide (transmembrane proteins), or no signal peptide and no TMD(s) (other proteins). The latter category (other) comprises all intracellular, cytosolic proteins which are not associated to membrane-bound organelles. The category transmembrane proteins could further be subdivided into plasma membrane proteins (with secretory signal peptide) or intracellular, membrane-bound proteins (no secretory signal peptide).

2.2.8.6 Pathway enrichment analysis and network generation

Pathway enrichment analysis was performed with Genomatix Generanker using ElDorado and Literature Mining Database version 02-2016 (http://www.genomatix.de, Genomatix Software, Munich, Germany). For Glyco-PAL experiments, all identified proteins were uploaded to the web-based tool to confirm enrichment of the cell surface compartment; no background protein list was submitted. For instance, the number of proteins annotated as *plasma membrane* (Gene ontology (GO) cellular components) was extracted from the Generanker results table. If a specific GO term of interest was not enriched in a Glyco-PAL data set, the number of respective member proteins was assessed with an in-house script developed by Marcel Blindert. This script retrieves GO term information on biological processes, cellular components and molecular functions for the respective protein accessions from a GO term text file downloaded from http://www.ensembl.org via BioMart data mining tool.

Pathway enrichment analysis of significantly regulated proteins in the clinical sample sets was also performed with Genomatix Generanker (Version 02-2016). Inclusion criteria for proteins in Generanker analysis were $q<0.05$ (i.e. FDR<5%) and quantification with ≥2 unique peptides. For the analysis of proteins in the DIA data sets, an additional criteria of ≥1.5-fold or ≤0.67-fold regulation (CD4+ T cell subset) and ≥1.3-fold or ≤0.77-fold regulation (CD4-depleted fraction) was applied. All protein identifications (≥2 unique peptides) in the respective data set were used as a background gene list. Gene names of input proteins were uploaded, and the corresponding overrepresented pathways and processes are listed in Tables 11 and 14. To generate protein networks the Genomatix Pathway System (GePS) or the web-based STRING tool (STRING consortium 2017, https://string-db.org) were used.

2.2.8.7 Visualization of data

The heatmaps generated by hierarchical cluster analyses for DIA data of the cellular subsets were exported from Spectronaut using the default settings (Manhattan Distance; Ward's method in linkage strategy) with Z-scoring function enabled. The hierarchical

cluster analysis for DIA serum data was performed using the Perseus software platform, version 1.5.6.0 (Tyanova et al., 2016). Here, normalized protein abundances of all identified proteins were first log2 transformed and then standardized with Z-scoring prior to heatmap generation using the default settings (Euclidean distance, average in linkage strategy). Principal component analysis (PCA) for DDA data of CD4+ T cells (set I) was exported from Progenesis QI software using all possible features (peptide signals). The PCA for DIA serum data was generated with Perseus (Version 1.5.6.0) using log2 transformed normalized protein abundances of all identified proteins. All scientific figures were created using either Excel (Version 2010, Microsoft Corporation, Redmond, USA) or GraphPad Prism (Version 6, GraphPad Software, La Jolla, USA). Adobe Illustrator CS3 (Version 13, Adobe Systems, San José, USA) was used for figure adaption and design.

2.2.9 Western Blot Analysis

For Immunblot validation of select candidate proteins, residual lysates of CD4+ T cell and CD4-depleted cell samples were pooled for antibody probing. All samples used for western blot analysis were obtained from patients from clinical sample set II. CD4+ T cell samples were pooled from 23 patients with T1D and from 23 control subjects; CD4-depleted cell samples were pooled from 28 patients with T1D and from 29 control subjects, respectively. Pooled protein lysates were processed using the Pierce SDS-PAGE Sample Prep Kit (#89888, Thermo Fisher) in accordance with the manufacturer's instructions. Subsequently, the protein concentration in the pooled lysates was measured using a BCA assay (#23225, Thermo Fisher), and 10 to 20 µg of total protein lysate was adjusted to a final concentration of 1x Laemmli buffer (50 mM Tris (pH 6.8), 1% SDS, 10% glycerol, 50 mM β-mercaptoethanol, and bromphenol blue). The samples were boiled for 5 min and loaded onto a precast gradient gel (#4561084, Bio-Rad) and separated by SDS-PAGE at 100-150 V. The proteins were then transferred for 60 min at 100V to a PVDF membrane using a wet tank blotting system (#1703930, Bio-Rad). The membranes were blocked with 5% (w/v) milk powder (#170-6404, Bio-Rad) in TBS containing 0.1% Tween 20 (TBST) for 1 hour at RT, then washed 3 times in TBST for 15 min each. The membranes were then incubated overnight at 4 °C in the following primary antibodies: mouse anti-ERAP2 (1:500, #MAB3830, R&D Systems), mouse anti-GAPDH (1:10000, #MAB374, Merck Millipore), rabbit anti-GZMA (1:500, #4928S, Cell Signaling), rabbit anti-MPO (1:500, #14569S, Cell Signaling), rabbit anti-NFATC2 (1:1000, #5861S, Cell Signaling), mouse anti-PRF1 (1:500, #ab47225, Abcam), rabbit anti-PRTN3 (1:500, #ab133613, Abcam), rabbit anti-RNASE3 (1:250, #ab135840, Abcam), rabbit anti-RPL7 (1:1000, #C358295, LifeSpan BioSciences) and rabbit anti-

ZAP70 (1:1000, #3165S, Cell Signaling). After washing three times in TBST for 20 min each, the membranes were incubated in the secondary antibody (peroxidase-conjugated goat anti-rabbit/mouse IgG, 1:10000, #111-036-045 (rabbit) or #115-036-062 (mouse), Jackson ImmunoResearch) for 1 hour at RT. After washing three times in TBST for 20 min each, the membranes were developed using 1 ml ECL Select Western Blotting Detection Reagent (#RPN2235, GE Healthcare), and the signal was captured using Fusion Fx7 (Vilber Lourmat). If the target candidate protein had the same size (kDa) as a housekeeping protein (GAPDH or RPL7), the probing of the housekeeper was performed after peroxidase quenching in NaN_3. To this end, membrane-bound peroxidase conjugate was inactivated by incubation in TBST containing 15 mM NaN_3 and 5% (w/v) milk for 1 hour at RT. The membrane was washed 5 times for 15 min each, and the blocking, antibody incubation and detection procedure was repeated as described above.

3 Results

3.1 Evaluation of sample preparation strategies for blood cell proteomics

The first aim of this thesis was to establish a SOP for proteomic profiling of clinical PBMC samples which had been placed in long-term biobank storage. The SOP needed to be applicable to quantity- and quality-limited blood cells yet achieving high sensitivity and robustness when DDA LC-MS/MS was performed on an Orbitrap XL mass spectrometer. All experiments during SOP development were performed with limited numbers of freshly isolated blood cells from healthy donors simulating restricted cell count.

3.1.1 Surfaceome profiling with glycocapture proteomics (Glyco-PAL)

Sample preparation strategies for proteomic profiling of (blood) cells can be divided into mainly two categories: whole-proteome and subcellular fractionation approaches. The latter are based on enrichment of a specific cellular compartment prior to LC-MS/MS thus enabling a much higher resolution of its proteomic composition. Cell surface proteins constitute approximately 30% of the whole proteome (Stevens and Arkin, 2000) and are of pivotal interest to clinical proteomics since they play a crucial role in cell communication and signal transduction. Besides their potential to translate into targets for drug discovery, they have also become prime targets for biomarker discovery being easily accessible by antibody-based methods like fluorescence activated cell sorting (FACS). Monitoring the cell surface compartment in high-resolution was therefore of central importance when evaluating different protocols for blood cell proteomics. First, the feasibility of applying Glyco-PAL to freshly isolated CD4+ T cells was explored in order to identify and relatively quantify the surfaceome.

Two independent glycocapture experiments (referred to as test 1 and 2) were carried out using only one million freshly isolated CD4+ T cells each. Downscaling the cellular starting material was a major change to the original protocol which otherwise relied on several million cells as input (Graessel et al., 2015). Results from test 1 and 2 were compared to a pilot experiment which was performed a year before the start of this thesis as a proof-of-principle experiment based on the same downscaled protocol. In total 129 and 114 proteins could be identified in test 1 and 2, respectively. In comparison, 137 proteins were identified in the pilot experiment. Since glycocapturing mainly targets glycosylated plasma membrane proteins, the presence of TMD containing proteins was analyzed with Phobius software. The proportion of integral membrane proteins accounted for 16% in the pilot experiment and >30% in test 1 and 2 (Fig. 16A). Next, the

amount of plasma membrane proteins among all identifications was evaluated by GO annotation (Fig. 16B). The number of annotated proteins was >70 in both test 1 and 2 indicating successful enrichment of the cell surface compartment (corresponding to enrichment factors of 58% and 64%). Pathway enrichment analysis using Generanker further demonstrated that proteins belonging to prominent T cell signaling pathway, e.g. *T cell receptor* or *Interleukin 2*, were enriched in both tests and to a certain extent also in the pilot experiment (Tab. 8). In summary, the conducted experiments demonstrated the feasibility of applying glycocapture proteomics to one million CD4+ T cells achieving enrichment scores of approximately 60%. Nonetheless, a substantial amount of identified proteins were of intracellular origin and can be considered as "background" carried over from the cytosol. The discrepancy between the pilot experiment and test run 1 and 2 can be attributed to multiple factors such as differences in blood donor, experimental operator, number of replicates, MS performance as well as the large time span in between the experiments.
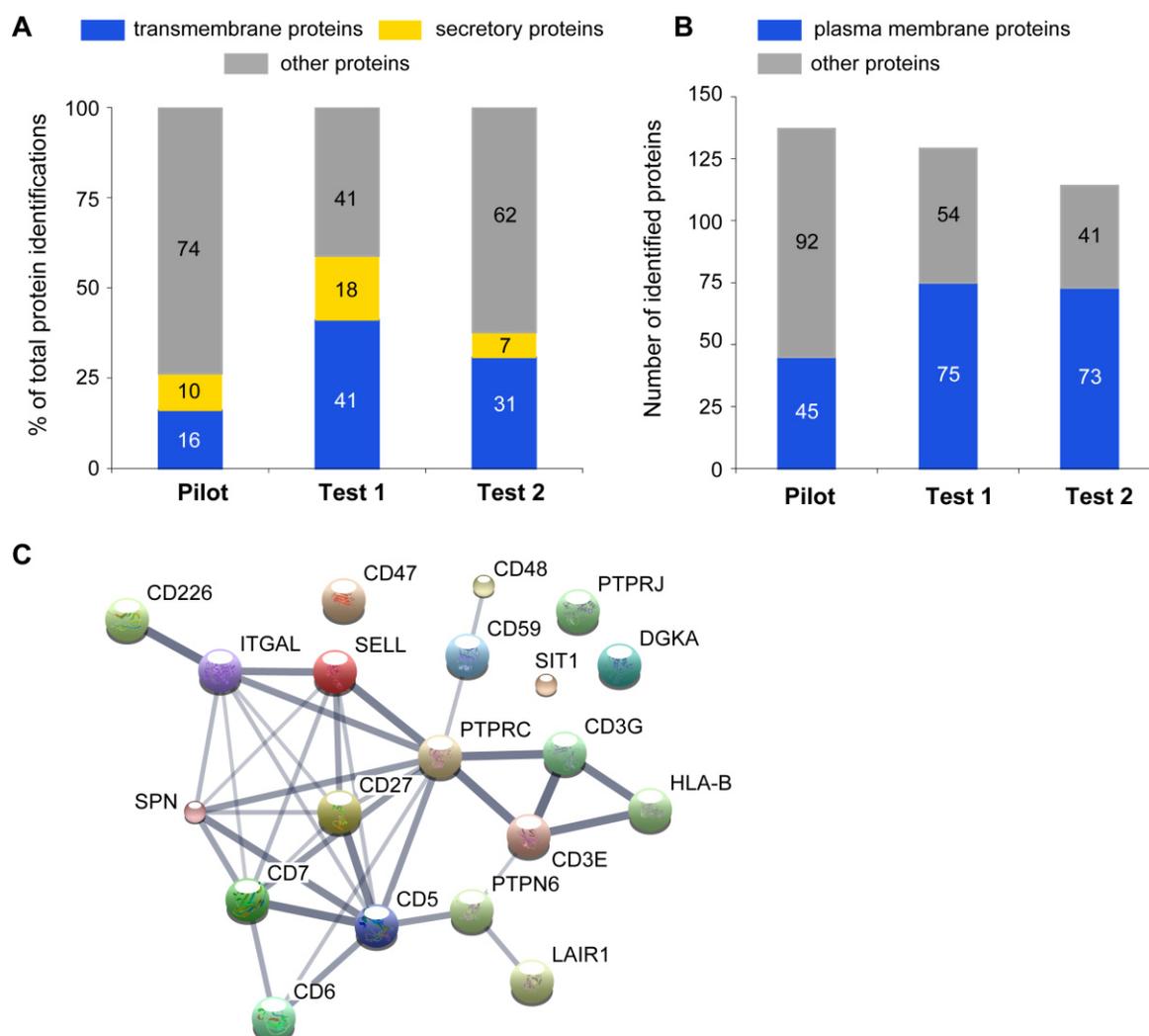
Figure 16 Surfaceome profiling of 1 million CD4+ T cells using glycocapture proteomics.

(A) Proportion of transmembrane, secretory and other proteins in the pilot experiment and two subsequent independent glycocapture experiments (test 1 and 2) as predicted by Phobius combined transmembrane and signal peptide predictor. (B) GO annotation reveals the absolute number of plasma membrane proteins in all three experiments. All proteins annotated as GO *plasma membrane (part)* and GO *cell surface* were combined in the category plasma membrane proteins. (C) STRING network of all identified proteins (20) which were annotated as *T cell receptor* in the Generanker analysis in one or more of experiments. GO, gene ontology.

Table 8 Gene Ontology (GO) and pathway enrichment analysis in CD4+ T cell protein sets generated by Glyco-PAL LC-MS/MS.
proteins refers to the number of observed proteins which belong to a given enriched pathway; p-value, enrichment score.

Gene Ontology/Pathway	Pilot (137)		Test 1 (129)		Test 2 (114)	
	p-value	# proteins	p-value	# proteins	p-value	# proteins
plasma membrane	not enriched		$2.19e^{-12}$	73	$1.88e^{-14}$	70
cell surface	$5.19e^{-05}$	17	$1.86e^{-19}$	35	$1.48e^{-15}$	29
Integrin	not enriched		$1.53e^{-19}$	25	$6.93e^{-21}$	26
T cell receptor	not enriched		$4.11e^{-08}$	14	$1.20e^{-05}$	11
Interleukin 2	not enriched		$3.07e^{-07}$	11	$2.37e^{-06}$	10
CD2	$1.12e^{-04}$	6	$2.71e^{-11}$	11	$1.89e^{-07}$	8
CD28	not enriched		$3.74e^{-05}$	7	$3.50e^{-05}$	7
Lymphocyte specific protein tyrosine kinase	$1.46e^{-03}$	6	$7.92e^{-06}$	8	$6.63e^{-05}$	7
Zeta chain (TCR) associated protein kinase 70	$8.26e^{-04}$	6	$3.04e^{-04}$	6	$2.87e^{-04}$	6
Protein tyrosine phosphatase, receptor type	not enriched		$3.95e^{-04}$	7	$5.48e^{-05}$	8
Focal adhesion kinase 1	not enriched		$5.96e^{-04}$	9	$5.52e^{-04}$	9

3.1.2 Evaluation of universal sample preparation methods

Protein expression can be explored across various cellular compartments using whole-proteome profiling approaches. The intended protocols do not rely on large amounts of viable cells for subproteome enrichment but can rather be performed with smaller amounts of frozen cells. The latter is an important advantage over the Glyco-PAL protocol by significantly reducing the handling time of viable cells. To this end, I compared the state-of-the-art method FASP to the more recently published iST strategy (Kulak et al., 2014) for profiling of PBMCs and CD4+ T cells. The iST method other than the FASP protocol employs peptide fractionation subsequent to protein digest to enhance protein and peptide identification rates. Both protocols were assessed using 10 µg whole cell lysate corresponding to approximately 170,000 PBMCs and 250,000 CD4+ T cells, respectively.

For PBMCs, LC-MS/MS resulted in the identification of 1244 proteins using FASP compared to 1693 identified proteins with the iST method. The iST method slightly

Results

outperformed FASP by identifying >600 proteins in addition to the 1035 overlapping identifications (Fig. 17A, top left). However, on peptide level this trend was not as clear as on protein level (Fig. 17A, bottom left). For CD4+ T cells, the best LC-MS/MS results were achieved with the iST method indicated by higher protein and peptide identification rates (Fig. 17A, right). In detail, 1584 proteins could be identified with the iST method including 587 proteins which were not identified by FASP. Additionally, iST clearly outperformed FASP regarding the total number of identified unique peptides (6192/4390). Next, proteomic profiling of CD4+ T cells using Glyco-PAL and whole-proteome approaches (FASP, iST) were compared concerning identified plasma membrane proteins (Fig. 17B). As a measurement for identified plasma membrane proteins absolute numbers of predicted proteins containing ≥1 TMD(s) and a secretory signal peptide were compared (Phobius). Intriguingly, absolute numbers of predicted plasma membrane proteins were comparable in all four experiments indicating that glycocapture proteomics with one million CD4+ T cells did not enable a deeper profiling of the surfaceome than conventional methods. Taken together, proteomic profiling of blood cells achieved best results with the iST method which was subsequently established as SOP of choice for biobanked clinical samples.
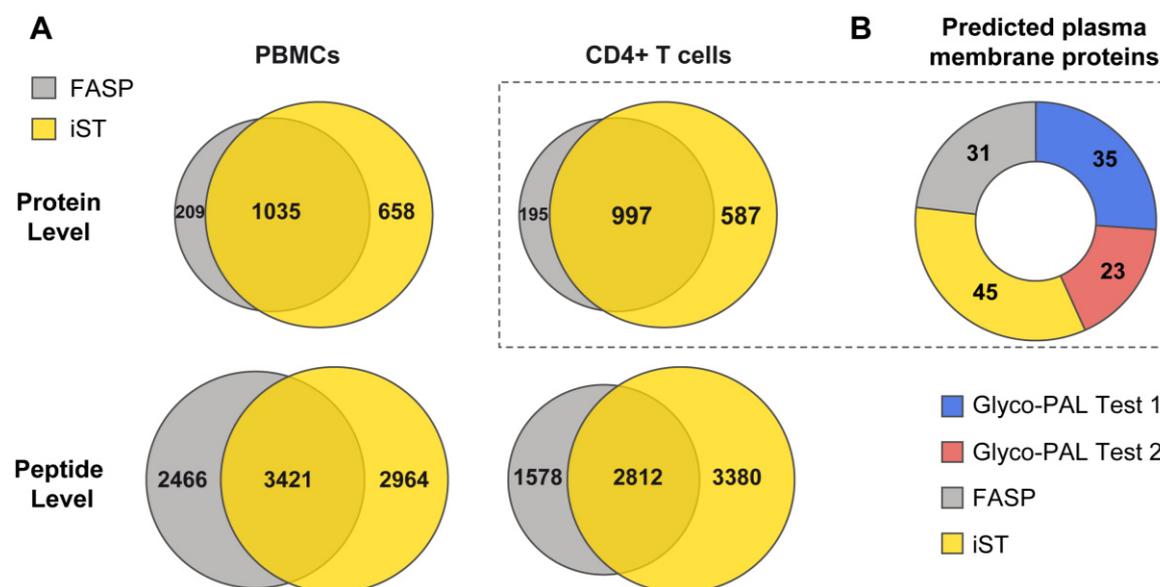
Figure 17 Comparison of proteomic profiling of PBMCs and CD4+ T cells using FASP and iST sample preparation.
(A) Subjecting 10 µg PBMC and CD4+ T cell protein lysate to FASP- and iST-based sample preparation with subsequent LC-MS/MS resulted in the identification of over 1000 proteins and more than 4000 unique peptides each. Direct comparison of both sample preparation strategies revealed that iST was slightly more sensitive with a higher number of total proteins identified and a greater amount of unique peptides detected. (B) Comparison of Phobius predicted plasma membrane proteins in the CD4+ T cell data sets (dashed rectangular). Predicted plasma membrane proteins were classified as containing ≥1 transmembrane domain(s) and a secretory signal peptide.

Before applying the iST method to a first set of clinical samples, the protocol was evaluated regarding further downscaling of cellular input material and technical reproducibility using CD4+ T cells. Since biobank storage adversely affects the viability of cells, a highly sensitive SOP applicable to minimal amounts of protein lysate was desired. To this end, a reproducibility test consisting of five independent technical replicates was performed using only 5 µg protein lysate (corresponding to approximately 125,000 cells). In this case, technical reproducibility referred to the entire iST workflow including cell lysate generation, iST processing and label-free LC-MS/MS analyses. The five replicates were obtained from separate CD4+ T cell lysates of identical biological origin.

Label-free quantitative LC-MS/MS analysis revealed a high correlation of measured protein abundances across all replicates ($\text{Ø}R^2$=0.96, Fig. 18). Halving the amount of processed CD4+ T cell lysate still resulted in the identification of over 1000 proteins, although a minor decline in protein (1194 vs. 1584) and peptide (4141 vs. 6192) identification sensitivity was apparent in comparison to 10 µg runs.
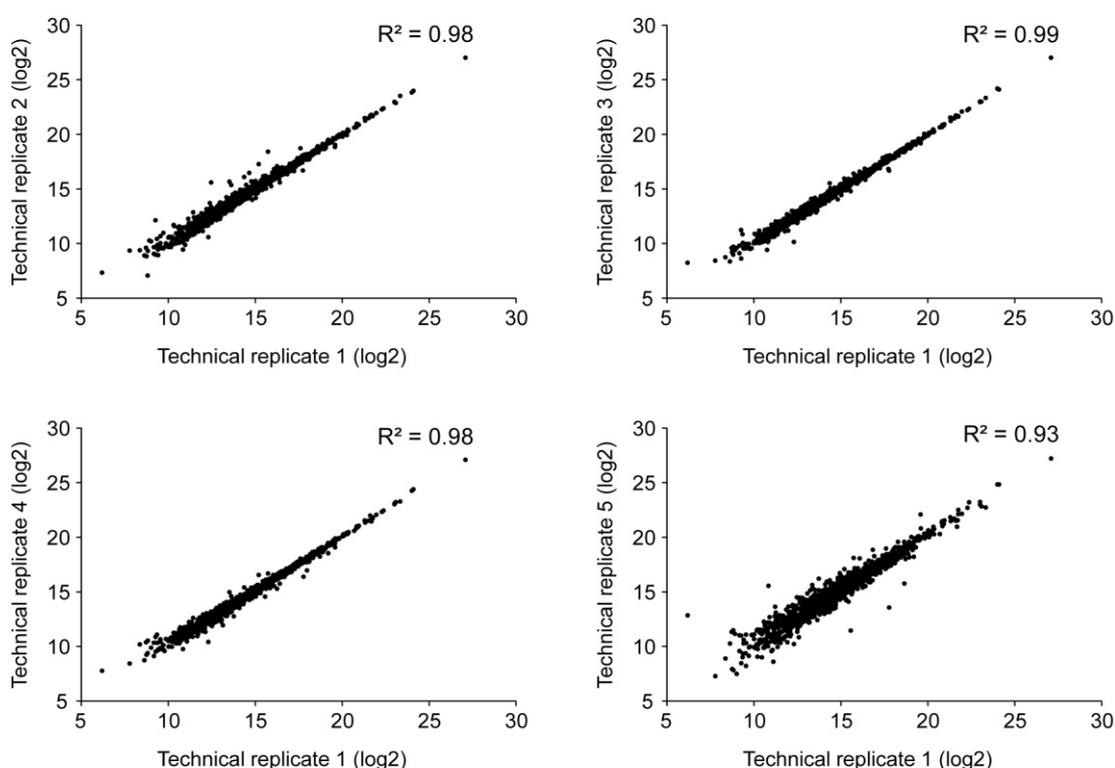
Figure 18 Quantitative reproducibility of the iST workflow across 5 independent approaches processing 5 µg CD4+ T cell lysate.
Label-free LC-MS/MS analysis resulted in the identification and relative quantification of 1194 proteins, corresponding protein abundance values highly correlated across all replicates.

Results

Exemplarily depicted are log values of protein abundance units measured in replicate 1 plotted against the respective values in replicate 2-5. R^2, coefficient of determination.

Plotting coefficient of variation (CV) values against their respective observed frequency revealed a log normal distribution displaying on average a workflow-induced measurement variation of 22% (Fig. 19A). The entire dynamic range of measured protein expression encompassed 6 orders of magnitude; as expected there was a tendency towards more robust measurement in highly abundant proteins, such as actin (ACTB), accounting for a CV of only 5% (Fig. 19B).
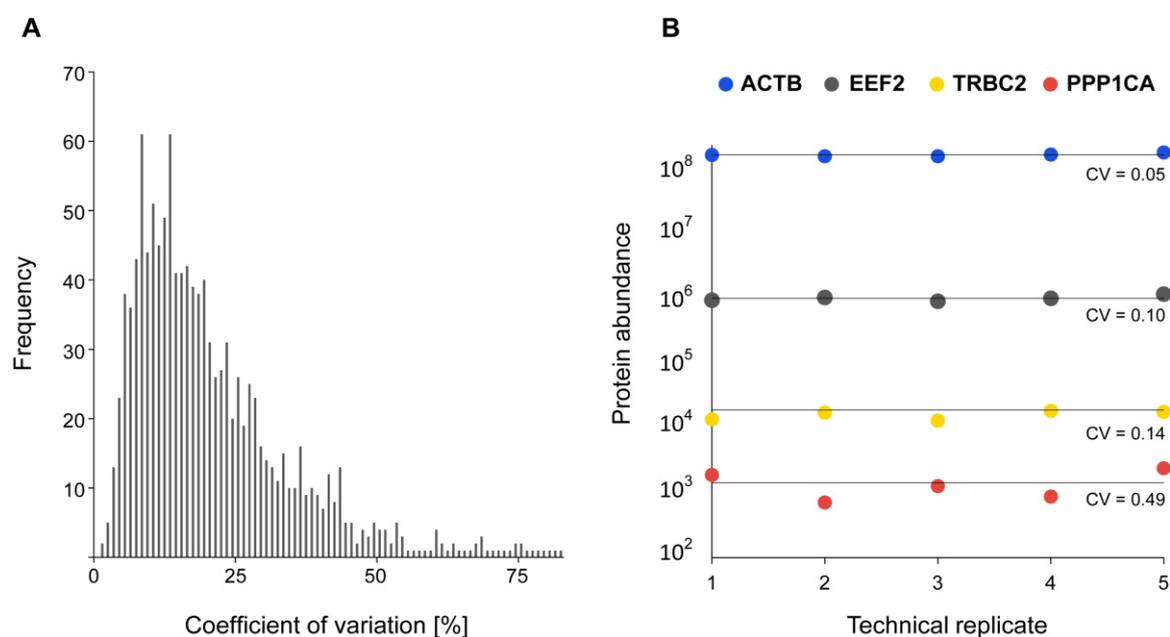
Figure 19 Coefficient of variation (CV) distribution across all quantified proteins and MS signals of four proteins representing the dynamic range of measured protein expression.
(A) CV scores were calculated for each quantified protein and plotted against their observed frequency; on average a CV of 22% was registered. (B) The dynamic range of measured protein expression is displayed by four proteins representing high, medium and low abundance. ACTB, actin beta; EEF2, eukaryotic translation elongation factor 2; TRBC2, T cell receptor beta constant 2; PPP1CA, protein phosphatase 1, catalytic subunit, α isozyme. CV, coefficient of variation.

Taken together, the iST method proved sensitive identification and reproducible quantification of >1000 cellular proteins when applied to limited cell counts of PBMCs and CD4+ T cells.

3.1.3 Purity of CD4+ T cell preparations

Before applying the iST method to patient-derived blood cells, different CD4+ T cell isolation strategies using MACS were evaluated. Purification of CD4+ T cells from PBMCs *via* MACS can be achieved by either positive or negative selection. In positive sorting, magnetic beads coated with anti-CD4 antibodies are used to specifically bind and enrich CD4+ T cells. In negative mode, CD4+ T cells are left untouched while all

CD4- cells present in a PBMC preparation are depleted by an antibody cocktail. Both positive and negative sorting where evaluated for CD4 purity with freshly isolated PBMCs from healthy blood donors (Fig. 20). Flow cytometry data gathered from four independent positive and two independent negative sorting experiments revealed that CD4+ T cell preparations achieved on median 90% purity in positive and 66% purity in negative mode, respectively. Consequently, MACS using positive CD4+ T cell isolation was incorporated into the SOP due to its significantly higher rate of purity.
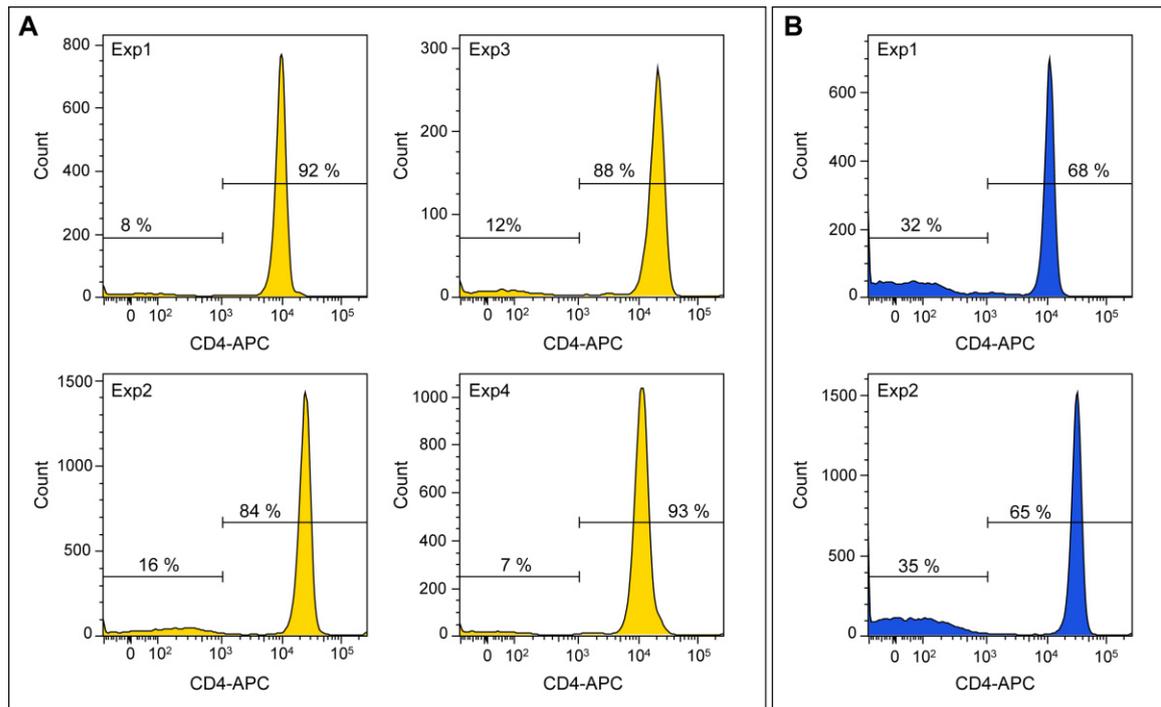
Figure 20 Purity of CD4+ T cell preparations using positive and negative magnetic activated cell sorting.
Quality control flow cytometry assessing CD4 expression in CD4+ T cell preparations obtained by positive (A) and negative (B) magnetic activated cell sorting (MACS). PBMCs from healthy blood donors were subjected to the CD4+ T cell isolation workflow using the positive or negative ("untouched") isolation mode. CD4+ T cells were stained with the CD4-APC antibody and CD4 expression was measured using flow cytometry. The CD4+ T cell fraction obtained by positive sorting was on median 90% pure, while the CD4+ T cell fraction isolated with the untouched strategy was on median 66% pure. Exp1-4, independent experiment 1-4.

3.2 Proteomic profiling of CD4+ T cells

An overview of the proteomics workflow including all label-free LC-MS/MS experiments performed with CD4+ T cells isolated from patient- and control-derived PBMC samples is provided in Fig. 21. Two independent clinical sample sets were processed consecutively; in both experiments biobanked PBMC samples of the T1D and control groups were thawed and enriched for CD4+ T cells *via* MACS (positive mode). The first set of CD4+ T cell samples, encompassing 23 newly diagnosed patients with T1D and 31 healthy controls, was processed with the iST SOP followed by DDA LC-MS/MS (Orbitrap XL).

Results

The second set of CD4+ T cell samples, consisting of 30 patients with T1D and 30 healthy controls, was processed with FASP due to its better compatibility with the then available DIA LC-MS/MS workflow. The second sample set was measured in both DDA and DIA LC-MS/MS (both on Q Exactive HF). The general characteristics of the study population are summarized in Tab. 1.
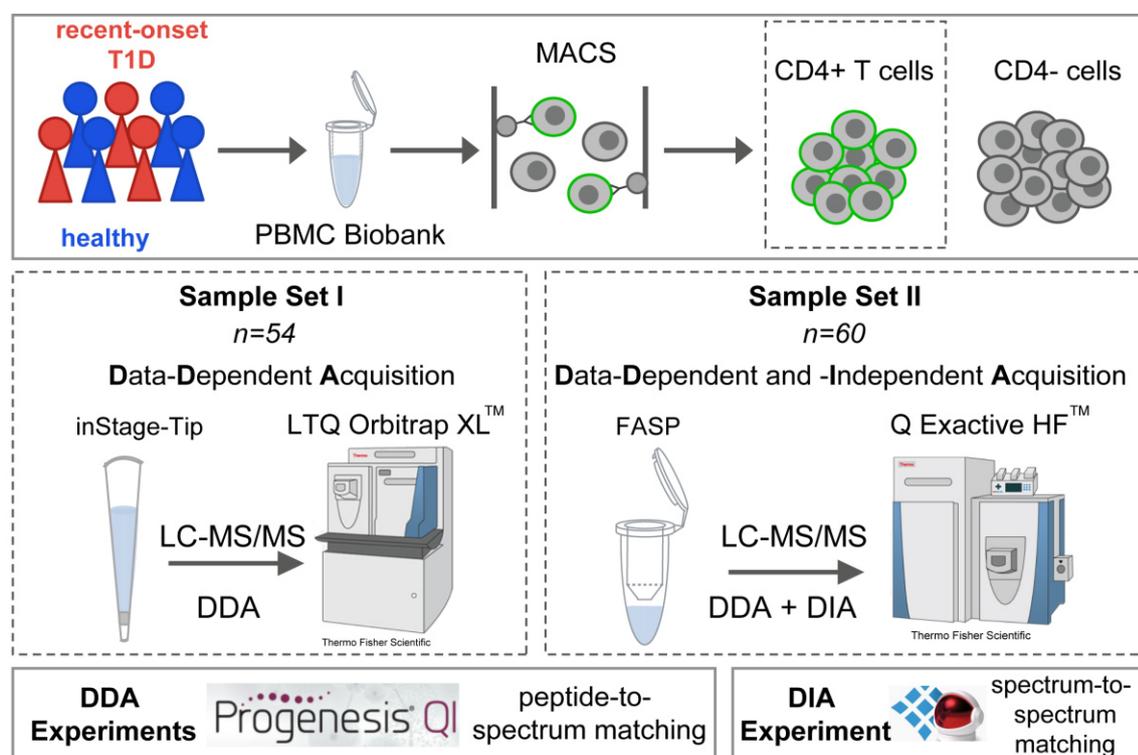
Figure 21 Overview of LC-MS/MS experiments conducted on CD4+ T cells isolated from biobanked PBMCs from sample set I and II.
Patient- and control-derived CD4+ T cells were isolated by MACS positive sorting from biobanked PBMC samples. The first set of CD4+ T cell samples (54) was processed with the iST SOP followed by DDA LC-MS/MS. The second set of CD4+ T cell samples (60) was processed with FASP and first measured in DDA and subsequently in DIA LC-MS/MS. Data processing was performed with Progenesis QI software for DDA experiments and with Spectronaut for the DIA experiment. DDA, data-dependent acquisition; DIA, data-independent acquisition; MACS, magnetic activated cell sorting; *n*, sample size.

3.2.1 Data-dependent acquisition (DDA) LC-MS/MS

3.2.1.1 CD4+ T cells from sample set I

After ensuring good quality of both CD4+ T cell preparation purity and quantitative reproducibility of the iST SOP, biobanked PBMC samples of the T1D and control groups from sample set I were thawed and enriched for CD4+ T cells *via* MACS (positive mode). After collection of the CD4+ T cell fraction, cells were counted, lysed and protein content was determined using BCA assay. Plotting CD4+ T cell count against the obtained protein yield revealed a linear correlation (R^2=0.84, Fig. 22). On average, CD4+ T cell

count was 455,000 (Ø 16 µg protein) influenced by the number of cryopreserved cells per vial and the number of vials processed per patient. Subsequently, all samples yielding at least 8 µg protein lysate were processed for DDA LC-MS/MS with the iST SOP; the processed cell number per sample (~250,000) reflected the target cell count simulated during SOP development. The iST method was conducted with peptide-based fractionation into three distinct pools to enhance proteome coverage.
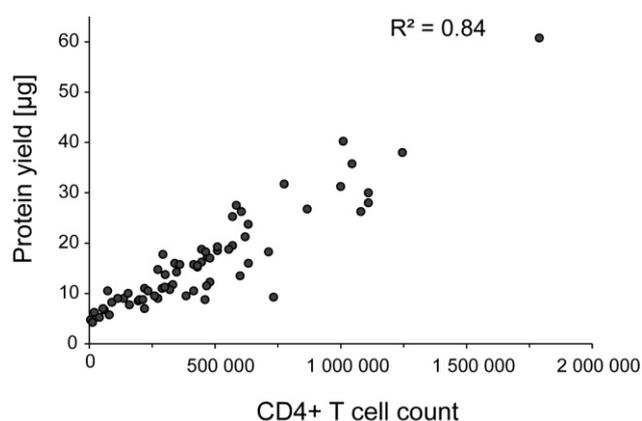
Figure 22 Protein yield in CD4+ T cell preparations derived from PBMC sample set I.
Biobanked PBMC samples were subjected to MACS positive isolation. The resulting CD4+ T cell fraction was counted, lysed and protein content was determined using BCA assay. CD4+ T cell count and protein yield correlated highly. R^2, coefficient of determination.

Using label-free protein quantification based on peptide precursor intensity, 3248 proteins including 1656 proteins with ≥2 unique (i.e. proteotypic) peptides could be quantified. One-peptide-based protein identifications accounted for 49% of all identifications and were excluded from T1D candidate protein examination in the subsequent statistical analyses. To determine significant differences in protein expression between the T1D patients and the control group, an unpaired Student's t-test was performed. The analysis revealed that 164 proteins were differentially expressed ($p<0.05$); 25 of these proteins had ≥2-fold difference in expression (Fig. 23A). Next, multiple hypothesis testing was corrected for using SAM in order to comply with a FDR of <5% (Fig. 23B); the analysis revealed that 35 proteins upregulated in the T1D group met this criterion.

Transforming and plotting protein abundance data in principle component space allows for separation of samples according to abundance variation. Thus, uniform changes in the proteomic composition of samples within a given phenotypic group would lead to cluster formation. The analyzed CD4+ T cell samples did not cluster according to disease status using all possible features in a PCA (Fig 23C).
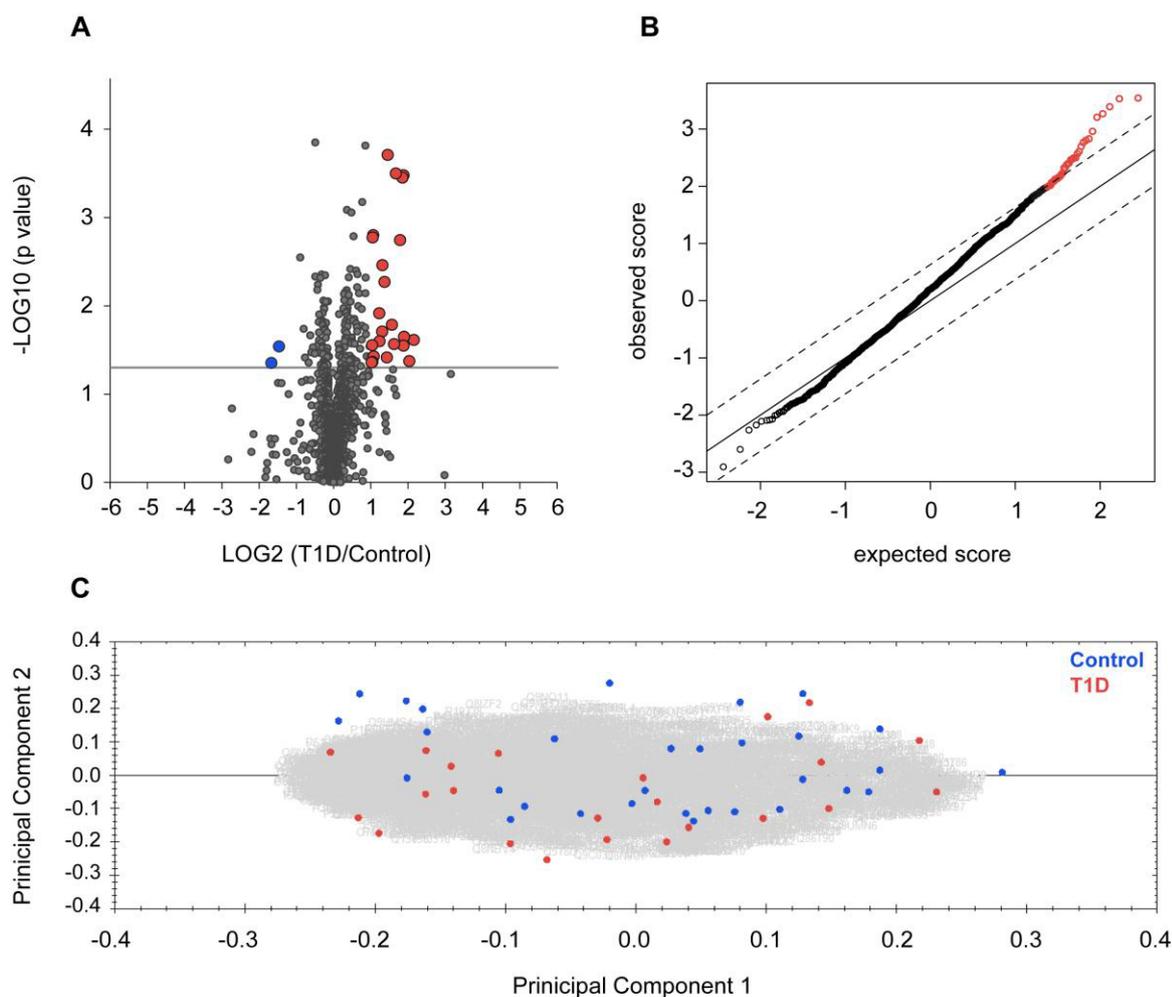
Figure 23 Differential protein expression in CD4+ T cells of sample set I measured in DDA LC-MS/MS.
(A) Volcano plot illustrating the fold change of expressed proteins (≥2 unique peptides) between the T1D and control groups plotted against respective statistical significance before correcting for multiple testing. The horizontal gray line indicates the p-value cut-off of 0.05. Proteins with a mean fold change ≥2 (in either direction) and p<0.05 are shown in red (higher levels in the T1D group) and blue (lower levels in the T1D group). (B) SAM scatter plot showing significantly regulated proteins applying the FDR threshold of 5%. Proteins that were significantly higher in the T1D group are shown in red. No cut-off was applied to the fold change in expression. (C) PCA of all samples based on all features indicates no group-specific separation. SAM, significance analysis of microarrays; FDR, false discovery rate; PCA, principal component analysis.

Pathway enrichment analysis of the 35 significantly different proteins (q<0.05) identified a cellular network consisting of 13 upregulated proteins in the T1D group (Fig. 24A), including inflammatory response proteins such as Elastase (ELANE), Myeloblastin (PRTN3), Myeloperoxidase (MPO), and S100 calcium-binding protein A9 (S100A9) (*Inflammatory*; enrichment score p=9.9e-04). The analysis further revealed that many of the investigated proteins are commonly found in the extracellular compartment, indicating a potential spill-over effect from the plasma (*Extracellular space*; p=8.8e-04). This hypothesis is supported by the enrichment of proteins annotated as *Serum Specimen* (p=1.5e-04), including Alpha 2-HS glycoprotein (AHSG) and Ribonuclease 3

(RNASE3), as well as the presence of neutrophil granule proteins such as MPO (*Neutrophil*; p=1.3e-03). The protein levels of five candidate markers that exemplify the T1D-associated signature in the cellular network are plotted in Fig. 24B; the remaining 30 significantly different proteins are plotted in Fig. 25.

Next, the overall abundance of these differentially regulated proteins was plotted against their respective ranking in the dynamic range of protein expression. The analysis revealed that the majority of network proteins were in the top third of measured protein abundance (Fig. 24C-D), indicating their prominent presence in the CD4+ T cell compartment.
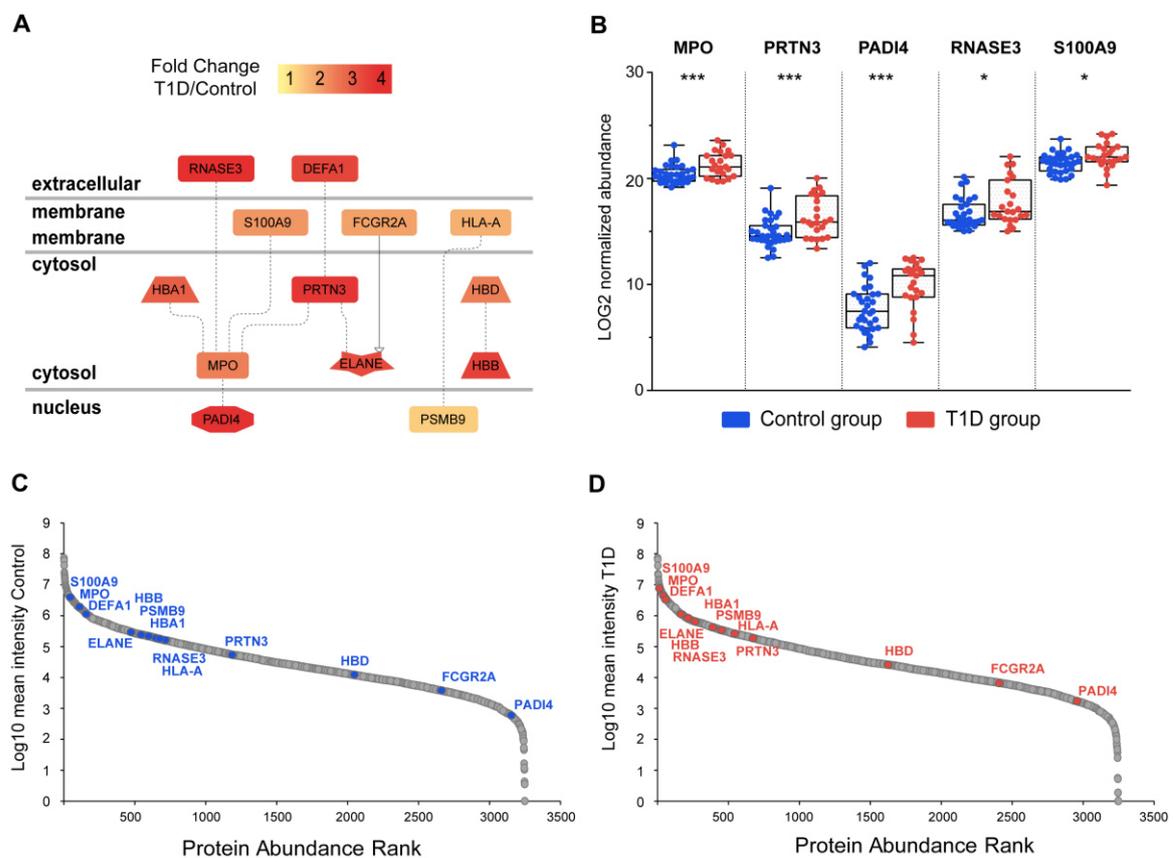
Figure 24 Visualization of differential protein expression in CD4+ T cells of sample set I measured with DDA LC-MS/MS.
(A) Genomatix network of 13 significantly regulated proteins (≥2 unique peptides, q<0.05) connected at the function level. Dotted lines represent interactions via co-citation; the solid line between FCGR2A and ELANE indicates an interaction confirmed by expert curation. The color legend indicates the log2 mean fold change of all cluster proteins. (B) Box plot summarizing the differential levels of expression of select candidate proteins between patients with T1D and the control group. The boxes represent the interquartile range (IQR), and the median value is shown as a solid black line within the box. The upper and lower whiskers represent the maximum and minimum value, respectively. *q<0.05 and ***q<0.001. (C and D) Dynamic range of proteins measured in the control group and T1D groups. Select proteins of the network depicted in (A) are highlighted with their respective abundance for control (blue) and T1D group (red).
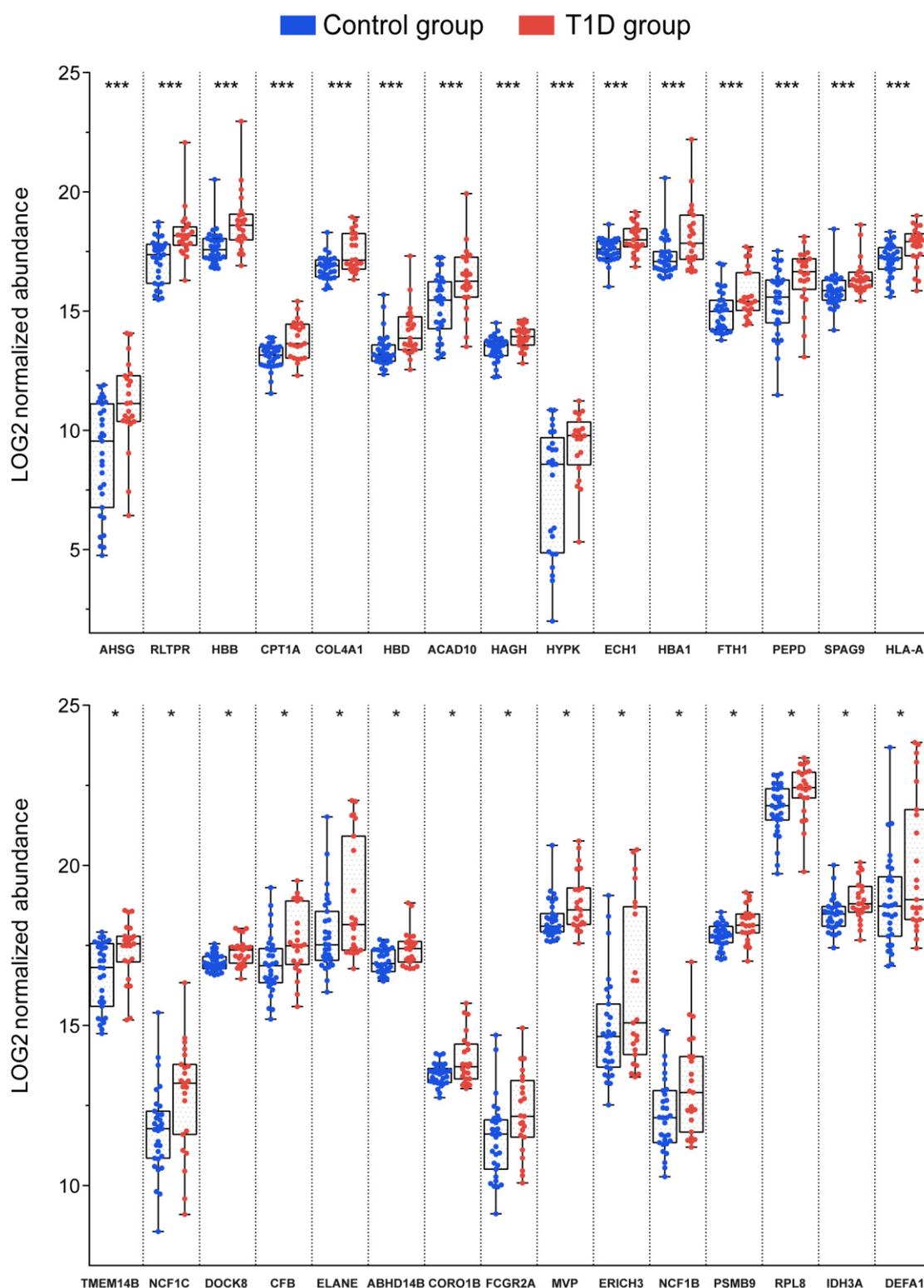
Figure 25 Box plots of proteins with significantly different expression in CD4+ T cells of sample set I.
Assessment of differential CD4+ T cell protein expression between the T1D and control groups using SAM resulted in 35 significant proteins (FDR<5%) for sample set I. All 30 candidate proteins that were not shown in Figure 24B are shown here. If the abundance of a protein for a patient or control was "0", the data point was omitted from the plot. The boxes represent the interquartile range (IQR), and the median value is shown as a solid black line within the box. The upper and lower whiskers represent the maximum and minimum value, respectively. *q<0.05 and ***q<0.001. SAM, significance analysis of microarrays; FDR, false discovery rate.

3.2.1.2 CD4+ T cells from sample set II

The next step was to confirm the discovered changes in protein expression by profiling a second, independent set of CD4+ T cell samples. The CD4+ T cell isolation procedure described for sample set I was repeated for 60 additional biobanked PBMC samples. Resulting CD4+ T cell pellets were processed for LC-MS/MS with FASP other than the iST SOP for two reasons: first the newly acquired Q Exactive HF machine resulted in much better identification rates than the Orbitrap XL making iST-based peptide fractionation dispensable and secondly, FASP produced good results with only one peptide fraction – a much more suitable setting for the desired DIA LC-MS/MS experiment. CD4+ T cell samples obtained from sample set II were then initially profiled using DDA LC-MS/MS which is described in the following section and later with DIA LC-MS/MS (see section 3.2.2).

Label-free protein quantification for CD4+ T cell samples measured in DDA mode resulted in the identification of 3546 proteins encompassing 2484 proteins with ≥2 unique peptides. Protein identification rates were comparable to sample set I with 1883 proteins found in both data sets indicating high reproducibility (Fig. 26A, left). The proportion of one-peptide-hits could be significantly reduced in sample set II attributable to the enhanced MS performance with the next-generation mass spectrometer (Fig. 26A, right; see also Fig. 21).

To determine significant differences in protein expression between both groups, an unpaired Student's t-test was performed. The analysis revealed that 347 proteins were differentially expressed ($p<0.05$); 56 of these proteins had ≥2-fold difference in expression (Fig. 26B). Next, multiple hypothesis testing was corrected for using SAM in order to comply with a FDR of <5% (Fig. 26C); the analysis revealed that 74 proteins (65 up- and 9 downregulated) met this criterion. Similar to the observations from sample set I, the majority of candidate proteins were upregulated in the T1D group (see also Fig. 23).
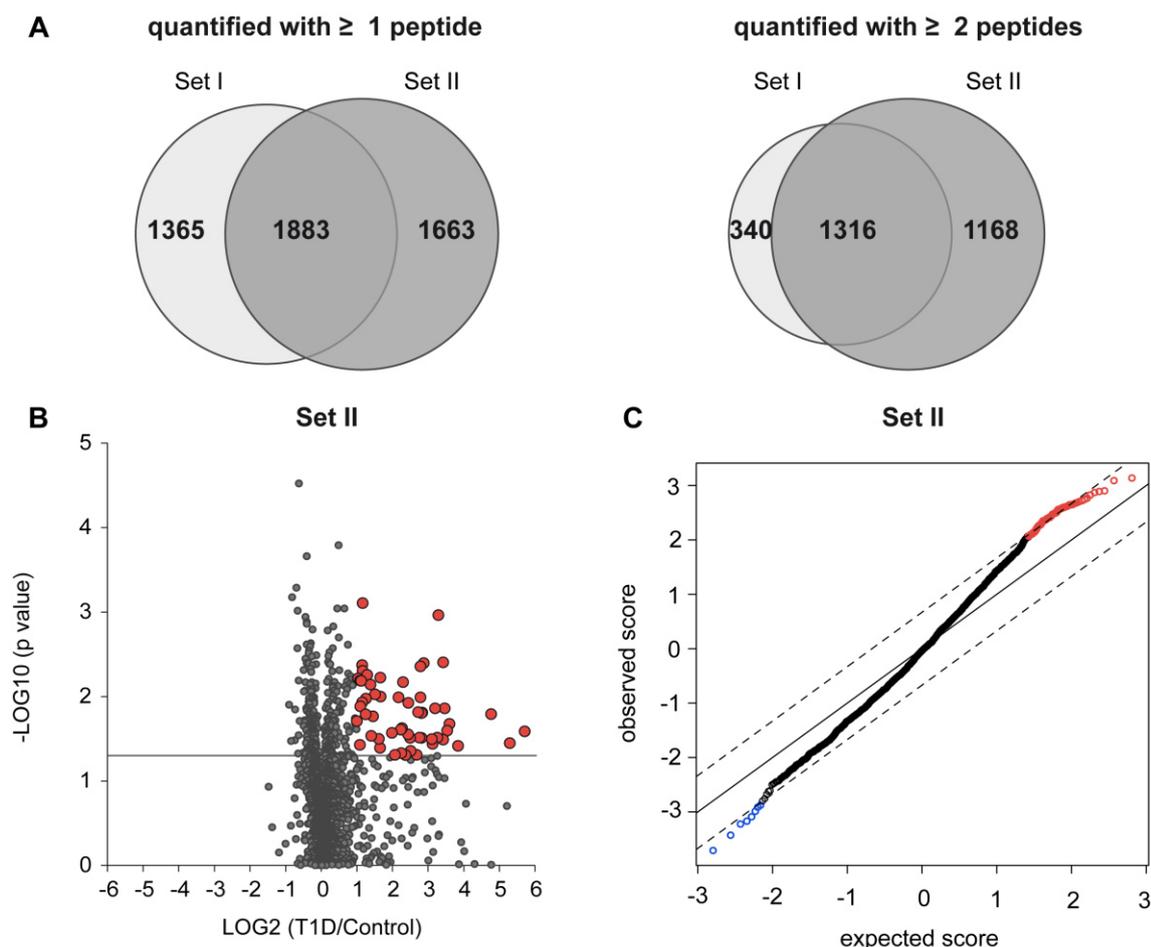
Figure 26 Protein identifications in CD4+ T cells of sample sets I and II and differential protein expression in sample set II.
(A) Total number of identified proteins from CD4+ T cells that were relatively quantified with ≥ 1 unique peptide (left panel) or ≥ 2 unique peptides (right panel) for sample sets I and II. (B) Volcano plot illustrating the fold change of expressed proteins (≥2 unique peptides) between the T1D and control groups plotted against respective statistical significance before correcting for multiple testing in sample set II. The horizontal gray line indicates the p-value cut-off of 0.05. Proteins with a mean fold change ≥2 (in either direction) and p<0.05 are shown in red (higher levels in the T1D group) and blue (lower levels in the T1D group). Three data points (p>0.05) were omitted from the graphic due to axis truncation. (B) SAM scatter plot showing significantly regulated proteins applying the FDR threshold of 5% for sample set II. Proteins that were significantly higher in the T1D group are shown in red. No cut-off was applied to the fold change in expression. SAM, significance analysis of microarrays; FDR, false discovery rate.

An overview of all proteins which were differentially regulated (p<0.05) in both sample sets is provided in Tab. 9 comprising 16 up- and 2 downregulated proteins. These 18 proteins are dominated by myeloid-derived (inflammatory) mediators such as Defensin α1 (DEFA1), Grancalcin (GCA), Proteoglycan 2 (PRG2), MPO, PRTN3 and RNASE3. The overlap further comprises several proteins associated with peptide and/or antigen processing such as Cathepsin S (CTSS), Aminopeptidase B (RNPEP) and Dipeptidyl peptidase 7 (DPP7). Interestingly, also Beta-arrestin-1 (ARRB1) was slightly upregulated in the T1D group – a multifunctional adaptor protein which has been linked to the survival of autoreactive CD4+ T cells in patients with multiple sclerosis (Shi et al., 2007). In

contrast, Glutathione S-transferase kappa 1 (GSTK1) was mildly downregulated in both sample sets; this protein is a peroxisomal enzyme implicated in cellular detoxification.

Table 9 Overlap of proteins regulated with p<0.05 between T1D and control groups in CD4+ T cells of clinical sample sets I and II.
All proteins were quantified with ≥ 2 unique peptides in both data sets. The calculated fold change of T1D/control is indicated as mean and [median]. The Phobius column refers to the results of a combined transmembrane topology and signal peptide predictor (Y). Proteins highlighted in bold are candidate proteins regulated with q<0.05 in both data sets. TMD, transmembrane domain.

Accession ID	Protein name	Gene Symbol	Phobius		Sample Set I		Sample Set II	
			TMD	SP	p-value	Fold change	p-value	Fold change
Q7Z5R6	Amyloid beta A4 precursor protein-binding family B member 1-interacting protein	APBB1IP	0	0	0.026	1.4 [1.7]	0.023	1.1 [1.1]
P49407	Beta-arrestin-1	ARRB1	0	0	0.037	2.1 [1.0]	0.015	1.4 [1.4]
Q9UNZ5	Leydig cell tumor 10 kDa protein homolog	C19orf53	0	0	0.012	0.8 [0.9]	0.022	0.8 [0.9]
P25774	Cathepsin S	CTSS	0	Y	0.016	1.5 [1.4]	0.035	1.6 [1.4]
P59665	Neutrophil defensin 1	DEFA1	0	Y	0.027	3.1 [1.1]	0.037	2.1 [3.7]
Q9UHL4	Dipeptidyl peptidase 7	DPP7	0	Y	0.043	1.5 [1.8]	0.009	1.4 [1.2]
Q03001	Dystonin	DST	0	0	0.031	1.6 [1.7]	0.027	1.1 [1.0]
P28676	Grancalcin	GCA	0	0	0.028	2.1 [1.8]	0.026	1.7 [2.6]
Q9Y2Q3	Glutathione S-transferase kappa 1	GSTK1	0	0	0.005	0.8 [0.8]	0.038	0.9 [0.9]
P68871	Hemoglobin subunit beta	HBB	0	0	0.000	3.2 [2.0]	0.033	1.2 [2.6]
P46940	Ras GTPase-activating-like protein	IQGAP1	0	0	0.036	1.2 [1.0]	0.019	1.2 [1.2]
P05164	**Myeloperoxidase**	**MPO**	**0**	**Y**	**0.002**	**2.1 [1.9]**	**0.005**	**1.7 [2.2]**
Q92614	Unconventional myosin-XVIIIa	MYO18A	0	0	0.029	1.2 [1.1]	0.017	1.2 [1.3]
P13727	Bone marrow proteoglycan	PRG2	0	Y	0.025	2.4 [1.2]	0.010	1.3 [2.7]
P24158	**Myeloblastin**	**PRTN3**	**0**	**Y**	**0.002**	**3.5 [2.5]**	**0.007**	**2.6 [2.2]**
P12724	**Eosinophil cationic protein**	**RNASE3**	**0**	**Y**	**0.022**	**3.7 [1.8]**	**0.007**	**1.1 [5.4]**
Q9H4A4	Aminopeptidase B	RNPEP	0	0	0.007	1.2 [1.3]	0.045	1.7 [1.1]
Q99536	Synaptic vesicle membrane protein VAT-1 homolog	VAT1	0	0	0.011	1.3 [1.2]	0.003	1.3 [1.3]

Among those 18 differentially regulated proteins (p<0.05) only three candidates (MPO, PRTN3 and RNASE3) reached statistical significance after FDR<5% (q<0.05) adjustment (Fig. 27). All three candidates were myeloid-derived inflammatory response proteins, confirming this finding in CD4+ T cell preparations of both sample sets.
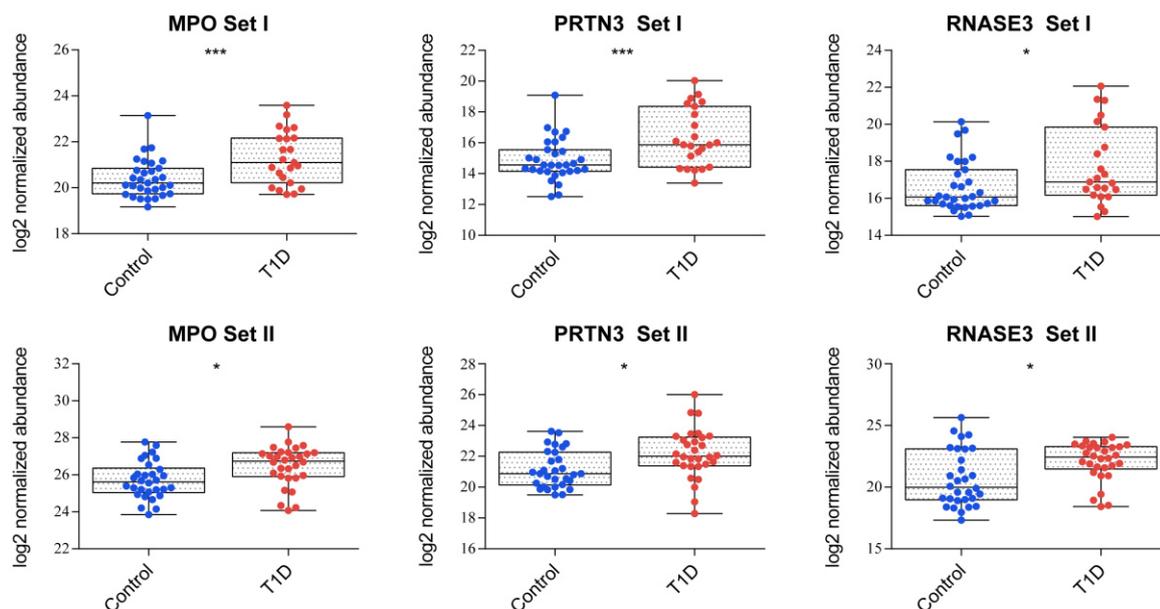
Figure 27 Boxplot of Proteinase 3 (PRTN3), Myeloperoxidase (MPO) and Ribonuclease-3 (RNASE3) protein expression in CD4+ T cell preparations of both clinical sample sets.
PRTN3, MPO and RNASE3 were the only three proteins that were significantly regulated in both data sets before and after FDR adjustment. The box represents the interquartile range (IQR); the median is shown as solid black line within the box. Whiskers extend from minimum to maximum of all data points. *q<0.05 and ***q<0.001.

3.2.2 Data-independent acquisition (DIA) LC-MS/MS

3.2.2.1 Comparison of proteome coverage in DDA and DIA data of CD4+ T cells

The DDA LC-MS/MS profiling experiment conducted on CD4+ T cells derived from sample set II identified a clear proteomic signature associated with T1D pathogenesis – confirming the main finding from sample set I. However, monitoring a subset of 2484 proteins with ≥2 unique peptides may not necessarily capture the entire range of in vivo protein dynamics, which reflects the common problem of "undersampling" in shotgun proteomics. To maximize our coverage and sensitivity, the very same CD4+ T cell samples were profiled again using a novel data-independent approach.

First, a comprehensive custom-made spectral library was generated using various human samples, including PBMCs, CD4+ T cells, and neutrophils; where possible, subcellular fractionation was used in order to increase the coverage of proteins in the nucleus and plasma membrane (Graessel et al., 2015; Zeng et al., 2009). The resulting spectral library included 11,403 protein groups and 432,673 peptide precursors (Tab. 7).

CD4+ T cell samples from sample set II were subsequently measured using DIA LC-MS/MS. Using Spectronaut software to obtain label-free quantification, 8160 proteins were identified with at least 1 unique peptide and 6377 proteins with ≥2 unique peptides,

representing a 2-fold increase in proteins identified using the DIA approach compared to the DDA approach (Fig. 28A). Specifically, 4914 more proteins were identified in addition to the 3246 proteins that were previously identified with the DDA approach (Fig. 28B). With respect to proteins identified with ≥2 unique peptides in the DDA approach, a recovery rate of 97% was achieved using the DIA approach, indicating high reproducibility. On average, the 3246 overlapping proteins were quantified based on 15 unique peptides each, reflecting an average increase of 8 peptides compared to DDA-based quantification (Fig. 28C).

Next, in silico Phobius analysis was used to categorize the identified proteins based on their subcellular localization (Fig. 28D). Overall, the distribution of proteins was similar between the DDA and DIA approaches; specifically, the majority of proteins were cytosolic, and approximately 25-30% were either membrane-bound or secreted proteins. The proportion of proteins containing transmembrane domains localizing either to the plasma membrane or to intracellular membranes could be slightly enhanced with DIA compared to DDA (1612 versus 528 proteins, representing 20% and 15% of total protein identifications, respectively). In particular, for examining the recovery rate of canonical surface markers in CD4+ T cells in the DDA and DIA approaches, I selected 19 classic proteins belonging primarily to the TCR complex and associated signaling pathways, and measured their proteomic coverage indicated by the number of identified unique peptides (Tab. 10). Four of these 19 marker proteins (TRGC1, CD28, NFATC1 and NFACT3) were identified only using the DIA approach. Furthermore, DIA also increased the peptide coverage of 13 additional proteins, including nuclear factor of activated T cells (encoded by NFATC2) and nuclear factor kappa-B (NFKB1/2).

Table 10 Prototypic CD4+ T cell markers and their respective proteome coverage indicated by the number of identified unique peptides in the DDA and DIA experiments for CD4+ T cell sample set II.
The T cell receptor complex, downstream signaling, and associated proteins are presented as example markers. - = protein not identified.

Accession ID	Protein name	Gene Symbol	Proteome Coverage (unique peptides)	
			DDA	DIA
P04234		CD3D	4	6
P07766	T cell surface glycoprotein CD3 δ/ε/γ chain	CD3E	6	8
P09693		CD3G	4	7
P01848		TRAC	2	3
P01850	T cell receptor alpha/beta/gamma chain C region	TRBC1	1	-
A0A5B9		TRBC2	2	1
P0CF51		TRGC1	-	1
P20963	T cell surface glycoprotein CD3 zeta chain	CD247	4	6
P06729	T cell surface antigen CD2	CD2	4	10
P01730	T cell surface glycoprotein CD4	CD4	8	12
P10747	T cell-specific surface glycoprotein CD28	CD28	-	2
P06239	Tyrosine-protein kinase LCK	LCK	15	20

Accession ID	Protein name	Gene Symbol	Proteome Coverage (unique peptides)	
			DDA	DIA
P43403	Tyrosine-protein kinase ZAP-70	ZAP70	22	34
O43561	Linker for activation of T cells family member 1/2	LAT	1	4
O95644		NFATC1	-	4
Q13469	Nuclear factor of activated T cells, cytoplasmic 1/2/3	NFATC2	4	22
Q12968		NFATC3	-	3
P19838	Nuclear factor NF-kappa-B p105/p100 subunit	NFKB1	21	33
Q00653		NFKB2	13	28

Finally, the refined CD4+ T cell proteome was compared with recently published studies of CD4+ T cells at the whole-proteome and surfaceome levels. The Human Proteome Map (HPM) project (Kim et al., 2014) profiled various human tissues, including CD4+ T cells obtained from three healthy donors. The HPM project used extensive peptide pre-fractionation (24 fractions) combined with DDA-based 1-hour gradient LC-MS/MS in order to achieve analytical depth. A qualitative comparison of proteins identified revealed nearly 6000 overlapping proteins, corresponding to an excellent rate of recovery (i.e., 72%) in my experiment (Fig. 28E). Importantly, the analysis also revealed a considerable amount of non-overlap between my approach and the HPM study with respect to CD4+ T cells. This may be attributed to differences in workflow, including the MS method used, the extent of peptide pre-fractionation, gradient length, and/or biological variability. Nevertheless, this comparison highlights the robust improvement in the coverage of CD4+ T cell proteins using the DIA approach, with only two hours of measuring time required per sample.

In addition, I compared the data set to a recently published surface atlas of naïve and activated CD4+ T cells obtained from healthy donors (Graessel et al. (2015); this atlas encompasses 229 proteins that map to 222 unique accessions (Fig. 28F). This surface atlas was generated using a combined omics approach with a flow cytometry marker panel in addition to LC-MS/MS–based surfaceome profiling. Strikingly, 73% of the published surface atlas was recovered using the DIA approach, without the need to use additional antibody-based methods, even though I used only a small fraction of the number of cells (250,000 cells compared to 8,000,000 cells used to generate the surface atlas). This suggests that using DIA LC-MS/MS can replace complex surfaceome profiling methods, which are often applicable only to large numbers of viable cells.
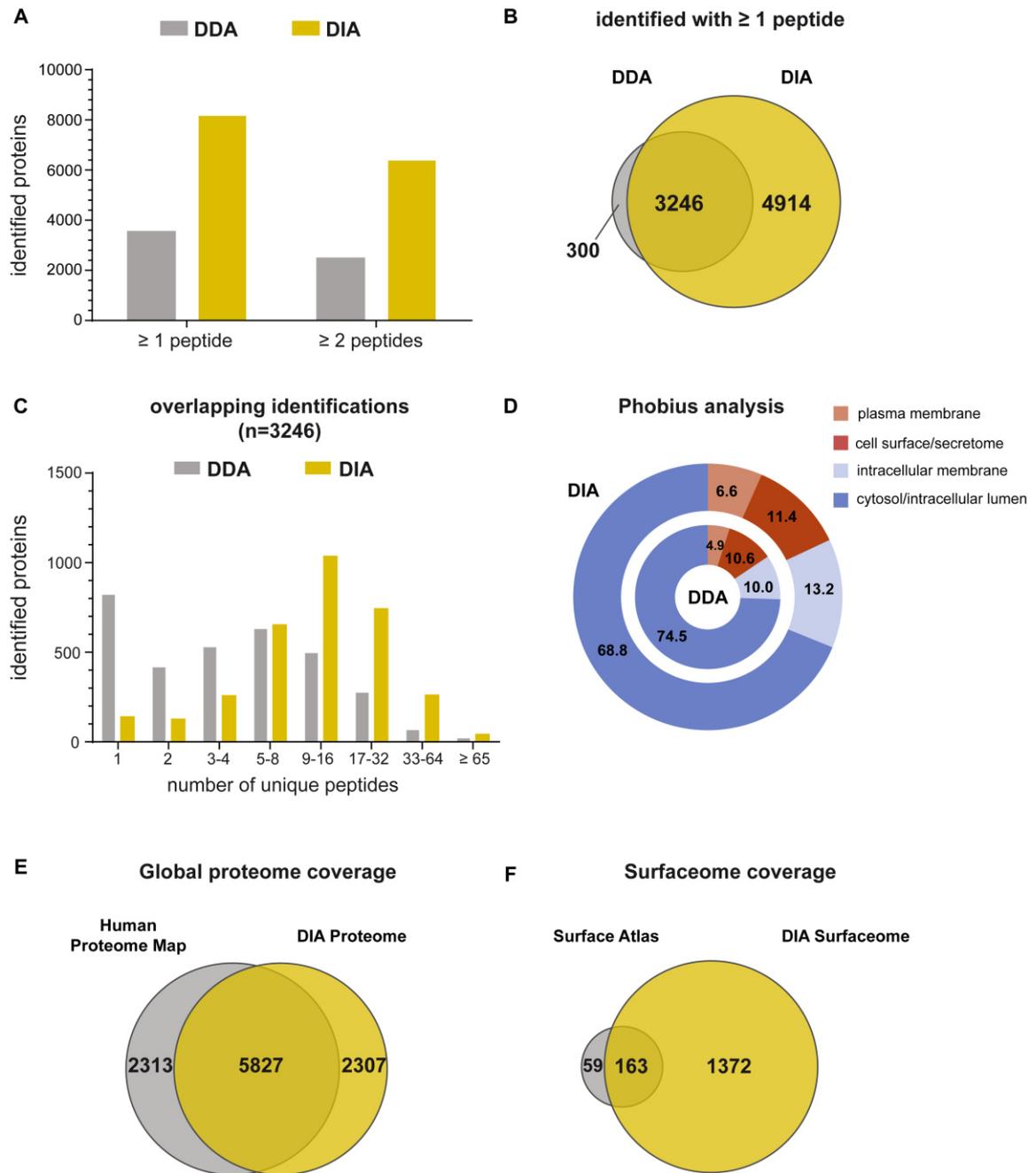
Figure 28 Comparison of the DDA and DIA data obtained from CD4+ T cell sample set II.
(A) Number of proteins identified in the DDA and DIA experiments with ≥1 or ≥2 unique peptides.
(B) Venn diagram showing the proteins identified in the DDA and/or DIA experiments (≥1 unique peptide). (C) Summary of peptide per protein coverage for overlapping protein identifications. (D) Overview of the subcellular localization of all proteins identified in the DDA and DIA experiments. Subcellular localization was determined using Phobius combined transmembrane topology and signal peptide prediction. (E) Venn diagram showing the CD4+ T cell proteins identified in the DIA experiment and the Human Proteome Map (HPM) project. (F) Venn diagram showing the surfaceome of naïve and activated CD4+ T cells in the published surface atlas (Graessel et al., 2015) and our DIA surfaceome. The DIA surfaceome includes the entire overlap of proteins identified with the surface atlas and all plasma membrane/cell surface proteins identified in the DIA study (Phobius prediction).

Results

Taken together, these results indicate that DIA LC-MS/MS can provide a comprehensive, high-resolution immunoproteomic profile of human CD4+ T cells, which is a key prerequisite for differentiating patient groups based on disease-associated proteomic signatures.

3.2.2.2 Differential protein expression in CD4+ T cell DIA data

Having obtained a high-resolution CD4+ T cell proteome, the next step was to perform hierarchical cluster analysis of global protein expression in order to examine whether T1D onset affected the global composition of the CD4+ T cell proteome, leading to a T1D-associated phenotype (Fig. 29). The resulting heatmap revealed that patient clustering did not reflect either a disease-specific or gender-specific phenotype, indicating that global biological variations had a stronger influence in the study population than proteomic changes accompanied by T1D onset. The average (median) CV in the control group was 48% (33%), and thus considerably lower than the average (median) CV in the T1D group accounting for a value of 56% (38%), respectively.
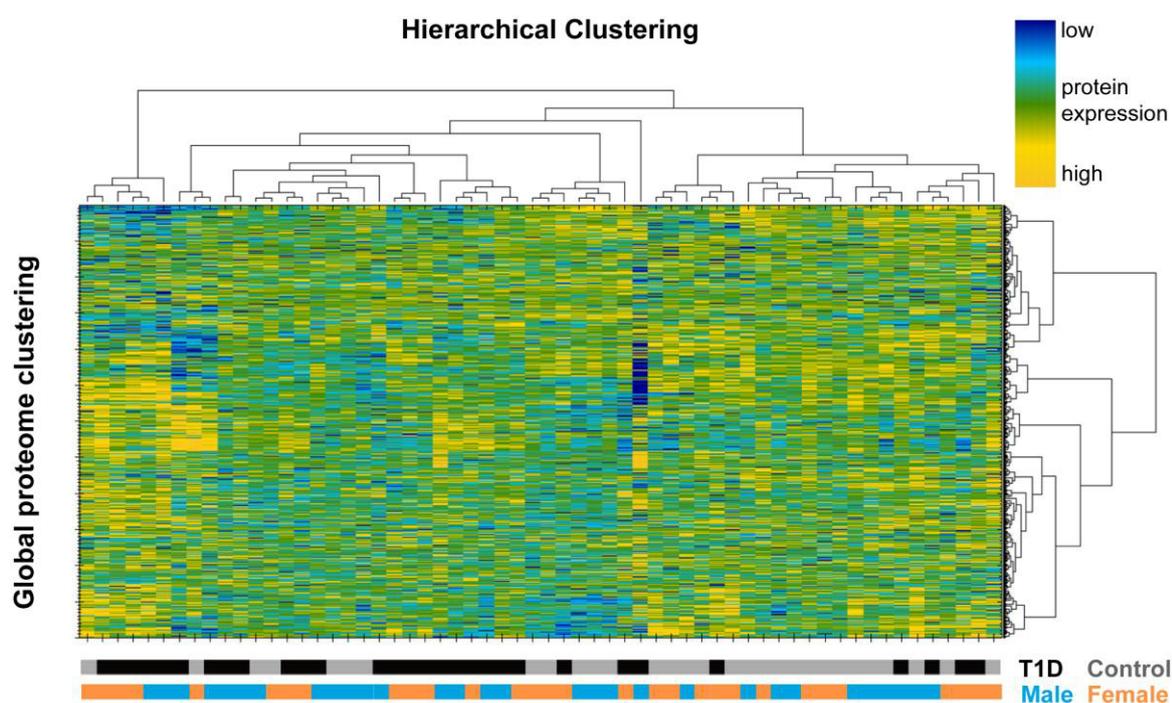
Figure 29 Hierarchical cluster analysis of the global CD4+ T cell proteome measured using DIA LC-MS/MS.
Hierarchical cluster analysis of the global CD4+ T cell proteome encompassing 8160 proteins and 60 subjects (30 patients with T1D and 30 control subjects) revealed no disease-specific clustering. Moreover, no sex-specific clustering was observed.

Specifically, proteins that were differentially regulated (q<0.05, where q represents the FDR-adjusted p-value) between the healthy and T1D state accounted for 11% of all proteins identified with ≥2 unique peptides, corresponding to 689 proteins (Fig. 30A).

With respect to ≥2-fold regulated proteins, the vast majority (46) of proteins were upregulated during disease onset and only 2 proteins were downregulated (Fig. 30A), similar to the observations from the DDA measurements (Fig. 23A-B).
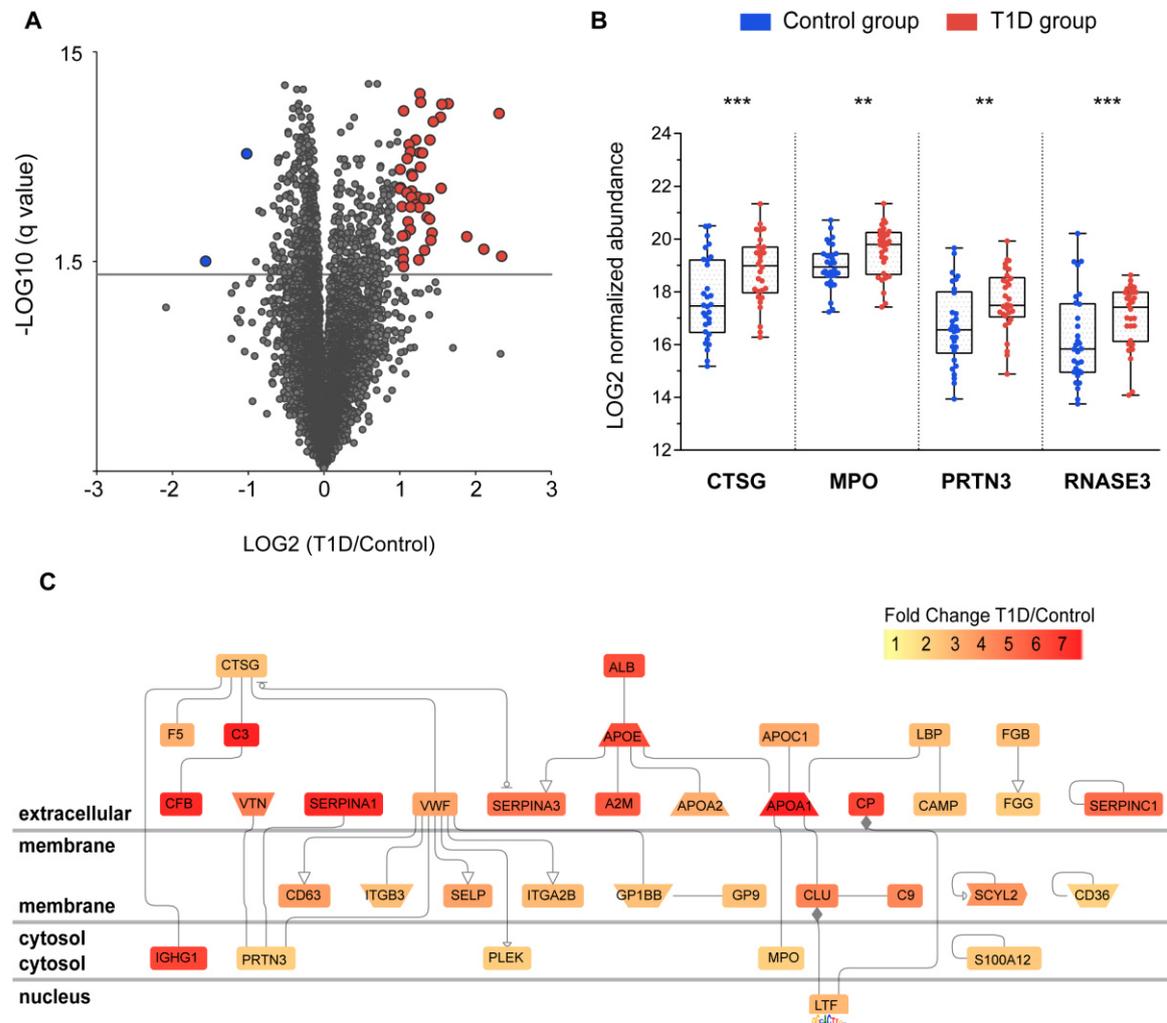
Figure 30 Differential protein expression in CD4+ T cells measured using DIA LC-MS/MS.
(A) Volcano plot showing the fold difference in expressed proteins (≥2 unique peptides) between the T1D and control groups plotted against the respective statistical significance. The horizontal gray line indicates the q-value cut-off at 0.05. Regulated proteins exhibiting a mean fold change of ≥2 and a q-value <0.05 are shown in red (higher levels in the T1D group) and blue (lower levels in the T1D group). (B) Protein expression of select inflammatory response proteins in patients with T1D and in the control group. The boxes represent the interquartile range (IQR), and the median value is shown as a solid black line within the box. The upper and lower whiskers indicate the maximum and minimum values, respectively. **q<0.01 and ***q<0.001. (C) Cellular network depicting proteins that differed significantly between the T1D and control groups measured using DIA LC-MS/MS. Proteins identified with ≥2 unique peptides, q<0.05, and ≥1.5-fold or ≤0.67-fold difference were included in the Genomatix Pathway System (GePS) analysis. Shown are 36 of the 110 input proteins connected at the validated regulatory level. The solid lines connecting various proteins indicate interactions confirmed by expert curation. The color legend indicates the highest log2 mean fold change of all cluster proteins.

Next, I determined the overrepresented signal transduction pathways and GO terms in order to identify disease onset-associated molecular patterns (Tab. 11). Similar to the

DDA analysis (sample set I), only the differentially expressed proteins quantified with ≥2 unique peptides were assessed in order to ensure reliability. Further, for input, only proteins that were regulated by at least 1.5-fold were included, resulting in 110 candidate proteins. All 110 candidate proteins were imported into the Genomatix Pathway System, and a cellular network encompassing 36 proteins was generated (Fig. 30C). The generated network highlights the predominant presence of extracellular proteins among the candidate proteins.

One of the most overrepresented signal transduction pathway was *Inflammatory* (p=1.6e-10), supporting the role of inflammatory response proteins during T1D onset (Tab. 11). The protein abundance of four inflammatory response markers (Cathepsin G (CTSG), MPO, PRTN3, and RNASE3) is shown in Fig. 30B. The inflammatory phenotype is further supported by the modulation of stress response proteins (*Response to stress*, p=4.4e-14), pro-inflammatory pathways (*Interleukin-6*, p=7.1e-05; *Matrix metalloproteinase*, p=1.8e-04; *Interleukin 1*, p=1.6e-03; *Toll like receptor*, p=1.4e-03), complement activation (*Alternative complement*, p=2.1e-12; *Classical complement*, p=4.5e-07), and platelet activation (*Platelet degranulation*, p=1.3e-18). In addition, the overrepresentation of integrin-associated signaling (*Integrin*, p=2.8e-03) and protein tyrosine kinase 2 beta signaling (*PTK2B*, p=1.7e-04) suggests that increased leukocyte migration and extravasation accompanies auto-inflammatory conditions.

The top hits in cellular components and tissues include *Extracellular space* (p=2.6e-37), *Secretory granule* (p=1.4e-17), *Plasma Specimen* (p=3.2e-22), and *Neutrophil* (p=2.6e-13), strongly suggesting that the majority of these regulated secretory proteins spilled over from localized regions of high concentration in close proximity to CD4+ T cells, as previously hypothesized based on our DDA analyses. The disease onset-associated proteomic pattern is completed by the inclusion of proteins associated with *Thrombospondin 1* and *Low density lipoprotein (LDL) receptor-related protein* signaling. Many aspects of the T1D-associated proteomic signature were observed across all analyses, indicating that the analysis method used did not influence the overall biological signature, and reflecting the stability of the signature between cohort sample sets. In summary, a distinct T1D-associated inflammatory phenotype was observed; presumably derived from the extracellular environment of CD4+ T cells attributed primarily to plasma proteins and circulating mediators of innate immunity.

Table 11 Pathway enrichment analysis was performed for all CD4+ T cell proteins that differed significantly between the T1D and control groups measured using DIA LC-MS/MS. All proteins identified with ≥2 unique peptides and ≥1.5-fold or ≤0.67 difference at q<0.05 were used as the input (n=110). The majority of proteins (n=106) were upregulated in the T1D group. Top hits of overrepresented pathways and functions are listed with their respective Genomatix enrichment scores (*p*-value).

Pathways enriched in CD4+ T cell compartment

Signal transduction	*p-value*	*Biological processes*	*p-value*
▪ Alternative complement	2.06E-12	▪ protein activation cascade	3.77E-32
▪ Inflammatory	1.62E-10	▪ hummoral immune response	4.26E-25
▪ Classical complement	4.51E-07	▪ complement activation	1.45E-22
▪ Thrombospondin 1	3.96E-05	▪ defense response	8.98E-22
▪ ATP binding cassette, sub family a (ABC1)	4.20E-05	▪ platelet degranulation	1.33E-18
▪ Low density lipoprotein receptor related protein	5.82E-05	▪ B cell mediated immunity	1.37E-16
▪ Interleukin 6 (inteferon, beta 2)	7.12E-05	▪ inflammatory response	7.07E-15
▪ PTK2B protein tyrosine kinase 2 beta	1.71E-04	▪ phagocytosis	1.29E-14
▪ Matrix metalloproteinase	1.82E-04	▪ blood coagulation	1.51E-14
▪ Calcium	5.72E-04	▪ innate immune response	3.09E-14
▪ Immune	7.54E-03	▪ response to stress	4.42E-14
▪ Coagulation factor II (thrombin) receptor	1.34E-03		
▪ Toll like receptor	1.44E-03	*Cellular components*	*p-value*
▪ Interleukin 1	1.57E-03	▪ blood microparticle	2.51E-41
▪ Very low density lipoprotein receptor	1.77E-03	▪ extracellular space	2.57E-37
▪ Tissue inhibitor of metalloproteinase	1.82E-03	▪ secretory granule	1.41E-17
▪ Vascular endothelial growth factor	2.58E-03	▪ extracellular vesicle	6.32E-17
▪ Integrin	2.78E-03	▪ immunoglobulin complex, circulating	1.85E-14
▪ Phospholipase A2	5.42E-03	▪ platelet alpha granule lumen	4.27E-13
▪ Formyl peptide receptor 1	5.51E-03	▪ cell surface	1.02E-12
▪ Nitric oxide	7.21E-03	▪ intracellular vesicle	4.49E-09
▪ Interleukin 18 (interferon gamma inducing factor)	8.39E-03	▪ plasma membrane	3.17E-08
		▪ high-density lipoprotein particle	1.50E-07
Molecular function	*p-value*	▪ fibrinogen complex	5.34E-06
▪ immunoglobulin receptor binding	4.16E-13	▪ extracellular matrix	2.73E-04
▪ serine-type endipeptidase activity	6.87E-12		
▪ receptor binding	1.85E-09	*Tissues*	*p-value*
▪ antigen binding	2.99E-07	▪ Plasma specimen	3.19E-22
▪ phosphatidylcholine-sterol-O-acyltransferase activator activity	4.01E-07	▪ Blood platelets	2.35E-14
▪ lipase inhibitor activity	5.73E-07	▪ Monocytes	9.73E-14
▪ enzyme inhibitor activity	6.69E-07	▪ Neutrophil	2.55E-13
▪ serine-type endopeptidase inhibitor activity	1.03E-06	▪ Blood clot	2.64E-12
▪ high-density lipoprotein particle receptor binding	4.89E-06		
▪ glycosylaminoglycan binding	6.88E-06		

Additionally, differential protein regulation was investigated in the previously described CD4+ T cell marker panel (Tab. 12, see also Tab. 10). Here, I observed predominantly a mild downregulation during T1D onset, possibly indicating CD4+ T cell immune dampening and/or exhaustion. Proteins with significant lower levels for the T1D group in the DIA approach included CD2, CD3D/E, CD247, CD4, CD28, LCK, ZAP70, NFATC2 and NFKB2. For the DDA data (sample set II), only NFATC2 reached statistical significance (q<0.05).

Table 12 Differential expression of prototypic CD4+ T cell markers between the T1D and control groups in the DDA and DIA experiments for sample set II.
The indicated fold change was calculated as the mean [median] ratio of T1D to control. "-", value was either not determined (i.e. protein not identified) or not included in the statistical analysis (only in the case of one-peptide identification in DDA data). Values printed in bold refer to proteins which were significantly different in expression between both groups (q<0.05).

Accession ID	Protein name	Gene symbol	DDA experiment		DIA experiment	
			q-value	Fold change	q-value	Fold change
P04234		CD3D	0.226	0.8 [0.9]	**0.000**	**0.9 [0.8]**
P07766	T cell surface glycoprotein CD3 δ/ε/γ chain	CD3E	0.285	0.9 [0.9]	**0.014**	**0.9 [0.9]**
P09693		CD3G	0.448	1.0 [0.9]	0.155	0.9 [1.0]
P01848		TRAC	0.386	1.0 [0.9]	0.374	0.9 [0.9]
P01850		TRBC1	-	1.3 [1.1]	-	-
A0A5B9	T cell receptor alpha/beta chain C region	TRBC2	0.156	0.8 [0.7]	0.655	0.9 [0.9]
P0CF51		TRGC1	-	-	0.532	0.9 [0.8]
P20963	T cell surface glycoprotein CD3 zeta chain	CD247	0.156	0.8 [0.8]	**0.005**	**0.8 [0.9]**
P06729	T cell surface antigen CD2	CD2	0.156	0.8 [0.8]	**0.000**	**0.8 [0.8]**
P01730	T cell surface glycoprotein CD4	CD4	0.156	0.8 [0.8]	**0.000**	**0.9 [0.9]**
P10747	T cell-specific surface glycoprotein CD28	CD28	-	-	**0.031**	**0.9 [0.9]**
P06239	Tyrosine-protein kinase Lck	LCK	0.331	0.9 [0.9]	**0.000**	**0.9 [0.8]**
P43403	Tyrosine-protein kinase ZAP-70	ZAP70	0.156	0.8 [0.8]	**0.000**	**0.9 [0.9]**
O43561	Linker for activation of T cells family member 1/2	LAT	-	0.6 [0.5]	0.498	1.0 [1.1]
O95644		NFATC1	-	-	0.640	1.2 [1.2]
Q13469	Nuclear factor of activated T cells, cytoplasmic 1/2/3	NFATC2	**0.040**	**0.6 [0.4]**	**0.000**	**0.8 [0.8]**
Q12968		NFATC3	-	-	0.447	1.0 [1.1]
P19838	Nuclear factor NF-kappa-B p105/p100	NFKB1	0.566	1.0 [1.0]	0.145	1.0 [0.9]
Q00653	subunit	NFKB2	0.331	1.3 [1.1]	**0.001**	**0.9 [0.9]**

The broad but mild downregulation observed for several T cell receptor and associated signaling proteins provoked me to investigate the proportion of mildly downregulated proteins in more detail. Pathway enrichment analysis was conducted for all candidate proteins (q<0.05) which were quantified with ≥2 unique peptides and which were at least 1.2-fold downregulated, corresponding to 65 input proteins. The top hit for signal transduction pathways was *T cell receptor* (p= 5.40e-05), confirming the observations from the canonical T cell marker panel (Tab. 12). Interestingly, the GO categories cellular component and biological processes revealed a broad downregulation of proteins associated with nucleocytoplasmic transport; the top hit in the former and latter category were *nuclear envelope* (p=5.2e-08) and *nucleocytoplasmic transport* (p=2.1e-07) (Fig. 31).
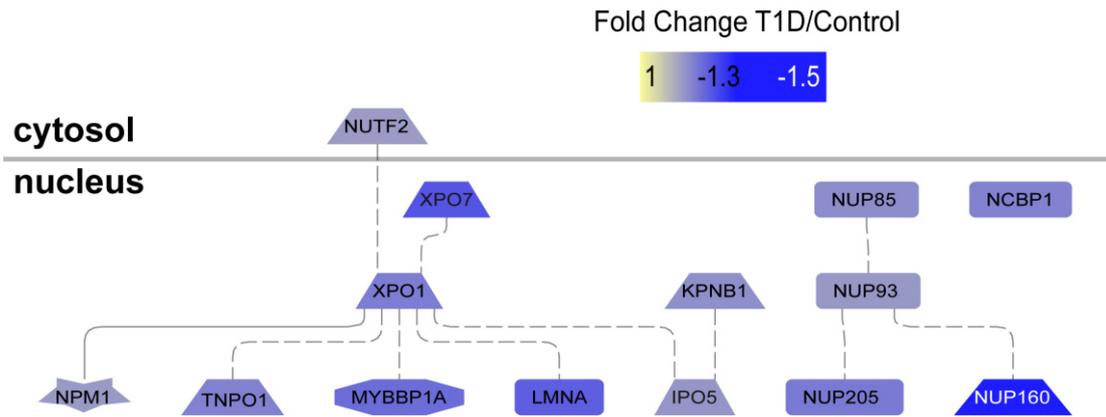
Figure 31 Pathway enrichment analysis reveals a downregulation of nucleocytoplasmic transport proteins in the T1D group for the CD4+ T cell DIA data.
Cellular network depicting proteins implicated in *nucleocytoplasmic transport* (GO biological processes) that differed significantly between the T1D and control groups measured using DIA LC-MS/MS. Proteins identified with ≥2 unique peptides, q<0.05, and ≤0.85-fold difference were included in the Genomatix Pathway System (GePS) analysis. Shown are 14 of the 65 input proteins connected at the function word level. The solid lines connecting various proteins indicate interactions confirmed by expert curation; the dotted lines represent protein associations based on co-citation. The color legend indicates the highest log2 mean fold change of all cluster proteins.

3.2.3 Western Blot validation of candidate proteins for CD4+ T cell samples

Finally, I performed western blot analysis to validate the MS results with respect to NFATC2 and tyrosine-protein kinase ZAP-70 (ZAP70) and three myeloid granule-specific inflammatory response proteins (MPO, PRTN3, and RNASE3). NFATC2 and ZAP70 were chosen as example proteins from the canonical CD4+ T cell marker panel. For both proteins significantly lower levels were measured in the DIA analysis (Tab. 12). GAPDH was chosen as a housekeeping protein, as it had virtually no biological variation in all analyses across both sample sets.

For these experiments, the residual CD4+ T cell lysates from sample set II obtained from 23 T1D patients and from 23 control group children were pooled, and the resulting signals were quantified relative to GAPDH (Fig. 32). Western blots confirmed that MPO, PRTN3, and RNASE3 were upregulated in CD4+ T cells in children with newly diagnosed T1D compared to healthy controls. In the case of NFATC2, immunoblotting confirmed a minimal downregulation in the T1D group as indicated by the MS analyses. With respect to ZAP70, the western blot showed a distinctly lower signal confirming reduced protein expression in the T1D group.
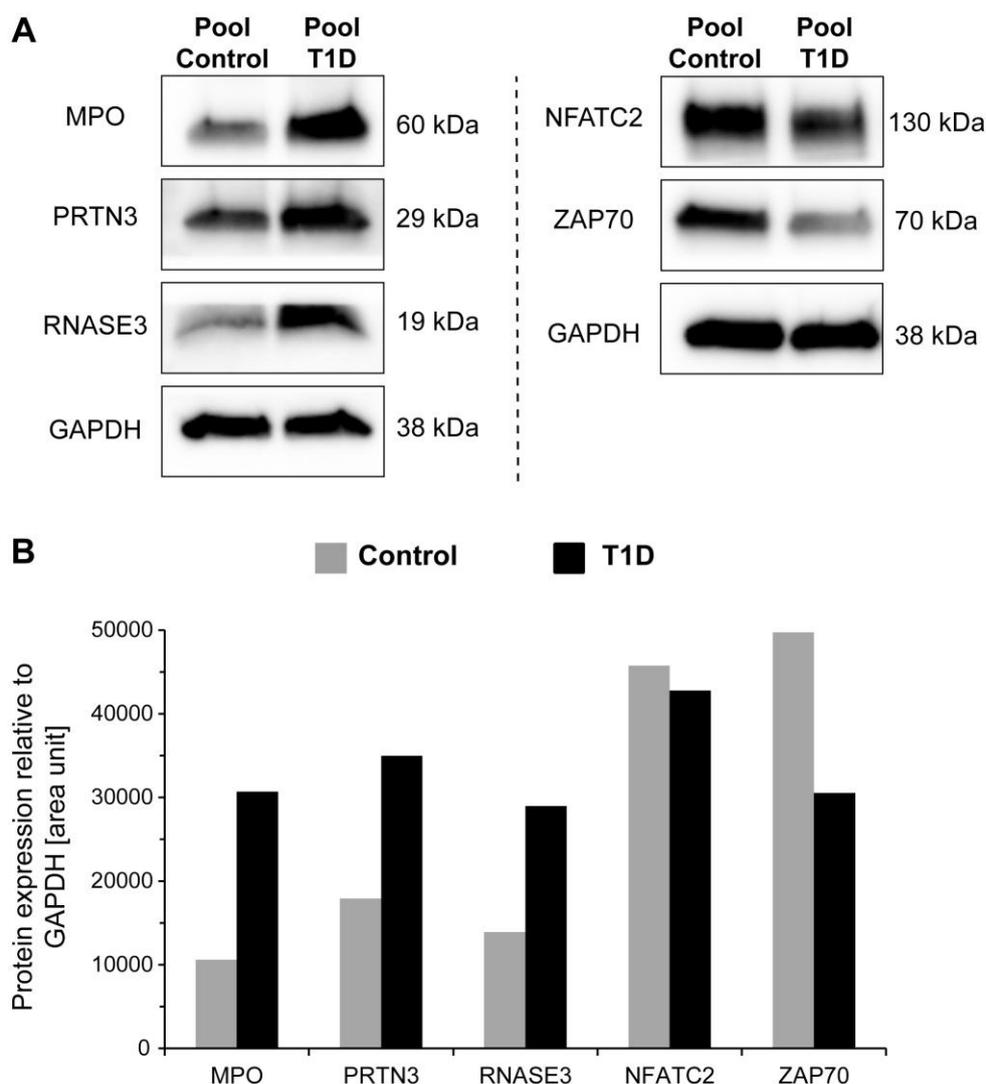
Figure 32 Western blot analysis for validating candidate proteins differentially expressed in the T1D group in CD4+ T cell lysates.
(A) CD4+ T cell lysates from sample set II were pooled from 23 patients with T1D and from 23 healthy controls and immunoblotted for MPO, PRTN3, and RNASE3 as exemplary inflammatory response proteins. NFATC2 and ZAP70 were chosen from the panel of canonical T cell markers. GAPDH was measured as a loading control. (B) The bands shown in (A) were quantified using ImageJ and expressed relative to GAPDH. The dotted line indicates two independent immunoblot experiments.

3.3 Proteomic profiling of the CD4-depleted cell fraction

After enrichment of CD4+ T cells from biobanked PBMCs the flow-through of the cell isolation procedure containing the CD4-depleted cell pool was sampled and analyzed *via* LC-MS/MS. An overview of proteomic profiling experiments performed with this residual, heterogeneous cell fraction is provided in Fig. 33. As described previously for the CD4+ T cell samples, the first set of samples comprised 23 patients with T1D and 31 healthy controls, while the second set contained 60 samples (30 from each group). In correspondence to the CD4+ T cell fraction, the first set was processed with the iST SOP

and acquired in DDA mode; while the second sample set was subsequently digested with the FASP protocol and measured in both DDA and DIA mode. In contrast to the CD4+ T cell experiments, all LC-MS/MS measurements were carried out using the then available next-generation mass spectrometer QExactive HF. The general characteristics of the study population are summarized in Tab. 1.
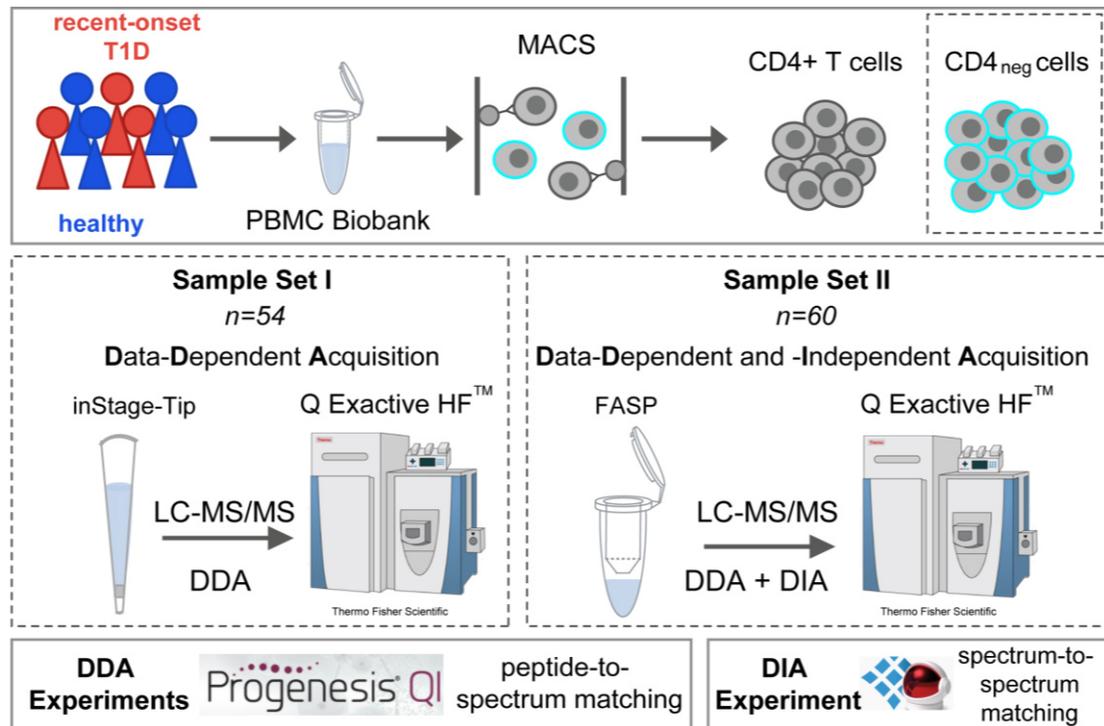
Figure 33 Overview of LC-MS/MS experiments conducted on the residual CD4-depleted cell pool derived from biobanked PBMCs from sample set I and II.
After patient- and control-derived PBMCS were thawed and enriched for CD4+ T cells, the residual CD4-depleted cell pool was sampled for proteomic profiling by LC-MS/MS. The first set of CD4-depleted cell samples (54) was processed with the iST SOP followed by DDA LC-MS/MS. The second set of CD4-depleted cell samples (60) was processed with FASP and first measured in DDA and subsequently in DIA LC-MS/MS. All data sets were acquired on Q Exactive HF machinery. Data processing was performed with Progenesis QI software for DDA experiments and with Spectronaut for the DIA experiment. DDA, data-dependent acquisition; DIA, data-independent acquisition; MACS, magnetic activated cell sorting; *n*, sample size.

3.3.1 Data-dependent acquisition (DDA) LC-MS/MS

3.3.1.1 The CD4-depleted cell fraction from sample set I and II

Proteomic profiling of the CD4-depleted cell pool samples was the experimental counterpart to CD4+ T cell phenotyping. The aim was to obtain a holistic picture of peripheral blood-derived cellular protein patterns ─ other than CD4+ T cell derived ─ associated with the onset of T1D. In the following section the DDA profiling results from sample sets I and II are directly compared, which in this case was adequate, since all technical settings apart from sample preparation were similar.

Label-free protein quantification with Progenesis QI resulted in 3421 protein identifications with ≥ 1 unique peptide in sample set I compared to 3748 proteins in sample set II; overlapping protein identifications amounted to 2329 proteins (Fig. 34A). Taking only protein identifications based on ≥2 unique peptides into consideration, the overall number of proteins detected were 2045 and 2564, respectively (Fig. 34B). These results indicate that protein identification rates were highly reproducible, independent of the sample preparation strategy or study population.

In harmony with all analyses conducted on CD4+ T cell data, all proteins which were quantified with a single unique peptide were subsequently excluded from statistical analyses. To determine significant differences in protein expression between the healthy and the T1D state, an unpaired Student's t-test was performed. As a result, 340 proteins (~17%) were differentially expressed ($p<0.05$) in sample set I; 68 of these proteins had ≥2-fold difference in expression (Fig. 34C). In comparison, 250 proteins (~10%) exhibited differential regulation ($p<0.05$) in sample set II with 47 proteins being ≥2-fold changed in expression (Fig. 34D). Notably, 37 proteins were found to be differentially regulated ($p<0.05$) in both sample sets of which 16 were up- and 21 downregulated in the T1D group (Tab. 13). Remarkably, many granulocyte-derived inflammatory mediators were differentially regulated in both CD4-depleted cell data sets highlighting that the observed phenomena was not restricted to the CD4+ T cell subset. In fact, if all differentially regulated proteins ($p<0.05$) listed for the CD4+ T cell subset (DDA measurements) in Tab. 9 and the corresponding counterpart for the CD4-depleted cell pool listed in Tab. 13 are compared, there are only five proteins overlapping: MPO, PRTN3, RNASE3, DEFA1 and GCA – all five proteins are known to be most likely neutrophil-derived if not neutrophil-specific (GO annotation).
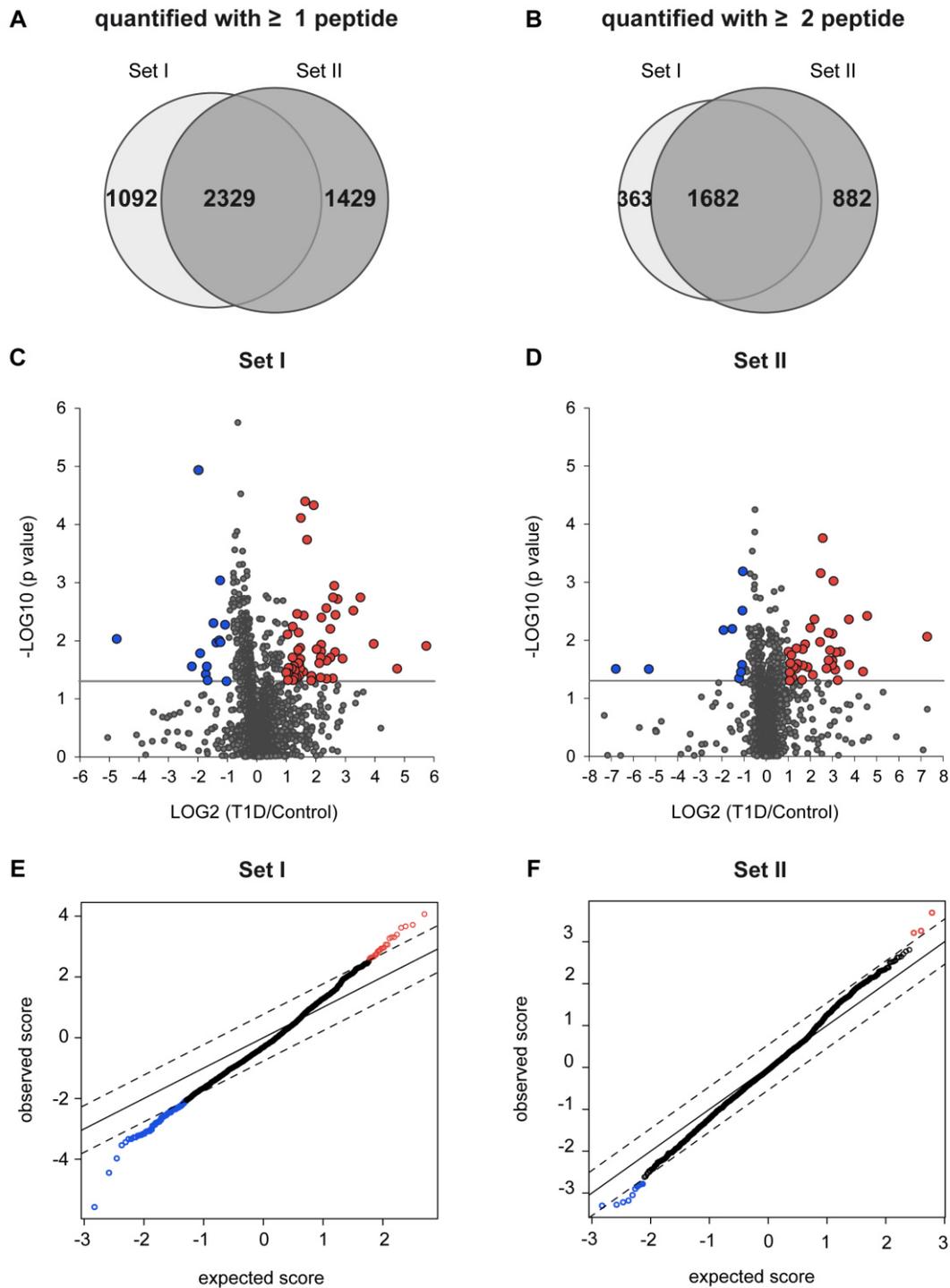
Figure 34 Comparison of protein identifications and differential protein expression in DDA data of the CD4-depleted cell fraction from sample set I and II.
(A+B) Total number of identified proteins from study part I versus II that were relatively quantified with ≥ 1 unique peptide or ≥ 2 unique peptides. (C+D) Volcano plot for sample set I and II illustrating the fold change of expressed proteins (≥2 unique peptides) between the T1D and control groups plotted against respective statistical significance before correcting for multiple testing. The horizontal gray line indicates the p-value cut-off of 0.05. Proteins with a mean fold change ≥2 (in either direction) and p<0.05 are shown in red (higher levels in the T1D group) and blue (lower levels in the T1D group). One data point has been omitted from (C) due to y-axis truncation. (E+F) SAM scatter plot showing significantly regulated proteins applying FDR threshold of 5%. Proteins that were significantly higher or lower in the T1D group are shown in red and blue. No cut-off was applied to the fold change in expression. SAM, significance analysis of microarrays; FDR, false discovery rate.

Results

Table 13 Overlap of proteins regulated with p<0.05 between T1D and control groups in the CD4-depleted cell fraction of clinical sample sets I and II.
All proteins were quantified with ≥ 2 unique peptides in both data sets. The calculated fold change of T1D/control is indicated as mean and [median]. The Phobius column refers to the results of a combined transmembrane topology and signal peptide (SP) predictor (Y). Proteins highlighted in bold are candidate proteins regulated with q<0.05 in both data sets. TMD, transmembrane domain; SP, secretory signal peptide.

Accession ID	Protein name	Gene Symbol	Phobius		Sample Set I		Sample Set II	
			TMD	SP	p-value	Fold change	p-value	Fold change
P49748	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADVL	0	0	0.039	1.3 [1.5]	0.033	1.1 [1.1]
Q9BPX5	Actin-related protein 2/3 complex subunit 5-like protein	ARPC5L	0	0	0.004	0.8 [0.7]	0.034	0.8 [0.8]
P20160	Azurocidin	AZU1	0	Y	0.025	4.3 [1.4]	0.024	1.5 [4.3]
P49913	Cathelicidin antimicrobial peptide	CAMP	0	Y	0.048	2.1 [3.0]	0.043	1.5 [2.3]
O00299	Chloride intracellular channel protein 1	CLIC1	0	0	0.022	0.9 [0.9]	0.027	0.9 [0.9]
P53634	Dipeptidyl peptidase 1	CTSC	0	Y	0.008	0.7 [0.7]	0.014	0.7 [0.7]
P08311	Cathepsin G	CTSG	0	Y	0.049	3.6 [1.9]	0.005	1.9 [5.8]
P56202	Cathepsin W	CTSW	0	Y	0.007	0.6 [0.6]	0.032	0.8 [0.8]
P59665	Neutrophil defensin 1	DEFA1	0	Y	0.011	15.6 [5.3]	0.036	1.5 [3.4]
O60496	Docking protein 2	DOK2	0	0	0.005	0.6 [0.6]	0.009	0.6 [0.5]
P60981	Destrin	DSTN	0	0	0.039	0.8 [0.8]	0.032	0.8 [0.9]
P60228	Eukaryotic translation initiation factor 3 subunit E	EIF3E	0	0	0.002	0.7 [0.4]	0.046	0.9 [0.9]
P28676	Grancalcin	GCA	0	0	0.033	2.5 [1.7]	0.011	1.6 [1.9]
Q8WWP7	GTPase IMAP family member 1	GIMAP1	1	0	0.002	0.7 [0.7]	0.019	0.9 [0.8]
P22749	**Granulysin**	**GNLY**	**0**	**Y**	**0.019**	**0.6 [0.4]**	**0.003**	**0.5 [0.4]**
P12544	**Granzyme A**	**GZMA**	**0**	**Y**	**0.001**	**0.6 [0.3]**	**0.001**	**0.7 [0.7]**
P20718	**Granzyme H**	**GZMH**	**0**	**Y**	**0.001**	**0.4 [0.5]**	**0.001**	**0.5 [0.4]**
P49863	Granzyme K	GZMK	0	Y	0.000	0.3 [0.2]	0.044	0.8 [0.6]
P51124	Granzyme M	GZMM	0	Y	0.001	0.6 [0.5]	0.006	0.6 [0.6]
P07305	Histone H1.0	H1F0	0	0	0.005	1.5 [1.4]	0.016	1.4 [1.2]
P02042	Hemoglobin subunit delta	HBD	0	0	0.006	5.6 [4.2]	0.030	1.5 [3.4]
P80188	Neutrophil gelatinase-associated lipocalin	LCN2	0	Y	0.040	2.7 [2.2]	0.050	1.5 [3.7]
P02788	Lactotransferrin	LTF	0	Y	0.044	2.4 [2.2]	0.007	1.8 [6.3]
P22894	Neutrophil collagenase	MMP8	0	Y	0.004	1.1 [11.8]	0.033	1.8 [2.4]
P05164	Myeloperoxidase	MPO	0	Y	0.03	2.3 [1.5]	0.038	1.4 [2.1]
P42785	Lysosomal Pro-X carboxypeptidase	PRCP	0	Y	0.040	0.8 [0.8]	0.049	0.8 [0.9]
P24158	Myeloblastin	PRTN3	0	Y	0.006	2.3 [2.6]	0.020	1.7 [3.3]
P61289	Proteasome activator complex subunit 3	PSME3	0	0	0.032	0.8 [0.9]	0.044	0.8 [0.8]

Accession ID	Protein name	Gene Symbol	Phobius		Sample Set I		Sample Set II	
			TMD	SP	p-value	Fold change	p-value	Fold change
Q14761	Protein tyrosine phosphatase receptor type C-associated protein	PTPRCAP	1	Y	0.016	0.7 [0.8]	0.026	0.8 [0.7]
P49023	Paxillin	PXN	0	0	0.001	0.6 [0.6]	0.036	0.8 [0.5]
Q684P5	Rap1 GTPase-activating protein 2	RAP1GAP2	0	0	0.011	0.4 [0.3]	0.011	0.8 [0.8]
P10153	Non-secretory ribonuclease	RNASE2	0	Y	0.024	2.7 [37.9]	0.016	1.3 [2.5]
P12724	Eosinophil cationic protein	RNASE3	0	Y	0.015	2.6 [9.2]	0.009	1.4 [5.8]
Q00325	Phosphate carrier protein, mitochondrial	SLC25A3	2	0	0.017	0.8 [0.8]	0.019	0.8 [0.8]
Q9Y5X3	Sorting nexin-5	SNX5	0	0	0.027	1.6 [1.2]	0.009	1.2 [1.2]
O95292	Vesicle-associated membrane protein-associated protein B/C	VAPB	1	0	0.035	0.9 [0.7]	0.013	0.8 [0.8]
P43403	**Tyrosine-protein kinase ZAP-70**	**ZAP70**	**0**	**0**	**0.004**	**0.7 [0.6]**	**0.001**	**0.7 [0.7]**

Next, SAM was applied to both CD4-depleted cell data sets to correct for multiple testing with a threshold of FDR<5% (q<0.05) resulting in 139 (set I) and 13 (set II) candidate proteins (Fig. 34E-F). In both sample sets, the majority of these candidate proteins had lower expression levels in the T1D group. Four proteins were significantly regulated at q<0.05 in both data sets: Granulysin (GNLY), Granzyme A and H (GZMA, GZMH) and ZAP70 — all four candidate proteins were downregulated in the T1D group (Fig. 35). Granzymes and GNLY are prominent effector molecules of cytotoxic lymphocytes such as natural killer (NK) cells and cytotoxic T lymphocytes (CTLs). The mild downregulation of ZAP70 was further in harmony with observations from CD4+ T cell data (Tab. 12, Fig. 32).

In summary, proteomic profiling of CD4-depleted cell samples from sample set I and II using DDA LC-MS/MS indicated two phenomena: on the one hand the upregulation of myeloid-derived inflammatory mediators in T1D patients likewise observed in CD4+ T cell data; and on the other hand the downregulation of several cytotoxic lymphocyte proteins. In the following section the DIA LC-MS/MS data of the CD4-depleted cell samples from set II will be presented in detail to confirm these observations.
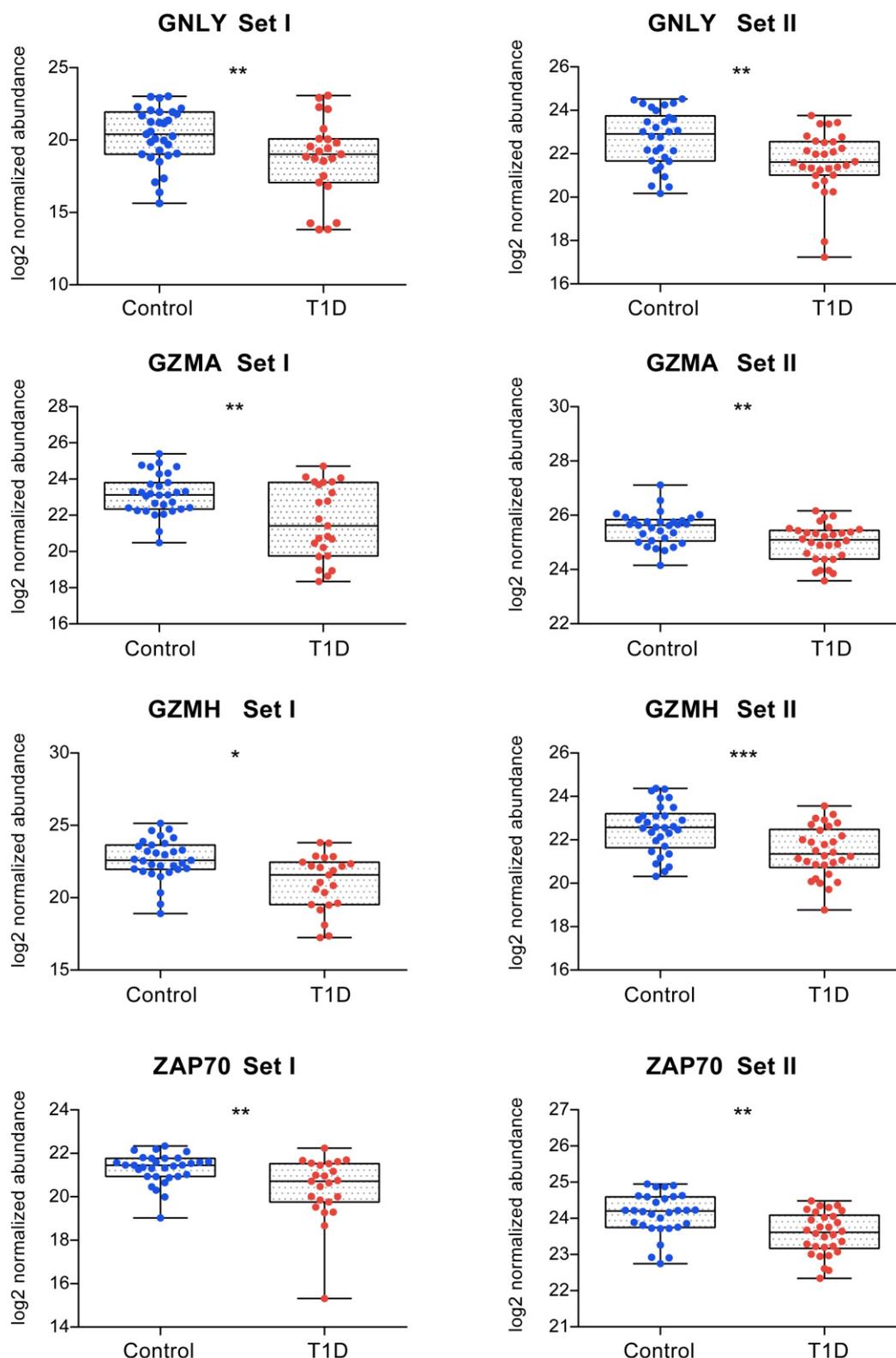
Figure 35 Boxplot of Granulysin (GNLY), Granzyme A and H (GZMA/H) and Tyrosine-protein kinase ZAP-70 (ZAP70) protein expression in the CD4-depleted cell fractions of both clinical sample sets.
GNLY, GZMA, GZMH and ZAP70 were the only four proteins that were significantly regulated in both data sets before and after FDR adjustment. The box represents the interquartile range (IQR); the median is shown as solid black line within the box. Whiskers extend from minimum to maximum of all data points. *q<0.05, **q<0.01 and ***q<0.001. FDR, false discovery rate.

3.3.2 Data-independent acquisition (DIA) LC-MS/MS

3.3.2.1 Comparison of proteome coverage in DDA and DIA data of CD4-depleted cells

Having profiled all CD4-depleted cell samples in DDA mode, the next step was to investigate the residual cell pool proteome of sample set II in DIA mode to overcome "undersampling" and maximize proteome coverage. For protein identification the same spectral library as described for the CD4+ T cell subset was used – encompassing >11000 human proteins with focus on the blood cell proteome (Tab. 7). In the following section the results from the DIA LC-MS/MS experiment are compared to the DDA profiling results of the same samples to evaluate the performance of the respective acquisition modes.

Using Spectronaut software to obtain label-free quantification, 8932 proteins were identified with at least 1 unique peptide and 7248 proteins with ≥2 unique peptides in the DIA experiment, representing a 2-fold increase in protein identifications compared to the DDA approach (Fig. 36A). In detail, 5540 more proteins were identified in addition to the 3392 proteins that were previously identified in the DDA experiment (Fig. 36B). On average, the 3392 overlapping proteins were quantified based on 17 unique peptides each, reflecting an average increase of 10 peptides compared to DDA-based quantification (Fig. 36C).

Using in silico Phobius analysis to categorize the identified proteins based on their subcellular localization revealed a similar distribution in the DDA and DIA data sets (Fig. 36D). Approximately 70% of the proteome was derived from the cytosol and 30% of proteins were either membrane-bound or secreted proteins. The proportion of proteins containing transmembrane domains localizing either to the plasma membrane or to intracellular membranes could be mildly enhanced with DIA compared to DDA — likewise observed for the CD4+ T cell subset (1881 versus 664 proteins, representing 21% and 18% of total protein identifications, respectively).

Next, proteome coverage indicated by the number of identified unique peptides was compared between DDA and DIA data with respect to select key blood cell markers for the following populations: CD8+ T cells (CD2/CD247/CD3/CD8), B cells (CD19 and CD20, also known as MS4A1), Natural killer (NK) cells (CD56 aka NCAM1), monocytes (CD14/CD33) and dendritic cells (CD11c aka ITGAX). Two of the listed proteins (CD19

and NCAM1) were only identified in the DIA approach; while peptide coverage could be distinctly enhanced for the remaining 11 proteins (Fig. 36E).
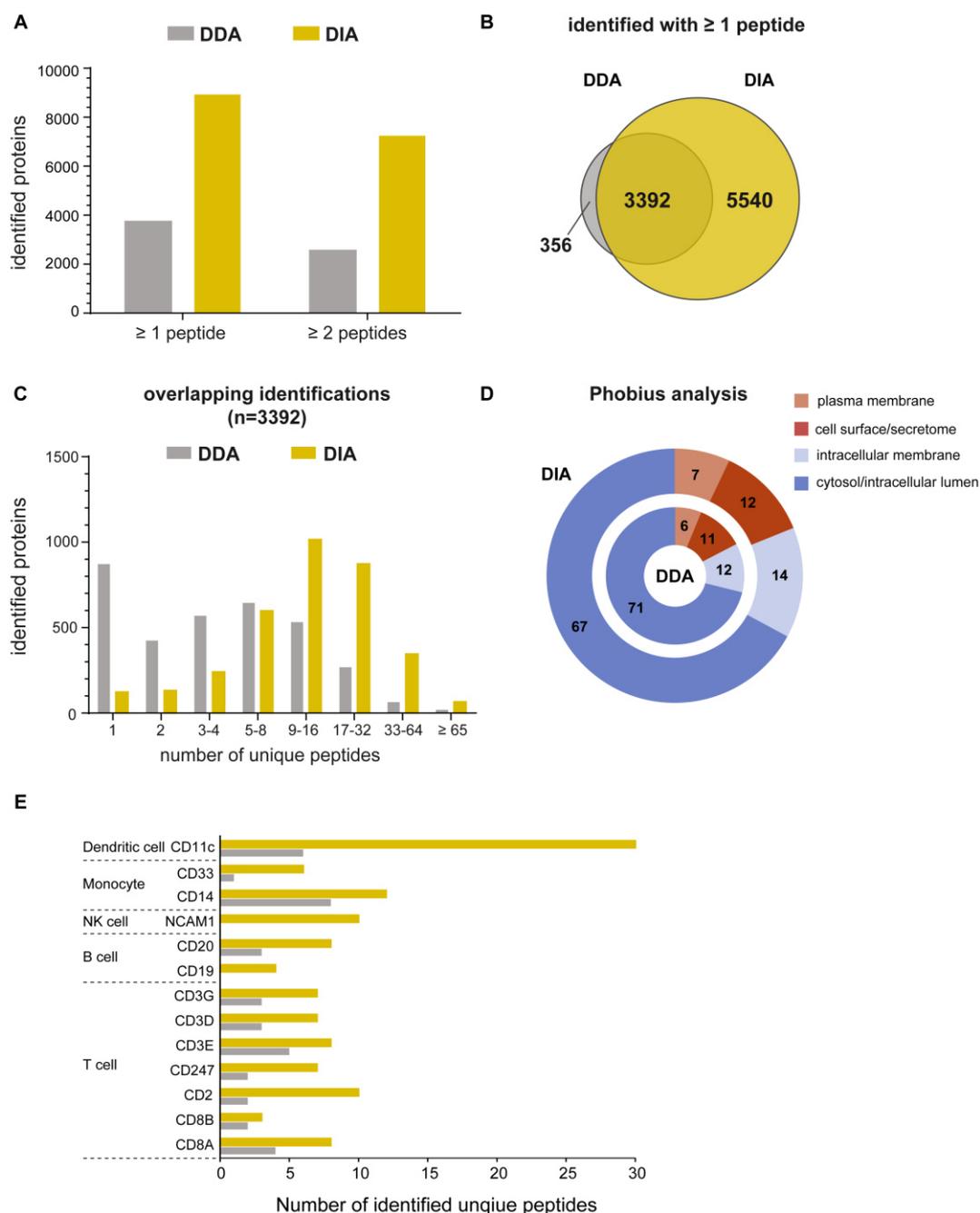
Figure 36 Comparison of the DDA and DIA data obtained from CD4-depleted cell sample set II.
(A) Number of proteins identified in the DDA and DIA experiments with ≥1 or ≥2 unique peptides. (B) Venn diagram showing the proteins identified in the DDA and/or DIA experiments (≥1 unique peptide). (C) Summary of peptide per protein coverage for overlapping protein identifications. (D) Overview of the subcellular localization of all proteins identified in the DDA and DIA experiments. Subcellular localization was determined using Phobius combined transmembrane topology and signal peptide prediction. (E) Peptide coverage of select key blood cell markers for dendritic cells, monocytes, NK cells, B cells and T cells in the DDA and DIA data.

3.3.2.2 Differential protein expression in CD4-depleted cell DIA data

The direct comparison of DDA and DIA data for the CD4-depleted cell fraction provided again compelling evidence that DIA LC-MS/MS outperformed shotgun proteomics with respect to protein identification rates and peptide per protein coverage. The next step was to analyze differential protein expression between the healthy and the T1D groups in the high-resolution DIA data.

First, protein expression data was hierarchically clustered to examine whether T1D onset affected the global composition of the CD4-depleted cell fraction proteome leading to a T1D-associated phenotype (Fig. 37). The resulting heatmap revealed that patient clustering did neither reflect a disease-specific nor a gender-specific phenotype. The same observation was previously reported for the CD4+ T cell subset and highlights once again that global biological variations had a stronger influence in the study population than proteomic changes accompanied by T1D onset. The average (median) CV values in the control and patient groups were 53% (38%) and 56% (41%).
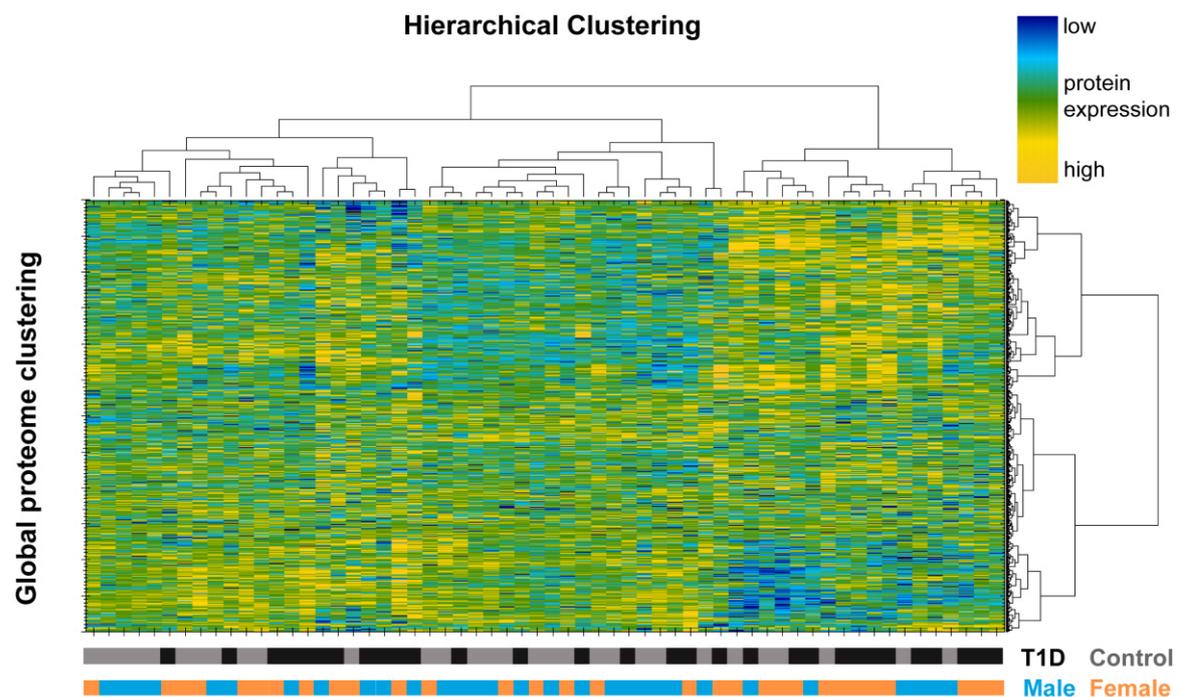
Figure 37 Hierarchical cluster analysis of the global CD4-depleted cell proteome measured using DIA LC-MS/MS.
Hierarchical cluster analysis of the global CD4+ T cell proteome encompassing 8932 proteins and 60 subjects (30 patients with T1D and 30 control subjects) revealed no disease-specific or sex-specific clustering.

In detail, proteins that were significantly changed in expression (q<0.05) between the healthy and T1D state accounted for 6% of all proteins identified with ≥2 unique

peptides, corresponding to 466 proteins (Fig. 38A). This candidate protein pool consisted of 278 proteins with higher and 188 proteins with lower expression levels in the T1D group, respectively. All proteins which exhibited at least 2-fold regulation (31) were upregulated in the T1D group, similar to the observations from the DIA LC-MS/MS measurement of the CD4+ T cell subset.
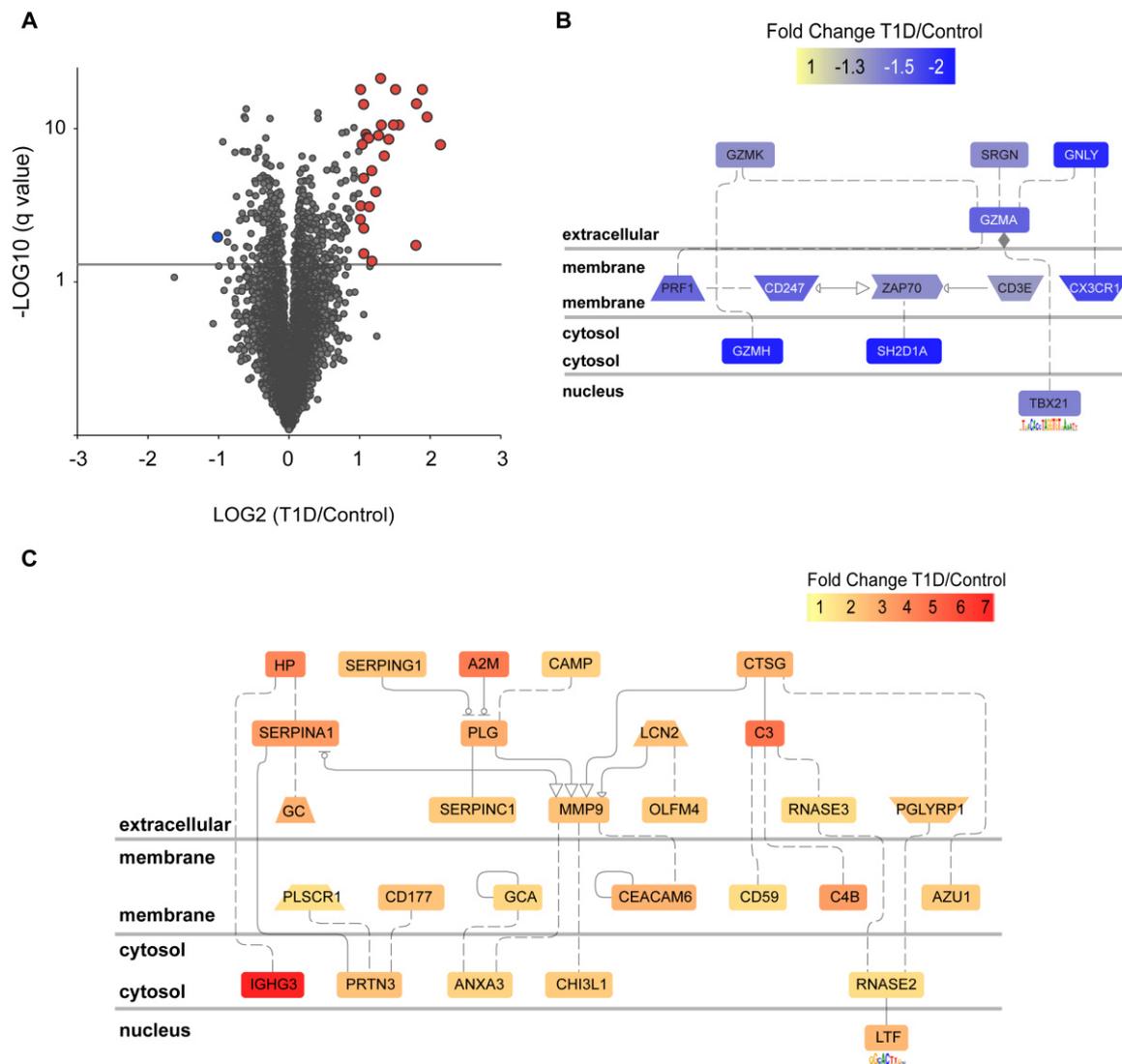
Figure 38 Differential protein expression in the CD4-depleted cell pool measured using DIA LC-MS/MS.
(A) Volcano plot showing the fold difference in expressed proteins (≥2 unique peptides) between the T1D and control groups plotted against the respective statistical significance. The horizontal gray line indicates the q-value cut-off at 0.05. Regulated proteins exhibiting a mean fold change of ≥2 and a q-value <0.05 are shown in red (higher levels in the T1D group) and blue (lower levels in the T1D group). (B+C) Cellular networks for the GO tissue categories *Natural killer cell* (B) and *Neutrophil* (C) depicting proteins that differed significantly between the T1D and control groups measured using DIA LC-MS/MS. Proteins identified with ≥2 unique peptides, q<0.05, and ≥1.3-fold or ≤0.77-fold difference were included in the Genomatix Pathway System (GePS) analysis. Shown are 12 (B) and 28 (C) of the 108 input proteins connected at the function word level. The solid lines connecting various proteins indicate interactions confirmed by expert curation; the dotted lines represent protein associations based on co-citation. The color legend indicates the highest log2 mean fold change of all cluster proteins.

To get a deeper understanding of disease onset-associated molecular patterns, overrepresented signal transduction pathways and GO terms were investigated. All differentially expressed proteins (q<0.05) quantified with ≥2 unique peptides and regulated by at least 1.3-fold (corresponding to 108 proteins) were analyzed using Genomatix Generanker. Input proteins which were upregulated during T1D onset (78) were examined separately from the downregulated candidate proteins (30) to facilitate information extraction from the results Tab. 14.

Pathway enrichment analysis for the upregulated candidate proteins revealed enrichment of many cellular signaling pathways and GO terms that had already been reported for the CD4+ T cell subset (see also Tab. 11). This cellular phenotype was coined by proteins associated with *Inflammatory* signaling (p=6.4e-08), *Alternative and Classical complement* activation (p=9.8e-11 and 3.0e-08), *Matrix metalloproteinase* (p=1.4e-04), *Interleukin 6* (p=4.5e-03) and *Toll like receptor* signaling (p=6.1e-3). In addition, the overrepresentation of *CXC receptor 1* (p=8.6e-04) and *Chemokine ligand 2 (CCL2)* (p=1.0e-03) signaling indicated enhanced leukocyte migration and extravasation. The top hits in the category cellular components (*blood microparticle*, p=7.1e-44; *extracellular space*, p=1.0e-37 and *secretory granule*, p=8.6e-12) further implied that the spatial origin of the majority of these signature proteins was predicted to be extracellular – either secreted or vesicle-bound. These findings highlight the hypothesis that enhanced protein levels of innate immune mediators, e.g. neutrophil-derived or complement proteins, were enriched in the entire patient-derived PBMC sample rather than being restricted to the CD4+ T cell subset in particular.

Next, the cellular pathways underrepresented during T1D onset were examined; this analysis was based on the input of the 30 downregulated candidate proteins (≥1.3-fold). In general, proteins which had lower expression levels in the T1D group were associated with multiple T cell related signaling pathways such as *Interleukin 2* (p=5.9e-6), *T cell receptor* (p=1.3e-4) and *NFAT cells* signaling (p=1.8e-3). In particular, several effector molecules of cytotoxic lymphocytes were profoundly downregulated during T1D onset resulting in enrichment of *Granzyme B/A* signaling (p=4.8e-07 and 2.5e-05), the molecular function *cytolysis* (p=1.1e-04) and enrichment of *Natural killer cells* (p=1.8e-13) and *Cytotoxic T lymphocytes* (p=5.5e-10) in the category tissues. This observation confirmed the previously reported findings from the DDA measurements (see 3.3.1). Moreover, several pathways closely linked to cell growth, proliferation and differentiation were enriched in the control group such as *Protein tyrosine phosphatase* (PTP) (p=1.8e-

03), *C terminal SRC kinase* (p=3.4e-03) and T*yrosine protein kinase FYN* signaling (3.9e-03) indicating a possible immune cell dampening or exhaustion during T1D onset.

Subsequently, for each of the two analyses (enrichment in T1D versus control groups) the most overrepresented cell population, according to the GO category tissue, was visualized as cellular protein network. The generated network in Fig. 38B highlights the underrepresentation of cytotoxic lymphocyte-derived proteins during T1D onset (as example *Natural killer cells*, p=1.8e-13). In contrast the network depicted in Fig. 38C emphasizes the enrichment of neutrophil-derived proteins in the T1D group (*Neutrophil*, p=1.7e-19). Protein abundance data for twelve proteins which exemplify either the myeloid-derived or the cytotoxic lymphocyte signature is depicted in Fig. 39A and Fig. 39B, respectively.

Table 14 Pathway enrichment analysis was performed for all CD4-depleted cell fraction proteins that differed significantly between the T1D and control groups measured using DIA LC-MS/MS.
All proteins identified with ≥2 unique peptides and ≥1.3-fold or ≤0.77 difference at q<0.05 were used as the input (n=108). The analysis was conducted separately for upregulated (78) and downregulated (30) proteins, referred to as enriched in T1D and control groups, respectively. Top hits of overrepresented pathways and functions are listed with their respective Genomatix enrichment scores (*p*-value).

Enriched in T1D group		Enriched in control group	
Signal transduction	*p-value*	*Signal transduction*	*p-value*
▪ Alternative complement	9.77E-11	▪ Granzyme B	4.79E-07
▪ Classical complement	2.95E-08	▪ Interleukin 2	5.87E-06
▪ Inflammatory	6.40E-08	▪ Granzyme A	2.50E-05
▪ Immune	1.65E-06	▪ Activating receptor associated to DAP10	3.32E-05
▪ Low density lipoprotein receptor related protein	6.09E-06	▪ T cell receptor	1.30E-04
▪ Tissue inhibitor of metalloproteinase	6.13E-05	▪ Cytotoxic T lymphocyte associated protein 4	2.56E-04
▪ Matrix metalloproteinase	1.39E-04	▪ T cell receptor CD3 complex	8.75E-04
▪ Chemokine (CXC motif) receptor 1	8.61E-04	▪ Protein tyrosine phosphatase, non-receptor type	1.08E-03
▪ Chemokine (CC motif) ligand 2	1.01E-03	▪ Lymphocyte specific protein tyrosine kinase	1.10E-03
▪ Very low density lipoprotein receptor	1.05E-03	▪ Cyclic ADP ribose hydrolase (CD38)	1.66E-03
▪ Second messenger sphingosine 1 phosphate	1.13E-03	▪ Nuclear factor of activated T cells, cytoplasmic, calcineurin dependent	1.81E-03
▪ Vitamin D (1,25 dihydroxyvitamin D3) receptor	1.69E-03	▪ Protein tyrosine phosphatase, receptor type	1.88E-03
▪ Interleukin 6 (interferon, beta 2)	4.50E-03	▪ CD2	2.28E-03
▪ Toll like receptor	6.09E-03	▪ Chemokine (CXC motif) receptor 3	2.63E-03
▪ Endocytic	6.86E-03	▪ C terminal SRC kinase	3.37E-03
▪ Calcium	8.48E-03	▪ Interleukin 4	3.55E-03
▪ Proteolytic	9.08E-03	▪ Tyrosine protein kinase FYN	3.85E-03
Molecular function	*p-value*	*Molecular function*	*p-value*
▪ serine-type endopeptidase activity	1.79E-16	▪ endopeptidase activity	5.62E-04
▪ immunoglobulin receptor binding	1.61E-10	▪ phosphatidylinositol bisphosphate kinase activity	2.32E-03
▪ antigen binding	1.92E-08	▪ transmembrane receptor activity	3.25E-03

Enriched in T1D group		Enriched in control group	
▪ enzyme inhibitor activity	4.04E-07	▪ histone pre-mRNA stem-loop binding	7.64E-03
▪ glycosaminoglycan binding	4.30E-07	▪ prostaglandin-F synthase activity	7.64E-03
▪ endopeptidase inhibitor activity	4.74E-07	▪ indanol dehydrogenase activity	7.64E-03
▪ receptor binding	1.33E-05	▪ histone pre-mRNA DCP binding	7.64E-03
▪ lipoprotein particle receptor binding	2.27E-05	▪ vascular endothelial growth factor binding	7.64E-03
▪ complement binding	2.27E-05	▪ geranylgeranyl reductase activity	7.64E-03
▪ lipase inhibitor activity	2.01E-04	▪ G-protein coupled receptor activity	8.67E-03

Cellular components	*p-value*	*Cellular components*	*p-value*
▪ blood microparticle	7.08E-44	▪ T cell receptor complex	8.17E-07
▪ extracellular space	9.96E-38	▪ immunological synapse	2.19E-04
▪ extracellular exosome	2.30E-21	▪ vacuolar lumen	3.51E-03
▪ secretory granule	8.57E-12		
▪ immunoglobulin complex, circulating	2.47E-11		
▪ cytoplasmic, membrane-bounded vesicle	6.28E-09		
▪ cell surface	8.13E-09		
▪ platelet alpha granule lumen	1.09E-08		
▪ cell periphery	1.36E-08		
▪ plasma membrane	1.83E-08		

Biological processes	*p-value*	*Biological processes*	*p-value*
▪ humoral immune response	5.11E-28	▪ cellular defense response	9.63E-08
▪ protein activation cascade	9.46E-27	▪ positive regulation of calcium-mediated signaling	3.19E-05
▪ complement activation	1.29E-24	▪ immune response	4.03E-05
▪ defense response	3.25E-17	▪ cytolysis	1.06E-04
▪ immunoglobulin mediated immune response	4.44E-17	▪ regulation of non-canonical Wnt signaling pathway	2.68E-04
▪ B cell mediated immunity	7.07E-17	▪ signal complex assembly	3.74E-04
▪ endocytosis	4.47E-15	▪ macrophage activation involved in immune response	3.74E-04
▪ vesicle-mediated transport	1.12E-14	▪ negative thymic T cell selection	4.97E-04
▪ leukocyte mediated immunity	2.09E-14	▪ positive regulation of leukocyte mediated immunity	5.98E-04
▪ immune effector process	2.97E-13	▪ positive regulation of alpha-beta T cell proliferation	6.37E-04

Tissues	*p-value*	*Tissues*	*p-value*
▪ Plasma specimen	2.02E-20	▪ Natural killer cells	1.84E-13
▪ Neutrophil	1.68E-19	▪ Cytotoxic T lymphocytes	5.53E-10
▪ Inflammatory exudate	4.94E-11	▪ Lymphocyte	9.55E-10
▪ Granulocyte	7.83E-11	▪ Lymphokine activated killer cells	3.85E-09
▪ Leukocytes	6.49E-10	▪ CD8 positive T lymphocytes	3.83E-08

A

B
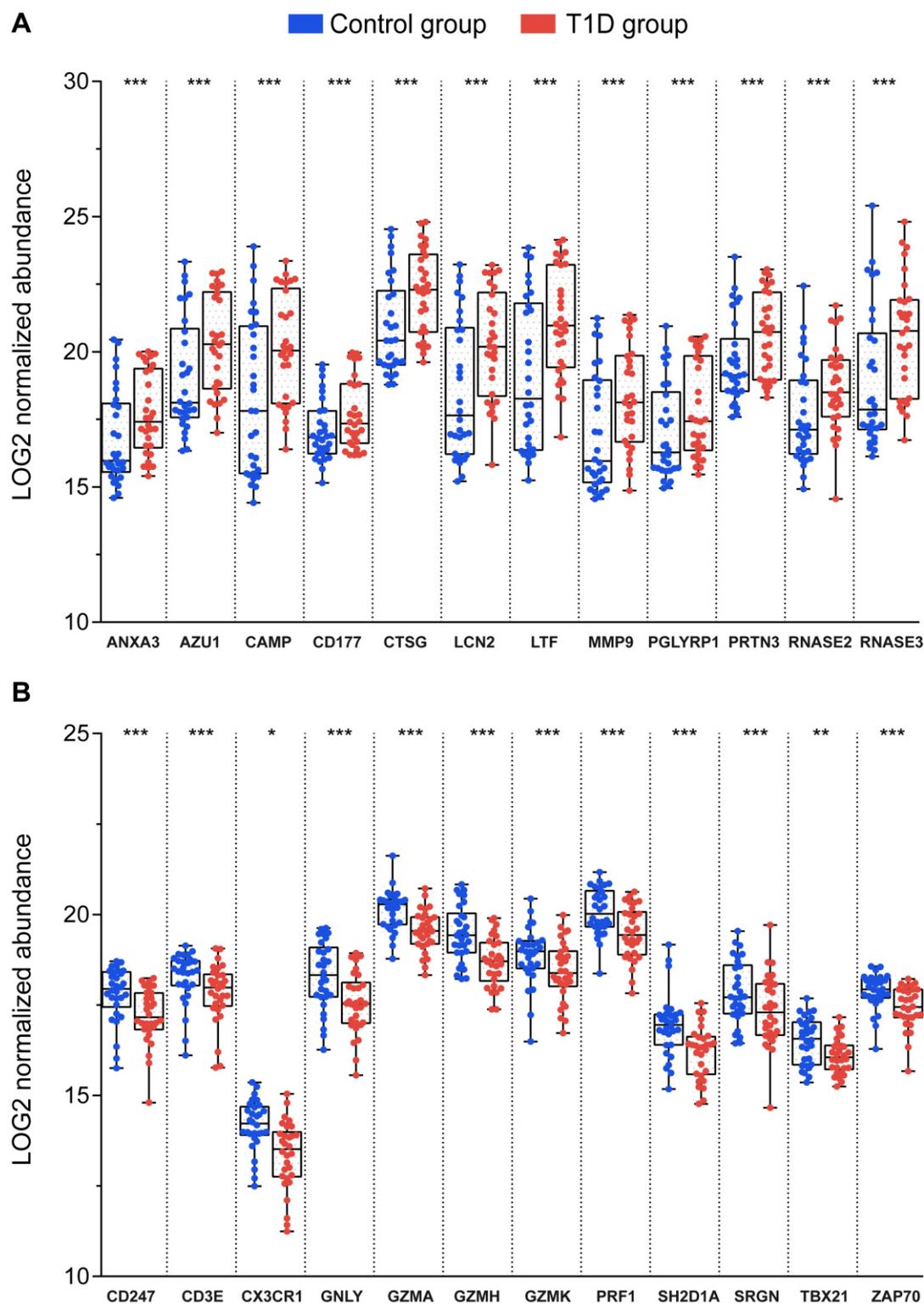
Figure 39 Box plots of candidate proteins with significantly different expression in the CD4-depleted cell fraction measured with DIA LC-MS/MS.
Pathway enrichment analysis revealed overrepresentation of neutrophil-derived proteins (A) during disease onset; while cytotoxic lymphocyte derived proteins (B) were reduced in abundance in the T1D group. Depicted are two panels of twelve proteins each with higher or lower expression in the T1D group which exemplify the signatures of both cell populations (*Neutrophil*, p=1.7e-19 and *Natural killer cells*, p=1.84e-13). The boxes represent the interquartile range (IQR), and the median value is shown as a solid black line within the box. The upper and lower whiskers represent the maximum and minimum value, respectively. *q<0.05, q<0.01** and ***q<0.001.

Taken together, pathway enrichment analysis of the CD4-depleted cell fraction DIA proteome confirmed the two main findings from the DDA profiling results: on the one hand the upregulation of numerous inflammatory proteins closely linked to innate immune activation during T1D onset; and on the other hand the downregulation of cytotoxic lymphocyte proteins originating from NK cells and/or CTLs.

In correspondence to the CD4+ T cell subset, differential protein expression was further investigated in a panel of classical blood cell markers for the distinct PBMC subpopulations in the CD4-depleted cell fraction (Tab. 15). To this end, several cluster of differentiation (CD) molecules were monitored including key markers for CD8+ T cells (CD2/CD3/CD8), B cells (CD19/20), NK cells (CD56 and CD16 aka FCGR3A), monocytes (CD14/33) and dendritic cells (CD11c). The key markers CD123 (aka IL3RA) for dendritic cells and NRC1 for NK cells were not identified by proteomic profiling. The panel further included proteins belonging to the T cell receptor complex and associated signaling molecules (e.g. CD247 or ZAP70). Importantly, many of these key markers are not exclusively but predominantly expressed by the specific cell population.

Proteins exhibiting significantly lower expression levels in the T1D group included CD3D/E, CD247, CD2, LCK, ZAP70 and NFATC2. For the DDA data, only ZAP70 reached statistical significance (q<0.05). These observations coincided with differential protein expression reported for the CD4+ T cell marker panel (see Tab. 12). Notably, the classical CD8+ T cell markers CD8A and CD8B as well as NK cell markers CD56 and CD16 did not change significantly in expression. Moreover, key blood cell markers of the B cell, monocyte and dendritic cell population were also not affected.

Table 15 Differential expression of key blood cell markers for T cells, B cells, NK cells, monocytes and dendritic cells between the T1D and control groups in the DDA and DIA experiments for sample set II.
Listed are key markers of the T cell receptor complex and associated signaling pathway as well as classical markers of the B cell (CD19/20), NK cell (CD56), monocyte (CD14/33) and dendritic cell populations (CD11c). The indicated fold change was calculated as the mean [median] ratio of T1D to control. "-", value was either not determined (i.e. protein not identified) or not included in the statistical analysis (only in the case of one-peptide identification in DDA data). Values printed in bold refer to proteins which were significantly different in expression between both groups (q<0.05).

Accession ID	Protein name	Gene symbol	DDA experiment		DIA experiment	
			q-value	Fold change	q-value	Fold change
P04234		CD3D	0.150	0.7 [0.6]	**0.004**	**0.9 [0.8]**
P07766	T cell surface glycoprotein CD3 δ/ε/γ chain	CD3E	0.226	0.6 [0.4]	**0.000**	**0.8 [0.7]**
P09693		CD3G	0.150	0.8 [0.8]	0.713	1.2 [1.0]

Results

Accession ID	Protein name	Gene symbol	DDA experiment		DIA experiment	
			q-value	Fold change	*q*-value	Fold change
P01848	T cell receptor alpha/beta chain C region	TRAC	-	-	0.399	1.1 [1.0]
A0A5B9		TRBC2	-	-	0.732	1.0 [0.8]
P0CF51		TRGC1	0.848	0.9 [0.8]	0.151	0.8 [0.8]
B7Z8K		TRDC	-	-	0.418	0.8 [1.0]
P20963	T cell surface glycoprotein CD3 zeta chain	CD247	0.382	0.8 [0.6]	**0.000**	**0.7 [0.6]**
P06729	T cell surface antigen CD2	CD2	0.150	0.6 [0.6]	**0.034**	**0.9 [0.8]**
P01732	T cell surface glycoprotein CD8 α/β chain	CD8A	0.426	0.8 [0.6]	0.203	0.9 [0.8]
P10966		CD8B	0.583	0.8 [0.4]	0.637	1.0 [0.9]
P10747	T cell-specific surface glycoprotein CD28	CD28	-	-	0.307	0.4 [0.8]
P06239	Tyrosine-protein kinase LCK	LCK	0.848	1.1 [1.0]	**0.000**	**0.9 [0.8]**
P43403	Tyrosine-protein kinase ZAP-70	ZAP70	**0.034**	**0.7 [0.7]**	**0.000**	**0.8 [0.7]**
O43561	Linker for activation of T cells family member 1/2	LAT	-	-	0.096	0.9 [0.9]
O95644	Nuclear factor of activated T cells, cytoplasmic 1/2/3	NFATC1	-	-	0.676	1.1 [1.4]
Q13469		NFATC2	0.790	0.8 [0.9]	**0.001**	**0.8 [0.7]**
Q12968		NFATC3	-	-	0.398	1.0 [1.1]
P15391	B-lymphocyte antigen CD19	CD19	-	-	0.118	1.2 [1.2]
P11836	B-lymphocyte antigen CD20	MS4A1	0.830	1.0 [1.0]	0.094	1.1 [1.1]
P13591	Neural cell adhesion molecule 1 (CD56)	NCAM1	-	-	0.343	0.9 [0.8]
P08637	Low affinity immunoglobulin gamma Fc region receptor III-A (CD16)	FCGR3A	0.475	0.8 [0.6]	0.295	0.8 [0.8]
P08571	Monocyte differentiation antigen CD14	CD14	0.848	1.1 [1.1]	0.742	1.1 [1.2]
P20138	Myeloid cell surface antigen CD33	CD33	-	1.0 [0.5]	0.688	1.5 [0.9]
P20702	Integrin alpha-X (CD11c)	ITGAX	0.426	0.8 [0.7]	0.613	1.0 [1.2]

3.3.3 Western Blot validation of candidate proteins for CD4-depleted cell samples

To confirm the differential proteomic signature discovered by proteomic profiling of the CD4-depleted cell fraction proteome, western blot analyses were carried out. MPO, PRTN3 and RNASE3 were selected for antibody probing to exemplify the myeloid-derived inflammatory signature overrepresented at T1D onset. GZMA and perforin 1 (PRF1) were chosen as effector molecules of the cytotoxic lymphocyte response which was found to be markedly downregulated in the T1D group. Moreover endoplasmic reticulum aminopeptidase 2 (ERAP2), NFATC2 and ZAP70 were chosen as example proteins from the panel of 30 distinctly downregulated candidates (q<0.05, ≤0.77-fold). RPL7 was chosen as a housekeeping protein, as it had virtually no biological variation in all analyses across both sample sets.

For immunoblotting experiments, the remaining CD4-depleted cell fraction lysates from sample set II obtained from 28 T1D patients and from 29 control group children were pooled, and the resulting signals were quantified relative to RPL7 (Fig. 40). Western

blots confirmed that MPO, PRTN3, and RNASE3 were upregulated in the CD4-depleted cell fraction of the T1D pool compared to healthy controls. However, in the case of PRTN3, western blot signals revealed only a mild upregulation in the T1D pool; whereas MS signals proclaimed a much stronger regulation (~2-fold). With respect to cytotoxic effector molecules GZMA and PRF1, immunoblotting confirmed a distinct downregulation in the T1D group as indicated by the MS analyses. Specifically regarding GZMA, the western blot quantification revealed a >2-fold reduction in protein expression in the T1D pool — exceeding the MS signal prediction. For the remaining three proteins ERAP2, NFATC2 and ZAP70 western blotting confirmed reduced expression in the T1D group as predicted by MS analyses. Particularly in the case of NFATC2 western blot signals indicated a >2-fold reduction in protein expression in the pooled CD4-depleted cell fractions of the T1D group. In summary, all reported western blot signals confirmed the respective expression trends as predicted by the corresponding MS signals.
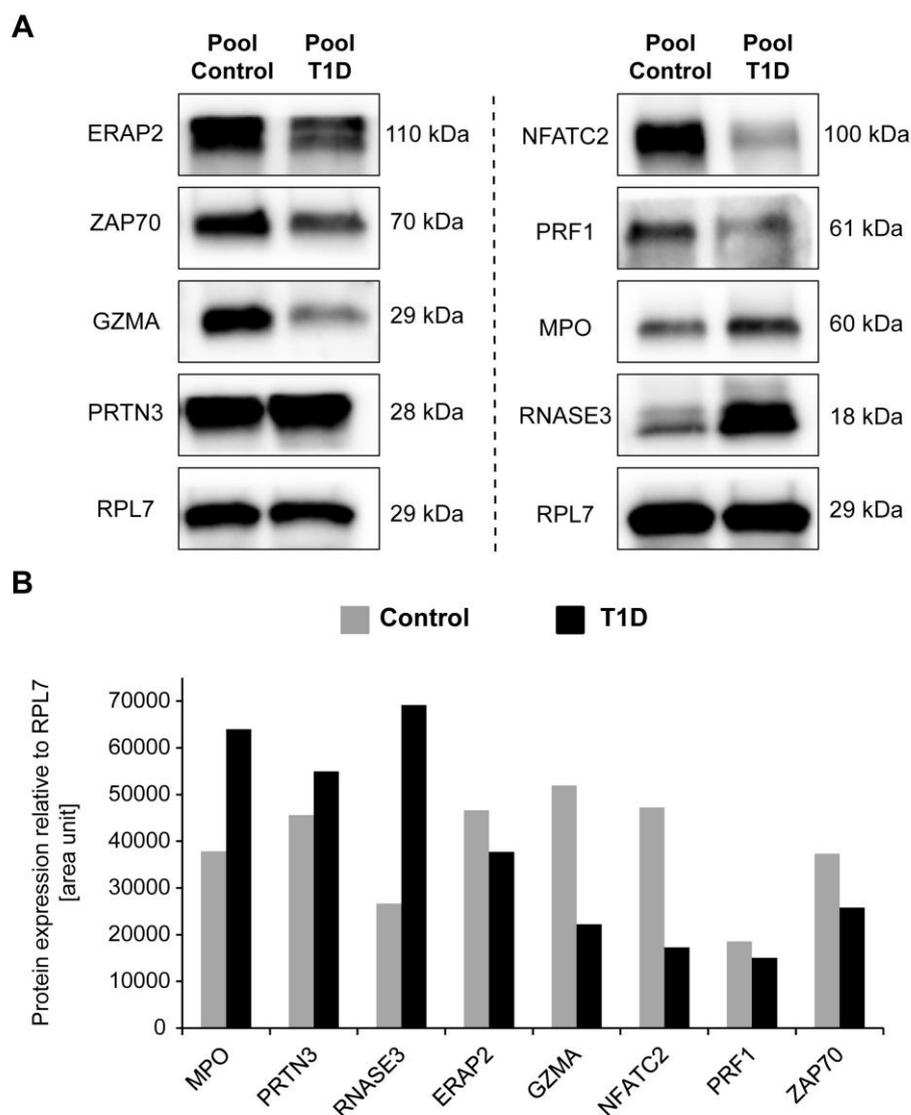
Figure 40 Western blot analysis for validating candidate proteins differentially expressed in the T1D group in the CD4-depleted cell fraction.
(A) CD4-depleted cell fraction lysates from sample set II were pooled from 28 patients with T1D and from 29 healthy controls and immunoblotted for MPO, PRTN3, and RNASE3 as exemplary inflammatory response proteins. ERAP20, GZMA, NFATC2, PRF1 and ZAP70 were chosen from the panel of downregulated candidate proteins (q<0.05). GZMA and PRF1 are effector molecules of cytotoxic lymphocytes. RPL7 was measured as a loading control. (B) The bands shown in (A) were quantified using ImageJ and expressed relative to RPL7. The dotted line indicates two independent immunoblot experiments.

3.4 Analyses of correlation between signature protein abundances and clinical parameters

Next, the relationship between protein abundance levels of selected signature proteins and clinical parameters was assessed for the T1D study population using correlation analyses. Quantitative proteome data for sample set II was obtained from the in-depth DIA experiments (CD4+ T cells and CD4-depleted cells).

For the CD4+ T cell data set, abundances of selected inflammatory response proteins were correlated to clinical endpoints HbA1c and C-peptide levels as well as patient age at disease onset. The selected candidate proteins included MPO, PRTN3 and RNASE3 which were consistently changed (q<0.05) throughout both sample sets and acquisition methods. Additionally, the inflammatory and neutrophil-associated protein ELANE was added to the panel which was previously implicated in the context of T1D biomarker research (Wang et al., 2014). As a result, no significant correlations could be observed between protein levels and the measured clinical endpoints (HbA1c and C-peptide) or patient age across both sample sets (Tab. 16). Scatter plots illustrating the relationship between the selected signature proteins and the clinical parameters or patient age are depicted in Fig. 41 — exemplified for the T1D subject population in sample set II. Similarly, significant correlations were also not observed for the selected inflammatory response proteins when protein abundance data was obtained from the CD4-depleted cell profiling (data not shown).

Table 16 Correlation between inflammatory protein abundances and clinical parameters HbA1c, C-peptide levels and patient age for both sample sets.
Pearson or Spearman correlation analyses were computed for log2-transformed protein abundance levels with clinical parameters HbA1c (%), fasting C-peptide levels (ng/ml) or patient age (years). Quantitative proteome data for sample set II was obtained from the DIA LC-MS/MS profiling. No statistically significant correlations were observed. The obtained p-values were adjusted for multiple testing using Bonferroni correction. r, correlation coefficient.

	Set I			Set II (DIA)		
	r	p	p-adjusted	r	p	p-adjusted
RNASE3						
HbA1c (%)	0.078	0.758	1.000	-0.099	0.603	1.000
fasting C-peptide levels (ng/ml)	-0.141	0.602	1.000	-0.438	0.037	0.110
age at sampling (years)	-0.299	0.166	0.499	0.038	0.843	1.000
MPO						
HbA1c (%)	0.229	0.361	1.000	-0.074	0.697	1.000
fasting C-peptide levels (ng/ml)	-0.137	0.613	1.000	0.008	0.971	1.000
age at sampling (years)	-0.268	0.216	0.647	-0.029	0.878	1.000
PRTN3						
HbA1c (%)	0.201	0.423	1.000	-0.236	0.210	0.630
fasting C-peptide levels (ng/ml)	-0.271	0.310	0.930	-0.273	0.207	0.622
age at sampling (years)	-0.358	0.094	0.281	0.043	0.822	1.000
ELANE						
HbA1c (%)	0.067	0.792	1.000	-0.234	0.213	0.639
fasting C-peptide levels (ng/ml)	-0.117	0.665	1.000	-0.400	0.059	0.177
age at sampling (years)	-0.343	0.109	0.326	0.013	0.945	1.000
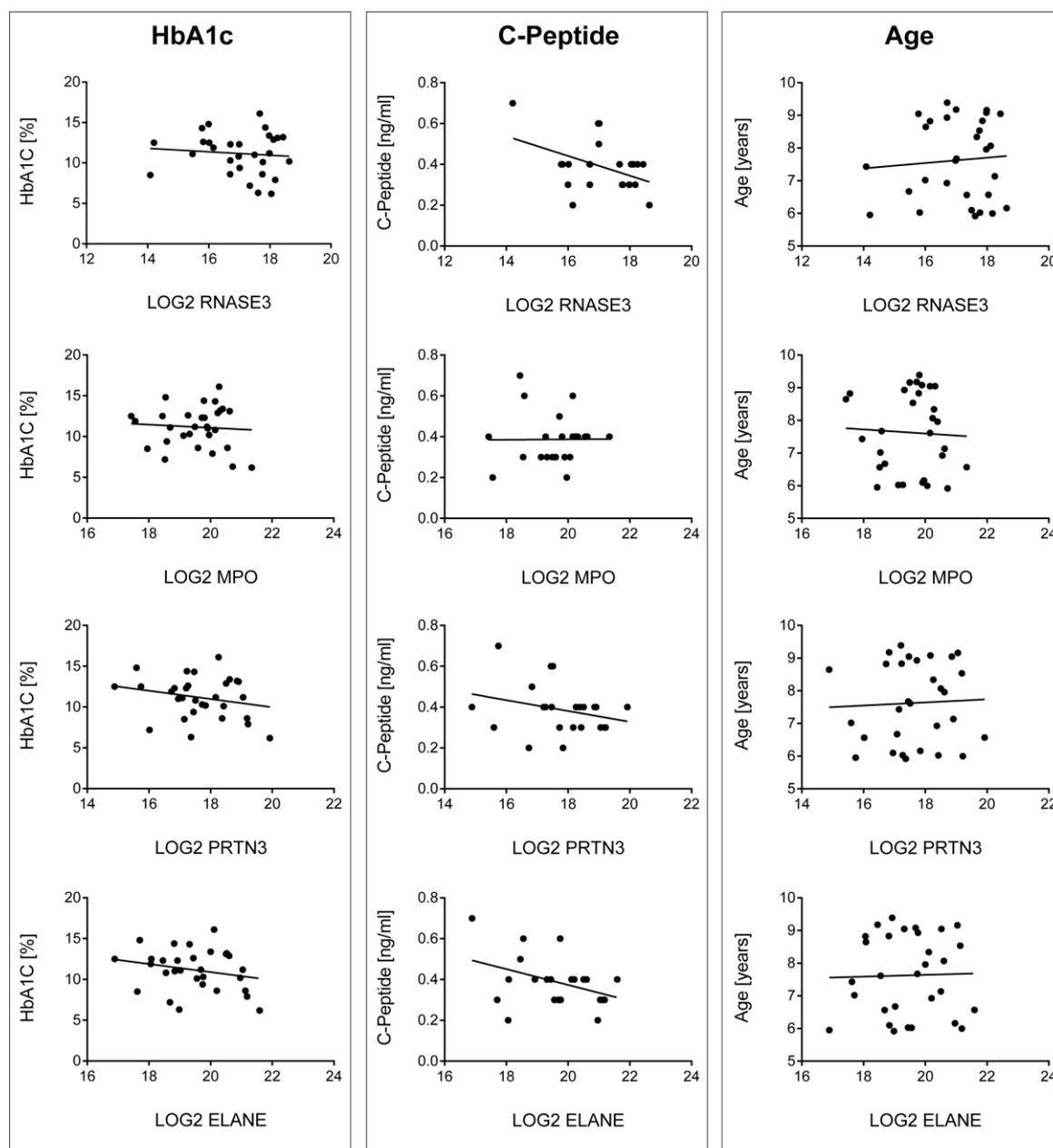
Figure 41 Correlation between RNASE3, MPO, PRTN3 and ELANE protein levels measured using DIA LC-MS/MS with HbA1c, fasting C-peptide levels and age for patients with T1D in sample set II.
Pearson or Spearman correlation analyses were computed for log2-transformed protein abundance levels with clinical parameters HbA1c (%), fasting C-peptide levels (ng/ml) or patient age (years). Data on HbA1c and age was available for all patients, C-peptide measurements for 23 patients, respectively. Obtained p-values were adjusted for multiple testing using Bonferroni correction. No statistically significant correlations were observed. *r*, correlation coefficient.

Correlation analyses were also computed between selected signature proteins in the CD4-depleted cell pool and clinical parameters (HbA1c, C-peptide) or patient age. The assessed proteins were GZMA, GZMH, GNLY and ZAP70 which are key markers of cytotoxic lymphocytes and which were significantly changed (q<0.05) across both sample sets and acquisition methods. Again, no significant correlations could be

observed between any of the assessed cytotoxic lymphocyte proteins and clinical measurements HbA1C and C-peptide or patient age (Tab. 17). Scatter plots visualizing the relationship between protein abundance levels and the assessed parameters are shown in Fig. 42 for sample set II.

Table 17 Correlation between cytotoxic lymphocyte protein abundances and clinical parameters HbA1c, C-peptide levels and patient age for both sample sets.
Pearson or Spearman correlation analyses were computed for log2-transformed protein abundance levels with clinical parameters HbA1c (%), fasting C-peptide levels (ng/ml) or patient age (years). No statistically significant correlations were observed. The obtained p-values were adjusted for multiple testing using Bonferroni correction. *r*, correlation coefficient.

	Set I			Set II (DIA)		
	r	*p*	*p*-adjusted	r	*p*	*p*-adjusted
GZMA	-0.377	0.123	0.369	-0.079	0.678	1.000
HbA1c (%)	0.249	0.352	1.000	0.023	0.916	1.000
fasting C-peptide levels (ng/ml)	0.408	0.053	0.159	0.182	0.335	1.000
age at sampling (years)						
GZMH	-0.034	0.893	1.000	-0.049	0.798	1.000
HbA1c (%)	0.071	0.793	1.000	0.212	0.331	0.993
fasting C-peptide levels (ng/ml)	0.389	0.067	0.201	0.109	0.566	1.000
age at sampling (years)						
GNLY	0.246	0.324	0.973	-0.074	0.697	1.000
HbA1c (%)	-0.371	0.157	0.470	0.326	0.129	0.387
fasting C-peptide levels (ng/ml)	0.113	0.608	1.000	0.270	0.150	0.449
age at sampling (years)						
ZAP70	-0.119	0.637	1.000	0.027	0.887	1.000
HbA1c (%)	0.139	0.608	1.000	-0.098	0.655	1.000
fasting C-peptide levels (ng/ml)	0.471	0.023	0.070	0.229	0.224	0.672
age at sampling (years)	-0.377	0.123	0.369	-0.079	0.678	1.000
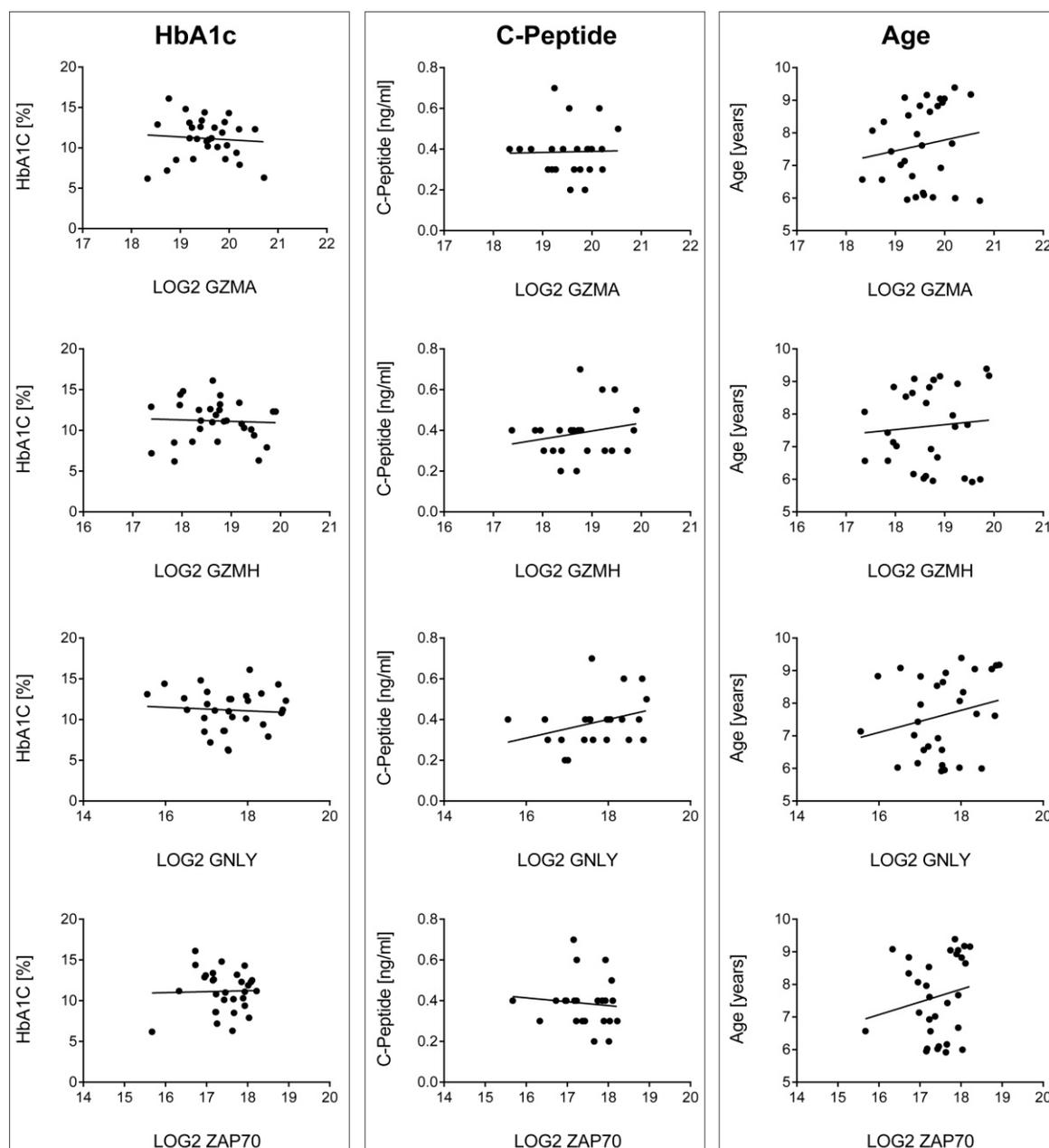
Figure 42 Correlation between GZMA, GZMH, GNLY and ZAP70 protein levels measured using DIA LC-MS/MS and HbA1c, fasting C-peptide levels and age for patients with T1D in sample set II.
Pearson or Spearman correlation analyses were computed for log2-transformed protein abundance levels with clinical parameters HbA1c (%), fasting C-peptide levels (ng/ml) or patient age. Data on HbA1c measurements and age was available for all patients, C-peptide measurements for 23 patients with T1D, respectively. Obtained p-values were adjusted for multiple testing using Bonferroni correction. No statistically significant correlations were observed. *r*, correlation coefficient.

3.5 Proteomic profiling of clinical serum samples

Proteomic phenotyping of peripheral CD4+ T cells and the residual CD4-depleted cell fraction revealed a proteomic signature enriched at T1D onset which was closely linked to the extracellular milieu. Specifically, pathway enrichment analysis indicated that a large proportion of candidate proteins (differentially abundant at q<0.05) were expected to be circulating in the blood plasma (*Plasma specimen*, p=3.2e-22 for CD4+ T cells; p=2.0e-20 for CD4-depleted cells). This inflammatory plasma signature can be broken down into proteins associated with innate immunity such as neutrophil-derived proteins (e.g. MPO, PRTN3 and RNASE3) and complement proteins (e.g. complement factor B (CFB)) as well as apolipoproteins (e.g. APOE) and hemostasis/coagulation-associated proteins (e.g. plasma protease C1 inhibitor (SERPING1)).

In order to complement the obtained proteomic fingerprint of PBMCs with its extracellular counterpart, a set of 100 serum samples was profiled using DIA LC-MS/MS. The aim was (i) to investigate whether the discovered PBMC signature was equally observable in serum and (ii) to discover novel proteomic patterns associated with T1D onset. The serum sample set consisted of 50 healthy control samples and 50 samples derived from newly diagnosed patients with T1D. Both sexes were equally distributed in the study population; however samples were not matched for age. The median age was 13.6 in the control group and 10.4 in the T1D group, respectively. The general characteristics of the study population used for serum profiling are summarized in Tab. 2. Serum samples were analyzed using DIA LC-MS/MS with a tailored serum/plasma spectral library encompassing 715 proteins and 27057 peptide precursors (Tab. 7).

Using Spectronaut to obtain label-free protein quantification, 354 proteins were identified with at least 1 unique peptide and 225 proteins with ≥2 unique peptides. First, technical reproducibility of the MS performance and subsequent Spectronaut quantification was evaluated in a pooled quality control (QC) sample (Fig. 43). The QC pool consisted of serum digests derived from ten randomly selected patients and controls (five per group) and was measured before and after a sequence of 25, 50, 75 and 100 serum samples. CV values were then calculated for the quality QC pool and the control and T1D groups. Notably, the QC pool achieved an average (median) CV value of 19% (10%) indicating exceptionally good technical reproducibility (Fig. 43). With respect to biological variation, the average (median) CV accounted for 51% (36%) in the control group and 57% (39%) in the T1D group indicating distinct degree of inter-individual variation (Fig. 43).
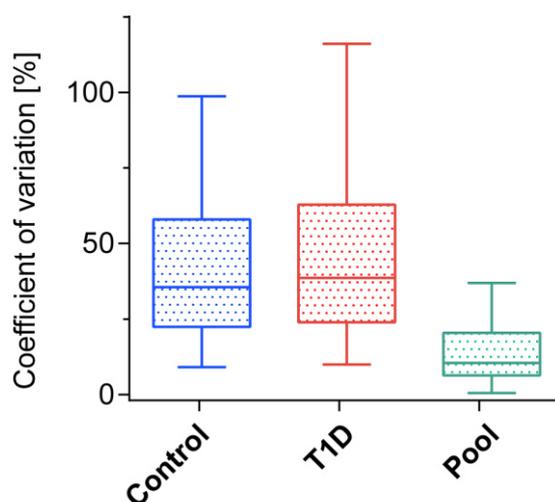
Figure 43 Coefficient of variation (CV) distribution in serum samples analyzed with DIA LC-MS/MS.
CVs were calculated for the control group, the T1D group and a pooled quality control (QC) sample. The QC pool was run before the first sample (1st), in-between (after the 25th, 50th and 75th samples) and after the last sample (100th); hence CV calculation for the pooled sample was based on 5 technical replicates. The box plot is depicted in Tukey style, i.e. the boxes represent the interquartile range (IQR), and the median value is shown as a solid line within the box. The upper and lower whiskers extend to the most extreme data point within 1.5xIQR from the edge of the box.

Next, hierarchical cluster analysis and PCA of global protein expression were conducted to assess whether proteomic changes accompanying T1D onset resulted in disease-specific sample clustering (Fig. 44). Notably, both analysis methods revealed a clear trend for a phenotypic separation of samples, although a distinct amount of biological variation was measured in the serum samples (Fig. 43). In contrast to the PBMC proteomes (Fig. 29 and 37), this observation implied that uniform changes in the composition of the serum proteome took place at T1D onset. This finding is supported by the previous discovery that a large proportion of differential protein expression in the PBMC proteome is closely linked to the extracellular space.

To determine differential protein expression (≥2 unique peptides) between T1D and control groups, ANCOVA was performed correcting for age as covariate. The resulting p-values were corrected for multiple hypothesis testing using the Benjamini Hochberg method (Benjamini and Hochberg, 1995). As a result, 111 proteins were significantly changed in expression (q<0.05); 64 proteins had higher expression levels and 47 proteins had lower expression levels in the T1D group, respectively. Notably, only 33 proteins were changed at least 1.3-fold in expression of which 16 were upregulated and 17 downregulated in the T1D group.
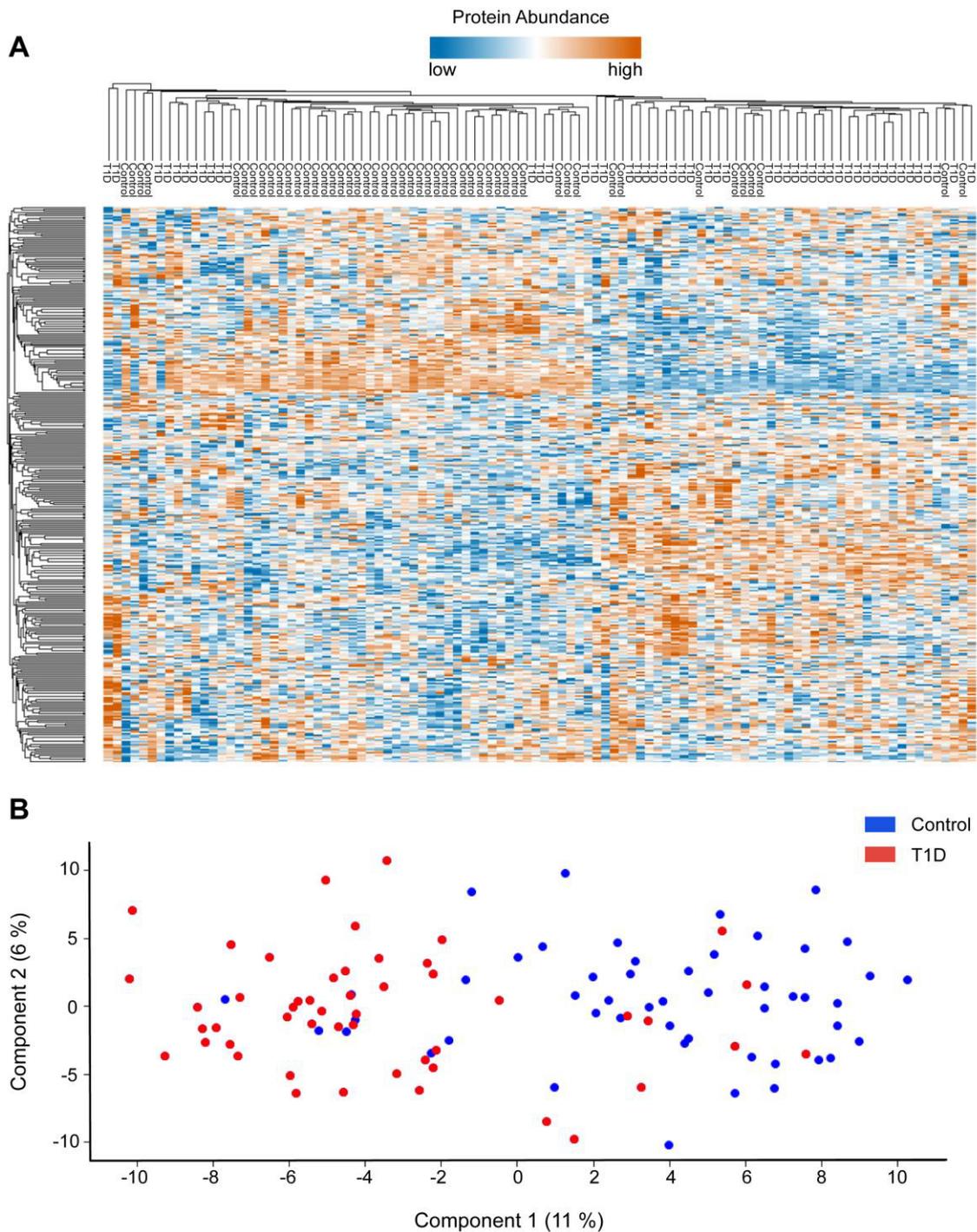
Figure 44 Hierarchical cluster analysis and principal component analysis (PCA) for the global serum proteome measured using DIA LC-MS/MS.
Hierarchical cluster analysis (A) and principal component analysis (B) of the global serum proteome encompassing 354 proteins and 100 subjects (50 patients with T1D and 50 control subjects) revealed a clear tendency towards clustering and separation based on disease status.

To visualize differential protein expression, a protein network encompassing 47 of the 111 differentially regulated proteins (q<0.05) was generated using GePS (Fig. 45). The protein network provides evidence for the upregulation of several components of the

complement system (e.g. C4A, C8G or C4BPA) and members of the apolipoprotein family (e.g. APOA4, APOB, APOC2 or APOE).
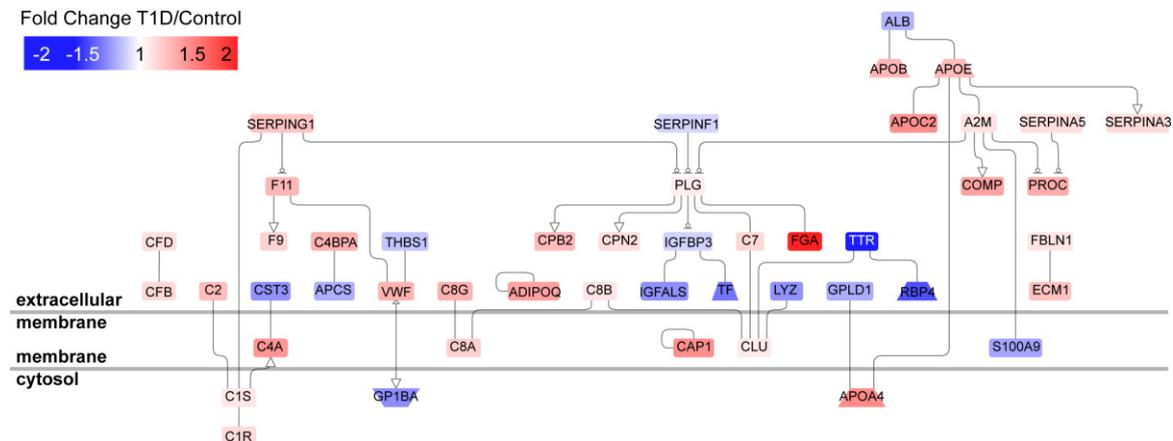
Figure 45 Genomatix network of differentially expressed proteins in the serum proteome measured with DIA LC-MS/MS.
Genomatix network depicting proteins that differed significantly between the T1D and control groups measured using DIA LC-MS/MS. Proteins identified with ≥2 unique peptides and q<0.05 were uploaded to the Genomatix Pathway System (GePS) tool; shown are 47 of the 111 input proteins connected at the validated regulatory level. The solid lines connecting various proteins indicate interactions confirmed by expert curation. The color legend indicates the highest log2 mean fold change of all cluster proteins.

The exact overlap of significantly regulated proteins (q<0.05) in the serum data and the DIA LC-MS/MS cellular data is listed in Tab. 18 for CD4+ T cells and in Tab. 19 for the CD4-depleted cell fraction, respectively. For CD4+ T cells, the table includes 29 overlapping proteins, however in only 17 cases the trend for upregulation or downregulation at T1D onset (mean and median fold change) was in harmony between both data sets. Among the proteins which were upregulated in both the CD4+ T cell data and the serum data were members of the apolipoprotein family (APOC1, APOC3, APOE and CLU), several protease inhibitors (A2M, ITIH2, ITIH3, SERPING1 and SERPINA3), components of the complement system (CF and C4BPA) and hemostasis-associated proteins (FGA and VWF).

Table 18 Overlap of proteins regulated with q<0.05 between T1D and control groups in peripheral CD4+ T cell preparations and serum samples measured using DIA-LC-MS/MS.
All proteins were quantified with ≥ 2 unique peptides in both data sets. The calculated fold change of T1D/control is indicated as mean and [median]. The Phobius column refers to the results of a combined transmembrane topology and signal peptide predictor (Y). TMD, transmembrane domain. Fold change trends for upregulation in the T1D group are shaded in red, whereas downregulation is shaded in blue. Conflicting fold changes (i.e. mean and median fold change differed in a positive or negative trend) are shaded in grey and no changes in regulation were left clear.

Accession ID	Protein name	Gene Symbol	Phobius		CD4+ T cell data		Serum data	
			TMD	SP	q-value	Fold change	q-value	Fold change
P01023	Alpha-2-macroglobulin	A2M	0	Y	0.002	4.6 [1.3]	0.000	1.1 [1.1]
P02768	Serum albumin	ALB	0	Y	0.000	5.1 [2.0]	0.000	0.8 [0.9]
P02760	Protein AMBP	AMBP	0	Y	0.038	2.2 [1.1]	0.000	1.0 [1.0]
P02654	Apolipoprotein C-I	APOC1	0	Y	0.041	2.2 [1.7]	0.000	1.3 [1.3]
P02656	Apolipoprotein C-III	APOC3	0	Y	0.013	1.8 [2.4]	0.000	1.2 [1.1]
P02649	Apolipoprotein E	APOE	0	Y	0.000	5.2 [2.4]	0.000	1.3 [1.2]
P02747	Complement C1q subcomponent subunit C	C1QC	0	Y	0.004	1.7 [1.1]	0.012	0.9 [0.9]
P04003	C4b-binding protein alpha chain	C4BPA	0	Y	0.000	3.1 [2.0]	0.000	1.3 [1.3]
P07360	Complement component C8 gamma chain	C8G	0	Y	0.022	1.5 [0.9]	0.000	1.3 [1.3]
P49913	Cathelicidin antimicrobial peptide	CAMP	0	Y	0.000	1.8 [3.4]	0.012	0.8 [0.7]
P00751	Complement factor B	CFB	0	Y	0.000	7.1 [2.4]	0.011	1.1 [1.1]
P10909	Clusterin	CLU	0	Y	0.000	3.1 [4.6]	0.027	1.1 [1.1]
P00450	Ceruloplasmin	CP	0	Y	0.010	5.6 [2.1]	0.000	1.1 [1.1]
P23142	Fibulin-1	FBLN1	0	Y	0.002	1.8 [1.4]	0.001	1.1 [1.1]
P02671	Fibrinogen alpha chain	FGA	0	Y	0.019	1.6 [1.4]	0.001	2.1 [1.4]
P06396	Gelsolin	GSN	0	Y	0.000	1.6 [1.6]	0.000	0.9 [0.9]
Q86YZ3	Hornerin	HRNR	0	0	0.000	1.3 [1.6]	0.001	1.6 [1.8]
P01877	Ig alpha-2 chain C region	IGHA2	0	0	0.001	1.9 [0.8]	0.034	1.2 [1.2]
P01859	Ig gamma-2 chain C region	IGHG2	0	0	0.000	7.2 [2.4]	0.008	0.8 [0.8]
P01780	Immunoglobulin heavy variable 3-7	IGHV3-7	0	Y	0.024	0.5 [1.4]	0.029	0.9 [0.9]
P01834	Ig kappa chain C region	IGKC	0	0	0.000	3.7 [2.3]	0.000	0.8 [0.8]
P19823	Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2	0	Y	0.000	1.6 [1.2]	0.001	1.1 [1.1]
Q06033	Inter-alpha-trypsin inhibitor heavy chain H3	ITIH3	0	Y	0.024	1.1 [1.2]	0.000	1.3 [1.3]
P06702	Protein S100-A9	S100A9	0	0	0.031	1.4 [1.5]	0.029	0.8 [0.7]
P01011	Alpha-1-antichymotrypsin	SERPINA3	0	Y	0.023	3.5 [1.2]	0.022	1.1 [1.1]
P05155	Plasma protease C1 inhibitor	SERPING1	0	Y	0.001	4.8 [1.7]	0.000	1.2 [1.2]
P07996	Thrombospondin-1	THBS1	0	Y	0.006	1.4 [1.7]	0.013	0.9 [0.8]
Q9Y490	Talin-1	TLN1	0	0	0.000	1.3 [1.2]	0.001	1.5 [1.5]
P04275	von Willebrand factor	VWF	0	Y	0.037	2.3 [1.2]	0.021	1.3 [1.3]

Results

The overlap of significantly regulated proteins (q<0.05) in the serum data and the DIA LC-MS/MS CD4-depleted cell data included 25 proteins of which 13 exhibited uniform trends in both data sets with regards to upregulation or downregulation at T1D onset. Proteins with harmonizing fold changes comprised again mainly apolipoproteins (APOC3, APOE and CLU) and complement factors (C4BPA; C8A and CFB).

Table 19 Overlap of proteins regulated with q<0.05 between T1D and control groups in the peripheral CD4-depleted cell fraction and serum samples measured using DIA-LC-MS/MS.
All proteins were quantified with ≥ 2 unique peptides in both data sets. The calculated fold change of T1D/control is indicated as mean and [median]. The Phobius column refers to the results of a combined transmembrane topology and signal peptide predictor (Y). TMD, transmembrane domain. Fold change trends for upregulation in the T1D group are shaded in red, whereas downregulation is shaded in blue. Conflicting fold changes (i.e. mean and median fold change differed in a positive or negative trend) are shaded in grey and no changes in regulation were left clear.

Accession ID	Protein name	Gene Symbol	Phobius		CD4-depleted data		Serum data	
			TMD	SP	q-value	Fold change	q-value	Fold change
P04217	Alpha-1B-glycoprotein	A1BG	0	Y	0.047	1.3 [1.2]	0.000	1.1 [1.1]
P01023	Alpha-2-macroglobulin	A2M	0	Y	0.000	3.0 [1.2]	0.000	1.1 [1.1]
P02768	Serum albumin	ALB	0	Y	0.000	1.7 [1.4]	0.000	0.8 [0.9]
P02760	Protein AMBP	AMBP	0	Y	0.024	3.7 [1.1]	0.000	1.0 [1.0]
P02656	Apolipoprotein C-III	APOC3	0	Y	0.014	1.5 [1.6]	0.000	1.2 [1.1]
P02649	Apolipoprotein E	APOE	0	Y	0.000	3.6 [2.8]	0.000	1.3 [1.2]
P02747	Complement C1q subcomponent subunit C	C1QC	0	Y	0.011	1.4 [1.0]	0.012	0.9 [0.9]
P04003	C4b-binding protein alpha chain	C4BPA	0	Y	0.000	1.8 [2.3]	0.000	1.3 [1.3]
P07357	Complement component C8 alpha chain	C8A	0	Y	0.002	2.1 [1.2]	0.000	1.2 [1.2]
P00915	Carbonic anhydrase 1	CA1	0	0	0.001	1.1 [1.5]	0.013	0.8 [0.8]
P49913	Cathelicidin antimicrobial peptide	CAMP	0	Y	0.000	1.5 [4.7]	0.012	0.8 [0.7]
P00751	Complement factor B	CFB	0	Y	0.007	2.1 [1.4]	0.011	1.1 [1.1]
P10909	Clusterin	CLU	0	Y	0.000	2.0 [1.2]	0.027	1.1 [1.1]
P00450	Ceruloplasmin	CP	0	Y	0.011	3.3 [2.0]	0.000	1.1 [1.1]
P23142	Fibulin-1	FBLN1	0	Y	0.002	1.9 [1.7]	0.001	1.1 [1.1]
P01877	Ig alpha-2 chain C region	IGHA2	0	0	0.031	2.5 [1.3]	0.034	1.2 [1.2]
P01859	Ig gamma-2 chain C region	IGHG2	0	0	0.000	7.5 [1.4]	0.008	0.8 [0.8]
P01834	Ig kappa chain C region	IGKC	0	0	0.000	3.6 [1.4]	0.000	0.8 [0.8]
P19823	Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2	0	Y	0.008	1.4 [0.9]	0.001	1.1 [1.1]
P00747	Plasminogen	PLG	0	Y	0.002	2.0 [1.2]	0.029	1.1 [1.1]
P05155	Plasma protease C1 inhibitor	SERPING1	0	Y	0.025	1.7 [0.9]	0.000	1.2 [1.2]
P09486	Secreted potein acidic and cysteine rich	SPARC	0	Y	0.002	0.8 [0.7]	0.000	1.1 [1.1]
P02787	Serotransferrin	TF	0	Y	0.000	4.8 [2.3]	0.000	0.7 [0.7]
P07996	Thrombospondin-1	THBS1	0	Y	0.017	0.8 [0.7]	0.013	0.9 [0.8]

Accession ID	Protein name	Gene Symbol	Phobius		CD4-depleted data		Serum data	
			TMD	SP	q-value	Fold change	q-value	Fold change
P04220	Ig mu heavy chain disease protein	not defined	0	0	0.036	2.4 [1.4]	0.000	0.7 [0.7]

Notably, circulating myeloid-derived (granule) proteins were largely underrepresented in the entire serum data, for instance DEFA1, PRG2 and RNASE3 were not detected at all, although corresponding spectra were contained in the spectral library. MPO and PRTN3 were only identified as one-peptide-hits with corresponding abundance ranks of 160/184 and 313/326 (control/T1D groups), respectively. Although one-peptide-hits were not included in ANCOVA analysis to enhance data validity, both proteins exhibited on average lower levels in the T1D group, contrarily to observations in the PBMC data. Intriguingly, two innate inflammatory mediators (S100A9 and cathelicidin antimicrobial peptide (CAMP)) which achieved reasonable proteome coverage in the serum data (covered with 7 and 2 unique peptides) and which I found to be upregulated in the cellular data (see Fig. 24, 30C, 38C, 39A and Tab. 13) exhibited rather significant downregulation (Tab. 18-19). This finding is surprising and contrary to the expectations resulting from PBMC profiling.

Next, I was particularly interested in the candidate biomarkers (q<0.05) with the highest fold changes, since only a small proportion of significantly regulated proteins exhibited a distinct up- or downregulation in the serum data. In Fig. 46 protein abundance was plotted for 15 proteins (≥2 unique peptides) which were regulated by at least 1.5-fold (mean or median) corresponding to six proteins with higher and 9 proteins with lower expression levels in T1D. Potential contaminants such as Hornerin (HRNR) and Ig-chains were not considered for the plot. Among the six proteins enriched in the T1D serum proteome were apolipoprotein A4 (APOA4), adenylyl cyclase-associated protein 1 (CAP1), fibrinogen alpha chain (FGA), hepatocyte growth factor-like protein (MST1), NCAM1 (CD56) and talin 1 (TLN1). The top nine downregulated proteins were conserved oligomeric Golgi complex subunit 5 (COG5), Fetuin-B (FETUB), plastin-2 (LCP1), profilin-1 (PFN1), retinol-binding protein 4 (RBP4), SH3 domain-binding glutamic acid-rich-like protein 3 (SH3BGRL3), transthyretin (TTR), Vasorin (VASN) and Villin-like protein (VILL).

In summary, serum profiling at T1D onset confirmed the previously observed overrepresentation of several complement factors and members of the apolipoprotein family. However, in-depth monitoring of myeloid-derived innate mediators, which were

key signature proteins in the cellular data, was not possible suggesting very low abundance or absence from the serum.
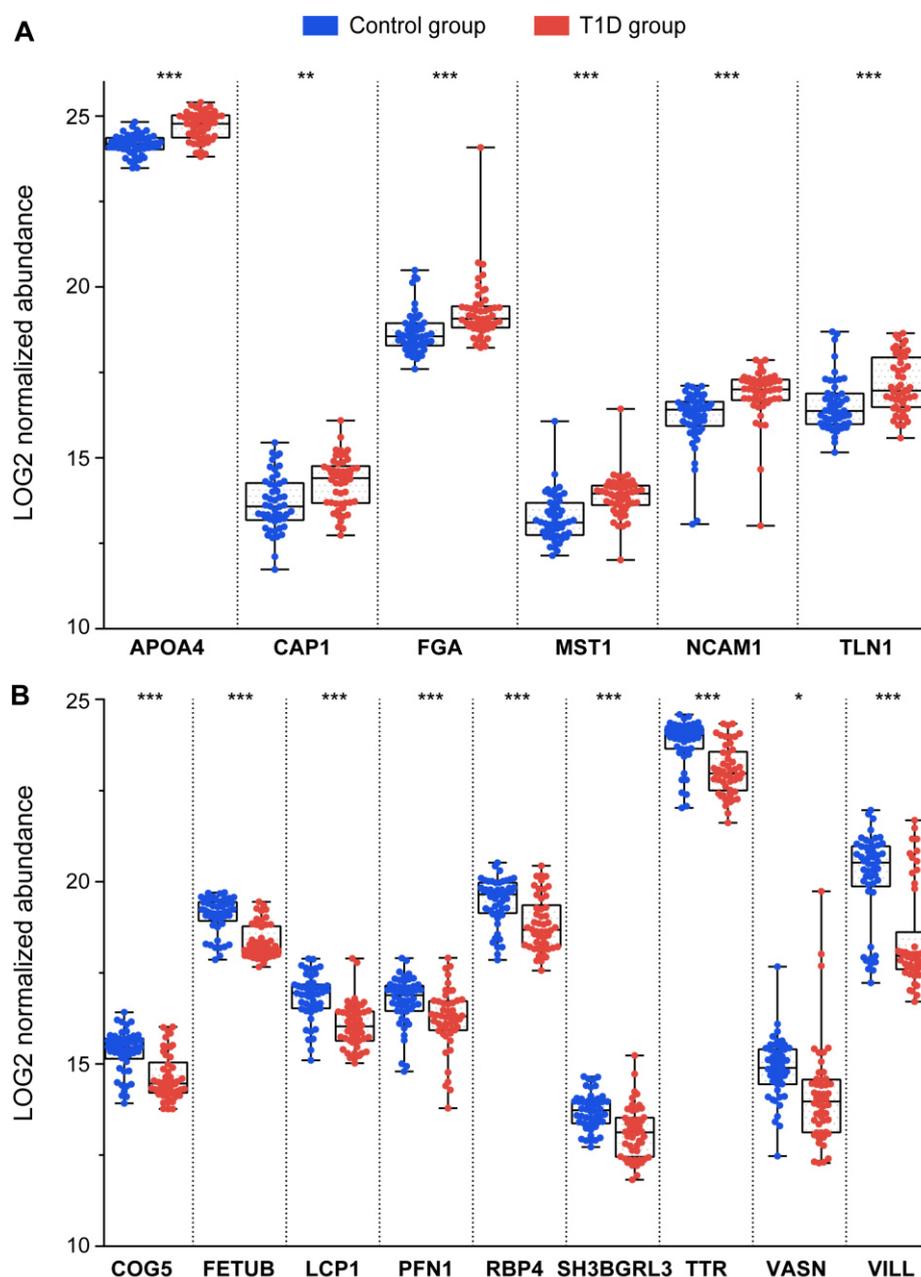
Figure 46 Box plots of proteins with significantly different expression between T1D and control groups in serum samples measured with DIA LC-MS/MS.
Differential protein expression between T1D and control groups in the serum proteome accounted for 111 proteins (q<0.05). Depicted here are the top six proteins with higher (A) and the top nine proteins with lower expression in the T1D group (B). All proteins were quantified with ≥2 unique peptides, exhibiting at least a mean or median regulation of ≥1.5-fold (q<0.05). The boxes represent the interquartile range (IQR), and the median value is shown as a solid black line within the box. The upper and lower whiskers represent the maximum and minimum value, respectively. *q<0.05, **q<0.01 and ***q<0.001.

4 Discussion

4.1 MS sample preparation and data acquisition strategies

4.1.1 Evaluation of sample preparation strategies for blood cell proteomics

Evaluation of immune intervention trials in T1D relies on metabolic endpoints, and progress in the field has been hampered by the lack of biomarkers capable of monitoring immune and metabolic progression of T1D (Lebastchi and Herold, 2012a). A promising strategy for discovering novel disease-associated signatures is proteomic phenotyping of the highly accessible PBMC population. Hence, the aim of the first stage of this thesis was to evaluate different proteomic sample preparation techniques for their applicability to limited numbers of primary human CD4+ T cells and PBMCs. The reasoning behind "simulating" working conditions with reduced cell count was the intended translation of these protocols to biobanked PBMC samples which are finite in quantity and exhibit reduced viability after thawing. All LC-MS/MS experiments conducted during the first stage of this thesis employed the DDA mode on the Orbitrap XL platform.

4.1.1.1 Investigation of Cell Surface Capturing with the Glyco-PAL strategy

First, qualitative profiling of the cell surface compartment (surfaceome) applying the Glyco-PAL strategy was assessed (Fig. 15). Cell surface proteins function as the primary signaling mediators at the cellular interface to the microenvironment being responsible for information sensing and triggering of subsequent immune reactions (Bock et al., 2012). The surfaceome further orchestrates cellular communication and is vital for correct functioning of distinct immunocyte subsets; hence differential cell surface protein expression could possibly accompany an auto-inflammatory situation such as T1D development. Additionally, beyond circulatory proteins, cell surface proteins are prime biomarker candidates being easily accessible for tailored clinical screening assays such as conventional FACS analysis or novel CyTOF technology (Bendall et al., 2011). The Glyco-PAL strategy introduced by Zeng et al. (2009) enables a simple approach for introducing molecular biotin tags into glycoproteins on the surface of living cells. The strategy was chosen because it combines the stringent selectivity of periodate-mediated oxidation of sialic acids with aniline-catalyzed oxime ligation under mild experimental conditions (neutral pH) reportedly maintaining high cell viability (Zeng et al., 2009).

Glyco-PAL surfaceome profiling applied to one million freshly isolated CD4+ T cells resulted in the identification of more than 100 proteins achieving enrichment factors of approximately 60% for GO annotated plasma membrane proteins (Fig. 16B) and 30-40% for integral plasma membrane proteins, respectively (Fig. 16A). In fact, several key

proteins involved in T cell signaling processes such as the *T cell receptor* were enriched in pathway enrichment analysis (Fig. 16C and Tab. 8). Although these experiments proved in principle the feasibility of applying Glyco-PAL to limited numbers (1×10^6) of primary CD4+ T cells, the absolute numbers of identified plasma membrane proteins were not satisfactory (~70 proteins, Fig 16B). This is substantiated by the fact that that roughly one quarter of human protein-coding genes are predicted to contain membrane domains, and the estimated size of the immunological surfaceome accounts for approximately 2383 proteins according to Uniprot protein database annotation (Bock et al., 2012; Fagerberg et al., 2010).

The insufficient number of encountered protein identifications is in part attributable to the limitation of the input cell count; in comparison Graessel et al. (2015) performed a comparable experiment applying the Glyco-PAL strategy to eight million naïve CD4+ T cells from four healthy donors. Graessel and colleagues subsequently reported the identification of 242 proteins including 173 N-glycoproteins as indicated by GO or CD annotation, corresponding to an enrichment factor of 72%. Importantly, both the Glyco-PAL experiments presented in this thesis as well as the work published by Graessel and coworkers employed the same HPLC-MS platform (Orbitrap XL), yet the reduction in enriched surfaceome proteins against background identifications accounted for approximately ~10% (60% vs. 72%) or ~100 proteins (70 vs. 173) in absolute numbers, respectively. In contrast, the Glyco-PAL strategy was recently applied to a human alveolar adenocarcinoma cell line (A549) to investigate the effects of cigarette smoke exposure resulting in the identification of >1000 potential surfaceome proteins using the Orbitrap XL platform (Mossina et al., 2017) (personal communication J. Merl-Pham). This finding highlights that beyond the amount of cellular starting material, the performance of the Glyco-PAL technique is likely affected by additional cellular properties such as adherent (A549) or suspension (T cells) culture conditions, the culturing environment (unstimulated vs. stimulated) and the intrinsic complexity of the surfaceome. Although Zeng et al. (2009) proved that the aminooxy-biotin derivate is membrane-impermeable and the biotin-labeled glycoproteins are further not endocytosed at 4 °C, it is likely assumable that in situations of reduced cell viability (e.g. after biobanking) membrane porosity might lead to enhanced labeling of intracellular sialyated glycoproteins found in subcellular organelles, thus enhancing the already prominent background identifications. Additionally, the coverage of the surfaceome observed in the Glyco-PAL experiments likely reflects the more abundant cell surface proteins due to the intensity-based selection of peptide precursor fragmentation in DDA-based profiling (Michalski et al., 2011). In turn, it is highly questionable if proteomic changes accompanying T1D onset

would manifest in this limited panel of surface proteins. In consequence, the translation of the Glyco-PAL strategy to patient-derived biobanked blood cell samples was not pursued in favor of whole-proteome profiling approaches without subcellular fractionation (iST, FASP). Notably, both methods achieved a comparable identification rate for integral plasma membrane proteins making the complex sample preparation employed in Glyco-PAL dispensable (Fig 17B). Here, the crucial advantage was that both the FASP and the iST method achieved these protein identifications with one quarter of the cellular input material while at the same time immensely reducing the handling time of viable cells — a decisive factor when working with limited sample material. The decision to abandon the translation of surfaceome profiling to biobanked samples was further substantiated by colleagues already working with comparable PBMC samples from the biobank who estimated the realistic number of obtainable CD4+ T cells from a sample to range around a minimum of 250,000 (personal communication C. Daniel).

4.1.1.2 Exploration of FASP and iST as alternative sample preparation methods

The insufficient results obtained for surfaceome profiling of CD4+ T cells urged me to pursue simpler proteomic sample preparation methods for PBMC profiling characterized by minimal handling time of viable cells and applicability to limited cell numbers such as the new benchmark of 250,000 CD4+ T cells. To this end, I decided to evaluate protocols aimed at whole-proteome profiling without preceding subcellular fractionation. Here, the crucial advantage is that once the desired cell population (e.g. CD4+ T cells) is isolated, it can either be pelleted or directly lysed and further be stored at -80 °C for an intermediate time period until further processing.

The sample preparation methods which were chosen for evaluation were FASP and the iST method which were both applied to snap-frozen cell pellets of CD4+ T cells and PBMCs of distinct cell count using a Urea-based lysis buffer (Kulak et al., 2014; Wisniewski et al., 2009). FASP is the standard method employed in our research group due to its simple and straight-forward protocol, its applicability to a wide range of detergent-containing lysis buffers and its compatibility with medium-throughput processing (approximately 50 samples can be handled by one operator per day). FASP takes advantage of a common, commercially-available ultrafiltration device (30kDa cut-off) functioning as a "proteomic reactor" for protein concentration, buffer exchange, chemical modification and protein digestion, thus reducing sample handling while at the same time enabling depletion of substances interfering with downstream LC-MS/MS analysis (Wisniewski et al., 2009). FASP was compared to the more recently described iST strategy which is a one-pot reaction thought to radically reduce contamination,

sample loss and workflow-induced PTMs (Kulak et al., 2014). The advantage of the iST method is that it principally allows to integrate several sample processing steps in one, such as cell lysis in a buffer already containing the reduction and alkylation agent (TCEP and CAA). Additionally, the reaction chamber is a low-cost pipette tip which is equipped with a plug of chromatographic material such as SDB-RPS functioning first as a barrier during processing, and secondly as a filtration and separation medium during elution (Kulak et al., 2014). This chromatographic plug further enables peptide-based fractionation depending on properties like hydrophobicity and polarity which cannot be directly incorporated into the FASP protocol. Specifically, SDB-RPS chromatographic material was chosen because Kulak and coworkers reported that a three-fraction SDB-RPS approach resulted in higher peptide identification rates for HeLa cells than a six-fraction SAX approach. A general advantage of the iST method over FASP is that its five manual steps can be performed in less than 30 min (excluding digestion). On the other hand, a disadvantage of iST was founded in the fact that the reaction device was not commercially available during the first two years of this thesis making manual construction necessary. This construction process included the punching of the SDB-RPS disk with a needle and positioning of the plug in a reproducible fashion in the pipette tip, thus potentially introducing workflow variation and/or contamination. By 2016, the iST method became commercially available with the introduction of the PreOmics Kit which was used for serum profiling in this thesis in a high-throughput 96-well format.

4.1.1.2.1 Sensitivity

Proteomic profiling of predefined counts of CD4+ T cells (~250,000) and PBMCs (~170,000) processed with the FASP and iST methods revealed that the latter achieved overall slightly better protein and peptide identification rates (Fig. 17A). The increased proteome coverage achieved with iST was likely attributable to the employed peptide fractionation into three distinct pools compared to single-shot elution in FASP. A similar experiment was performed in collaboration with Grosche et al. (2016) to compare whole-proteome profiling of murine retinal tissue processed with either the iST method or a subcellular fractionation approach with FASP-based protein digest (surfaceome, cytosol, nucleus). Subsequently, Grosche et al. (2016) reported 3077 identified proteins in the iST experiment compared to 1649 identified proteins in the subcellular fractionation approach, respectively. Notably, Grosche and colleagues detected 1754 proteins exclusively using the iST method compared to 326 exclusive proteins in the subcellular fractionation approach, highlighting the high sensitivity achieved with iST-based proteomic profiling. In consequence, it was decided to further pursue the iST over the FASP method for translation of proteomic profiling to biobanked samples. By reducing

the complexity of the sample with distinct peptide fractions, it was anticipated to improve peptide resolution and identification, thus counteracting the low sampling speed of the Orbitrap XL instrument. Here, a greater analytical depth was chosen over time- and cost-effectiveness, given the much higher MS run time required for iST samples.

4.1.1.2.2 Reproducibility

Next, technical reproducibility of the iST process including cell pellet lysis, device construction, iST processing and label-free quantitative LC-MS/MS analysis was determined across 5 replicates. The average CV accounted for 22% and measured protein abundances were highly correlating across replicates with a mean coefficient of determination (R^2) of 0.96 (Fig. 18-19). Similarly, Kulak et al. (2014) also reported excellent quantitative reproducibility between quadruplicate measurements of iST processed HeLa cells with values of R^2 corresponding to 0.97. If technical reproducibility was considered only for the label-free quantitative LC-MS/MS part, and thus in isolation from the remaining iST processing steps, average CVs of 10% were routinely observed for repeated injections of the same sample into the HPLC system as reported in collaboration with Grosche et al. (2016) on the same MS instruments. In summary, the documented technical reproducibility for the entire iST workflow was excellent being composed of roughly 10% iST processing variation and 10% label-free quantitative LC-MS/MS variation, the latter emphasizes the presence of robust HPLC performance and high alignment quality. In general, the semi-stochastic nature of peptide sampling in DDA LC-MS/MS hampers reproducibility which means that in repeated measurements of the same sample a different population of precursors will be selected for fragmentation (Tabb et al., 2010). However, by employing MS1-based quantification, as performed here, this problem can be alleviated taking advantage of peptide feature matching across sample runs with accurate alignment of MS1 XICs, resulting in excellent CV values. In comparison, Ahsan et al. (2017) investigated the quantitative reproducibility of label-free DDA LC-MS/MS analysis for TiO_2 enriched phosphopeptides from Jurkat T cells to evaluate distinct LC gradients and analytical column constructions. Ahsan and coworkers compared ion chromatogram peak areas of an exogenously spiked standard peptide across 5 replicates and found CV values to be <20%. This example highlights that in the field of label-free quantitative proteomics a process variation of approximately <20% is commonly desirable, independent of employed sample preparation and HPLC-MS setup. In conclusion, the observed average CV of ~20% for the iST workflow was highly acceptable when considering that it contained all steps possibly introducing process variation comprising cell lysis, iST construction and processing, and quantitative label-free LC-MS/MS analysis. This was even more valid, since in my analysis the average CV

was calculated across >1000 of quantified proteins rather than by evaluating a single indicator peptide.

4.1.2 Data-dependent versus data-independent acquisition mass spectrometry

After processing of the first clinical sample with the iST method, which was specifically chosen to counteract low sampling efficiency of the Orbitrap XL platform, new instrumentation, namely Q Exactive HF became available (Fig. 11), which expanded the analytical limits for proteomic profiling of complex samples. The Q Exactive HF is a state-of-the-art instrument which is characterized by advanced quadrupole technology for improved precursor selection and transmission as well as an ultra-high-field mass Orbitrap analyzer enabling a doubling of resolution or acquisition speed compared to conventional Orbitrap analyzers (Scheltema et al., 2014). With availability of the Q Exactive HF instruments in our lab, the sample preparation method for CD4+ T cell/CD4-depleted cell profiling of sample set II was changed in favor of single-shot elution FASP. This protocol change was implemented for two reasons: first the enhanced performance of the new MS instrument already enabled high analytical depth for samples without peptide pre-fractionation, thus saving a great amount of MS measurement time. Secondly, the aim was to achieve greater analytical profiling depth by applying MS2 spectra acquisition in data-independent mode rather than by employing peptide pre-fractionation. Data acquired with all-ion fragmentation workflows such as DIA is continuous in time and fragment ion intensity, hereby increasing data dimensionality in comparison to DDA profiling where full fragment ion intensity scans are only recorded at defined time points (see Fig. 12) (Rost et al., 2014b). In consequence, DIA profiling has the great potential to overcome several of the main drawbacks of DDA profiling such as semi-stochastic and irreproducible peptide precursor selection, "undersampling" and long instrument cycle times (Li et al., 2009; Liu et al., 2004; Michalski et al., 2011; Navarro et al., 2016).

When PBMC samples were profiled on the Q Exactive HF platform in a comparative experiment employing both DDA and DIA LC-MS/MS successively, a marked increase in proteome coverage could be achieved in the DIA mode (Fig. 28A-C and 36A-C). For DIA profiling of the CD4+ T cell subset, identified proteins accumulated to a total number of 8160 (6377 with ≥ 2 unique peptides) representing a 2.3-fold (2.6-fold for ≥ 2 unique peptides) increase in protein identifications compared to DDA-based profiling (Fig. 28A). Accordingly, DIA profiling of the CD4-depleted cell fraction resulted in 8932 identified proteins (7248 with ≥ 2 unique peptides) corresponding to a 2.4-fold (2.8-fold for ≥ 2 unique peptides) increase in analytical depth (Fig. 36A). Besides the considerable

improvement observed with regards to the qualitative readout, the quantification of overlapping protein identifications was also enhanced indicated by higher peptide per protein coverage (Fig. 28C, 36C/E, and Tab. 10). Specifically, the peptide per protein coverage for proteins identified with both acquisition methods was increased by an average of 8 (CD4+ T cells) and 10 (CD4-depleted cells) peptides, respectively (Fig. 28C and 36C). Importantly, this comparative experiment was performed by profiling the exact same samples, and thus independent of employed sample preparation, HPLC-MS platform or biological variation. Naturally, the downstream software used for MS data analysis was different with the application of Progenesis QI for the DDA data and Spectronaut for the DIA data.

The much higher protein identification rates observed in the DIA approaches (CD4+ T cells/CD4-depleted cells) are primarily attributable to overcoming intensity-based peptide precursor selection by comprehensively fragmenting all detectable peptide precursors of a selected mass range in predefined, overlapping isolation windows. The DDA mode is significantly flawed by the employed top 10 method in which only the ten most abundant peptide precursors from the survey scan are chosen for fragmentation. In detail, this phenomenon known as "undersampling" generally results in only 16% of the eluting peptides being targeted for fragmentation, and only 60% of these peptides can be successfully identified at the 1% FDR threshold (Michalski et al., 2011). Another potential drawback of the DDA strategy employed here is founded in the concept of simultaneous protein identification and quantification in which a quantification-relevant survey scan is directly coupled to subsequent MS2 scans relevant for peptide identification. In general, quantification precision is improved by higher numbers of survey scans across an LC peak for optimal reconstruction of XICs; however maximal analytical depth is achieved by accumulating as many MS2 spectra as possible (Bantscheff et al., 2012). In consequence, maximizing one factor (e.g. analytical depth) usually sacrifices the other (e.g. quantification precision) highlighting that finding the right balance is crucial (Radulovic et al., 2004). DIA-based profiling overcomes this limitation by uncoupling of parallel peptide identification and quantification.

Bruderer et al. (2015) compared DDA and DIA acquisition modes in single MS runs using stable human cell line protein extracts (HEK-293) and demonstrated the superiority of the DIA LC-MS/MS approach. Specifically, Bruderer and coworkers reported on average 28610 identified peptides per measurement in their DIA experiment compared to an average of 17547 peptides in their DDA approach, respectively. The authors attribute the average surplus of 60% more peptide identifications to the comprehensive

nature of precursor fragmentation in DIA and to their comprehensive spectral library. Accordingly, the improved proteome coverage observed for blood cell DIA profiling in this thesis has similarly benefited from the comprehensive meta human spectral library used for spectrum-to-spectrum matching. Notably, the construction of this peptide spectral library required DDA LC-MS/MS profiling of a diverse array of human tissues eventually resulting in an accumulation of spectra corresponding to 432,673 different peptide precursors and 11403 unique protein groups. To date, this is one of the most in-depth human peptide spectral libraries generated — larger than the human repository recently published by Rosenberger et al. (2014) which comprised spectra for 210,921 peptide precursors and 10524 protein groups after import into Spectronaut. Notably, the meta human spectral library used in this thesis contained as little as 701 protein entries based on a single peptide reflecting only 6% of all entries; in comparison the peptide library described by Rosenberger and coworkers comprised 1252 single hits (12%).

One potential limitation of the DDA/DIA comparison employed in this thesis is that protein identifications are not compared between repeated injections of single runs but between the accumulated protein identifications resulting from matching-between-runs for the entire sample set. The match-between-runs function was enabled for both DDA and DIA data analyses, yet it is difficult to compare due to the immense differences in underlying MS2 data acquisition. In Progenesis QI, first alignment of different LC-MS runs is performed by two-dimensional overlay of m/z versus retention time. Secondly, aligned runs are transformed into an aggregate data set enabling consistent peptide peak picking which results in a single map of peptide ions for each sample. Differences in the area under the peak for XICs (MS1) are then exploited to evaluate quantitative difference and protein identity is revealed if corresponding MS2 spectra information is available in at least one sample. Accordingly, matching-between-runs was also enabled in the DIA workflow with the "q-value sparse" mode, i.e. a particular peptide signal was considered for identification and quantification across all samples when identified with corresponding MS2 spectra at a FDR below 1% (q<0.01) in a single run. As previously outlined, acquisition of MS2 spectra is continuous in DIA whereas discontinuous in DDA, thus DIA data intrinsically harbors much more MS2 spectra available for peptide assignment leading to a much higher number of transferred peptide identifications (Fig. 12). Notably, by fine tuning of the match-between-runs function in the DIA mode the operator may greatly influence overall peptide and protein identifications. For instance, the operator may apply a q-value percentile setting which means that only peptide signals accurately assigned with high quality MS2 spectra (q<0.01) in a predefined percentile of all samples (e.g. 25%) are considered for identification and quantification

across all runs. For instance, if a q-value percentile filter of 25% was applied to the CD4+ T cell DIA data analysis, protein identifications would have dropped from 8160 to 5751, with the corresponding increase in proteome coverage compared to the respective DDA experiment dropping from 2.4-fold to 1.6-fold. However, for DIA-based blood cell profiling employed in this thesis, the more flexible "sparse" mode was chosen in favor of greater analytical depth and better identification of low-abundant proteins which from an empirical point of view are often lost when data filtering is set too stringently. This decision was further substantiated by the fact that protein identifications based on a single peptide ranged from 17% to 22% in the DIA analyses of CD4-depleted and CD4+ T cells, respectively. In comparison, one-peptide-based protein identifications ranged between 31-32% for the counterpart experiments performed in DDA mode (see Fig. 28A and 36A). As a precaution for avoiding reporting of false positive identifications one-peptide-based protein identifications were excluded from any downstream statistical analysis aimed at extraction of differentially regulated candidate proteins — independent of DDA or DIA profiling mode.

Beyond improved proteome coverage, another principal advantage resulting from continuous MS2 spectra acquisition in DIA is that peptide quantification can be based on fragment ion level (MS2), as employed in this thesis, rather than on the precursor level (MS1). In fact, the DIA workflow described in this thesis uses more than 90% of the acquisition time for generating quantification-relevant MS2 spectra, thus improving quantification precision. In contrast, a conventional DDA workflow uses only approximately 25% of the total acquisition time for quantification-relevant MS1 spectra (Bruderer et al., 2015). Additionally, peptide quantification on MS2 level is less susceptible to interferences because the likelihood of simultaneous interferences for all fragment ions is low (Bruderer et al., 2015). In a comparable study, Muntel et al. (2015) investigated urinary biomarkers identifying twice as many peptides and proteins with half of the CV using the DIA approach compared to conventional shotgun proteomics. Likewise, Bruderer et al. (2015) found that CVs of proteomic profiles measured in DIA were 53% lower than the respective profiles generated with shotgun proteomics. Accordingly, when comparing overlapping CD4+ T cell protein identifications in DDA and DIA data sets (Fig. 28B-C), a halving of the respective average CV values was observable for T1D (40% vs. 80%) and control groups (32% vs. 61%) (data not shown).

The comprehensive CD4+ T cell proteome created by DIA profiling in this thesis achieved >70% recovery of known CD4+ T cell expression, corresponding to ~6000 proteins, as published in the human proteome map project (Kim et al., 2014) (Fig. 28E).

Discussion

Additionally, the CD4+ T cell DIA data further recovered >70% of a publicly available CD4+ T cell surface atlas published by Graessel et al. (2015) (Fig. 28F). This finding emphasizes good coverage of proteins which are commonly only accessible by specialized methods such as Glyco-PAL and antibody-based profiling — both of which were employed for generation of the surface atlas (Graessel et al., 2015). The reasonable good coverage of the plasma membrane compartment likely stems from the integration of samples generated by Glyco-PAL into the design of the meta human spectral library (Zeng et al., 2009). This finding highlights that applying DIA LC-MS/MS to a small number of cells (e.g. 250,000 CD4+ T cells) can effectively complement and possibly even replace complex surfaceome profiling methods, which are often only applicable to large numbers of viable cells.

In summary, both the comparative profiling of CD4+ T cells/CD4-depleted cells presented in this thesis as well as studies by Bruderer et al. (2015) and Muntel et al. (2015) demonstrate the superiority of DIA over DDA LC-MS/MS profiling with regards to analytical depth and quantification precision. Yet, central advantages of the DDA workflow are its comprehensive and established software analysis pipelines which have been developed over more than two decades making shotgun proteomics the gold standard in the field (Hu et al., 2016). In contrast, post-analysis in the DIA workflow is rather laborious if peptide spectral libraries are not readily available and need to be tailored to the biological question. In consequence, the intermediate future will show whether DDA will yield to DIA for a broad application to biological systems. On the way, the proteomic community needs to establish a consensus on still uncleared issues such as statistical control of peptide and protein error rates in large-scale data-independent workflows (Rosenberger et al., 2017). Last but not least, it needs to be appreciated that neither with DDA nor with DIA, proteomics has yet advanced to the stage where every single protein present in a sample can be reliably and reproducibly quantified.

4.2 Proteomic profiling of CD4+ T cells and CD4-depleted cells in recent-onset T1D

The causative agent that triggers islet cell autoimmunity and onset of T1D remains poorly understood. In this thesis, the first blood cell-based proteomic study of recent-onset T1D using state-of-the-art LC-MS/MS was carried out. The primary focus was put on the CD4+ T helper cell compartment, which has historically been considered as a key player in mediating β-cell destruction (Haskins and Cooke, 2011; Roep, 2003). All patient- and control-derived PBMC samples were cryopreserved in a biobank, showing the applicability of the presented profiling approaches for analyzing quantity- and/or quality-limited samples (see also Fig. 22). This factor is particularly relevant in terms of

the increasing interest in clinical proteomics to provide solutions to the growing need for personalized medicine (Duarte and Spencer, 2016; Hughes and Morin, 2017; Yoo et al., 2017). In total, two clinical PBMC sample sets containing 114 pediatric donors were subfractionated into CD4+ T cells and the CD4-depleted cell pool and profiled by a combination of data-dependent and data-independent LC-MS/MS approaches to obtain a comprehensive proteomic fingerprint of the PBMC population in recent-onset T1D (Fig. 21 and 31).

One central finding was that proteomic changes underlying T1D onset did not translate to a global, proteome-wide scale as indicated by principal component or hierarchical cluster analyses. In fact, patient group-specific clustering was neither observed for CD4+ T cells nor for the CD4-depleted cell fraction across both clinical sample sets (Fig. 23C, 29, 37). This observation can be attributed to different factors: (i) first a pathologic condition such as T1D does not necessarily result in changes affecting the global proteome of a broad range of cell types (present in PBMCs). Here, a potential drawback of monitoring complex cell populations such as pan CD4+ T cells and the heterogeneous pool of CD4-depleted cells is the likeliness to overlook changes affecting only a small subset of the monitored cells. For example, autoreactive, diabetogenic T cells are likely prime candidates for harboring phenotypic features differing from those of benign T cells. Indeed, several studies provided evidence that autoreactive T cells exhibit pro-inflammatory polarization in T1D and a regulatory phenotype in healthy subjects (Arif et al., 2004; Chujo et al., 2013; van Lummel et al., 2014). However, autoreactive T cells are only present at extremely low frequencies in the circulation (Pugliese, 2017). Specifically, their residence is expected to be in or around the pancreas, and their presence in the circulation is probably limited to distinct periods of activation and expansion highlighting their limited practicality to function as biomarker discovery tool in the periphery (Bonifacio, 2015). (ii) Secondly, the lack of disease-specific sample clustering as observed in this thesis for all cellular profiles emphasizes the rather high degree of inter-individual variation in global protein abundance, accounting for mean protein CV values of approximately 50% in the healthy control group (56% for T1D group). The observed inter-individual biological variation likely stems from differences in genotype, composition of cellular subsets, environmental influences or physiological effects such as diet and stress. Accordingly, heterogeneity in human peripheral blood cell expression is a common observation being the topic of investigation in several gene expression profiling studies (Eady et al., 2005; Radich et al., 2004; Whitney et al., 2003). Despite the issue of general heterogeneity in human samples, highly overlapping and statistically significant

Discussion

differences in protein abundances between T1D and control children could be established in this thesis.

4.2.1 Recent-onset T1D is associated with an inflammatory proteomic signature

Using both DDA and DIA MS approaches, I obtained proteomic profiles of global CD4+ T cell and CD4-depleted cell expression and identified a clear inflammatory phenotype in patients with T1D. Strikingly, this phenotype is closely connected to an inflammatory extracellular milieu indicated by the profound enrichment of proteins predicted to prevail and function in secretory granules, the extracellular space and the blood plasma compartment, respectively (Tab. 11 and 14). A primary characteristic of the inflammatory signature is the strong association with the myeloid cell lineage as evidenced by overrepresentation of proteins linked to the neutrophil and to a lesser extent also the monocyte cell types (summarized in Fig. 24A-B, 27, 30B-C, 38C, 39A and Tab. 9, 11, 13, 14). Additionally, proteins implicated in platelet and complement activation (e.g. CFB or SERPING1) as well as apolipoproteins (e.g. APOE) are key features of the observed proteomic pattern (Fig. 25, 30C, 38C). Notably, this inflammatory phenotype is distinctly visible throughout both investigated cellular compartments, i.e. CD4+ T cells and CD4-depleted cells, and further independent of employed sample preparation, MS data acquisition or platform, and study cohort sample set, hence represents a very stable finding. In summary, enhanced protein levels of (circulating) innate immune mediators, e.g. neutrophil-derived or complement proteins, were present in the entire patient-derived PBMC sample suggesting that innate immune activity was altered during the onset of overt metabolic T1D.

Although T1D has largely been attributed to aberrant adaptive immunity, increasing evidence suggests that innate inflammation also plays a role in the underlying pathogenesis (Cabrera et al., 2016b; Kolb and von Herrath, 2017). An elevated or dysregulated state of innate inflammation present in patients with T1D has previously been observed in multiple PBMC gene expression profiling studies and in a global serum-based proteomic analysis (Jin et al., 2013; Kaizer et al., 2007; Zhang et al., 2013a). Specifically, Kaizer et al. (2007) observed elevated transcript levels of *IL1B* in PBMCs from patients with recent-onset T1D, and Stechova et al. (2012) further identified differences in IL-1 signaling among T1D subjects, healthy relatives and unrelated healthy controls. Many of the signature transcripts described by Kaizer and colleagues such as *IL1B*, *CXCL1*, *EGR2/3* and *TREM1* were not identified in the present proteomic analysis, likely due to their low protein abundance, making a qualitative comparison difficult. Yet, PEA revealed that IL-1 signaling was elevated in the CD4+ T cell compartment of

patients with T1D suggesting that several signature proteins are closely linked to this pathway (Tab. 11).

Several studies suggest that monocytes in particular give rise to the elevated state of innate inflammation in T1D by exhibiting functional abnormalities evidenced by increased production of pro-inflammatory IL1-ß, IL-6, superoxide anion, and cyclooxygenase-2 (PTGS2) (Devaraj et al., 2006; Litherland et al., 2003; Plesner et al., 2002). Accordingly, increased gene expression levels of C-C motif chemokine ligands 2 and 7 (*CCL2*, *CLL7*) which are implicated in chemotaxis and cell migration were reported in monocytes derived from subjects with T1D (Beyan et al., 2010; Padmos et al., 2008). Although both CCL2 and CCL7 were not identified in the proteomic profiles described in this thesis, CCL2 signaling was among the overrepresented pathways in the CD4-depleted cell pool of the T1D group which harbors the peripheral monocyte population (Tab. 14). Interestingly, the overexpression of myeloid-derived inflammatory genes in T1D was further established in a large-scale study investigating PBMC samples from almost 2000 study participants (Jin et al., 2013). Importantly, several of the key signature genes reported by Jin and coworkers were also found to be significantly upregulated in the T1D group in the present proteomic investigation including S100A8 (CD4- cells), S100A9 (CD4+ T cells, Fig. 24A-B), MNDA and CD74 (both fractions).

At least in the case of the CD4+ T cell preparations, the presence of myeloid-derived inflammatory signature proteins is unexpected; however peripheral monocytes do express the CD4 molecule albeit at much lower frequencies than T cells (Filion et al., 1990). Additionally, endogenous CD4 expression has also been reported for neutrophils, the most abundant representative of the myeloid cell population in peripheral blood (Biswas et al., 2003). In consequence, it cannot be entirely precluded that myeloid cells were present in these CD4+ cell preparations. However, overall cell counts of neutrophils and monocytes have been reported to be downregulated at T1D onset (Harsunen et al., 2013; Valle et al., 2013). This finding would argue against a frequency-induced overrepresentation of these cells in the CD4+ or CD4-depleted cell fractions. Moreover, significant protein abundance changes between T1D and control groups were not observed for key surface markers CD14 and CD33 of the monocyte population (see Tab. 15) (Simmons et al., 1989; Ulyanova et al., 1999). Nonetheless, pro-inflammatory myeloid cell polarization as indicated by higher expression of inflammatory mediators does not need to be coupled to cell count and could possibly even exist in a situation of perturbed leukocyte homeostasis as observed by Harsunen et al. (2013).

Discussion

Interestingly, protein abundance levels of circulating CD14 were found to be significantly higher in the T1D group in the serum profiling (q-value 0.008, 1.2-fold higher). In fact, CD14 does not only exist as a GPI-anchored plasma membrane protein but may also enter the extracellular space by cleavage from the cell surface or release from intracellular storage pools (Bazil and Strominger, 1991; Bufler et al., 1995; Durieux et al., 1994). In fact, elevated protein levels of soluble CD14 have been reported in a broad spectrum of pathologic conditions including viral infections (HBV, HCV and HIV), chronic heart failure, pediatric inflammatory lung diseases, and rheumatoid arthritis (Anker et al., 1997; Bas et al., 2004; Lien et al., 1998; Marcos et al., 2010; Mendez-Lagares et al., 2013). Intriguingly, Shive et al. (2015) recently postulated that soluble CD14 is a nonspecific marker of monocyte activation. Using in vitro PBMC cultures, the authors demonstrated that the release of soluble CD14 is induced upon activation with various pro-inflammatory stimuli such as cytokines (IL-1ß, IL-6) or different TLR ligands including lipopolysaccharide, flagellin and CpG oligodeoxynucleotides. In conclusion, elevated circulating levels of CD14 support the concept of an increased myeloid-derived innate inflammatory state in T1D.

Monocytes account for 2-8% of blood leukocytes and thus represent a rather small cell population in peripheral blood with regards to absolute frequencies (Hansen and Netter, 2014). However, the myeloid-derived inflammatory signature proteins observed in the present proteomic investigation were among the more abundant proteins indicating their prominent presence in the cell preparations (see Fig. 24C-D). This finding confers the possibility that other representatives of the myeloid cell lineage could be the source of the observed signature, in particular neutrophils which represent 50-70% of the entire blood leukocyte population (Hansen and Netter, 2014).

4.2.2 Neutrophil granulocytes as potential carriers of the inflammatory signature

Neutrophils belong to the group of polymorphonuclear leukocytes – the most abundant cell type in human peripheral blood – which play a central role in first-line defense against invading pathogens. Neutrophils are produced in bone marrow and are subsequently released into the bloodstream as terminally differentiated, dormant cells (Bainton et al., 1971; Borregaard, 2010). Upon activation, neutrophils are recruited to sites of inflammation and kill microorganisms using antimicrobial agents liberated from granules (e.g. peptides and proteases) or generated by metabolic activation such as reactive oxygen species (ROS) (Borregaard, 2010; Mayadas et al., 2014). Traditionally, neutrophils have been regarded as short-lived, non-mitotic cells producing a variety of pro-inflammatory mediators which guide the initiation and amplification of the

inflammatory response; in consequence their deployment is tightly controlled under physiologic conditions to avoid organ damage (Carmona-Rivera and Kaplan, 2013; Weiss et al., 1981).

The core inflammatory signature observed in the T1D group in the present proteomic investigation had a strong connection to the neutrophil cell type which was repeatedly listed among the top hits in pathway enrichment analyses throughout both investigated cellular compartments – independent of employed MS strategy or platform (Tab. 11 and 14, Fig. 38C). Strikingly, the only three signature proteins which achieved statistical significance (q<0.05) throughout all CD4+ T cell profiling experiments – independent of sample set or acquisition method – were key effector proteins of neutrophils: MPO, PRTN3 and RNASE3 (Fig. 27 and 30B, Tab. 9). Interestingly, these signature proteins also exhibited higher levels in the CD4-depleted cell fraction of the T1D group (Tab. 13). Additionally, higher protein abundance levels of MPO, PRTN3 and RNASE3 could be confirmed for the T1D group in both CD4+ T cell and CD4-depleted cell lysates using immunoblotting (Fig. 32 and 40). Furthermore, several other proteins strongly implicated in neutrophil function were frequently observed at higher levels in the T1D group as evidenced in profiling data from both cellular compartments across the entire study population, exemplified by proteins like Azurocidin 1 (AZU1), CAMP, CTSG, DEFA1, ELANE, GCA or S100A9 (see Fig. 24A-B, 25, 30B-C, Tab. 9 and 13).

MPO serves as a critical enzyme in the formation of ROS necessary for oxygen-dependent microbicidal activity by catalyzing the production of hypochlorous acid (Nauseef et al., 1988). MPO is further the primary component of neutrophil azurophilic granules constituting approximately 25% of the granule protein and 5% of the total neutrophil protein, respectively (Segal, 2005). Detection of MPO is commonly used as a marker of neutrophil tissue infiltration as exemplified by the investigation of pancreatic tissue from the nPOD biobank (Lundberg et al., 2017). Impaired control of MPO activity can lead to tissue damage; in fact MPO has been implicated in the pathology of a broad range of disorders such as atherosclerosis, cystic fibrosis, multiple sclerosis or rheumatoid arthritis (Daugherty et al., 1994; Kettle et al., 2004; Malle et al., 2000; Nagra et al., 1997; Podrez et al., 1999; Weiss, 1989). Other key signature proteins including PRTN3, ELANE and CTSG belong to the family of neutrophil serinproteases which are also stored in azurophilic granules where they function in the non-oxidative pathway of intracellular and extracellular pathogen destruction (Korkmaz et al., 2010). In addition to their antimicrobial activity, neutrophil serine proteases function in extracellular matrix degradation and proteolytic processing of chemo- and cytokines highlighting their role as

specific modulators of the immune response (Pham, 2006). Lastly, RNASE3 (aka eosinophil cationic protein) is an antimicrobial protein found in eosinophils and neutrophils which destroys pathogens by disrupting lipid membranes through the formation of voltage-insensitive, non-selective transmembrane pores (Monteseirin and Vega, 2008; Sur et al., 1999; Young et al., 1986). RNASE3 exhibits a low ribonuclease activity which is non-essential for its cytotoxic capacity (Rosenberg, 1995). Notably, RNASE3 cytotoxicity is not limited to parasites and bacteria, for instance Navarro et al. (2008) demonstrated that RNASE3 is capable of inducing cell death in human cell lines *via* a caspase-3-like activity. The potent cytotoxic function of RNASE3 has further been reported to affect human bronchial epithelium; in fact the protein has been extensively studied as biomarker of airway inflammation (Chang et al., 2010; Koh et al., 2007).

The overrepresentation of potentially neutrophil-derived inflammatory proteins in patient-derived PBMC samples, let alone CD4+ T cell preparations, was unexpected and raises the question of their precise tempo-spatial origin. Specifically, neutrophils are not expected to be present in large numbers in Ficoll-extracted PBMC samples because they segregate as bottom sediment in the cell isolation procedure due to their higher density. However, the Ficoll-based isolation procedure is rather a depletion of these cells and cannot guarantee 100% purity. In summary, there are several possible theories which could explain the presence of the observed neutrophil signature proteins in the present PBMC profiling data:

      (i)   low density granulocytes

      (ii)   increased degranulation in plasma (spill-over effect)

      (iii)   co-precipitation of neutrophil extracellular traps (NETs)

4.2.2.1 Low density granulocytes

Indeed, there are several studies which suggest that auto-inflammatory conditions are accompanied by higher frequencies of low density granulocytes (LDGs) which accumulate in the PBMC interphase ring during the Ficoll isolation procedure (Bennett et al., 2003; Carmona-Rivera and Kaplan, 2013; Hacbarth and Kajdacsy-Balla, 1986). Specifically, in systemic lupus erythematosus (SLE) the presence of LDGs in PBMC preparations was ascertained and found to correlate with disease activity including enhanced prevalence of skin involvement or vasculitis (Denny et al., 2010; Pavon et al., 2012; Zhang et al., 2017b). LDGs are believed to be a group of low buoyancy, immature neutrophils which are released early from the bone marrow (Bennett et al., 2003). Upon

stimulation, LDGs excessively secrete a variety of pro-inflammatory mediators including type I IFN, TNF-α, and IFN-γ which can promote tissue damage (Denny et al., 2010). In comparison to normal density neutrophils, LDGs further exhibit a strikingly higher capacity to form NETs which induce endothelial cell damage (Carmona-Rivera and Kaplan, 2013; Denny et al., 2010; Villanueva et al., 2011). With regards to T1D, an involvement of LDGs has yet to be established. Although the hypothetical presence of LDGs could be an explanation for the overrepresentation of neutrophil-derived inflammatory proteins in the patient-derived PBMC sample in the first place, this phenomenon would not explain their presence in pure CD4+ T cell preparation (see also Fig. 20). In consequence, other theories which could lead to potential explanations for the observed phenomenon need to be explored such as the degranulation (ii) and NETosis hypothesis (iii).

4.2.2.2 Degranulation hypothesis

Increased neutrophil activity at disease onset might lead to increased degranulation in the circulation and thus increase the abundance of neutrophil granule proteins in patient plasma. In turn, increased plasma protein levels could then lead to a spill-over effect into the final PBMC sample or a selective deposition on the surface of PBMCs. The degranulation hypothesis is supported by a study that found increased serum levels of circulating ELANE and PRTN3 proteins in patients with T1D, particularly in patients who were diagnosed within one year (Wang et al., 2014). Moreover, Wang et al. (2014) studied the NOD mouse and found increased levels of circulating ELANE/PRTN3 activity well before the onset of hyperglycemia and diabetes, and this activity gradually declined after the onset of overt diabetes, thus suggesting a close correlation with β-cell autoimmunity. Interestingly, I could not establish a significant correlation between measured protein levels of the key signature proteins MPO, PRTN3, RNASE3 or ELANE with either HbA1c or fasting C-peptide levels in the investigated T1D study population (Tab. 16, Fig. 41). Likewise, Wang and colleagues could neither establish an association between elevated circulating levels of ELANE or PRTN3 with the severity of hyperglycemia (HbA1c or fasting blood glucose) or fasting C-peptide levels. However, Wang and coworkers demonstrated that circulating levels and enzymatic activities of ELANE and PRTN3 were increased progressively with increased numbers of detected autoantibodies (GADA, IA2A and Znt8A). Notably, the measured titers of autoantibodies GADA and IA2A correlated positively with protein levels of ELANE and PRNT3, and their enzymatic activities (Wang et al., 2014). Thus, the elevated levels of inflammatory response proteins in recent-onset T1D do not seem to be a consequence of impaired glycemic control per se but could likely be associated to the underlying disease process.

Discussion

Additionally, Stoikou et al. (2017) investigated circulating levels of MPO and ELANE in the context of diabetes and found significantly higher levels in T1D, T2D and gestational diabetes as compared to healthy controls; notably levels were highest in T1D subjects. Interestingly, elevated plasma levels of neutrophil azurophilic enzymes MPO, PRTN3 and ELANE were also found in pediatric patients diagnosed with diabetic ketoacidosis as compared to insulin-controlled patients with T1D (Woo et al., 2016). Interestingly, Woo et al. (2016) could neither establish a correlation between protein levels of any of the three proteins with either blood glucose levels or glycated hemoglobin further supporting that neutrophil activation is likely not a mere bystander effect triggered by hyperglycemia. Elevated levels of neutrophil azurophilic enzymes in patients with diabetic ketoacidosis as opposed to insulin-controlled patients with T1D likely reflect the influence of disease duration as demonstrated by Wang and coworkers who equally reported significantly higher levels of circulating PRTN3 and ELANE in patients with shorter disease duration (< 1 year) compared to those with longer disease duration (1-5 or >5 years) (Wang et al., 2014).

However, increased circulating levels of neutrophil granule proteins in T1D could not be established by a different research group in a follow-up study to Wang et al. (2014). Specifically, Qin et al. (2016) investigated serum levels of PRTN3 and ELANE and could not confirm higher proteins levels in patients with T1D as compared to healthy controls. Quite the contrary, Qin and coworkers reported significantly reduced levels in patients with T1D, especially those within 3 years of diagnosis, consistent with the known reduction in neutrophil counts in recent-onset T1D (Valle et al., 2013). In consequence, the presence or absence of higher circulating levels of neutrophil inflammatory proteins in T1D is still a conundrum since no clear consensus could yet be established in the community. Possibly the best example for this conundrum is a two-phase study carried out by Zhi et al. (2011) who first found significantly higher protein levels of MPO in patients with T1D (3.5-fold) using a 2D LC-MS/MS discovery approach and subsequently reported significantly lower levels of MPO (2-fold) in the validation stage using a Luminex assay. Unexpectedly, I also could not confirm significantly higher levels of neutrophil granule proteins in serum of patients with recent-onset T1D using DIA LC-MS/MS. However, several proteins of interest (e.g. DEFA1, ELANE and RNASE3) were not identified in this validation experiment; moreover the two key signature proteins MPO and PRTN3 were only identified based on a single peptide which strongly aggravated to draw a reliable conclusion with regards to differential abundance levels; albeit the tendency for those two peptides was towards lower expression in T1D. Additionally, serum protein levels for two myeloid-derived inflammatory proteins (S100A9 and CAMP),

which were previously reported with higher expression in cellular data, were found to be significantly lower in the T1D group (Fig. 24B, Tab. 13, 18 and 19). Although proteome coverage for key inflammatory signature proteins was insufficient in the serum profiling experiment, the quantitative results for S100A9 and CAMP were unexpected and partly deprived the degranulation hypothesis (spill-over effect) of its argumentative basis. In consequence, the conflicting observations for circulating levels of these proteins across numerous different studies raise the possibility that neutrophil inflammatory protein abundance could be influenced by additional factors such as the type of sample analyzed, the method of sampling or the assay for protein detection. Specifically, the formation of NETs could be an explanation for this phenomenon.

4.2.2.3 NETosis hypothesis

The increased quantities of neutrophil granule proteins measured in the PBMC preparations by LC-MS/MS may also be explained by the NETosis hypothesis (iii). Specifically, the release of NETs might cause a physical interaction between neutrophils and other circulating blood cells, resulting in co-precipitation into the final sample. With this hypothesis, the underlying assumption is that T1D patient-derived neutrophils have an increased capacity to undergo NETosis, a process in which uncondensed chromatin lined with granule proteins is extruded from the cell as an antimicrobial defense mechanism (Brinkmann et al., 2004). NET formation enables neutrophils to deposit high concentrations of serine proteases in the extracellular space in order to degrade virulence factors (Brinkmann et al., 2004; Pham, 2006). In addition to helping trap pathogens such as bacteria, this extracellular fibrous matrix can also interact with other blood components, including erythrocytes and platelets. Notably, Fuchs et al. (2012) demonstrated that NETs provide an exclusive scaffold for platelet and erythrocyte adhesion, thereby promoting fibrin formation and stimulating (red) thrombus formation. Interestingly, Urban et al. (2009) investigated the qualitative composition of NETs by LC-MS/MS and found primarily neutrophil granule proteins (ELANE, Lactotransferrin (LTF), AZU1, CTSG, MPO, PRTN3, Lysozyme (LYZ), and DEFA1/A3), cytoplasmic proteins (S100A8-9/12) as wells as members of the histone family among the major components. Intriguingly, almost all of the granule proteins which Urban and colleagues identified as integral proteomic constituents of NETs were among the key myeloid-derived signature proteins observed across PBMC profiling experiments in this thesis (Fig. 24A-B, 25, 27, 30B-C, 28, 39A and Tab. 9, 13). The NETosis hypothesis is supported by a recent report highlighting that neutrophils derived from diabetic patients exhibit an increased intrinsic capacity for NET formation (Wong et al., 2015). In consequence, the NETosis hypothesis could not only explain the presence of increased levels of neutrophil granule proteins but

might also serve as an explanation for the marked increase in erythrocyte proteins (e.g., hemoglobin) and platelet-derived proteins in PBMC data (see Fig. 24A, 30C, 38 and Tab. 9, 13). In fact, a commonly accepted marker of NETosis are circulating DNA-MPO complexes which Wang et al. (2014) reported to be elevated in patients with T1D. Accordingly, Stoikou et al. (2017) measured circulating cell-free DNA (nucleosomes) as alternative NETosis marker and found elevated levels in patients with diabetes (T1D, T2D and gestational diabetes) as compared to healthy controls. As previously outlined, several groups reported that neutrophil counts are reduced prior to and during the onset of T1D (Harsunen et al., 2013; Valle et al., 2013). Wang et al. (2014) confirmed this finding and attributed it at least partially to increased NETosis.

A prerequisite for NET formation is the citrullination of arginine residues in histones by the enzyme Protein-arginine deiminase type-4 (PADI4), thereby initiating chromatin decondensation. PADI4 was recently reported to be increased in neutrophils in patients with diabetes (Wong et al., 2015). In general, NETosis is associated with changes in posttranslational modifications such as methylation, acetylation, and citrullination, which may represent an important source of neo-autoantigens, thereby promoting the generation of autoantibodies (Knight et al., 2012). In particular, a growing body of evidence supports a pathogenic role for citrullinated autoantigens in triggering autoimmune responses in a variety of autoimmune disorders. For example, a recent study found that citrullinated glucose-regulated protein 78 (GRP78) is a novel autoantigen in T1D (Dwivedi and Radic, 2014). Indeed, the measured protein levels of PADI4 were distinctly higher in the T1D group in both the DDA and DIA data sets across both cellular compartments, indicating a possible role for excessive citrullination in the pathogenesis of T1D (see also Fig. 24B).

Intriguingly, netting neutrophils have recently been attributed a key role in the initiation process of T1D in young NOD mice (Diana et al., 2013). Specifically, Diana and coworkers highlighted that physiological ß-cell death leads to the recruitment and activation of a cellular triad including innate-like B-1a cells, neutrophils and plasmacytoid dendritic cells (pDCs). B-1a cells and netting neutrophils then co-operate to activate pDCs *via* a TLR9-MyD88-dependent pathway leading to IFN-α production in pancreatic islets (Diana et al., 2013). As a result of the immune cell crosstalk, diabetogenic T cell responses are initiated eventually leading to T1D development (Diana et al., 2013). Interestingly, Allen et al. (2009) found that IFN-α producing pDCs are overrepresented in the periphery of patients close to T1D onset; a finding which likely supports the model of T1D initiation proposed by Diana et al. (2013). The pDC subpopulation is predominantly

an effector of the antiviral immune response mediating the secretion of large amounts of IFN-α and IFN-ß (Reizis et al., 2011). However, pDCs also promote a variety of autoimmune diseases including psoriasis and SLE (Gilliet et al., 2008; Swiecki and Colonna, 2010). With regards to T1D, a type 1 IFN-regulated gene expression signature was reported to associate with disease onset and autoantibody positivity in the periphery (Reynier et al., 2010). Furthermore, transient IFN-driven gene expression signatures were reported in children prior to seroconversion (Ferreira et al., 2014; Kallionpaa et al., 2014). Ferreira and colleagues further demonstrated that the observed IFN signature was temporally linked to a recent history of viral respiratory infection. The observations of IFN-induced gene expression signatures preceding and accompanying T1D onset could potentially be the result of excessive IFN-α production by pDCs in the context of an antiviral response and/or immune cell crosstalk by B1-a cells and netting neutrophils as proposed in the model by Diana et al. (2013). Intriguingly, IFN-α treatment of patients with viral infections or leukemia reportedly associates with increased incidence of T1D highlighting a pathogenic role for IFN-α in T1D development (Fabris et al., 1998; Guerci et al., 1994). While the contribution of NETosis to the pathogenesis of human T1D remains elusive, NET formation has been shown to play an important role in several other autoimmune diseases due to the resulting extracellular deposition of self-molecules (Amulic et al., 2012). Specifically, in SLE autoantibodies are directed against DNA, histones, and neutrophil granule proteins (Lande et al., 2011; Naegele et al., 2012; Villanueva et al., 2011).

With regards to the observed NETosis-associated signature proteins in this thesis, a possible explanation could be a selective enrichment of NET material in the cellular preparations due to the sticky properties — potentially leading to entanglement of PBMCs, platelets and erythrocytes. On the other hand, when proteomic profiling was performed on serum, these structures could have been either depleted due to cellular deposition/entanglement or inaccessible to proteolytic digest. This hypothesis would be supported by the tendency of selected NETosis proteins towards lower expression in the T1D group (e.g. MPO, PRTN3, S100A9 and CAMP, see also Tab. 18 and 19). In fact, the previously outlined proteomic study investigating the proteomic composition of NETs carried out a DNAse digest prior to proteolytic digest — potentially to enable access to NET-associated proteins (Urban et al., 2009). Serum samples which were processed in the presented DIA profiling were not subjected to DNAse digest but only briefly boiled; however temperatures under 100 °C do not lead to DNA degradation (Karni et al., 2013). In contrast, all cellular lysates were not only boiled but also sonicated in order to

fragment all cellular DNA to obtain a homogenous protein lysate thus having potentially enabled proteolytic access to NET structures.

In summary, a growing body of evidence supports a pathogenic role for neutrophil activation, NETosis and the associated circulating inflammatory proteins in the pathogenesis of T1D. In retrospective, the precise origin of the neutrophil-associated inflammatory signature proteins identified in PBMC profiling in this thesis cannot be determined with 100% certainty; however a close connection to altered neutrophil activity seems likely. The concept of altered neutrophil activity is further supported by a recent proteomics study which reported a marked upregulation of several myeloid-derived inflammatory mediators (PRTN3, MPO and LTF) in postmortem pancreatic tissue from donors with T1D and non-diabetic donors with T1D-associated autoantibodies compared to healthy controls (Burch et al., 2015). Additionally, Burch et al. (2015) found RNASE3 and ELANE to be exclusively higher expressed in prediabetic pancreas, highlighting the potential role of these factors at the prediabetic stage. Specifically, Burch and coworkers reported that RNASE3 was the protein which exhibited the most pronounced upregulation (18-fold) of all proteins with significant abundance changes in prediabetic pancreatic tissue. Strikingly, RNASE3 is also one of the key signature proteins observed in the presented PBMC profiling experiments (Fig. 24A-B, 27, 30B, 32, 38C, 39-40 and Tab. 9, 13). In consequence, these findings highlight a potential pathologic involvement of RNASE3 in T1D development, given the potent cytotoxic capacities of this protein, a direct damage to ß-cells cannot be ruled out. Lastly, it cannot be precluded that certain populations of lymphocytes, e.g. CD4+ T cells, are capable of expressing myeloid-typical proteins under certain circumstances. In fact, MPO expression — albeit at low concentrations — has recently been discovered for human peripheral B, CD4+ and CD8+ T lymphocytes (Okada et al., 2016). Specifically, Okada and coworkers found lymphocyte MPO levels to positively associate with viral infections (HIV, HCV). However, this unexpected finding currently relies on the results of a single research study and needs to be further investigated for establishment.

4.2.3 T cell markers and cytotoxic lymphocyte proteins are downregulated at T1D onset

4.2.3.1 Immune-dampening effect

Although the PBMC profiling data establish a clear link between innate inflammatory mediators and the onset of T1D, the precise role of T cells, in particular CD4+ T cells, in this process remains poorly understood. With regards to canonical T cell surface markers, the DIA profiling experiments revealed a mild but significant reduction of protein

levels in the T1D group. Lower protein levels were consistently observed for both CD4+ T cell and CD4-depleted cell preparations regarding the selected canonical T cell marker panel mainly implicated in T cell receptor signaling (Tab. 12 and 15). This dampened T cell phenotype was also revealed when pathway enrichment analysis was performed with significantly downregulated candidate proteins in the CD4-depleted cell fraction revealing an underrepresentation of proteins associated with multiple T cell related signaling pathways as well as cell growth and proliferation-associated pathways in the T1D group (see Tab. 14). For instance, significantly lower protein levels were measured for CD2, CD3D, CD3E, CD247, LCK, ZAP70 and NFATC2 in patients with T1D across both cell fractions (Tab. 12 and 15). Additionally, CD4+ T cell profiling further revealed significantly lower protein levels of CD4 and CD28 in the T1D group (Tab. 12). Notably, statistical significance (q<0.05) was only achieved in the DIA profiling experiments for the majority of these canonical T cell proteins possibly due to higher peptide per protein coverage and thus improved quantification accuracy (see Tab. 10 and Fig. 36E). Yet, tendencies towards lower abundances of canonical T cell proteins were also observed in the corresponding DDA experiments, albeit not achieving statistical significance. One exception is NFATC2 which was measured with significantly lower levels in both DDA and DIA experiments of patient-derived CD4+ T cells from sample set II (Tab. 12). Another example is ZAP70 which was found with significantly lower protein levels in both DDA and DIA profiling approaches of patient-derived CD4-depleted cells from sample set II (Tab. 15). In consequence, NFATC2 and ZAP70 were chosen as target proteins for antibody probing in immunoblots to confirm the observed proteomic signature. As a result, lower protein abundance levels were confirmed for ZAP70 in pooled protein lysates of patient-derived CD4+ T cells and CD4-depleted cells (Fig. 32 and 40). Additionally, western blots revealed a marked reduction of NFACT2 protein levels in CD4-depleted cells and a minimal reduction in CD4+ T cells, respectively (Fig. 32 and 40).

The nominal yet significant downregulation of several canonical T cell markers in the DIA data sets suggest a quantitative and/or functional impairment of the (CD4+) T cell compartment. Although the amount of digested protein was equal across all samples and inter-sample normalization was based on all signals, it cannot be precluded that numeric counts of (CD4+) T cells were different between T1D and control groups, thus contributing to differential sample composition. Specifically, T cell count could have been reduced in the T1D group in both cellular fractions, while myeloid-derived proteins were elevated leading to divergent protein composition in the sample. In fact, reduced numbers of circulating lymphocytes preceding and during the onset of T1D have been

reported in the periphery, indicating a perturbation in leukocyte homeostasis which could partially explain the observed signature (Harsunen et al., 2013). A general reduction in peripheral lymphocyte counts could be a consequence of impaired hematopoiesis or maturation, increased peripheral cell death or damage, and tissue infiltration. Indeed, lymphopenia is a common feature of autoimmune diseases which although widely studied remains poorly understood (Khoruts and Fraser, 2005; Merayo-Chalico et al., 2016). The proposed causes for autoimmune lymphopenia range from increased apoptosis to reduced thymic output, to viral infections (Sheu et al., 2014). Interestingly, a general immune dampening effect characterized by low spontaneous cytokine secretion was reported for PBMCs isolated from patients with newly diagnosed T1D compared to healthy controls (Ryden et al., 2014). The authors speculated that this apparently suppressed phenotype could be a consequence of immune exhaustion following a period of strong immune activation preceding the onset of T1D (Ryden et al., 2014). This idea is closely connected to the perception that the bigger part of ß-cells is destroyed once overt metabolic disease sets on, suggesting an attenuation of autoimmune reactivity. Accordingly, Orban et al. (2007) examined CD4+ T cells from patients with newly diagnosed T1D and found a reduction in the mRNA expression of CD4+ T cell-specific genes, including genes that affect key immune functions (e.g. CD2 and CD3) and the cell cycle (e.g. LCK), thus suggesting reduced CD4+ T helper cell function. The authors attributed this phenotype at least partially to the reduced expression of histone deacetylase (HDAC1) in patients with T1D and the subsequent reduction in epigenetic modification. Although the presented CD4+ T cell profiling data largely support this postulated "hypo-responsive" phenotype, a direct involvement of HDAC1 or epigenetic mechanisms in particular cannot be confirmed. Specifically, marginally reduced abundance levels of HDAC1 were measured in the T1D group in DIA data, however not reaching statistical significance (fold change 0.95, q-value 0.064). A reduced T helper cell function may further be supported by the mild yet significant downregulation of several proteins functioning in nucleocytoplasmic trafficking of proteins and RNAs such as exportins (XPO-1, XPO-2, and XPO-7) or nucleoporins (NUP85, NUP93, NUP160, NUP205) (Fig. 31). Notably, significant lower levels were also observed for the majority of nucleocytoplasmic transport proteins in the CD4-depleted cell compartment indicating a global effect. For example, XPO-1 (also known as CRM1) serves as a gatekeeper for cell survival and ribosomal biogenesis (Ishizawa et al., 2015). Interestingly, a marked reduction of the expression of nucleocytoplasmic transport genes such as importins, exportins, karyopherins and Ran GTPase cycle-related factors has been identified as distinct characteristic of cellular senescence (Kim et al., 2010). Specifically, impaired nucleocytoplasmic trafficking of signaling molecules in senescent cells ultimately renders

them hypo-responsive to external stimuli such as growth factors or apoptotic stress (Kim et al., 2012). With respect to CD4+ T cells, protein levels of the transcriptional repressor NF-X1 (NFX1) were significantly upregulated in patients with T1D (DIA data). NFX1 has been implicated as a repressor of human telomerase thus potentially contributing to cellular senescence (Gewin et al., 2004; Xu et al., 2008). A senescence-like cell phenotype could potentially reflect a higher proportion of long-lived memory T cells in patient-derived PBMC samples during T1D onset. Intriguingly, a cardinal feature of patients with T1D is the presence of autoreactive, antigen-experienced T cells with a memory phenotype (Ehlers and Rigby, 2015; Monti et al., 2007; Oling et al., 2012; Skowera et al., 2015). Notably, the process of premature CD4+ T cell aging characterized by a reduced naïve/memory ratio has recently been implicated in the development of T1D in the NOD mouse model (Sheu et al., 2014).

4.2.3.2 Reduced expression of cytotoxic lymphocyte proteins

The second major proteomic pattern which was consistently observed with statistical significance throughout all CD4-depleted cell profiling experiments were the markedly reduced levels of cytotoxic lymphocyte proteins in the T1D group. Cytotoxic lymphocytes encompass both the NK cell and the CD8+ T cell (CTL) populations which induce target cell death through cytotoxic effector activity to remove abnormal or malignant cells, and to eliminate virus-infected cells (Chavez-Galan et al., 2009). When cytotoxic lymphocytes conjugate to a target cell, secretory granules traffic to the immunological synapse at the cell surface and release cytotoxic effector cargo including PRF1, granzymes and GNLY (Voskoboinik et al., 2015). Subsequently, PRF1 forms large transmembrane pores enabling granzymes and GNLY to access the cytosol of the target cell (Voskoboinik et al., 2015). GNLY is tumoricidal and broadly antimicrobial exhibiting potent cytolytic activity by compromising membrane integrity or activation of caspase 3 signaling (Kaspar et al., 2001; Krensky and Clayberger, 2009; Stenger et al., 1998). On the other hand, granzymes are a distinct family of serine proteases which upon translocation into the target cell cleave intracellular proteins eventually resulting in apoptosis (Bots and Medema, 2006).

The core of this signature are four proteins which were measured with significantly (q<0.05) lower protein levels across both sample sets and data acquisition methods: GZMA, GZMH, GNLY and ZAP70 (Fig. 35, Fig. 39B). The signature is extended by additional cytotoxic effector mediators including GZMM, GZMK and PRF1 which were found to be reduced at T1D onset (Tab. 13, Fig. 39B). This finding is further supported by pathway enrichment analysis performed with significantly downregulated proteins in

Discussion

CD4-depleted cell data (DIA) revealing an underrepresentation of granzyme signaling, cytolysis effectors, and the two cell types NK cells and CTLs in the T1D group (see Tab. 14). Notably, enrichment scores were observed with lowest p-values for NK cells although the majority of implicated proteins are expressed by both cell types (see also Fig. 38B, 39B) (Kim et al., 2014). For instance, ZAP70 is not only an essential kinase in T cell signaling but is also expressed by NK cells (Long et al., 2013; Wang et al., 2010). The cytotoxic effector molecules GZMA and PRF1 were chosen as target proteins for antibody probing in pooled CD4-depleted cell lysates. Using immunoblotting, distinctly lower protein levels of PRF1 and GZMA could be confirmed in the T1D group; this effect was particularly prominent for GZMA (Fig. 40). In line with the observations presented in this thesis, significantly depressed gene expression levels have also been communicated for three cytotoxic lymphocyte molecules (GZMB, PRF1, FasL) in patients with longstanding T1D compared to healthy controls (Han et al., 2011). Additionally, GNLY was also among the key signature proteins identified with lower expression levels in patients with T1D in the large-scale PBMC expression profiling reported by Jin et al. (2013).

The presence of the outlined cytotoxic lymphocyte signature could once again be partially explained by the overall reduction of lymphocyte counts present in the period of disease onset (Harsunen et al., 2013). Secondly, a decreased frequency of NK cells and CTLs in peripheral blood of patients with T1D may simply reflect the trafficking of these cells to the affected tissue where they possibly contribute to autoimmune destruction of ß-cells. Diabetogenic CD8+ CTLs have been proposed as key effector cells in mediating ß-cell death *via* the perforin/granzyme pathway or Fas/FasL interaction (Kreuwel et al., 1999; Thomas et al., 2010). A direct involvement of NK cells at the ß-cell level is supported by the fact that infiltrating NK cells have been detected in pancreatic islets of T1D patients (Dotta et al., 2007). On the other hand, the observed signature could also reflect the resolution of a period of autoimmune activation by peripheral T cell death or transitioning into memory cell phenotype associated with an attenuation of T cell effector activity as previously outlined in the immune dampening effect (see also 4.2.3.1) (Masopust and Schenkel, 2013; Wherry and Kurachi, 2015).

With regards to NK cells in particular, there are several studies which have reported a numeric deficiency or functional abnormality in peripheral blood of patients with T1D (Herold et al., 1984; Lorini et al., 1994; Negishi et al., 1986). Importantly, Rodacki et al. (2007) informed about a numeric reduction in peripheral NK cells marking the onset period of T1D. Rodacki and colleagues further observed a mild reduction in the

expression of killer activating receptor NKG2D in all patients with T1D, irrespective of disease duration, indicating reduced NK cell activity. Additionally, a defect in function and signaling of NKG2D has also been reported in children suffering from T1D (Qin et al., 2011). Finally, diminished peripheral NK cell counts have further been reported in latent autoimmune diabetes in adults (LADA) patients (Akesson et al., 2010). These findings highlight that T1D is characterized by numeric and functional impairment of the NK cell compartment which could have potentially manifested in the observed proteomic signature. Although autoimmune disorders are primarily mediated by T cells and B cells, NK cells have been implicated in the induction and/or persistence of inappropriate adaptive immune responses (Fogel et al., 2013). In many instances of autoimmune diseases such as SLE or rheumatoid arthritis, a reduction in the number of NK cells along with decreased cytotoxic function has been observed (Aramaki et al., 2009; Mandal and Viswanathan, 2015; Yabuhara et al., 1996). Notably, there is a growing body of evidence indicating that NK cells can also prevent and limit adaptive (auto-) immune responses via killing of autologous myeloid and lymphoid cells (Lünemann et al., 2009). If this unique role in the maintenance of cell homeostasis is impaired a persistent activation of T cells and macrophages might be the consequence, thus, the decreased expression of cytotoxic lymphocyte proteins could have contributed to the inability to maintain normal levels of peripheral tolerance which is essential for protection from autoimmune disease.

GZMA, one of the key signature proteins observed in this thesis, is particularly interesting in the context of T1D, likely harboring biomarker potential. In accordance with the results presented in this thesis, transcript and protein expression levels of GZMA were recently reported to be markedly downregulated in children with T1D compared to healthy controls (Hamel et al., 2016). Strikingly, a recent study in the NOD mouse highlighted that GZMA deficiency broke immune tolerance and promoted the development of T1D in a type I IFN-dependent manner (Mollah et al., 2017). Specifically, GZMA deficiency resulted in an elevated incidence of T1D associated with an accumulation of ssDNA in immune cells and the induction of an IFN response in pancreatic islets (Mollah et al., 2017). The authors postulated that ssDNA acts as an endogenous signal which promotes local type I IFN production. In consequence, Mollah and coworkers suggest that GZMA has a non-cytolytic role in limiting autoimmunity by functioning in sensing and degradation of intracellular ssDNA. Mollah and colleagues identified double positivity for ssDNA and GZMA expression only in cells of the innate immune system, namely NK cells and dendritic cells. The authors further propose that a defect in the removal of aberrant cytosolic nucleic acids attributed at least in part to

reduced GZMA expression could be implicated in the development of T1D in humans (Mollah et al., 2017). This finding highlights an important role for GZMA in maintaining immune tolerance and the GZMA-associated signature observed in this study could likely reflect the break of just this delicate balance.

4.3 Proteomic phenotyping of serum samples in recent-onset T1D

The human plasma/serum proteome likely harbors several thousand proteins, thus representing a reservoir rich in biological and diagnostic information (Schwenk et al., 2017). However, analyzing the human/plasma proteome in-depth represents an enormous analytical challenge due to the vast dynamic range of protein abundance. Although several research groups have approached protein biomarker discovery in T1D by serum profiling using shotgun or targeted proteomics, the data presented in this thesis reflect the first approach exploring DIA LC-MS/MS. Due to the difficult nature of serum as a sample per se, technical reproducibility of the label-free quantitative LC-MS/MS set-up was tightly controlled during the 100 sample LC-MS sequence. Using a QC pool which was measured 5 times across the sequence in regular intervals, an average CV of 19% (median 10%) could be achieved applying the Spectronaut q-value 25% percentile setting (Fig. 43). In contrast, if the more flexible match-between-runs function q-value sparse was enabled, the average CV value would have dropped to 39% (median 21%) which urged me to choose the more stringent 25% percentile setting. Using this approach, overall serum protein identifications from 50 patients with new-onset T1D and 50 controls accumulated to 354 proteins (225, ≥ 2 unique peptides) using a tailored spectral library encompassing ~700 protein entities (50% recovery). With regards to T1D and control groups, the measured average (median) CV values accounted for 51% (36%) and 57% (39%), respectively (Fig. 43). This finding is consistent with a recent plasma proteomics study that found substantial variability in protein abundance among a large cohort of 116 sets of adult twins (Liu et al., 2015). Similarly, Geyer et al. (2016) measured 345 plasma proteins in ten individuals and found that only 19% of the protein abundances had a CV value <20%.

Despite the large extent of inter-individual variation present in human samples, principal component and hierarchical cluster analyses of the global serum proteome resulted in patient and control group-specific clustering (Fig. 44). Unlike PBMC data where no clustering effect was observed, this finding suggests that the serum proteome of patients is affected by uniform abundance changes during the onset period of T1D (see also Fig. 29 and 37). This hypothesis is strongly supported by the results obtained from PBMC profiling highlighting that the greater part of differentially regulated proteins between T1D

and control groups was closely connected to the extracellular space and serum/plasma specimen in particular (Tab. 11 and 14). Although there is no PCA data on the global serum or plasma proteome in previous T1D-related studies, a recent proteomic phenotyping of pancreatic tissue has also reported a clear separation of T1D and control groups based on the entire pancreatic proteome (Liu et al., 2016). However, one major limitation of the serum sample set is that patients and controls were not matched for age (Tab. 2). Interestingly, a recent study investigated the quantitative age-specific variability of the plasma proteome in healthy neonates, toddlers, young children and adults (Bjelosevic et al., 2017). Hierarchical clustering revealed well separated clusters corresponding to the experimental age groups on a proteome-wide scale, additionally >100 proteins exhibited significant differential expression levels across the age groups (Bjelosevic et al., 2017). Therefore, the group-specific clustering observed in the presented serum data may not entirely reflect proteomic differences attributed to T1D onset but could likely also reflect an age-specific effect. In consequence, analysis of differential protein abundance between T1D and control groups was performed with ANCOVA correcting for age as covariate resulting in 111 proteins achieving statistical significance (q<0.05). Notably, observed fold changes were very mild and only 15 candidate proteins exhibited at least 1.5-fold regulation in either direction (Fig. 46). Besides the distinct myeloid-derived inflammatory signature, PBMC data also revealed elevated protein abundance levels of complement, platelet and apolipoproteins which established the overall signature reflecting an altered extracellular milieu (Tab. 11 and 14). Indeed, the serum profiling experiment largely confirmed increased abundance levels of several complement proteins (CF, C4BPA, C8A, CFB), protease inhibitors (A2M, ITIH2/H3, SEPING1/A3), apolipoproteins (APOC1/C3, APOE, CLU) and hemostasis-associated proteins (FGA, VWF) previously identified in cellular data (Fig. 45, Tab. 18 and 19).

Increased complement activation in patient plasma suggests that T1D onset reflects an elevated state of systemic innate inflammation. Indeed, already three decades ago Sundsmo et al. (1985) informed about classical and alternative complement activation in new-onset T1D. A pathologic role for complement activation underlying T1D development has further been suggested from genetic association studies and the presence of complement-fixing autoantibodies in T1D sera (Bottazzo et al., 1980; Bottazzo et al., 1985; Hagglof et al., 1986; Mustonen et al., 1984). In addition, complement activation seems to be closely linked to the vascular complications of diabetes (Acosta et al., 2000; Hansen et al., 2004). Intriguingly, increased complement

Discussion

activation T1D has also been established by immunohistochemistry using organ donor pancreatic tissue (Rowe et al., 2013).

Currently, there are four major MS-based serum/plasma profiling studies investigating patients with T1D and two studies profiling prediabetic, seropositive subjects (Albrethsen et al., 2009; Metz et al., 2008; Moulder et al., 2015; von Toerne et al., 2017; Zhang et al., 2013a; Zhi et al., 2011). A recent review summarizing the efforts of these serum profiling approaches highlighted that the overlap of presented biomarker candidate panels is however marginal (Moulder et al., 2017). Qualitative and quantitative differences in the identified serum/plasma proteome likely stem from biological differences (inter-individual variation, ethnicity, disease stage) and methodical aspects including large differences in underlying sample preparation (e.g. protein depletion), MS method and platform as wells as data processing and post-analysis steps (e.g. statistics). Additionally, the previously outlined prominent effect of age on the plasma proteome might also have influenced the comparability of the different studies. Nevertheless, several of the proteins observed with significant differential protein abundance levels between T1D and control groups in the present DIA profiling had been identified in previous studies highlighting their consistency and underlining their biomarker potential. This qualitative overlap of recurrent candidate proteins is visualized in Tab. 20; importantly, only proteins are listed which achieved statistical significance and harmonizing fold changes in both this study and the comparative study.

Table 20 Overlap of protein biomarker candidates identified in the DIA serum profiling and current literature on MS-based serum/plasma proteomics in the context of (pre-) T1D.
The table lists only candidate proteins identified in the present DIA profiling and the current literature. In order to be listed the candidate biomarker needed to exhibit statistical significance and non-conflicting fold changes in both this study and the respective comparative study. The panel size refers to the number of key candidate proteins reported by the individual studies; the number depends largely on the applied statistics/criteria. The term higher/lower levels refer to protein abundance in the (pre-) T1D group. y, years; m, months; aab, autoantibody; -, either no proteins reported or no overlap between studies.

	Metz (2008)	Albrethsen (2009)	Zhi (2011)	Zhang (2013)	Moulder (2015)	v. Törne (2017)
study features	new-onset age <30 y	longitudinal post-onset 1, 6, 12 m ø age 9 y no controls	T1D patients duration not specified all age groups (0-90 y)	new-onset age <30 y	longitudinal prediabetic progressors pre/post seroconversion	prediabetic aab+ vs. aab-
sampling	worldwide	18 centers in 15 countries	central	worldwide	central	central
sample size	20	766	60+1987*	150	266	45[#]

	Metz (2008)	Albrethsen (2009)	Zhi (2011)	Zhang (2013)	Moulder (2015)	v. Törne (2017)
higher levels	APOA4 C1R LRG1 SERPINA6	APOC1 APOC3	ADIPOQ APOA4	SERPING1	-	BTD CP C8B ITIH2
lower levels	TF	-	-	AZGP1 GSN PGLYRP2 TTR	PFN1 after seroconversion	TTR
panel size	5 (30*)	2	21 (6*)	24	not specified	26

* Metz et al. (2008) reported a panel of 30 proteins at p<0.05 based on protein-level statistics and further reduced this panel to 5 proteins with additional peptide-based statistics and exclusion of one-peptide-hits. Zhi et al. (2011) had a sample size of 60 in the LC-MS/MS discovery approach and subsequently performed antibody-based validation (Luminex/ELISA) in a large-scale study including 1987 probands. The authors presented a panel of 21 candidate proteins (p<0.05, 1.5-fold cut-off) in the discovery stage from which 6 proteins were selected for validation.
\# Comparison with von Toerne et al. (2017) was based on the application phase of their two-stage study.

For instance, lower levels of transthyretin (TTR, t̲ransports t̲hyroxine and r̲etinol) were measured in patients with T1D in both the present DIA profiling (Fig. 45, 46B), by Zhi et al. (2011) and furthermore in prediabetic subjects as described by von Toerne et al. (2017) (Tab. 20). Indeed, TTR is synthesized within pancreatic islets and has a functional role in ß-cells promoting insulin release and protecting against ß-cell death (Jacobsson et al., 1990; Refai et al., 2005). According to the human proteome map, TTR expression is much stronger in pancreatic tissue than in liver implying that the source of the serum signature could likely stem from altered pancreatic TTR dynamics (Kim et al., 2014). Interestingly, Itoh et al. (1992) also observed lower TTR levels in sera of patients with newly-diagnosed T1D and suggested, based on subsequent immunohistochemistry, impaired TRR synthesis or storage in pancreatic ß-cells — but not α-cells — to underlie T1D development. Interestingly, among the top candidate proteins in the present DIA profiling was also retinol-binding protein 4 (RBP4) which exhibited markedly lower protein levels in T1D sera (Fig. 46B). RBP4 is the carrier of vitamin A (retinol) in human plasma, in which it circulates in complex with TTR. In the context of T1D, it was suggested that reduced levels of TTR lead to reduced TTR-RBP4 complex formation and thus increased glomerular filtering of RBP4 lowering its respective plasma levels (Pullakhandam et al., 2012). In fact, a lowering of both TTR and RBP4 serum levels in T1D has been reported in multiple occasions strengthening the consistency of this finding (Basu et al., 1989; Forga et al., 2016; Pullakhandam et al., 2012). In summary, the protective role of TTR in ß-cell apoptosis, the presence of lower TTR levels in the

prediabetic stage and the circumstance that TTR and RBP4 levels remain high in T2D underline a potential role for the TTR-RBP4 axis in T1D pathogenesis (Pullakhandam et al., 2012; Refai et al., 2005; von Toerne et al., 2017).

Additionally, members of the apolipoprotein family (e.g. APOC1/C3) were repeatedly among the significantly regulated proteins in both the PBMC data and the serum profiling exhibiting on average higher expression levels in the T1D group (Tab. 18 and 19). Accordingly, Albrethsen et al. (2009) reported that levels of APOC1 and APOC3 increase progressively in longitudinal samples of patients with T1D collected one, six and twelve months after diagnosis. Importantly, Juntti-Berggren et al. (2004) also found increased levels of APOC3 in T1D sera and further proved that APOC3 promotes Ca^{2+}-dependent ß-cell death in T1D. In the current DIA approach, the most prominently upregulated apolipoprotein in the T1D group was APOA4 (q<0.001, 1.5-fold, Fig. 46B) which was also found with higher abundance levels in studies from Metz et al. (2008) and Zhi et al. (2011) (Tab. 20). In contrast, lower apolipoprotein levels were reported for the prediabetic phase by Moulder et al. (2015) (APOC2/C4) and von Toerne et al. (2017) (APOA4/E/C4) indicating that apolipoprotein dynamics are very complex in T1D development and need further investigation to clarify their precise role. Interestingly, apolipoproteins have been attributed a broad functional role in viral infections (Grassi et al., 2016; Singh et al., 1999). For example, higher and lower serum levels of APOC3 have been associated with cleared versus active HCV infection (Rowell et al., 2012). In consequence, complex alterations in apolipoprotein metabolism could likely support the concept of a viral etiology in T1D.

Besides the qualitative overlap in candidate proteins with current literature (Tab.20), the DIA profiling experiment further revealed novel circulating signature proteins which could further be evaluated for their biomarker potential (see also Fig. 46). Here, one very promising candidate is hepatocyte growth factor-like protein (MST1, aka macrophage stimulating 1) which was the protein exhibiting the most distinct upregulation in the T1D group from the entire data set (median 1.8-fold, q<0.001, Fig. 46A). Interestingly, the protein was among the very low abundant proteins measured in serum with corresponding abundance ranks of 291 and 311 for T1D and control groups, respectively, and has to my knowledge not been reported as candidate protein in other T1D-related serum profiling studies (see Tab. 20). Early studies provided evidence that MST1 is a serum protein mainly secreted by hepatocytes which stimulates chemotaxis and phagocytosis in macrophages (Bezerra et al., 1993; Leonard and Skeel, 1976; Skeel et al., 1991). More recent studies have showed that MST1 is also a ubiquitously

expressed serine/threonine kinase in the Hippo signaling pathway implicated in numerous cellular functions such as morphogenesis, cell growth, stress response and apoptosis (Avruch et al., 2012; Ling et al., 2008). Specifically, MST1 induces stress kinase JNK and caspase-3 activation, thereby functioning in amplifying apoptotic signaling to promote cell death (Avruch et al., 2012). Strikingly, MST1 has recently been identified as critical regulator of ß-cell apoptosis and dysfunction and further been proposed as a novel therapeutic target for restoration of functional ß-cell mass in diabetes (Ardestani and Maedler, 2016; Ardestani et al., 2014). Specifically, Ardestani et al. (2014) demonstrated that MST1 was strongly activated in human and rodent ß-cells under diabetogenic conditions, positively correlating with ß-cell apoptosis. In turn, MST1 deficiency allowed the restoration of normoglycemia, ß-cell function and survival (Ardestani et al., 2014). The close connection of MST1 overexpression and impaired ß-cell function and survival confers the possibility that elevated levels of MST1 in T1D sera could likely stem from aberrant MST1 dynamics in the pancreatic islets. For instance, increased circulating MST1 levels could stem from ß-cell death or leakage or increased secretion.

Beyond MST1, the panel of potential biomarker candidates shown in Fig. 46 contains other interesting proteins such as the human resistin receptor CAP1 which mediates inflammatory action in monocytes (Lee et al., 2014) or TLN1. Interestingly, protein levels of TLN1 were found to be significantly higher in the T1D group in both CD4+ T cells and in serum (Tab. 18). Notably, a very recent study reported elevated circulatory levels of TLN1 in patients with multiple sclerosis in association with disease activity (Muto et al., 2017). With regards to CD4+ T cells, TLN1-dependent integrin activation is selectively required for T follicular helper cell differentiation (Meli et al., 2016). Accordingly, T follicular helper cells have recently been implicated in the pathology of islet autoimmunity and the frequency of this cell subset increases in the circulation of patients close to diagnosis, further persisting in long-standing T1D (Ferreira et al., 2015; Serr et al., 2016; Viisanen et al., 2017).

5 Perspectives

Taken together, the present investigation has established a precise proteomic landscape of recent-onset T1D in the periphery, including the phenotyping of CD4+ T cells, CD4-depleted cells, and serum. The application of next-generation MS platforms and novel data-independent strategies in addition to conventional shotgun proteomics enabled a progressive augmentation of analytical depth necessary for extraction of differential protein abundance in quantity-limited samples. The obtained proteomic profiles

represent a valuable repository of candidate proteins potentially relevant for T1D pathogenesis which could provide a basis for future investigation of PBMC function or biomarker discovery. The proteomic fingerprint of recent-onset T1D provides compelling evidence for distinct alterations in innate and adaptive immune function.

The primary insight from PBMC data is the predominant presence of innate inflammatory mediators during T1D onset characterized by myeloid-derived proteins and complement activation. The precise origin of this signature is unknown and can only be speculated; the prime candidate being netting neutrophils from the extracellular milieu. However, several studies investigating (circulating) NETosis-associated biomarkers in the context of T1D have produced conflicting results. Therefore, detailed follow-up studies quantifying NETs and investigating functional aspects of patient-derived neutrophils during distinct disease stages would be necessary to clarify these issues. Nevertheless, the presence of this precise signature supports a role for altered innate immunity accompanying T1D onset, either as a bystander effect, a consequence of aberrant adaptive immunity or as a potential pathogenic cause.

The central dogma portraying T1D as a T cell-mediated autoimmune disease has promoted detailed investigation of adaptive immunity but resulted in broad underinvestigation of innate immunity — a phenomenon known as "streetlight effect" (Battaglia and Atkinson, 2015). Notably, the potential pathologic role of neutrophils in disease formation has recently been highlighted as one of many factors which are still underinvestigated although there is a growing body of evidence supporting this concept (Battaglia, 2014; Battaglia and Atkinson, 2015; Diana et al., 2013; Valle et al., 2013). The perception that ß-cell destruction is primarily or exclusively carried by adaptive immunity has been challenged by several investigators (Cabrera et al., 2016b; Itoh and Ridgway, 2017; Kolb and von Herrath, 2017; Skog et al., 2013). For instance, Kolb and von Herrath (2017) argue to the extent that the disease process underlying T1D development is carried by innate immune reactivity, and that regulatory T cell development and ß-cell proliferation will be supported if a non-polarized functional state of innate immunity will be achieved. Future studies will reveal whether moving outside of the "streetlight" will yield valuable, new insights into the pathogenesis of T1D.

Besides innate inflammation, the second central discovery from the cellular profiles was the marked reduction of cytotoxic lymphocyte protein levels at T1D onset. This finding is also interesting and requires precise investigation to establish its cellular origin and underlying cause. In this context, the biomarker potential of GZMA should be

emphasized because it was recently implicated as an important factor in promoting immune tolerance (Mollah et al., 2017). The observed depressed protein levels of GZMA at T1D onset could likely reflect an underlying pathogenic process.

Finally, serum profiling has complemented the proteomic profiles of PBMCs with their extracellular counterpart confirming biomarker candidates proposed in current literature and introducing novel signatures. With respect to the clinical utility of circulating biomarkers, it is however unlikely that a single serum protein will be able to reflect the complex molecular networks and dynamics involved in a heterogeneous disease like T1D. Nevertheless, a promising strategy is to identify and monitor pancreatic islet products which mirror ongoing ß-cell damage or death. The identification of elevated levels of circulating MST1 at T1D onset could be a promising starting point for further evaluation, given the critical role of MST1 in regulating ß-cell dysfunction and death (Ardestani and Maedler, 2016).

6 References

American Diabetes Association. (2003). Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care *26 Suppl 1*, S5-20.

Abbas, A.K., Lichtman, A.H.H., and Pillai, S. (2014). Cellular and Molecular Immunology E-Book (Elsevier Health Sciences).

Achenbach, P., Hummel, M., Thumer, L., Boerschmann, H., Hofelmann, D., and Ziegler, A.G. (2013). Characteristics of rapid vs slow progression to type 1 diabetes in multiple islet autoantibody-positive children. Diabetologia *56*, 1615-1622.

Acosta, J., Hettinga, J., Fluckiger, R., Krumrei, N., Goldfine, A., Angarita, L., and Halperin, J. (2000). Molecular basis for a link between complement and the vascular complications of diabetes. Proceedings of the National Academy of Sciences of the United States of America *97*, 5450-5455.

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature *422*, 198-207.

Ahsan, N., Belmont, J., Chen, Z., Clifton, J.G., and Salomon, A.R. (2017). Highly reproducible improved label-free quantitative analysis of cellular phosphoproteome by optimization of LC-MS/MS gradient and analytical column construction. Journal of proteomics *165*, 69-74.

Akesson, C., Uvebrant, K., Oderup, C., Lynch, K., Harris, R.A., Lernmark, A., Agardh, C.D., and Cilio, C.M. (2010). Altered natural killer (NK) cell frequency and phenotype in latent autoimmune diabetes in adults (LADA) prior to insulin deficiency. Clinical and experimental immunology *161*, 48-56.

Akirav, E.M., Lebastchi, J., Galvan, E.M., Henegariu, O., Akirav, M., Ablamunits, V., Lizardi, P.M., and Herold, K.C. (2011). Detection of beta cell death in diabetes using differentially methylated circulating DNA. Proceedings of the National Academy of Sciences of the United States of America *108*, 19018-19023.

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

References

Allen, J.S., Pang, K., Skowera, A., Ellis, R., Rackham, C., Lozanoska-Ochser, B., Tree, T., Leslie, R.D., Tremble, J.M., Dayan, C.M.*, et al.* (2009). Plasmacytoid dendritic cells are proportionally expanded at diagnosis of type 1 diabetes and enhance islet autoantigen presentation to T-cells through immune complex capture. Diabetes *58*, 138-145.

Altelaar, A.F.M., Munoz, J., and Heck, A.J.R. (2013). Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet *14*, 35-48.

Amulic, B., Cazalet, C., Hayes, G.L., Metzler, K.D., and Zychlinsky, A. (2012). Neutrophil function: from mechanisms to disease. Annual review of immunology *30*, 459-489.

Anderson, N.L., and Anderson, N.G. (2002). The human plasma proteome: history, character, and diagnostic prospects. Molecular & cellular proteomics : MCP *1*, 845-867.

Anker, S.D., Egerer, K.R., Volk, H.D., Kox, W.J., Poole-Wilson, P.A., and Coats, A.J. (1997). Elevated soluble CD14 receptors and altered cytokines in chronic heart failure. Am J Cardiol *79*, 1426-1430.

Aramaki, T., Ida, H., Izumi, Y., Fujikawa, K., Huang, M., Arima, K., Tamai, M., Kamachi, M., Nakamura, H., Kawakami, A.*, et al.* (2009). A significantly impaired natural killer cell activity due to a low activity on a per-cell basis in rheumatoid arthritis. Modern rheumatology / the Japan Rheumatism Association *19*, 245-252.

Ardestani, A., and Maedler, K. (2016). MST1: a promising therapeutic target to restore functional beta cell mass in diabetes. Diabetologia *59*, 1843-1849.

Ardestani, A., Paroni, F., Azizi, Z., Kaur, S., Khobragade, V., Yuan, T., Frogne, T., Tao, W., Oberholzer, J., Pattou, F.*, et al.* (2014). MST1 is a novel regulator of apoptosis in pancreatic beta-cells. Nature medicine *20*, 385-397.

Arif, S., Tree, T.I., Astill, T.P., Tremble, J.M., Bishop, A.J., Dayan, C.M., Roep, B.O., and Peakman, M. (2004). Autoreactive T cell responses show proinflammatory polarization in diabetes but a regulatory phenotype in health. The Journal of clinical investigation *113*, 451-463.

Arpaia, N., Campbell, C., Fan, X., Dikiy, S., van der Veeken, J., deRoos, P., Liu, H., Cross, J.R., Pfeffer, K., Coffer, P.J.*, et al.* (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. Nature *504*, 451-455.

Atkinson, M.A., and Eisenbarth, G.S. (2001). Type 1 diabetes: new perspectives on disease pathogenesis and treatment. Lancet *358*, 221-229.

Atkinson, M.A., Eisenbarth, G.S., and Michels, A.W. (2014). Type 1 diabetes. Lancet *383*, 69-82.

Avruch, J., Zhou, D., Fitamant, J., Bardeesy, N., Mou, F., and Barrufet, L.R. (2012). Protein kinases of the Hippo pathway: regulation and substrates. Seminars in cell & developmental biology *23*, 770-784.

Bach, J.-F., and Chatenoud, L. (2012). The Hygiene Hypothesis: An Explanation for the Increased Frequency of Insulin-Dependent Diabetes. Cold Spring Harbor Perspectives in Medicine *2*, a007799.

Baekkeskov, S., Aanstoot, H.J., Christgau, S., Reetz, A., Solimena, M., Cascalho, M., Folli, F., Richter-Olesen, H., and De Camilli, P. (1990). Identification of the 64K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. Nature *347*, 151-156.

Bainton, D.F., Ullyot, J.L., and Farquhar, M.G. (1971). The development of neutrophilic polymorphonuclear leukocytes in human bone marrow. The Journal of experimental medicine *134*, 907-934.

Bantscheff, M., Lemeer, S., Savitski, M.M., and Kuster, B. (2012). Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Analytical and bioanalytical chemistry *404*, 939-965.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. Analytical and bioanalytical chemistry *389*, 1017-1031.

Barnett, A.H., Eff, C., Leslie, R.D., and Pyke, D.A. (1981). Diabetes in identical twins. A study of 200 pairs. Diabetologia *20*, 87-93.

Barrett, J.C., Clayton, D.G., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H.A., Julier, C., Morahan, G., Nerup, J., Nierras, C.*, et al.* (2009). Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nature genetics *41*, 703-707.

Bas, S., Gauthier, B.R., Spenato, U., Stingelin, S., and Gabay, C. (2004). CD14 is an acute-phase protein. Journal of immunology *172*, 4470-4479.

Basu, T.K., Tze, W.J., and Leichter, J. (1989). Serum vitamin A and retinol-binding protein in patients with insulin-dependent diabetes mellitus. The American journal of clinical nutrition *50*, 329-331.

Battaglia, M. (2014). Neutrophils and type 1 autoimmune diabetes. Current opinion in hematology *21*, 8-15.

Battaglia, M., and Atkinson, M.A. (2015). The Streetlight Effect in Type 1 Diabetes. Diabetes *64*, 1081-1090.

Bazil, V., and Strominger, J.L. (1991). Shedding as a mechanism of down-modulation of CD14 on stimulated human monocytes. Journal of immunology *147*, 1567-1574.

Bendall, S.C., Simonds, E.F., Qiu, P., Amir el, A.D., Krutzik, P.O., Finck, R., Bruggner, R.V., Melamed, R., Trejo, A., Ornatsky, O.I.*, et al.* (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science (New York, NY) *332*, 687-696.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) *57*, 289-300.

Bennett, L., Palucka, A.K., Arce, E., Cantrell, V., Borvak, J., Banchereau, J., and Pascual, V. (2003). Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. The Journal of experimental medicine *197*, 711-723.

Bennett, S.T., Wilson, A.J., Esposito, L., Bouzekri, N., Undlien, D.E., Cucca, F., Nistico, L., Buzzetti, R., Bosi, E., Pociot, F.*, et al.* (1997). Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. The IMDIAB Group. Nature genetics *17*, 350-352.

Bergholdt, R., Brorsson, C., Palleja, A., Berchtold, L.A., Floyel, T., Bang-Berthelsen, C.H., Frederiksen, K.S., Jensen, L.J., Storling, J., and Pociot, F. (2012). Identification of novel type 1 diabetes candidate genes by integrating genome-wide association data, protein-protein interactions, and human pancreatic islet gene expression. Diabetes *61*, 954-962.

Bern, M., Kil, Y.J., and Becker, C. (2012). Byonic: Advanced Peptide and Protein Identification Software. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al] *CHAPTER*, Unit13.20-Unit13.20.

Bevan, M.J. (2004). Helping the CD8(+) T-cell response. Nature reviews Immunology *4*, 595-602.

References

Beyan, H., Drexhage, R.C., van der Heul Nieuwenhuijsen, L., de Wit, H., Padmos, R.C., Schloot, N.C., Drexhage, H.A., and Leslie, R.D. (2010). Monocyte gene-expression profiles associated with childhood-onset type 1 diabetes and disease risk: a study of identical twins. Diabetes *59*, 1751-1755.

Beyerlein, A., Chmiel, R., Hummel, S., Winkler, C., Bonifacio, E., and Ziegler, A.G. (2014). Timing of gluten introduction and islet autoimmunity in young children: updated results from the BABYDIET study. Diabetes Care *37*, e194-195.

Bezerra, J.A., Witte, D.P., Aronow, B.J., and Degen, S.J. (1993). Hepatocyte-specific expression of the mouse hepatocyte growth factor-like protein. Hepatology *18*, 394-399.

Bilbao, A., Varesio, E., Luban, J., Strambio-De-Castillia, C., Hopfgartner, G., Muller, M., and Lisacek, F. (2015). Processing strategies and software solutions for data-independent acquisition in mass spectrometry. Proteomics *15*, 964-980.

Biswas, P., Mantelli, B., Sica, A., Malnati, M., Panzeri, C., Saccani, A., Hasson, H., Vecchi, A., Saniabadi, A., Lusso, P.*, et al.* (2003). Expression of CD4 on human peripheral blood neutrophils. Blood *101*, 4452-4456.

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

Bluestone, J.A., Herold, K., and Eisenbarth, G. (2010). Genetics, pathogenesis and clinical interventions in type 1 diabetes. Nature *464*, 1293-1300.

Bock, T., Bausch-Fluck, D., Hofmann, A., and Wollscheid, B. (2012). CD proteome and beyond - technologies for targeting the immune cell surfaceome. Front Biosci (Landmark Ed) *17*, 1599-1612.

Bonifacio, E. (2015). Predicting Type 1 Diabetes Using Biomarkers. Diabetes Care *38*, 989-996.

Borregaard, N. (2010). Neutrophils, from marrow to microbes. Immunity *33*, 657-670.

Bots, M., and Medema, J.P. (2006). Granzymes at a glance. J Cell Sci *119*, 5011-5014.

Bottazzo, G.F., Dean, B.M., Gorsuch, A.N., Cudworth, A.G., and Doniach, D. (1980). Complement-fixing islet-cell antibodies in type-I diabetes: possible monitors of active beta-cell damage. Lancet *1*, 668-672.

Bottazzo, G.F., Dean, B.M., McNally, J.M., MacKay, E.H., Swift, P.G., and Gamble, D.R. (1985). In situ characterization of autoimmune phenomena and expression of HLA molecules in the pancreas in diabetic insulitis. N Engl J Med *313*, 353-360.

Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D.S., Weinrauch, Y., and Zychlinsky, A. (2004). Neutrophil extracellular traps kill bacteria. Science (New York, NY) *303*, 1532-1535.

Brown, C.T., Davis-Richardson, A.G., Giongo, A., Gano, K.A., Crabb, D.B., Mukherjee, N., Casella, G., Drew, J.C., Ilonen, J., Knip, M.*, et al.* (2011). Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. PloS one *6*, e25792.

Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinovic, S.M., Cheng, L.Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C.*, et al.* (2015). Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Molecular & cellular proteomics : MCP *14*, 1400-1410.

Bruderer, R., Bernhardt, O.M., Gandhi, T., and Reiter, L. (2016). High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. Proteomics *16*, 2246-2256.

Buckner, J.H. (2010). Mechanisms of impaired regulation by CD4(+)CD25(+)FOXP3(+) regulatory T cells in human autoimmune diseases. Nature reviews Immunology *10*, 849-859.

Bufler, P., Stiegler, G., Schuchmann, M., Hess, S., Kruger, C., Stelter, F., Eckerskorn, C., Schutt, C., and Engelmann, H. (1995). Soluble lipopolysaccharide receptor (CD14) is released via two different mechanisms from human monocytes and CD14 transfectants. European journal of immunology *25*, 604-610.

Burch, T.C., Morris, M.A., Campbell-Thompson, M., Pugliese, A., Nadler, J.L., and Nyalwidhe, J.O. (2015). Proteomic Analysis of Disease Stratified Human Pancreas Tissue Indicates Unique Signature of Type 1 Diabetes. PloS one *10*, e0135663.

Busse, D., de la Rosa, M., Hobiger, K., Thurley, K., Flossdorf, M., Scheffold, A., and Hofer, T. (2010). Competing feedback loops shape IL-2 signaling between helper and regulatory T lymphocytes in cellular microenvironments. Proceedings of the National Academy of Sciences of the United States of America *107*, 3058-3063.

Buzzetti, R., Cernea, S., Petrone, A., Capizzi, M., Spoletini, M., Zampetti, S., Guglielmi, C., Venditti, C., and Pozzilli, P. (2011). C-peptide response and HLA genotypes in subjects with recent-onset type 1 diabetes after immunotherapy with DiaPep277: an exploratory study. Diabetes *60*, 3067-3072.

Cabrera, S.M., Chen, Y.G., Hagopian, W.A., and Hessner, M.J. (2016a). Blood-based signatures in type 1 diabetes. Diabetologia *59*, 414-425.

Cabrera, S.M., Henschel, A.M., and Hessner, M.J. (2016b). Innate inflammation in type 1 diabetes. Translational research : the journal of laboratory and clinical medicine *167*, 214-227.

Calderon, B., Suri, A., and Unanue, E.R. (2006). In CD4(+) T-Cell-Induced Diabetes, Macrophages Are the Final Effector Cells that Mediate Islet β-Cell Killing : Studies from an Acute Model. The American Journal of Pathology *169*, 2137-2147.

Campbell-Thompson, M., Fu, A., Kaddis, J.S., Wasserfall, C., Schatz, D.A., Pugliese, A., and Atkinson, M.A. (2016). Insulitis and beta-Cell Mass in the Natural History of Type 1 Diabetes. Diabetes *65*, 719-731.

Campbell, I.L., Oxbrow, L., West, J., and Harrison, L.C. (1988). Regulation of MHC protein expression in pancreatic beta-cells by interferon-gamma and tumor necrosis factor-alpha. Molecular endocrinology (Baltimore, Md) *2*, 101-107.

Carmona-Rivera, C., and Kaplan, M.J. (2013). Low-density granulocytes: a distinct class of neutrophils in systemic autoimmunity. Semin Immunopathol *35*, 455-463.

Chait, B.T. (2006). Chemistry. Mass spectrometry: bottom-up or top-down? Science (New York, NY) *314*, 65-66.

Chang, K.-C., Lo, C.-W., Fan, T.-c., Chang, M.D.-T., Shu, C.-W., Chang, C.-H., Chung, C.-T., Fang, S.-l., Chao, C.-C., Tsai, J.-J.*, et al.* (2010). TNF-α Mediates Eosinophil Cationic Protein-induced Apoptosis in BEAS-2B Cells. BMC Cell Biology *11*, 6-6.

Chavez-Galan, L., Arenas-Del Angel, M.C., Zenteno, E., Chavez, R., and Lascurain, R. (2009). Cell death mechanisms induced by cytotoxic lymphocytes. Cellular & molecular immunology *6*, 15-25.

References

Chen, G., Gharib, T.G., Huang, C.C., Taylor, J.M., Misek, D.E., Kardia, S.L., Giordano, T.J., Iannettoni, M.D., Orringer, M.B., Hanash, S.M.*, et al.* (2002). Discordant protein and mRNA expression in lung adenocarcinomas. Molecular & cellular proteomics : MCP *1*, 304-313.

Chimienti, F., Favier, A., and Seve, M. (2005). ZnT-8, a pancreatic beta-cell-specific zinc transporter. Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine *18*, 313-317.

Chmiel, R., Giannopoulou, E.Z., Winkler, C., Achenbach, P., Ziegler, A.G., and Bonifacio, E. (2015). Progression from single to multiple islet autoantibodies often occurs soon after seroconversion: implications for early screening. Diabetologia *58*, 411-413.

Chujo, D., Foucat, E., Nguyen, T.S., Chaussabel, D., Banchereau, J., and Ueno, H. (2013). ZnT8-Specific CD4+ T cells display distinct cytokine expression profiles between type 1 diabetes patients and healthy adults. PloS one *8*, e55595.

Clarke, T.B., Davis, K.M., Lysenko, E.S., Zhou, A.Y., Yu, Y., and Weiser, J.N. (2010). Recognition of peptidoglycan from the microbiota by Nod1 enhances systemic innate immunity. Nature medicine *16*, 228-231.

Clayton, D.G. (2009). Prediction and interaction in complex disease genetics: experience in type 1 diabetes. PLoS genetics *5*, e1000540.

Concannon, P., Chen, W.M., Julier, C., Morahan, G., Akolkar, B., Erlich, H.A., Hilner, J.E., Nerup, J., Nierras, C., Pociot, F.*, et al.* (2009). Genome-wide scan for linkage to type 1 diabetes in 2,496 multiplex families from the Type 1 Diabetes Genetics Consortium. Diabetes *58*, 1018-1022.

Cooper, J.D., Smyth, D.J., Smiles, A.M., Plagnol, V., Walker, N.M., Allen, J.E., Downes, K., Barrett, J.C., Healy, B.C., Mychaleckyj, J.C.*, et al.* (2008). Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. Nature genetics *40*, 1399-1401.

Coppieters, K.T., Dotta, F., Amirian, N., Campbell, P.D., Kay, T.W., Atkinson, M.A., Roep, B.O., and von Herrath, M.G. (2012a). Demonstration of islet-autoreactive CD8 T cells in insulitic lesions from recent onset and long-term type 1 diabetes patients. The Journal of experimental medicine *209*, 51-60.

Coppieters, K.T., Wiberg, A., and von Herrath, M.G. (2012b). Viral infections and molecular mimicry in type 1 diabetes. Apmis *120*, 941-949.

Cox, J., and Mann, M. (2007). Is proteomics the new genomics? Cell *130*, 395-398.

Cox, J., and Mann, M. (2008). 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., and Mann, M. (2011). Quantitative, high-resolution proteomics for data-driven systems biology. Annual review of biochemistry *80*, 273-299.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. Journal of proteome research *10*, 1794-1805.

Crevecoeur, I., Rondas, D., Mathieu, C., and Overbergh, L. (2015). The beta-cell in type 1 diabetes: What have we learned from proteomic studies? Proteomics Clinical applications *9*, 755-766.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

Dabelea, D. (2009). The accelerating epidemic of childhood diabetes. Lancet *373*, 1999-2000.

Dahlquist, G.G., Ivarsson, S., Lindberg, B., and Forsgren, M. (1995). Maternal enteroviral infection during pregnancy as a risk factor for childhood IDDM. A population-based case-control study. Diabetes *44*, 408-413.

Daugherty, A., Dunn, J.L., Rateri, D.L., and Heinecke, J.W. (1994). Myeloperoxidase, a catalyst for lipoprotein oxidation, is expressed in human atherosclerotic lesions. The Journal of clinical investigation *94*, 437-444.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proceedings of the National Academy of Sciences of the United States of America *107*, 14691-14696.

de Goffau, M.C., Luopajarvi, K., Knip, M., Ilonen, J., Ruohtula, T., Harkonen, T., Orivuori, L., Hakala, S., Welling, G.W., Harmsen, H.J.*, et al.* (2013). Fecal microbiota composition differs between children with beta-cell autoimmunity and those without. Diabetes *62*, 1238-1244.

Denny, M.F., Yalavarthi, S., Zhao, W., Thacker, S.G., Anderson, M., Sandy, A.R., McCune, W.J., and Kaplan, M.J. (2010). A distinct subset of proinflammatory neutrophils isolated from patients with systemic lupus erythematosus induces vascular damage and synthesizes type I IFNs. Journal of immunology *184*, 3284-3297.

Devaraj, S., Glaser, N., Griffen, S., Wang-Polagruto, J., Miguelino, E., and Jialal, I. (2006). Increased monocytic activity and biomarkers of inflammation in patients with type 1 diabetes. Diabetes *55*, 774-779.

Diana, J., Simoni, Y., Furio, L., Beaudoin, L., Agerberth, B., Barrat, F., and Lehuen, A. (2013). Crosstalk between neutrophils, B-1a cells and plasmacytoid dendritic cells initiates autoimmune diabetes. Nature medicine *19*, 65-73.

Doerr, A. (2015). DIA mass spectrometry. Nat Meth *12*, 35-35.

Dole, M., Mack, L.L., Hines, R.L., Mobley, R.C., Ferguson, L.D., and Alice, M.B. (1968). Molecular Beams of Macroions. The Journal of Chemical Physics *49*, 2240-2249.

Domon, B., and Aebersold, R. (2010). Options and considerations when selecting a quantitative proteomics strategy. Nat Biotech *28*, 710-721.

Dorfer, V., Pichler, P., Stranzl, T., Stadlmann, J., Taus, T., Winkler, S., and Mechtler, K. (2014). MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. Journal of proteome research *13*, 3679-3684.

Dotta, F., Censini, S., van Halteren, A.G., Marselli, L., Masini, M., Dionisi, S., Mosca, F., Boggi, U., Muda, A.O., Del Prato, S.*, et al.* (2007). Coxsackie B4 virus infection of beta cells and natural killer cell insulitis in recent-onset type 1 diabetic patients. Proceedings of the National Academy of Sciences of the United States of America *104*, 5115-5120.

Douglas, D.J., Frank, A.J., and Mao, D. (2005). Linear ion traps in mass spectrometry. Mass Spectrom Rev *24*, 1-29.

Duarte, T.T., and Spencer, C.T. (2016). Personalized Proteomics: The Future of Precision Medicine. Proteomes *4*.

Durieux, J.J., Vita, N., Popescu, O., Guette, F., Calzada-Wack, J., Munker, R., Schmidt, R.E., Lupker, J., Ferrara, P., Ziegler-Heitbrock, H.W.*, et al.* (1994). The two soluble forms of the lipopolysaccharide receptor, CD14: characterization and release by normal human monocytes. European journal of immunology *24*, 2006-2012.

References

Dwivedi, N., and Radic, M. (2014). Citrullination of autoantigens implicates NETosis in the induction of autoimmunity. Annals of the rheumatic diseases *73*, 483-491.

Eady, J.J., Wortley, G.M., Wormstone, Y.M., Hughes, J.C., Astley, S.B., Foxall, R.J., Doleman, J.F., and Elliott, R.M. (2005). Variation in gene expression profiles of peripheral blood mononuclear cells from healthy volunteers. Physiological genomics *22*, 402-411.

Ehlers, M.R., and Rigby, M.R. (2015). Targeting memory T cells in type 1 diabetes. Current diabetes reports *15*, 84.

Eizirik, D.L., Colli, M.L., and Ortis, F. (2009). The role of inflammation in insulitis and [beta]-cell loss in type 1 diabetes. Nature reviews Endocrinology *5*, 219-226.

Eizirik, D.L., Sammeth, M., Bouckenooghe, T., Bottu, G., Sisino, G., Igoillo-Esteve, M., Ortis, F., Santin, I., Colli, M.L., Barthson, J.*, et al.* (2012). The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. PLoS genetics *8*, e1002552.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature methods *4*, 207-214.

Elo, L.L., Mykkanen, J., Nikula, T., Jarvenpaa, H., Simell, S., Aittokallio, T., Hyoty, H., Ilonen, J., Veijola, R., Simell, T.*, et al.* (2010). Early suppression of immune response pathways characterizes children with prediabetes in genome-wide gene expression profiling. Journal of autoimmunity *35*, 70-76.

Endesfelder, D., Hagen, M., Winkler, C., Haupt, F., Zillmer, S., Knopff, A., Bonifacio, E., Ziegler, A.G., Zu Castell, W., and Achenbach, P. (2016). A novel approach for the analysis of longitudinal profiles reveals delayed progression to type 1 diabetes in a subgroup of multiple-islet-autoantibody-positive children. Diabetologia *59*, 2172-2180.

Eng, J.K., McCormack, A.L., and Yates, J.R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry *5*, 976-989.

Erener, S., Mojibian, M., Fox, J.K., Denroche, H.C., and Kieffer, T.J. (2013). Circulating miR-375 as a biomarker of beta-cell death and diabetes in mice. Endocrinology *154*, 603-608.

Erlich, H., Valdes, A.M., Noble, J., Carlson, J.A., Varney, M., Concannon, P., Mychaleckyj, J.C., Todd, J.A., Bonella, P., Fear, A.L.*, et al.* (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.

Evangelista, A.F., Collares, C.V., Xavier, D.J., Macedo, C., Manoel-Caetano, F.S., Rassi, D.M., Foss-Freitas, M.C., Foss, M.C., Sakamoto-Hojo, E.T., Nguyen, C.*, et al.* (2014). Integrative analysis of the transcriptome profiles observed in type 1, type 2 and gestational diabetes mellitus reveals the role of inflammation. BMC medical genomics *7*, 28.

Fabris, P., Betterle, C., Greggio, N.A., Zanchetta, R., Bosi, E., Biasin, M.R., and de Lalla, F. (1998). Insulin-dependent diabetes mellitus during alpha-interferon therapy for chronic viral hepatitis. Journal of hepatology *28*, 514-517.

Fagerberg, L., Jonasson, K., von Heijne, G., Uhlen, M., and Berglund, L. (2010). Prediction of the human membrane proteome. Proteomics *10*, 1141-1149.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science (New York, NY) *246*, 64-71.

Fennessy, M., Metcalfe, K., Hitman, G.A., Niven, M., Biro, P.A., Tuomilehto, J., and Tuomilehto-Wolf, E. (1994). A gene in the HLA class I region contributes to susceptibility to IDDM in the Finnish population. Childhood Diabetes in Finland (DiMe) Study Group. Diabetologia *37*, 937-944.

Ferraro, A., Socci, C., Stabilini, A., Valle, A., Monti, P., Piemonti, L., Nano, R., Olek, S., Maffi, P., Scavini, M.*, et al.* (2011). Expansion of Th17 cells and functional defects in T regulatory cells are key features of the pancreatic lymph nodes in patients with type 1 diabetes. Diabetes *60*, 2903-2913.

Ferreira, R.C., Guo, H., Coulson, R.M., Smyth, D.J., Pekalski, M.L., Burren, O.S., Cutler, A.J., Doecke, J.D., Flint, S., McKinney, E.F.*, et al.* (2014). A type I interferon transcriptional signature precedes autoimmunity in children genetically at risk for type 1 diabetes. Diabetes *63*, 2538-2550.

Ferreira, R.C., Simons, H.Z., Thompson, W.S., Cutler, A.J., Dopico, X.C., Smyth, D.J., Mashar, M., Schuilenburg, H., Walker, N.M., Dunger, D.B.*, et al.* (2015). IL-21 production by CD4+ effector T cells and frequency of circulating follicular helper T cells are increased in type 1 diabetes patients. Diabetologia *58*, 781-790.

Fiehn, O. (2002). Metabolomics--the link between genotypes and phenotypes. Plant molecular biology *48*, 155-171.

Filion, L.G., Izaguirre, C.A., Garber, G.E., Huebsh, L., and Aye, M.T. (1990). Detection of surface and cytoplasmic CD4 on blood monocytes from normal and HIV-1 infected individuals. J Immunol Methods *135*, 59-69.

Fisher, M.M., Watkins, R.A., Blum, J., Evans-Molina, C., Chalasani, N., DiMeglio, L.A., Mather, K.J., Tersey, S.A., and Mirmira, R.G. (2015). Elevations in Circulating Methylated and Unmethylated Preproinsulin DNA in New-Onset Type 1 Diabetes. Diabetes *64*, 3867-3872.

Floyel, T., Brorsson, C., Nielsen, L.B., Miani, M., Bang-Berthelsen, C.H., Friedrichsen, M., Overgaard, A.J., Berchtold, L.A., Wiberg, A., Poulsen, P.*, et al.* (2014). CTSH regulates beta-cell function and disease progression in newly diagnosed type 1 diabetes patients. Proceedings of the National Academy of Sciences of the United States of America *111*, 10305-10310.

Fogel, L.A., Yokoyama, W.M., and French, A.R. (2013). Natural killer cells in human autoimmune disorders. Arthritis research & therapy *15*, 216.

Forga, L., Bolado, F., Goñi, M.J., Tamayo, I., Ibáñez, B., and Prieto, C. (2016). Low Serum Levels of Prealbumin, Retinol Binding Protein, and Retinol Are Frequent in Adult Type 1 Diabetic Patients. Journal of diabetes research *2016*, 2532108.

Foulis, A.K., Farquharson, M.A., and Hardman, R. (1987). Aberrant expression of class II major histocompatibility complex molecules by B cells and hyperexpression of class I major histocompatibility complex molecules by insulin containing islets in type 1 (insulin-dependent) diabetes mellitus. Diabetologia *30*, 333-343.

Foulis, A.K., Liddle, C.N., Farquharson, M.A., Richmond, J.A., and Weir, R.S. (1986). The histopathology of the pancreas in type 1 (insulin-dependent) diabetes mellitus: a 25-year review of deaths in patients under 20 years of age in the United Kingdom. Diabetologia *29*, 267-274.

Foulis, A.K., and Stewart, J.A. (1984). The pancreas in recent-onset type 1 (insulin-dependent) diabetes mellitus: insulin content of islets, insulitis and associated changes in the exocrine acinar tissue. Diabetologia *26*, 456-461.

Fuchs, T.A., Brill, A., and Wagner, D.D. (2012). Neutrophil extracellular trap (NET) impact on deep vein thrombosis. Arteriosclerosis, thrombosis, and vascular biology *32*, 1777-1783.

Fuchtenbusch, M., Irnstetter, A., Jager, G., and Ziegler, A.G. (2001). No evidence for an association of coxsackie virus infections during pregnancy and early childhood with development

References

of islet autoantibodies in offspring of mothers or fathers with type 1 diabetes. Journal of autoimmunity *17*, 333-340.

Gale, E.A. (2005). Type 1 diabetes in the young: the harvest of sorrow goes on. Diabetologia *48*, 1435-1438.

Gamble, D.R., Kinsley, M.L., FitzGerald, M.G., Bolton, R., and Taylor, K.W. (1969). Viral Antibodies in Diabetes Mellitus. British Medical Journal *3*, 627-630.

Garcia-Contreras, M., Shah, S.H., Tamayo, A., Robbins, P.D., Golberg, R.B., Mendez, A.J., and Ricordi, C. (2017). Plasma-derived exosome characterization reveals a distinct microRNA signature in long duration Type 1 diabetes. Scientific reports *7*, 5998.

Garg, G., Tyler, J.R., Yang, J.H., Cutler, A.J., Downes, K., Pekalski, M., Bell, G.L., Nutland, S., Peakman, M., Todd, J.A.*, et al.* (2012). Type 1 diabetes-associated IL2RA variation lowers IL-2 signaling and contributes to diminished CD4+CD25+ regulatory T cell function. Journal of immunology *188*, 4644-4653.

Gepts, W. (1965). Pathologic anatomy of the pancreas in juvenile diabetes mellitus. Diabetes *14*, 619-633.

Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., and Gygi, S.P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proceedings of the National Academy of Sciences of the United States of America *100*, 6940-6945.

Gewin, L., Myers, H., Kiyono, T., and Galloway, D.A. (2004). Identification of a novel telomerase repressor that interacts with the human papillomavirus type-16 E6/E6-AP complex. Genes & development *18*, 2269-2282.

Geyer, P.E., Kulak, N.A., Pichler, G., Holdt, L.M., Teupser, D., and Mann, M. (2016). Plasma Proteome Profiling to Assess Human Health and Disease. Cell systems *2*, 185-195.

Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., Furlotte, N.*, et al.* (2011). Comparative analysis of proteome and transcriptome variation in mouse. PLoS genetics *7*, e1001393.

Gillespie, K.M., Bain, S.C., Barnett, A.H., Bingley, P.J., Christie, M.R., Gill, G.V., and Gale, E.A. (2004). The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes. Lancet *364*, 1699-1700.

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

Gilliet, M., Cao, W., and Liu, Y.J. (2008). Plasmacytoid dendritic cells: sensing nucleic acids in viral infection and autoimmune diseases. Nature reviews Immunology *8*, 594-606.

Giongo, A., Gano, K.A., Crabb, D.B., Mukherjee, N., Novelo, L.L., Casella, G., Drew, J.C., Ilonen, J., Knip, M., Hyoty, H.*, et al.* (2011). Toward defining the autoimmune microbiome for type 1 diabetes. The ISME journal *5*, 82-91.

Gottlieb, P.A., Quinlan, S., Krause-Steinrauf, H., Greenbaum, C.J., Wilson, D.M., Rodriguez, H., Schatz, D.A., Moran, A.M., Lachin, J.M., and Skyler, J.S. (2010). Failure to preserve beta-cell function with mycophenolate mofetil and daclizumab combined therapy in patients with new-onset type 1 diabetes. Diabetes Care *33*, 826-832.

Graessel, A., Hauck, S.M., von Toerne, C., Kloppmann, E., Goldberg, T., Koppensteiner, H., Schindler, M., Knapp, B., Krause, L., Dietz, K.*, et al.* (2015). A Combined Omics Approach to

Generate the Surface Atlas of Human Naive CD4+ T Cells during Early T-Cell Receptor Activation. Molecular & cellular proteomics : MCP *14*, 2085-2102.

Grassi, G., Di Caprio, G., Fimia, G.M., Ippolito, G., Tripodi, M., and Alonzi, T. (2016). Hepatitis C virus relies on lipoproteins for its life cycle. World journal of gastroenterology *22*, 1953-1965.

Graves, P.M., Rotbart, H.A., Nix, W.A., Pallansch, M.A., Erlich, H.A., Norris, J.M., Hoffman, M., Eisenbarth, G.S., and Rewers, M. (2003). Prospective study of enteroviral infections and development of beta-cell autoimmunity. Diabetes autoimmunity study in the young (DAISY). Diabetes research and clinical practice *59*, 51-61.

Graves, P.R., and Haystead, T.A.J. (2002). Molecular Biologist's Guide to Proteomics. Microbiology and Molecular Biology Reviews *66*, 39-63.

Greenbaum, C.J., Schatz, D.A., Cuthbertson, D., Zeidler, A., Eisenbarth, G.S., and Krischer, J.P. (2000). Islet cell antibody-positive relatives with human leukocyte antigen DQA1*0102, DQB1*0602: identification by the Diabetes Prevention Trial-type 1. J Clin Endocrinol Metab *85*, 1255-1260.

Gregg, B.E., Moore, P.C., Demozay, D., Hall, B.A., Li, M., Husain, A., Wright, A.J., Atkinson, M.A., and Rhodes, C.J. (2012). Formation of a human beta-cell population within pancreatic islets is set early in life. J Clin Endocrinol Metab *97*, 3197-3206.

Gregg, E.W., Li, Y., Wang, J., Burrows, N.R., Ali, M.K., Rolka, D., Williams, D.E., and Geiss, L. (2014). Changes in diabetes-related complications in the United States, 1990-2010. N Engl J Med *370*, 1514-1523.

Griffin, K.J., Thompson, P.A., Gottschalk, M., Kyllo, J.H., and Rabinovitch, A. (2014). Combination therapy with sitagliptin and lansoprazole in patients with recent-onset type 1 diabetes (REPAIR-T1D): 12-month results of a multicentre, randomised, placebo-controlled, phase 2 trial. Lancet Diabetes Endocrinol *2*, 710-718.

Grosche, A., Hauser, A., Lepper, M.F., Mayo, R., von Toerne, C., Merl-Pham, J., and Hauck, S.M. (2016). The Proteome of Native Adult Muller Glial Cells From Murine Retina. Molecular & cellular proteomics : MCP *15*, 462-480.

Guerci, A.P., Guerci, B., Levy-Marchal, C., Ongagna, J., Ziegler, O., Candiloros, H., Guerci, O., and Drouin, P. (1994). Onset of insulin-dependent diabetes mellitus after interferon-alfa therapy for hairy cell leukaemia. Lancet *343*, 1167-1168.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature biotechnology *17*, 994-999.

Hacbarth, E., and Kajdacsy-Balla, A. (1986). Low density neutrophils in patients with systemic lupus erythematosus, rheumatoid arthritis, and acute rheumatic fever. Arthritis Rheum *29*, 1334-1342.

Hagglof, B., Holmgren, G., Holmlund, G., Lindblom, B., Olaisen, B., and Teisberg, P. (1986). Studies of HLA, factor B (Bf), complement C2 and C4 haplotypes in type 1 diabetic and control families from northern Sweden. Human heredity *36*, 201-212.

Hamel, Y., Mauvais, F.X., Pham, H.P., Kratzer, R., Marchi, C., Barilleau, E., Waeckel-Enee, E., Arnoux, J.B., Hartemann, A., Cordier, C.*, et al.* (2016). A unique CD8(+) T lymphocyte signature in pediatric type 1 diabetes. Journal of autoimmunity *73*, 54-63.

Han, D., Leyva, C.A., Matheson, D., Mineo, D., Messinger, S., Blomberg, B.B., Hernandez, A., Meneghini, L.F., Allende, G., Skyler, J.S.*, et al.* (2011). Immune profiling by multiple gene expression analysis in patients at-risk and with type 1 diabetes. Clinical immunology *139*, 290-301.

References

Hansen, J.T., and Netter, F.H. (2014). Netter's Clinical Anatomy (Saunders/Elsevier).

Hansen, T.K., Tarnow, L., Thiel, S., Steffensen, R., Stehouwer, C.D., Schalkwijk, C.G., Parving, H.H., and Flyvbjerg, A. (2004). Association between mannose-binding lectin and vascular complications in type 1 diabetes. Diabetes *53*, 1570-1576.

Harjutsalo, V., Sjoberg, L., and Tuomilehto, J. (2008). Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study. Lancet *371*, 1777-1782.

Harsunen, M.H., Puff, R., D'Orlando, O., Giannopoulou, E., Lachmann, L., Beyerlein, A., von Meyer, A., and Ziegler, A.G. (2013). Reduced blood leukocyte and neutrophil numbers in the pathogenesis of type 1 diabetes. Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme *45*, 467-470.

Haskins, K., and Cooke, A. (2011). CD4 T cells and their antigens in the pathogenesis of autoimmune diabetes. Current opinion in immunology *23*, 739-745.

Heinonen, M.T., Moulder, R., and Lahesmaa, R. (2015). New Insights and Biomarkers for Type 1 Diabetes: Review for Scandinavian Journal of Immunology. Scandinavian journal of immunology *82*, 244-253.

Hermann, R., Knip, M., Veijola, R., Simell, O., Laine, A.P., Akerblom, H.K., Groop, P.H., Forsblom, C., Pettersson-Fernholm, K., and Ilonen, J. (2003). Temporal changes in the frequencies of HLA genotypes in patients with Type 1 diabetes--indication of an increased environmental pressure? Diabetologia *46*, 420-425.

Herold, K.C., Gitelman, S.E., Willi, S.M., Gottlieb, P.A., Waldron-Lynch, F., Devine, L., Sherr, J., Rosenthal, S.M., Adi, S., Jalaludin, M.Y.*, et al.* (2013). Teplizumab treatment may improve C-peptide responses in participants with type 1 diabetes after the new-onset period: a randomised controlled trial. Diabetologia *56*, 391-400.

Herold, K.C., Huen, A., Gould, L., Traisman, H., and Rubenstein, A.H. (1984). Alterations in lymphocyte subpopulations in type 1 (insulin-dependent) diabetes mellitus: exploration of possible mechanisms and relationships to autoimmune phenomena. Diabetologia *27 Suppl*, 102-105.

Herold, K.C., Usmani-Brown, S., Ghazi, T., Lebastchi, J., Beam, C.A., Bellin, M.D., Ledizet, M., Sosenko, J.M., Krischer, J.P., and Palmer, J.P. (2015). beta cell death and dysfunction during type 1 diabetes development in at-risk individuals. The Journal of clinical investigation *125*, 1163-1173.

Herrath, M., Peakman, M., and Roep, B. (2013). Progress in immune-based therapies for type 1 diabetes. Clinical and experimental immunology *172*, 186-202.

Holmberg, H., Wahlberg, J., Vaarala, O., and Ludvigsson, J. (2007). Short duration of breast-feeding as a risk-factor for beta-cell autoantibodies in 5-year-old children from the general population. The British journal of nutrition *97*, 111-116.

Honeyman, M.C., Stone, N.L., Falk, B.A., Nepom, G., and Harrison, L.C. (2010). Evidence for molecular mimicry between human T cell epitopes in rotavirus and pancreatic islet autoantigens. Journal of immunology *184*, 2204-2210.

Hu, A., Noble, W.S., and Wolf-Yadlin, A. (2016). Technical advances in proteomics: new developments in data-independent acquisition. F1000Research *5*, F1000 Faculty Rev-1419.

Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. Journal of mass spectrometry : JMS *40*, 430-443.

Hughes, C.S., and Morin, G.B. (2017). Using Public Data for Comparative Proteome Analysis in Precision Medicine Programs. Proteomics Clinical applications.

Hull, C.M., Peakman, M., and Tree, T.I.M. (2017). Regulatory T cell dysfunction in type 1 diabetes: what's broken and how can we fix it? Diabetologia.

Hummel, M., Durinovic-Bello, I., and Ziegler, A.G. (1996). Relation between cellular and humoral immunity to islet cell antigens in type 1 diabetes. Journal of autoimmunity *9*, 427-430.

Hunt, D.F., Buko, A.M., Ballard, J.M., Shabanowitz, J., and Giordani, A.B. (1981). Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. Biomedical mass spectrometry *8*, 397-408.

Hunt, D.F., Yates, J.R., 3rd, Shabanowitz, J., Winston, S., and Hauer, C.R. (1986). Protein sequencing by tandem mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America *83*, 6233-6237.

Hyoty, H., Hiltunen, M., Knip, M., Laakkonen, M., Vahasalo, P., Karjalainen, J., Koskela, P., Roivainen, M., Leinikki, P., Hovi, T.*, et al.* (1995). A prospective study of the role of coxsackie B and other enterovirus infections in the pathogenesis of IDDM. Childhood Diabetes in Finland (DiMe) Study Group. Diabetes *44*, 652-657.

Hyoty, H., and Taylor, K.W. (2002). The role of viruses in human diabetes. Diabetologia *45*, 1353-1361.

Hypponen, E., Laara, E., Reunanen, A., Jarvelin, M.R., and Virtanen, S.M. (2001). Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. Lancet *358*, 1500-1503.

In't Veld, P. (2011). Insulitis in human type 1 diabetes: The quest for an elusive lesion. Islets *3*, 131-138.

Insel, R.A., Dunne, J.L., Atkinson, M.A., Chiang, J.L., Dabelea, D., Gottlieb, P.A., Greenbaum, C.J., Herold, K.C., Krischer, J.P., Lernmark, A.*, et al.* (2015). Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. Diabetes Care *38*, 1964-1974.

Iribarne, J.V., and Thomson, B.A. (1976). On the evaporation of small ions from charged droplets. The Journal of Chemical Physics *64*, 2287-2294.

Ishizawa, J., Kojima, K., Hail, N., Jr., Tabe, Y., and Andreeff, M. (2015). Expression, function, and targeting of the nuclear exporter chromosome region maintenance 1 (CRM1) protein. Pharmacology & therapeutics *153*, 25-35.

Itoh, A., and Ridgway, W.M. (2017). Targeting innate immunity to downmodulate adaptive immunity and reverse type 1 diabetes. ImmunoTargets and Therapy *6*, 31-38.

Itoh, N., Hanafusa, T., Miyagawa, J., Tamura, S., Inada, M., Kawata, S., Kono, N., and Tarui, S. (1992). Transthyretin (prealbumin) in the pancreas and sera of newly diagnosed type I (insulin-dependent) diabetic patients. J Clin Endocrinol Metab *74*, 1372-1377.

Jacobsson, B., Carlstrom, A., Platz, A., and Collins, V.P. (1990). Transthyretin messenger ribonucleic acid expression in the pancreas and in endocrine tumors of the pancreas and gut. J Clin Endocrinol Metab *71*, 875-880.

Jin, Y., Sharma, A., Carey, C., Hopkins, D., Wang, X., Robertson, D.G., Bode, B., Anderson, S.W., Reed, J.C., Steed, R.D.*, et al.* (2013). The expression of inflammatory genes is upregulated in peripheral blood of patients with type 1 diabetes. Diabetes Care *36*, 2794-2802.

Joffre, O.P., Segura, E., Savina, A., and Amigorena, S. (2012). Cross-presentation by dendritic cells. Nature reviews Immunology *12*, 557-569.

References

Juntti-Berggren, L., Refai, E., Appelskog, I., Andersson, M., Imreh, G., Dekki, N., Uhles, S., Yu, L., Griffiths, W.J., Zaitsev, S., *et al.* (2004). Apolipoprotein CIII promotes Ca(2+)-dependent β cell death in type 1 diabetes. Proceedings of the National Academy of Sciences of the United States of America *101*, 10090-10094.

Kahn, H.S., Morgan, T.M., Case, L.D., Dabelea, D., Mayer-Davis, E.J., Lawrence, J.M., Marcovina, S.M., Imperatore, G., and Group, S.f.D.i.Y.S. (2009). Association of type 1 diabetes with month of birth among U.S. youth: The SEARCH for Diabetes in Youth Study. Diabetes Care *32*, 2010-2015.

Kaizer, E.C., Glaser, C.L., Chaussabel, D., Banchereau, J., Pascual, V., and White, P.C. (2007). Gene expression in peripheral blood mononuclear cells from children with diabetes. J Clin Endocrinol Metab *92*, 3705-3711.

Kall, L., Canterbury, J.D., Weston, J., Noble, W.S., and MacCoss, M.J. (2007a). Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nature methods *4*, 923-925.

Kall, L., Krogh, A., and Sonnhammer, E.L. (2004). A combined transmembrane topology and signal peptide prediction method. Journal of molecular biology *338*, 1027-1036.

Kall, L., Krogh, A., and Sonnhammer, E.L. (2007b). Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic acids research *35*, W429-432.

Kallionpaa, H., Elo, L.L., Laajala, E., Mykkanen, J., Ricano-Ponce, I., Vaarma, M., Laajala, T.D., Hyoty, H., Ilonen, J., Veijola, R., *et al.* (2014). Innate immune activity is detected prior to seroconversion in children with HLA-conferred type 1 diabetes susceptibility. Diabetes *63*, 2402-2414.

Kaprio, J., Tuomilehto, J., Koskenvuo, M., Romanov, K., Reunanen, A., Eriksson, J., Stengard, J., and Kesaniemi, Y.A. (1992). Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. Diabetologia *35*, 1060-1067.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Analytical chemistry *60*, 2299-2301.

Karni, M., Zidon, D., Polak, P., Zalevsky, Z., and Shefi, O. (2013). Thermal degradation of DNA. DNA and cell biology *32*, 298-301.

Kaspar, A.A., Okada, S., Kumar, J., Poulain, F.R., Drouvalakis, K.A., Kelekar, A., Hanson, D.A., Kluck, R.M., Hitoshi, Y., Johnson, D.E., *et al.* (2001). A distinct pathway of cell-mediated apoptosis initiated by granulysin. Journal of immunology *167*, 350-356.

Kassem, S.A., Ariel, I., Thornton, P.S., Scheimberg, I., and Glaser, B. (2000). Beta-cell proliferation and apoptosis in the developing normal human pancreas and in hyperinsulinism of infancy. Diabetes *49*, 1325-1333.

Keenan, H.A., Sun, J.K., Levine, J., Doria, A., Aiello, L.P., Eisenbarth, G., Bonner-Weir, S., and King, G.L. (2010a). Residual insulin production and pancreatic ss-cell turnover after 50 years of diabetes: Joslin Medalist Study. Diabetes *59*, 2846-2853.

Keenan, H.A., Sun, J.K., Levine, J., Doria, A., Aiello, L.P., Eisenbarth, G., Bonner-Weir, S., and King, G.L. (2010b). Residual Insulin Production and Pancreatic β-Cell Turnover After 50 Years of Diabetes: Joslin Medalist Study. Diabetes *59*, 2846-2853.

Kettle, A.J., Chan, T., Osberg, I., Senthilmohan, R., Chapman, A.L., Mocatta, T.J., and Wagener, J.S. (2004). Myeloperoxidase and protein oxidation in the airways of young children with cystic fibrosis. American journal of respiratory and critical care medicine *170*, 1317-1323.

Keymeulen, B., Walter, M., Mathieu, C., Kaufman, L., Gorus, F., Hilbrands, R., Vandemeulebroucke, E., Van de Velde, U., Crenier, L., De Block, C.*, et al.* (2010). Four-year metabolic outcome of a randomised controlled CD3-antibody trial in recent-onset type 1 diabetic patients depends on their age and baseline residual beta cell mass. Diabetologia *53*, 614-623.

Khoruts, A., and Fraser, J.M. (2005). A causal link between lymphopenia and autoimmunity. Immunology letters *98*, 23-31.

Kim, M.S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S.*, et al.* (2014). A draft map of the human proteome. Nature *509*, 575-581.

Kim, S.Y., Kang, H.T., Han, J.A., and Park, S.C. (2012). The transcription factor Sp1 is responsible for aging-dependent altered nucleocytoplasmic trafficking. Aging cell *11*, 1102-1109.

Kim, S.Y., Ryu, S.J., Ahn, H.J., Choi, H.R., Kang, H.T., and Park, S.C. (2010). Senescence-related functional nuclear barrier by down-regulation of nucleo-cytoplasmic trafficking gene expression. Biochemical and biophysical research communications *391*, 28-32.

Kimpimaki, T., Erkkola, M., Korhonen, S., Kupila, A., Virtanen, S.M., Ilonen, J., Simell, O., and Knip, M. (2001a). Short-term exclusive breastfeeding predisposes young children with increased genetic risk of Type I diabetes to progressive beta-cell autoimmunity. Diabetologia *44*, 63-69.

Kimpimaki, T., Kupila, A., Hamalainen, A.M., Kukko, M., Kulmala, P., Savola, K., Simell, T., Keskinen, P., Ilonen, J., Simell, O.*, et al.* (2001b). The first signs of beta-cell autoimmunity appear in infancy in genetically susceptible children from the general population: the Finnish Type 1 Diabetes Prediction and Prevention Study. J Clin Endocrinol Metab *86*, 4782-4788.

Knight, J.S., Carmona-Rivera, C., and Kaplan, M.J. (2012). Proteins derived from neutrophil extracellular traps may serve as self-antigens and mediate organ damage in autoimmune diseases. Frontiers in immunology *3*, 380.

Knip, M., and Hyöty, H. (2008). Environmental Determinants. In Epidemiology of Pediatric and Adolescent Diabetes (CRC Press), pp. 63-84.

Knip, M., Siljander, H., Ilonen, J., Simell, O., and Veijola, R. (2016). Role of humoral beta-cell autoimmunity in type 1 diabetes. Pediatric diabetes *17 Suppl 22*, 17-24.

Knip, M., and Simell, O. (2012). Environmental Triggers of Type 1 Diabetes. Cold Spring Harbor Perspectives in Medicine *2*, a007690.

Kockum, I., Lernmark, A., Dahlquist, G., Falorni, A., Hagopian, W.A., Landin-Olsson, M., Li, L.C., Luthman, H., Palmer, J.P., Sanjeevi, C.B.*, et al.* (1996). Genetic and immunological findings in patients with newly diagnosed insulin-dependent diabetes mellitus. The Swedish Childhood Diabetes Study Group and The Diabetes Incidence in Sweden Study (DISS) Group. Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme *28*, 344-347.

Koh, G.C., Shek, L.P., Goh, D.Y., Van Bever, H., and Koh, D.S. (2007). Eosinophil cationic protein: is it useful in asthma? A systematic review. Respiratory medicine *101*, 696-705.

Kohlbacher, O., Reinert, K., Gropl, C., Lange, E., Pfeifer, N., Schulz-Trieglaff, O., and Sturm, M. (2007). TOPP--the OpenMS proteomics pipeline. Bioinformatics *23*, e191-197.

Kolb, H., and von Herrath, M. (2017). Immunotherapy for Type 1 Diabetes: Why Do Current Protocols Not Halt the Underlying Disease Process? Cell metabolism *25*, 233-241.

Korkmaz, B., Horwitz, M.S., Jenne, D.E., and Gauthier, F. (2010). Neutrophil Elastase, Proteinase 3, and Cathepsin G as Therapeutic Targets in Human Diseases. Pharmacological Reviews *62*, 726-759.

References

Kostic, A.D., Gevers, D., Siljander, H., Vatanen, T., Hyotylainen, T., Hamalainen, A.M., Peet, A., Tillmann, V., Poho, P., Mattila, I., *et al.* (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell host & microbe *17*, 260-273.

Krensky, A.M., and Clayberger, C. (2009). Biology and clinical relevance of granulysin. Tissue antigens *73*, 193-198.

Kreuwel, H.T., Morgan, D.J., Krahl, T., Ko, A., Sarvetnick, N., and Sherman, L.A. (1999). Comparing the relative role of perforin/granzyme versus Fas/Fas ligand cytotoxic pathways in CD8+ T cell-mediated insulin-dependent diabetes mellitus. Journal of immunology *163*, 4335-4341.

Krogvold, L., Edwin, B., Buanes, T., Frisk, G., Skog, O., Anagandula, M., Korsgren, O., Undlien, D., Eike, M.C., Richardson, S.J., *et al.* (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.

Krogvold, L., Wiberg, A., Edwin, B., Buanes, T., Jahnsen, F.L., Hanssen, K.F., Larsson, E., Korsgren, O., Skog, O., and Dahl-Jorgensen, K. (2016). Insulitis and characterisation of infiltrating T cells in surgical pancreatic tail resections from patients at onset of type 1 diabetes. Diabetologia *59*, 492-501.

Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nature methods *11*, 319-324.

La Torre, D., Seppanen-Laakso, T., Larsson, H.E., Hyotylainen, T., Ivarsson, S.A., Lernmark, A., and Oresic, M. (2013). Decreased cord-blood phospholipids in young age-at-onset type 1 diabetes. Diabetes *62*, 3951-3956.

Lamb, M.M., Miller, M., Seifert, J.A., Frederiksen, B., Kroehl, M., Rewers, M., and Norris, J.M. (2015). The effect of childhood cow's milk intake and HLA-DR genotype on risk of islet autoimmunity and type 1 diabetes: the Diabetes Autoimmunity Study in the Young. Pediatric diabetes *16*, 31-38.

Lande, R., Ganguly, D., Facchinetti, V., Frasca, L., Conrad, C., Gregorio, J., Meller, S., Chamilos, G., Sebasigari, R., Riccieri, V., *et al.* (2011). Neutrophils activate plasmacytoid dendritic cells by releasing self-DNA-peptide complexes in systemic lupus erythematosus. Science translational medicine *3*, 73ra19.

Lebastchi, J., and Herold, K.C. (2012a). Immunologic and metabolic biomarkers of beta-cell destruction in the diagnosis of type 1 diabetes. Cold Spring Harb Perspect Med *2*, a007708.

Lebastchi, J., and Herold, K.C. (2012b). Immunologic and Metabolic Biomarkers of β-Cell Destruction in the Diagnosis of Type 1 Diabetes. Cold Spring Harbor Perspectives in Medicine *2*, a007708.

Lee, S., Lee, H.-C., Kwon, Y.-W., Lee, S.E., Cho, Y., Kim, J., Lee, S., Kim, J.-Y., Lee, J., Yang, H.-M., *et al.* (2014). Adenylyl Cyclase-Associated Protein 1(CAP1) is a Receptor for Human Resistin and Mediates Inflammatory Actions of Human Monocytes. Cell metabolism *19*, 484-497.

Lehmann, W.D. (2010). Protein Phosphorylation Analysis by Electrospray Mass Spectrometry: A Guide to Concepts and Practice (Royal Society of Chemistry).

Lehuen, A., Diana, J., Zaccone, P., and Cooke, A. (2010). Immune cell crosstalk in type 1 diabetes. Nature reviews Immunology *10*, 501-513.

Lenzen, S., Drinkgern, J., and Tiedge, M. (1996). Low antioxidant enzyme gene expression in pancreatic islets compared with various other mouse tissues. Free Radic Biol Med *20*, 463-466.

Leonard, E.J., and Skeel, A. (1976). A serum protein that stimulates macrophage movement, chemotaxis and spreading. Experimental cell research *102*, 434-438.

Lernmark, A., and Larsson, H.E. (2013). Immune therapy in type 1 diabetes mellitus. Nature reviews Endocrinology *9*, 92-103.

Leslie, R.D. (2010). Predicting adult-onset autoimmune diabetes: clarity from complexity. Diabetes *59*, 330-331.

Li, G.Z., Vissers, J.P., Silva, J.C., Golick, D., Gorenstein, M.V., and Geromanos, S.J. (2009). Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. Proteomics *9*, 1696-1719.

Lien, E., Aukrust, P., Sundan, A., Muller, F., Froland, S.S., and Espevik, T. (1998). Elevated levels of serum-soluble CD14 in human immunodeficiency virus type 1 (HIV-1) infection: correlation to disease progression and clinical events. Blood *92*, 2084-2092.

Lindley, S., Dayan, C.M., Bishop, A., Roep, B.O., Peakman, M., and Tree, T.I. (2005). Defective suppressor function in CD4(+)CD25(+) T-cells from patients with type 1 diabetes. Diabetes *54*, 92-99.

Ling, P., Lu, T.J., Yuan, C.J., and Lai, M.D. (2008). Biosignaling of mammalian Ste20-related kinases. Cellular signalling *20*, 1237-1247.

Lipponen, K., Gombos, Z., Kiviniemi, M., Siljander, H., Lempainen, J., Hermann, R., Veijola, R., Simell, O., Knip, M., and Ilonen, J. (2010). Effect of HLA Class I and Class II Alleles on Progression From Autoantibody Positivity to Overt Type 1 Diabetes in Children With Risk-Associated Class II Genotypes. Diabetes *59*, 3253-3256.

Litherland, S.A., She, J.X., Schatz, D., Fuller, K., Hutson, A.D., Peng, R.H., Li, Y., Grebe, K.M., Whittaker, D.S., Bahjat, K.*, et al.* (2003). Aberrant monocyte prostaglandin synthase 2 (PGS2) expression in type 1 diabetes before and after disease onset. Pediatric diabetes *4*, 10-18.

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

Liu, H., Sadygov, R.G., and Yates, J.R., 3rd (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Analytical chemistry *76*, 4193-4201.

Liu, Y., Buil, A., Collins, B.C., Gillet, L.C., Blum, L.C., Cheng, L.Y., Vitek, O., Mouritsen, J., Lachance, G., Spector, T.D.*, et al.* (2015). Quantitative variability of 342 plasma proteins in a human twin population. Molecular systems biology *11*, 786.

Long, E.O., Kim, H.S., Liu, D., Peterson, M.E., and Rajagopalan, S. (2013). Controlling NK Cell Responses: Integration of Signals for Activation and Inhibition. Annual review of immunology *31*, 10.1146/annurev-immunol-020711-075005.

Long, S.A., Rieck, M., Sanda, S., Bollyky, J.B., Samuels, P.L., Goland, R., Ahmann, A., Rabinovitch, A., Aggarwal, S., Phippard, D.*, et al.* (2012). Rapamycin/IL-2 combination therapy in patients with type 1 diabetes augments Tregs yet transiently impairs beta-cell function. Diabetes *61*, 2340-2348.

Lonnrot, M., Korpela, K., Knip, M., Ilonen, J., Simell, O., Korhonen, S., Savola, K., Muona, P., Simell, T., Koskela, P.*, et al.* (2000). Enterovirus infection as a risk factor for beta-cell autoimmunity in a prospectively observed birth cohort: the Finnish Diabetes Prediction and Prevention Study. Diabetes *49*, 1314-1318.

Lorini, R., Moretta, A., Valtorta, A., d'Annunzio, G., Cortona, L., Vitali, L., Bozzola, M., and Severi, F. (1994). Cytotoxic activity in children with insulin-dependent diabetes mellitus. Diabetes research and clinical practice *23*, 37-42.

References

Lottspeich, F. (1999). Proteome Analysis: A Pathway to the Functional Analysis of Proteins. Angewandte Chemie *38*, 2476-2492.

Ludvigsson, J., Cheramy, M., Axelsson, S., Pihl, M., Akerman, L., Casas, R., and Clinical, G.A.D.S.G.i.S. (2014). GAD-treatment of children and adolescents with recent-onset type 1 diabetes preserves residual insulin secretion after 30 months. Diabetes Metab Res Rev *30*, 405-414.

Ludvigsson, J., Krisky, D., Casas, R., Battelino, T., Castano, L., Greening, J., Kordonouri, O., Otonkoski, T., Pozzilli, P., Robert, J.J.*, et al.* (2012). GAD65 antigen therapy in recently diagnosed type 1 diabetes mellitus. N Engl J Med *366*, 433-442.

Lundberg, M., Seiron, P., Ingvast, S., Korsgren, O., and Skog, O. (2017). Insulitis in human diabetes: a histological evaluation of donor pancreases. Diabetologia *60*, 346-353.

Lünemann, A., Lünemann, J.D., and Münz, C. (2009). Regulatory NK-Cell Functions in Inflammation and Autoimmunity. Molecular Medicine *15*, 352-358.

Maahs, D.M., West, N.A., Lawrence, J.M., and Mayer-Davis, E.J. (2010). Chapter 1: Epidemiology of Type 1 Diabetes. Endocrinology and metabolism clinics of North America *39*, 481-497.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966-968.

Makarov, A. (2000). Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Analytical chemistry *72*, 1156-1162.

Malek, T.R., and Castro, I. (2010). Interleukin-2 receptor signaling: at the interface between tolerance and immunity. Immunity *33*, 153-165.

Malle, E., Waeg, G., Schreiber, R., Grone, E.F., Sattler, W., and Grone, H.J. (2000). Immunohistochemical evidence for the myeloperoxidase/H2O2/halide system in human atherosclerotic lesions: colocalization of myeloperoxidase and hypochlorite-modified proteins. Eur J Biochem *267*, 4495-4503.

Mandal, A., and Viswanathan, C. (2015). Natural killer cells: In health and disease. Hematology/oncology and stem cell therapy *8*, 47-55.

Manza, L.L., Stamer, S.L., Ham, A.J., Codreanu, S.G., and Liebler, D.C. (2005). Sample preparation and digestion for proteomic analyses using spin filters. Proteomics *5*, 1742-1745.

Marchand, L., Jalabert, A., Meugnier, E., Van den Hende, K., Fabien, N., Nicolino, M., Madec, A.M., Thivolet, C., and Rome, S. (2016). miRNA-375 a Sensor of Glucotoxicity Is Altered in the Serum of Children with Newly Diagnosed Type 1 Diabetes. Journal of diabetes research *2016*, 1869082.

Marcos, V., Latzin, P., Hector, A., Sonanini, S., Hoffmann, F., Lacher, M., Koller, B., Bufler, P., Nicolai, T., Hartl, D.*, et al.* (2010). Expression, regulation and clinical significance of soluble and membrane CD14 receptors in pediatric inflammatory lung diseases. Respiratory research *11*, 32.

Marrack, P., and Kappler, J.W. (2012). Do MHCII-presented neoantigens drive type 1 diabetes and other autoimmune diseases? Cold Spring Harb Perspect Med *2*, a007765.

Maslowski, K.M., Vieira, A.T., Ng, A., Kranich, J., Sierro, F., Yu, D., Schilter, H.C., Rolph, M.S., Mackay, F., Artis, D.*, et al.* (2009). Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. Nature *461*, 1282-1286.

Masopust, D., and Schenkel, J.M. (2013). The integration of T cell migration, differentiation and function. Nature reviews Immunology *13*, 309-320.

Mastrandrea, L., Yu, J., Behrens, T., Buchlis, J., Albini, C., Fourtner, S., and Quattrin, T. (2009). Etanercept treatment in children with new-onset type 1 diabetes: pilot randomized, placebo-controlled, double-blind study. Diabetes Care *32*, 1244-1249.

Mayadas, T.N., Cullere, X., and Lowell, C.A. (2014). The multifaceted functions of neutrophils. Annual review of pathology *9*, 181-218.

Meissner, F., and Mann, M. (2014). Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology. Nature immunology *15*, 112-117.

Meli, A.P., Fontes, G., Avery, D.T., Leddon, S.A., Tam, M., Elliot, M., Ballesteros-Tato, A., Miller, J., Stevenson, M.M., Fowell, D.J.*, et al.* (2016). The Integrin LFA-1 Controls T Follicular Helper Cell Generation and Maintenance. Immunity *45*, 831-846.

Mendez-Lagares, G., Romero-Sanchez, M.C., Ruiz-Mateos, E., Genebat, M., Ferrando-Martinez, S., Munoz-Fernandez, M.A., Pacheco, Y.M., and Leal, M. (2013). Long-term suppressive combined antiretroviral treatment does not normalize the serum level of soluble CD14. J Infect Dis *207*, 1221-1225.

Merayo-Chalico, J., Rajme-Lopez, S., Barrera-Vargas, A., Alcocer-Varela, J., Diaz-Zamudio, M., and Gomez-Martin, D. (2016). Lymphopenia and autoimmunity: A double-edged sword. Hum Immunol *77*, 921-929.

Metz, T.O., Qian, W.J., Jacobs, J.M., Gritsenko, M.A., Moore, R.J., Polpitiya, A.D., Monroe, M.E., Camp, D.G., 2nd, Mueller, P.W., and Smith, R.D. (2008). Application of proteomics in the discovery of candidate protein biomarkers in a diabetes autoantibody standardization program sample subset. Journal of proteome research *7*, 698-707.

Michalski, A., Cox, J., and Mann, M. (2011). More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. Journal of proteome research *10*, 1785-1793.

Molin, S., Merl, J., Dietrich, K.A., Regauer, M., Flaig, M., Letule, V., Saucke, T., Herzinger, T., Ruzicka, T., and Hauck, S.M. (2015). The hand eczema proteome: imbalance of epidermal barrier proteins. The British journal of dermatology *172*, 994-1001.

Mollah, Z.U., Quah, H.S., Graham, K.L., Jhala, G., Krishnamurthy, B., Francisca, M.D.J., Chee, J., Trivedi, P.M., Pappas, E.G., Mackin, L.*, et al.* (2017). Granzyme A-deficiency breaks immune tolerance and promotes Autoimmune Diabetes Through a Type I Interferon-Dependent Pathway. Diabetes.

Moltchanova, E.V., Schreier, N., Lammi, N., and Karvonen, M. (2009). Seasonal variation of diagnosis of Type 1 diabetes mellitus in children worldwide. Diabetic medicine : a journal of the British Diabetic Association *26*, 673-678.

Monteseirin, J., and Vega, A. (2008). Eosinophil cationic protein is not only a distinctive eosinophil protein. Thorax *63*, 185; author reply 185.

Monti, P., Scirpoli, M., Rigamonti, A., Mayr, A., Jaeger, A., Bonfanti, R., Chiumello, G., Ziegler, A.G., and Bonifacio, E. (2007). Evidence for in vivo primed and expanded autoreactive T cells as a specific feature of patients with type 1 diabetes. Journal of immunology *179*, 5785-5792.

Moran, A., Bundy, B., Becker, D.J., DiMeglio, L.A., Gitelman, S.E., Goland, R., Greenbaum, C.J., Herold, K.C., Marks, J.B., Raskin, P.*, et al.* (2013). Interleukin-1 antagonism in type 1 diabetes of recent onset: two multicentre, randomised, double-blind, placebo-controlled trials. Lancet *381*, 1905-1915.

References

Morgan, N.G., Leete, P., Foulis, A.K., and Richardson, S.J. (2014). Islet inflammation in human type 1 diabetes mellitus. IUBMB Life *66*, 723-734.

Mossina, A., Lukas, C., Merl-Pham, J., Uhl, F.E., Mutze, K., Schamberger, A., Staab-Weijnitz, C., Jia, J., Yildirim, A.O., Konigshoff, M.*, et al.* (2017). Cigarette smoke alters the secretome of lung epithelial cells. Proteomics *17*.

Moulder, R., Bhosale, S.D., Erkkila, T., Laajala, E., Salmi, J., Nguyen, E.V., Kallionpaa, H., Mykkanen, J., Vaha-Makila, M., Hyoty, H.*, et al.* (2015). Serum proteomes distinguish children developing type 1 diabetes in a cohort with HLA-conferred susceptibility. Diabetes *64*, 2265-2278.

Moulder, R., Bhosale, S.D., Lahesmaa, R., and Goodlett, D.R. (2017). The progress and potential of proteomic biomarkers for type 1 diabetes in children. Expert review of proteomics *14*, 31-41.

Muntel, J., Xuan, Y., Berger, S.T., Reiter, L., Bachur, R., Kentsis, A., and Steen, H. (2015). Advancing Urinary Protein Biomarker Discovery by Data-Independent Acquisition on a Quadrupole-Orbitrap Mass Spectrometer. Journal of proteome research *14*, 4752-4762.

Murri, M., Leiva, I., Gomez-Zumaquero, J.M., Tinahones, F.J., Cardona, F., Soriguer, F., and Queipo-Ortuno, M.I. (2013). Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. BMC medicine *11*, 46.

Mustonen, A., Knip, M., Huttunen, N.P., Puukka, R., Kaar, M.L., and Akerblom, H.K. (1984). Evidence of delayed beta-cell destruction in type 1 (insulin-dependent) diabetic patients with persisting complement-fixing cytoplasmic islet cell antibodies. Diabetologia *27*, 421-426.

Muto, M., Mori, M., Liu, J., Uzawa, A., Uchida, T., Masuda, H., Ohtani, R., Sugimoto, K., and Kuwabara, S. (2017). Serum soluble Talin-1 levels are elevated in patients with multiple sclerosis, reflecting its disease activity. Journal of neuroimmunology *305*, 131-134.

Naegele, M., Tillack, K., Reinhardt, S., Schippling, S., Martin, R., and Sospedra, M. (2012). Neutrophils in multiple sclerosis are characterized by a primed phenotype. Journal of neuroimmunology *242*, 60-71.

Nagra, R.M., Becher, B., Tourtellotte, W.W., Antel, J.P., Gold, D., Paladino, T., Smith, R.A., Nelson, J.R., and Reynolds, W.F. (1997). Immunohistochemical and genetic evidence of myeloperoxidase involvement in multiple sclerosis. Journal of neuroimmunology *78*, 97-107.

Nanjappa, V., Thomas, J.K., Marimuthu, A., Muthusamy, B., Radhakrishnan, A., Sharma, R., Ahmad Khan, A., Balakrishnan, L., Sahasrabuddhe, N.A., Kumar, S.*, et al.* (2014). Plasma Proteome Database as a resource for proteomics research: 2014 update. Nucleic acids research *42*, D959-965.

Nauseef, W.M., Olsson, I., and Arnljots, K. (1988). Biosynthesis and processing of myeloperoxidase--a marker for myeloid cell differentiation. European journal of haematology *40*, 97-110.

Navarro, P., Kuharev, J., Gillet, L.C., Bernhardt, O.M., MacLean, B., Rost, H.L., Tate, S.A., Tsou, C.C., Reiter, L., Distler, U.*, et al.* (2016). A multicenter study benchmarks software tools for label-free proteome quantification. Nature biotechnology *34*, 1130-1136.

Navarro, S., Aleu, J., Jiménez, M., Boix, E., Cuchillo, C.M., and Nogués, M.V. (2008). The cytotoxicity of eosinophil cationic protein/ribonuclease 3 on eukaryotic cell lines takes place through its aggregation on the cell membrane. Cellular and Molecular Life Sciences *65*, 324-337.

Negishi, K., Waldeck, N., Chandy, G., Buckingham, B., Kershnar, A., Fisher, L., Gupta, S., and Charles, M.A. (1986). Natural killer cell and islet killer cell activities in type 1 (insulin-dependent) diabetes. Diabetologia *29*, 352-357.

Nesvizhskii, A.I., Vitek, O., and Aebersold, R. (2007). Analysis and validation of proteomic data generated by tandem mass spectrometry. Nature methods *4*, 787-797.

Nielsen, L.B., Wang, C., Sorensen, K., Bang-Berthelsen, C.H., Hansen, L., Andersen, M.L., Hougaard, P., Juul, A., Zhang, C.Y., Pociot, F.*, et al.* (2012). Circulating levels of microRNA from children with newly diagnosed type 1 diabetes and healthy controls: evidence that miR-25 associates to residual beta-cell function and glycaemic control during disease progression. Experimental diabetes research *2012*, 896362.

Noble, J.A., Valdes, A.M., Cook, M., Klitz, W., Thomson, G., and Erlich, H.A. (1996). The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. Am J Hum Genet *59*, 1134-1148.

Noble, J.A., Valdes, A.M., Varney, M.D., Carlson, J.A., Moonsamy, P., Fear, A.L., Lane, J.A., Lavant, E., Rappner, R., Louey, A.*, et al.* (2010a). HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. Diabetes *59*, 2972-2979.

Noble, J.A., Valdes, A.M., Varney, M.D., Carlson, J.A., Moonsamy, P., Fear, A.L., Lane, J.A., Lavant, E., Rappner, R., Louey, A.*, et al.* (2010b). HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. Diabetes *59*, 2972-2979.

Norris, J.M., Beaty, B., Klingensmith, G., Yu, L., Hoffman, M., Chase, H.P., Erlich, H.A., Hamman, R.F., Eisenbarth, G.S., and Rewers, M. (1996). Lack of association between early exposure to cow's milk protein and beta-cell autoimmunity. Diabetes Autoimmunity Study in the Young (DAISY). Jama *276*, 609-614.

Norris, J.M., Yin, X., Lamb, M.M., Barriga, K., Seifert, J., Hoffman, M., Orton, H.D., Baron, A.E., Clare-Salzler, M., Chase, H.P.*, et al.* (2007). Omega-3 polyunsaturated fatty acid intake and islet autoimmunity in children at increased risk for type 1 diabetes. Jama *298*, 1420-1428.

Oikarinen, S., Martiskainen, M., Tauriainen, S., Huhtala, H., Ilonen, J., Veijola, R., Simell, O., Knip, M., and Hyöty, H. (2011). Enterovirus RNA in Blood Is Linked to the Development of Type 1 Diabetes. Diabetes *60*, 276-279.

Okada, S.S., de Oliveira, E.M., de Araujo, T.H., Rodrigues, M.R., Albuquerque, R.C., Mortara, R.A., Taniwaki, N.N., Nakaya, H.I., Campa, A., and Moreno, A.C. (2016). Myeloperoxidase in human peripheral blood lymphocytes: Production and subcellular localization. Cellular immunology *300*, 18-25.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., and Ahn, N.G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Molecular & cellular proteomics : MCP *4*, 1487-1502.

Oling, V., Reijonen, H., Simell, O., Knip, M., and Ilonen, J. (2012). Autoantigen-specific memory CD4+ T cells are prevalent early in progression to Type 1 diabetes. Cellular immunology *273*, 133-139.

Onengut-Gumuscu, S., Chen, W.M., Burren, O., Cooper, N.J., Quinlan, A.R., Mychaleckyj, J.C., Farber, E., Bonnie, J.K., Szpak, M., Schofield, E.*, et al.* (2015). Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nature genetics *47*, 381-386.

Ong, S.-E., and Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol *1*, 252-262.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Molecular & cellular proteomics : MCP *1*, 376-386.

References

Orban, T., Kis, J., Szereday, L., Engelmann, P., Farkas, K., Jalahej, H., and Treszl, A. (2007). Reduced CD4+ T-cell-specific gene expression in human type 1 diabetes mellitus. Journal of autoimmunity *28*, 177-187.

Oresic, M., Gopalacharyulu, P., Mykkanen, J., Lietzen, N., Makinen, M., Nygren, H., Simell, S., Simell, V., Hyoty, H., Veijola, R.*, et al.* (2013). Cord serum lipidome in prediction of islet autoimmunity and type 1 diabetes. Diabetes *62*, 3268-3274.

Oresic, M., Simell, S., Sysi-Aho, M., Nanto-Salonen, K., Seppanen-Laakso, T., Parikka, V., Katajamaa, M., Hekkala, A., Mattila, I., Keskinen, P.*, et al.* (2008). Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. The Journal of experimental medicine *205*, 2975-2984.

Overgaard, A.J., Kaur, S., and Pociot, F. (2016). Metabolomic Biomarkers in the Progression to Type 1 Diabetes. Current diabetes reports *16*, 127.

Padmos, R.C., Schloot, N.C., Beyan, H., Ruwhof, C., Staal, F.J., de Ridder, D., Aanstoot, H.J., Lam-Tse, W.K., de Wit, H., de Herder, C.*, et al.* (2008). Distinct monocyte gene-expression profiles in autoimmune diabetes. Diabetes *57*, 2768-2773.

Palmer, J., Asplin, C., Clemons, P., Lyen, K., Tatpati, O., Raghu, P., and Paquette, T. (1983). Insulin antibodies in insulin-dependent diabetics before insulin treatment. Science (New York, NY) *222*, 1337-1339.

Parikka, V., Nanto-Salonen, K., Saarinen, M., Simell, T., Ilonen, J., Hyoty, H., Veijola, R., Knip, M., and Simell, O. (2012). Early seroconversion and rapidly increasing autoantibody concentrations predict prepubertal manifestation of type 1 diabetes in children at genetic risk. Diabetologia *55*, 1926-1936.

Patterson, C.C., Dahlquist, G.G., Gyurus, E., Green, A., and Soltesz, G. (2009). Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. Lancet *373*, 2027-2033.

Paul, D.S., Teschendorff, A.E., Dang, M.A., Lowe, R., Hawa, M.I., Ecker, S., Beyan, H., Cunningham, S., Fouts, A.R., Ramelius, A.*, et al.* (2016). Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. Nature communications *7*, 13555.

Pavon, E.J., Garcia-Rodriguez, S., Zumaquero, E., Perandres-Lopez, R., Rosal-Vela, A., Lario, A., Longobardo, V., Carrascal, M., Abian, J., Callejas-Rubio, J.L.*, et al.* (2012). Increased expression and phosphorylation of the two S100A9 isoforms in mononuclear cells from patients with systemic lupus erythematosus: a proteomic signature for circulating low-density granulocytes. Journal of proteomics *75*, 1778-1791.

Payton, M.A., Hawkes, C.J., and Christie, M.R. (1995). Relationship of the 37,000- and 40,000-M(r) tryptic fragments of islet antigens in insulin-dependent diabetes to the protein tyrosine phosphatase-like molecule IA-2 (ICA512). The Journal of clinical investigation *96*, 1506-1511.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551-3567.

Perry, R.H., Cooks, R.G., and Noll, R.J. (2008). Orbitrap mass spectrometry: instrumentation, ion motion and applications. Mass Spectrom Rev *27*, 661-699.

Pescovitz, M.D., Greenbaum, C.J., Bundy, B., Becker, D.J., Gitelman, S.E., Goland, R., Gottlieb, P.A., Marks, J.B., Moran, A., Raskin, P.*, et al.* (2014). B-lymphocyte depletion with rituximab and beta-cell function: two-year results. Diabetes Care *37*, 453-459.

Petronis, A. (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature *465*, 721-727.

Pflueger, M., Seppanen-Laakso, T., Suortti, T., Hyotylainen, T., Achenbach, P., Bonifacio, E., Oresic, M., and Ziegler, A.G. (2011). Age- and islet autoimmunity-associated differences in amino acid and lipid metabolites in children at risk for type 1 diabetes. Diabetes *60*, 2740-2747.

Pham, C.T.N. (2006). Neutrophil serine proteases: specific regulators of inflammation. Nature reviews Immunology *6*, 541-550.

Phillips, J.M., Parish, N.M., Raine, T., Bland, C., Sawyer, Y., De La Peña, H., and Cooke, A. (2009). Type 1 Diabetes Development Requires Both CD4+ and CD8+ T cells and Can Be Reversed by Non-Depleting Antibodies Targeting Both T Cell Populations. The review of diabetic studies : RDS *6*, 97-103.

Plesner, A., Greenbaum, C.J., Gaur, L.K., Ernst, R.K., and Lernmark, A. (2002). Macrophages from high-risk HLA-DQB1*0201/*0302 type 1 diabetes mellitus patients are hypersensitive to lipopolysaccharide stimulation. Scandinavian journal of immunology *56*, 522-529.

Pociot, F., Akolkar, B., Concannon, P., Erlich, H.A., Julier, C., Morahan, G., Nierras, C.R., Todd, J.A., Rich, S.S., and Nerup, J. (2010). Genetics of Type 1 Diabetes: What's Next? Diabetes *59*, 1561-1571.

Pociot, F., and Lernmark, A. (2016). Genetic risk factors for type 1 diabetes. Lancet *387*, 2331-2339.

Podar, T., Solntsev, A., Karvonen, M., Padaiga, Z., Brigis, G., Urbonaite, B., Viik-Kajander, M., Reunanen, A., and Tuomilehto, J. (2001). Increasing incidence of childhood-onset type I diabetes in 3 Baltic countries and Finland 1983-1998. Diabetologia *44 Suppl 3*, B17-20.

Podrez, E.A., Schmitt, D., Hoff, H.F., and Hazen, S.L. (1999). Myeloperoxidase-generated reactive nitrogen species convert LDL into an atherogenic form in vitro. The Journal of clinical investigation *103*, 1547-1560.

Poy, M.N., Eliasson, L., Krutzfeldt, J., Kuwajima, S., Ma, X., Macdonald, P.E., Pfeffer, S., Tuschl, T., Rajewsky, N., Rorsman, P.*, et al.* (2004). A pancreatic islet-specific microRNA regulates insulin secretion. Nature *432*, 226-230.

Pozzilli, P., Maddaloni, E., and Buzzetti, R. (2015). Combination immunotherapies for type 1 diabetes mellitus. Nature reviews Endocrinology *11*, 289-297.

Pratt, J.M., Simpson, D.M., Doherty, M.K., Rivers, J., Gaskell, S.J., and Beynon, R.J. (2006). Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. Nat Protocols *1*, 1029-1043.

Pugliese, A. (2017). Autoreactive T cells in type 1 diabetes. The Journal of clinical investigation *127*, 2881-2891.

Pullakhandam, R., Palika, R., Ghosh, S., and Reddy, G.B. (2012). Contrasting effects of type 2 and type 1 diabetes on plasma RBP4 levels: the significance of transthyretin. IUBMB Life *64*, 975-982.

Qin, H., Lee, I.F., Panagiotopoulos, C., Wang, X., Chu, A.D., Utz, P.J., Priatel, J.J., and Tan, R. (2011). Natural killer cells from children with type 1 diabetes have defects in NKG2D-dependent function and signaling. Diabetes *60*, 857-866.

Qin, J., Fu, S., Speake, C., Greenbaum, C.J., and Odegard, J.M. (2016). NETosis-associated serum biomarkers are reduced in type 1 diabetes in association with neutrophil count. Clinical and experimental immunology *184*, 318-322.

Raab, J., Giannopoulou, E.Z., Schneider, S., Warncke, K., Krasmann, M., Winkler, C., and Ziegler, A.G. (2014). Prevalence of vitamin D deficiency in pre-type 1 diabetes and its association with disease progression. Diabetologia *57*, 902-908.

References

Radich, J.P., Mao, M., Stepaniants, S., Biery, M., Castle, J., Ward, T., Schimmack, G., Kobayashi, S., Carleton, M., Lampe, J.*, et al.* (2004). Individual-specific variation of gene expression in peripheral blood leukocytes. Genomics *83*, 980-988.

Radulovic, D., Jelveh, S., Ryu, S., Hamilton, T.G., Foss, E., Mao, Y., and Emili, A. (2004). Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. Molecular & cellular proteomics : MCP *3*, 984-997.

Rassi, D.M., Junta, C.M., Fachin, A.L., Sandrin-Garcia, P., Mello, S., Marques, M.M., Fernandes, A.P., Foss-Freitas, M.C., Foss, M.C., Sakamoto-Hojo, E.T.*, et al.* (2006). Metabolism genes are among the differentially expressed ones observed in lymphomononuclear cells of recently diagnosed type 1 diabetes mellitus patients. Annals of the New York Academy of Sciences *1079*, 171-176.

Refai, E., Dekki, N., Yang, S.N., Imreh, G., Cabrera, O., Yu, L., Yang, G., Norgren, S., Rossner, S.M., Inverardi, L.*, et al.* (2005). Transthyretin constitutes a functional component in pancreatic beta-cell stimulus-secretion coupling. Proceedings of the National Academy of Sciences of the United States of America *102*, 17020-17025.

Regnéll, S.E., and Lernmark, Å. (2013). The environment and the origins of islet autoimmunity and Type 1 diabetes. Diabetic medicine : a journal of the British Diabetic Association *30*, 155-160.

Reinders, J., and Sickmann, A. (2009). Proteomics: Methods and Protocols (Humana Press).

Reizis, B., Bunin, A., Ghosh, H.S., Lewis, K.L., and Sisirak, V. (2011). Plasmacytoid dendritic cells: recent progress and open questions. Annual review of immunology *29*, 163-183.

Rekers, N.V., von Herrath, M.G., and Wesley, J.D. (2015). Immunotherapies and immune biomarkers in Type 1 diabetes: A partnership for success. Clinical immunology *161*, 37-43.

Resic Lindehammer, S., Honkanen, H., Nix, W.A., Oikarinen, M., Lynch, K.F., Jonsson, I., Marsal, K., Oberste, S., Hyoty, H., and Lernmark, A. (2012). Seroconversion to islet autoantibodies after enterovirus infection in early pregnancy. Viral Immunol *25*, 254-261.

Rewers, M., and Ludvigsson, J. (2016). Environmental risk factors for type 1 diabetes. The Lancet *387*, 2340-2348.

Reynier, F., Pachot, A., Paye, M., Xu, Q., Turrel-Davin, F., Petit, F., Hot, A., Auffray, C., Bendelac, N., Nicolino, M.*, et al.* (2010). Specific gene expression signature associated with development of autoimmune type-I diabetes using whole-blood microarray analysis. Genes and immunity *11*, 269-278.

Richardson, S.J., Rodriguez-Calvo, T., Gerling, I.C., Mathews, C.E., Kaddis, J.S., Russell, M.A., Zeissler, M., Leete, P., Krogvold, L., Dahl-Jorgensen, K.*, et al.* (2016). Islet cell hyperexpression of HLA class I antigens: a defining feature in type 1 diabetes. Diabetologia *59*, 2448-2458.

Richardson, S.J., Willcox, A., Bone, A.J., Foulis, A.K., and Morgan, N.G. (2009). The prevalence of enteroviral capsid protein vp1 immunostaining in pancreatic islets in human type 1 diabetes. Diabetologia *52*, 1143-1151.

Roberts, K.G., and Mullighan, C.G. (2015). Genomics in acute lymphoblastic leukaemia: insights and treatment implications. Nature reviews Clinical oncology *12*, 344-357.

Robles, D.T., Eisenbarth, G.S., Wang, T., Erlich, H.A., Bugawan, T.L., Babu, S.R., Barriga, K., Norris, J.M., Hoffman, M., Klingensmith, G.*, et al.* (2002). Identification of Children with Early Onset and High Incidence of Anti-islet Autoantibodies. Clinical immunology *102*, 217-224.

Rodacki, M., Svoren, B., Butty, V., Besse, W., Laffel, L., Benoist, C., and Mathis, D. (2007). Altered natural killer cells in type 1 diabetic patients. Diabetes *56*, 177-185.

Rodriguez-Calvo, T., Ekwall, O., Amirian, N., Zapardiel-Gonzalo, J., and von Herrath, M.G. (2014). Increased immune cell infiltration of the exocrine pancreas: a possible contribution to the pathogenesis of type 1 diabetes. Diabetes *63*, 3880-3890.

Roep, B.O. (2003). The role of T-cells in the pathogenesis of Type 1 diabetes: from cause to cure. Diabetologia *46*, 305-321.

Roep, B.O., and Peakman, M. (2010). Surrogate end points in the design of immunotherapy trials: emerging lessons from type 1 diabetes. Nature reviews Immunology *10*, 145-152.

Roep, B.O., and Peakman, M. (2011). Diabetogenic T lymphocytes in human Type 1 diabetes. Current opinion in immunology *23*, 746-753.

Roep, B.O., and Tree, T.I.M. (2014). Immune modulation in humans: implications for type 1 diabetes mellitus. Nature reviews Endocrinology *10*, 229-242.

Roepstorff, P., and Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomedical mass spectrometry *11*, 601.

Roivainen, M., Rasilainen, S., Ylipaasto, P., Nissinen, R., Ustinov, J., Bouwens, L., Eizirik, D.L., Hovi, T., and Otonkoski, T. (2000). Mechanisms of coxsackievirus-induced damage to human pancreatic beta-cells. J Clin Endocrinol Metab *85*, 432-440.

Roncarolo, M.-G., and Battaglia, M. (2007). Regulatory T-cell immunotherapy for tolerance to self antigens and alloantigens in humans. Nature reviews Immunology *7*, 585-598.

Roncarolo, M.G., Battaglia, M., and Gregori, S. (2003). The role of interleukin 10 in the control of autoimmunity. Journal of autoimmunity *20*, 269-272.

Rosenberg, H.F. (1995). Recombinant human eosinophil cationic protein. Ribonuclease activity is not essential for cytotoxicity. J Biol Chem *270*, 7876-7881.

Rosenberger, G., Bludau, I., Schmitt, U., Heusel, M., Hunter, C.L., Liu, Y., MacCoss, M.J., MacLean, B.X., Nesvizhskii, A.I., Pedrioli, P.G.A.*, et al.* (2017). Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. Nature methods *14*, 921-927.

Rosenberger, G., Koh, C.C., Guo, T., Rost, H.L., Kouvonen, P., Collins, B.C., Heusel, M., Liu, Y., Caron, E., Vichalkovski, A.*, et al.* (2014). A repository of assays to quantify 10,000 human proteins by SWATH-MS. Scientific data *1*, 140031.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S.*, et al.* (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Molecular & cellular proteomics : MCP *3*, 1154-1169.

Rost, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S.M., Schubert, O.T., Wolski, W., Collins, B.C., Malmstrom, J., Malmstrom, L.*, et al.* (2014a). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nature biotechnology *32*, 219-223.

Rost, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S.M., Schubert, O.T., Wolski, W., Collins, B.C., Malmstrom, J., Malmstrom, L.*, et al.* (2014b). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotech *32*, 219-223.

Rowe, P., Wasserfall, C., Croker, B., Campbell-Thompson, M., Pugliese, A., Atkinson, M., and Schatz, D. (2013). Increased complement activation in human type 1 diabetes pancreata. Diabetes Care *36*, 3815-3817.

References

Rowell, J., Thompson, A.J., Guyton, J.R., Lao, X.Q., McHutchison, J.G., McCarthy, J.J., and Patel, K. (2012). Serum apolipoprotein C-III is independently associated with chronic hepatitis C infection and advanced fibrosis. Hepatology international *6*, 475-481.

Ryden, A., Ludvigsson, J., Fredrikson, M., and Faresjo, M. (2014). General immune dampening is associated with disturbed metabolism at diagnosis of type 1 diabetes. Pediatric research *75*, 45-50.

Samandari, N., Mirza, A.H., Nielsen, L.B., Kaur, S., Hougaard, P., Fredheim, S., Mortensen, H.B., and Pociot, F. (2017). Circulating microRNA levels predict residual beta cell function and glycaemic control in children with type 1 diabetes mellitus. Diabetologia *60*, 354-363.

Sarween, N., Chodos, A., Raykundalia, C., Khan, M., Abbas, A.K., and Walker, L.S. (2004). CD4+CD25+ cells controlling a pathogenic CD4 response inhibit cytokine differentiation, CXCR-3 expression, and tissue invasion. Journal of immunology *173*, 2942-2951.

Scheltema, R.A., Hauschild, J.-P., Lange, O., Hornburg, D., Denisov, E., Damoc, E., Kuehn, A., Makarov, A., and Mann, M. (2014). The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer. Molecular & cellular proteomics : MCP *13*, 3698-3708.

Schmidt, A., Kellermann, J., and Lottspeich, F. (2005). A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics *5*, 4-15.

Schmidt, M. (1902). Über die Beziehung der Langerhans'schen Inseln des Pankreas zum Diabetes mellitus. Munch Med Wochenschr *49*, 51–54.

Schneider, A., Rieck, M., Sanda, S., Pihoker, C., Greenbaum, C., and Buckner, J.H. (2008). The effector T cells of diabetic subjects are resistant to regulation via CD4+ FOXP3+ regulatory T cells. Journal of immunology *181*, 7350-7355.

Schwenk, J.M., Omenn, G.S., Sun, Z., Campbell, D.S., Baker, M.S., Overall, C.M., Aebersold, R., Moritz, R.L., and Deutsch, E.W. (2017). The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. Journal of proteome research.

Segal, A.W. (2005). How neutrophils kill microbes. Annual review of immunology *23*, 197-223.

Serr, I., Furst, R.W., Ott, V.B., Scherm, M.G., Nikolaev, A., Gokmen, F., Kalin, S., Zillmer, S., Bunk, M., Weigmann, B.*, et al.* (2016). miRNA92a targets KLF2 and the phosphatase PTEN signaling to promote human T follicular helper precursors in T1D islet autoimmunity. Proceedings of the National Academy of Sciences of the United States of America *113*, E6659-e6668.

Sherwin, R., and Jastreboff, A.M. (2012). Year in diabetes 2012: The diabetes tsunami. J Clin Endocrinol Metab *97*, 4293-4301.

Sheu, T.-T., Chiang, B.-L., Yen, J.-H., and Lin, W.-C. (2014). Premature CD4(+) T Cell Aging and Its Contribution to Lymphopenia-Induced Proliferation of Memory Cells in Autoimmune-Prone Non-Obese Diabetic Mice. PloS one *9*, e89379.

Shi, Y., Feng, Y., Kang, J., Liu, C., Li, Z., Li, D., Cao, W., Qiu, J., Guo, Z., Bi, E.*, et al.* (2007). Critical regulation of CD4+ T cell survival and autoimmunity by beta-arrestin 1. Nature immunology *8*, 817-824.

Shive, C.L., Jiang, W., Anthony, D.D., and Lederman, M.M. (2015). Soluble CD14 is a nonspecific marker of monocyte activation. AIDS (London, England) *29*, 1263-1265.

Shizuru, J.A., Taylor-Edwards, C., Banks, B.A., Gregory, A.K., and Fathman, C.G. (1988). Immunotherapy of the nonobese diabetic mouse: treatment with an antibody to T-helper lymphocytes. Science (New York, NY) *240*, 659-662.

Sibley, R.K., Sutherland, D.E., Goetz, F., and Michael, A.F. (1985). Recurrent diabetes mellitus in the pancreas iso- and allograft. A light and electron microscopic and immunohistochemical analysis of four cases. Laboratory investigation; a journal of technical methods and pathology *53*, 132-144.

Simmons, D.L., Tan, S., Tenen, D.G., Nicholson-Weller, A., and Seed, B. (1989). Monocyte antigen CD14 is a phospholipid anchored membrane protein. Blood *73*, 284-289.

Simmons, K.M., Gottlieb, P.A., and Michels, A.W. (2016). Immune Intervention and Preservation of Pancreatic Beta Cell Function in Type 1 Diabetes. Current diabetes reports *16*, 97.

Simmons, K.M., and Michels, A.W. (2015). Type 1 diabetes: A predictable disease. World Journal of Diabetes *6*, 380-390.

Singh, I.P., Chopra, A.K., Coppenhaver, D.H., Ananatharamaiah, G.M., and Baron, S. (1999). Lipoproteins account for part of the broad non-specific antiviral activity of human serum. Antiviral Res *42*, 211-218.

Skeel, A., Yoshimura, T., Showalter, S.D., Tanaka, S., Appella, E., and Leonard, E.J. (1991). Macrophage stimulating protein: purification, partial amino acid sequence, and cellular activity. The Journal of experimental medicine *173*, 1227-1234.

Skog, O., Korsgren, S., Melhus, A., and Korsgren, O. (2013). Revisiting the notion of type 1 diabetes being a T-cell-mediated autoimmune disease. Current opinion in endocrinology, diabetes, and obesity *20*, 118-123.

Skowera, A., Ladell, K., McLaren, J.E., Dolton, G., Matthews, K.K., Gostick, E., Kronenberg-Versteeg, D., Eichmann, M., Knight, R.R., Heck, S.*, et al.* (2015). beta-cell-specific CD8 T cell phenotype in type 1 diabetes reflects chronic autoantigen exposure. Diabetes *64*, 916-925.

Skyler, J.S. (2015). Immune therapy for treating type 1 diabetes: challenging existing paradigms. The Journal of clinical investigation *125*, 94-96.

Skyler, J.S., Krischer, J.P., Wolfsdorf, J., Cowie, C., Palmer, J.P., Greenbaum, C., Cuthbertson, D., Rafkin-Mervis, L.E., Chase, H.P., and Leschek, E. (2005). Effects of oral insulin in relatives of patients with type 1 diabetes: The Diabetes Prevention Trial--Type 1. Diabetes Care *28*, 1068-1076.

Smith, C.M., Wilson, N.S., Waithman, J., Villadangos, J.A., Carbone, F.R., Heath, W.R., and Belz, G.T. (2004). Cognate CD4(+) T cell licensing of dendritic cells in CD8(+) T cell immunity. Nature immunology *5*, 1143-1148.

Sojka, D.K., Huang, Y.-H., and Fowell, D.J. (2008). Mechanisms of regulatory T-cell suppression – a diverse arsenal for a moving target. Immunology *124*, 13-22.

Sorkio, S., Cuthbertson, D., Barlund, S., Reunanen, A., Nucci, A.M., Berseth, C.L., Koski, K., Ormisson, A., Savilahti, E., Uusitalo, U.*, et al.* (2010). Breastfeeding patterns of mothers with type 1 diabetes: results from an infant feeding trial. Diabetes Metab Res Rev *26*, 206-211.

Stechova, K., Kolar, M., Blatny, R., Halbhuber, Z., Vcelakova, J., Hubackova, M., Petruzelkova, L., Sumnik, Z., Obermannova, B., Pithova, P.*, et al.* (2012). Healthy first-degree relatives of patients with type 1 diabetes exhibit significant differences in basal gene expression pattern of immunocompetent cells compared to controls: expression pattern as predeterminant of autoimmune diabetes. Scandinavian journal of immunology *75*, 210-219.

Steck, A.K., Johnson, K., Barriga, K.J., Miao, D., Yu, L., Hutton, J.C., Eisenbarth, G.S., and Rewers, M.J. (2011). Age of islet autoantibody appearance and mean levels of insulin, but not GAD or IA-2 autoantibodies, predict age of diagnosis of type 1 diabetes: diabetes autoimmunity study in the young. Diabetes Care *34*, 1397-1399.

References

Steen, H., and Mann, M. (2004). The abc's (and xyz's) of peptide sequencing. Nat Rev Mol Cell Biol *5*, 699-711.

Stefan, M., Zhang, W., Concepcion, E., Yi, Z., and Tomer, Y. (2014). DNA methylation profiles in type 1 diabetes twins point to strong epigenetic effects on etiology. Journal of autoimmunity *50*, 33-37.

Stene, L.C., Oikarinen, S., Hyöty, H., Barriga, K.J., Norris, J.M., Klingensmith, G., Hutton, J.C., Erlich, H.A., Eisenbarth, G.S., and Rewers, M. (2010). Enterovirus Infection and Progression From Islet Autoimmunity to Type 1 Diabetes: The Diabetes and Autoimmunity Study in the Young (DAISY). Diabetes *59*, 3174-3180.

Stene, L.C., and Rewers, M. (2012). Immunology in the clinic review series; focus on type 1 diabetes and viruses: the enterovirus link to type 1 diabetes: critical review of human studies. Clinical and experimental immunology *168*, 12-23.

Stenger, S., Hanson, D.A., Teitelbaum, R., Dewan, P., Niazi, K.R., Froelich, C.J., Ganz, T., Thoma-Uszynski, S., Melian, A., Bogdan, C.*, et al.* (1998). An antimicrobial activity of cytolytic T cells mediated by granulysin. Science (New York, NY) *282*, 121-125.

Stevens, T.J., and Arkin, I.T. (2000). Do more complex organisms have a greater proportion of membrane proteins in their genomes? Proteins *39*, 417-420.

Stoikou, M., Grimolizzi, F., Giaglis, S., Schafer, G., van Breda, S.V., Hoesli, I.M., Lapaire, O., Huhn, E.A., Hasler, P., Rossi, S.W.*, et al.* (2017). Gestational Diabetes Mellitus Is Associated with Altered Neutrophil Activity. Frontiers in immunology *8*, 702.

Storey, J.D. (2002). A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology) *64*, 479-498.

Størling, J., and Pociot, F. (2017). Type 1 Diabetes Candidate Genes Linked to Pancreatic Islet Cell Inflammation and Beta-Cell Apoptosis. Genes *8*, 72.

Strimbu, K., and Tavel, J.A. (2010). What are Biomarkers? Current opinion in HIV and AIDS *5*, 463-466.

Sumpter, K.M., Adhikari, S., Grishman, E.K., and White, P.C. (2011). Preliminary studies related to anti-interleukin-1beta therapy in children with newly diagnosed type 1 diabetes. Pediatric diabetes *12*, 656-667.

Sundsmo, J.S., Papin, R.A., Wood, L., Hirani, S., Waldeck, N., Buckingham, B., Kershnar, A., Ascher, M., and Charles, M.A. (1985). Complement activation in type 1 human diabetes. Clinical immunology and immunopathology *35*, 211-225.

Sur, S., Glitz, D.G., Kita, H., Kujawa, S.M., Peterson, E.A., Weiler, D.A., Kephart, G.M., Wagner, J.M., George, T.J., Gleich, G.J.*, et al.* (1999). Localization of eosinophil-derived neurotoxin and eosinophil cationic protein in neutrophilic leukocytes. International archives of allergy and immunology *118*, 255-258.

Swiecki, M., and Colonna, M. (2010). Unraveling the functions of plasmacytoid dendritic cells during viral infections, autoimmunity, and tolerance. Immunol Rev *234*, 142-162.

Tabb, D.L., Vega-Montoto, L., Rudnick, P.A., Variyath, A.M., Ham, A.J., Bunk, D.M., Kilpatrick, L.E., Billheimer, D.D., Blackman, R.K., Cardasis, H.L.*, et al.* (2010). Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. Journal of proteome research *9*, 761-776.

Taflin, D.C., Ward, T.L., and Davis, E.J. (1989). Electrified droplet fission and the Rayleigh limit. Langmuir *5*, 376-384.

Tang, Q., Adams, J.Y., Tooley, A.J., Bi, M., Fife, B.T., Serra, P., Santamaria, P., Locksley, R.M., Krummel, M.F., and Bluestone, J.A. (2006). Visualizing regulatory T cell control of autoimmune responses in nonobese diabetic mice. Nature immunology *7*, 83-92.

Thomas, H.E., Trapani, J.A., and Kay, T.W. (2010). The role of perforin and granzymes in diabetes. Cell death and differentiation *17*, 577-585.

Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Analytical chemistry *75*, 1895-1904.

Thumer, L., Adler, K., Bonifacio, E., Hofmann, F., Keller, M., Milz, C., Munte, A., and Ziegler, A.G. (2010). German new onset diabetes in the young incident cohort study: DiMelli study design and first-year results. The review of diabetic studies : RDS *7*, 202-208.

Tian, Q., Stepaniants, S.B., Mao, M., Weng, L., Feetham, M.C., Doyle, M.J., Yi, E.C., Dai, H., Thorsson, V., Eng, J.*, et al.* (2004). Integrated genomic and proteomic analyses of gene expression in Mammalian cells. Molecular & cellular proteomics : MCP *3*, 960-969.

Todd, J.A., Acha-Orbea, H., Bell, J.I., Chao, N., Fronek, Z., Jacob, C.O., McDermott, M., Sinha, A.A., Timmerman, L., Steinman, L.*, et al.* (1988). A molecular basis for MHC class II--associated autoimmunity. Science (New York, NY) *240*, 1003-1009.

Todd, J.A., Walker, N.M., Cooper, J.D., Smyth, D.J., Downes, K., Plagnol, V., Bailey, R., Nejentsev, S., Field, S.F., Payne, F.*, et al.* (2007). Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nature genetics *39*, 857-864.

Tracy, S., Drescher, K.M., and Chapman, N.M. (2011). Enteroviruses and type 1 diabetes. Diabetes Metab Res Rev *27*, 820-823.

Trajkovski, M., Mziaut, H., Schubert, S., Kalaidzidis, Y., Altkruger, A., and Solimena, M. (2008). Regulation of insulin granule turnover in pancreatic beta-cells by cleaved ICA512. J Biol Chem *283*, 33719-33729.

Tsou, C.-C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A.-C., and Nesvizhskii, A.I. (2015). DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat Meth *12*, 258-264.

Tsou, C.C., Tsai, C.F., Teo, G.C., Chen, Y.J., and Nesvizhskii, A.I. (2016). Untargeted, spectral library-free analysis of data-independent acquisition proteomics data generated using Orbitrap mass spectrometers. Proteomics *16*, 2257-2271.

Turley, S., Poirot, L., Hattori, M., Benoist, C., and Mathis, D. (2003). Physiological beta cell death triggers priming of self-reactive T cells by dendritic cells in a type-1 diabetes model. The Journal of experimental medicine *198*, 1527-1537.

Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R., and Gordon, J.I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Science translational medicine *1*, 6ra14.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America *98*, 5116-5121.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nature methods *13*, 731-740.

References

Ulyanova, T., Blasioli, J., Woodford-Thomas, T.A., and Thomas, M.L. (1999). The sialoadhesin CD33 is a myeloid-specific inhibitory receptor. European journal of immunology *29*, 3440-3449.

Urban, C.F., Ermert, D., Schmid, M., Abu-Abed, U., Goosmann, C., Nacken, W., Brinkmann, V., Jungblut, P.R., and Zychlinsky, A. (2009). Neutrophil extracellular traps contain calprotectin, a cytosolic protein complex involved in host defense against Candida albicans. PLoS Pathog *5*, e1000639.

Vaarala, O., Paronen, J., Otonkoski, T., and Akerblom, H.K. (1998). Cow milk feeding induces antibodies to insulin in children--a link between cow milk and insulin-dependent diabetes mellitus? Scandinavian journal of immunology *47*, 131-135.

Valle, A., Giamporcaro, G.M., Scavini, M., Stabilini, A., Grogan, P., Bianconi, E., Sebastiani, G., Masini, M., Maugeri, N., Porretti, L.*, et al.* (2013). Reduction of circulating neutrophils precedes and accompanies type 1 diabetes. Diabetes *62*, 2072-2077.

van Belle, T.L., Coppieters, K.T., and von Herrath, M.G. (2011). Type 1 diabetes: etiology, immunology, and therapeutic strategies. Physiological reviews *91*, 79-118.

van Lummel, M., Duinkerken, G., van Veelen, P.A., de Ru, A., Cordfunke, R., Zaldumbide, A., Gomez-Tourino, I., Arif, S., Peakman, M., Drijfhout, J.W.*, et al.* (2014). Posttranslational modification of HLA-DQ binding islet autoantigens in type 1 diabetes. Diabetes *63*, 237-247.

van Lummel, M., van Veelen, P.A., Zaldumbide, A., de Ru, A., Janssen, G.M., Moustakas, A.K., Papadopoulos, G.K., Drijfhout, J.W., Roep, B.O., and Koning, F. (2012). Type 1 diabetes-associated HLA-DQ8 transdimer accommodates a unique peptide repertoire. J Biol Chem *287*, 9514-9524.

Vehik, K., Hamman, R.F., Lezotte, D., Norris, J.M., Klingensmith, G.J., Rewers, M., and Dabelea, D. (2008). Trends in High-Risk HLA Susceptibility Genes Among Colorado Youth With Type 1 Diabetes. Diabetes Care *31*, 1392-1396.

Venable, J.D., Dong, M.Q., Wohlschlegel, J., Dillin, A., and Yates, J.R. (2004). Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nature methods *1*, 39-45.

Viisanen, T., Ihantola, E.L., Nanto-Salonen, K., Hyoty, H., Nurminen, N., Selvenius, J., Juutilainen, A., Moilanen, L., Pihlajamaki, J., Veijola, R.*, et al.* (2017). Circulating CXCR5+PD-1+ICOS+ Follicular T Helper Cells Are Increased Close to the Diagnosis of Type 1 Diabetes in Children With Multiple Autoantibodies. Diabetes *66*, 437-447.

Villanueva, E., Yalavarthi, S., Berthier, C.C., Hodgin, J.B., Khandpur, R., Lin, A.M., Rubin, C.J., Zhao, W., Olsen, S.H., Klinker, M.*, et al.* (2011). Netting neutrophils induce endothelial damage, infiltrate tissues, and expose immunostimulatory molecules in systemic lupus erythematosus. Journal of immunology *187*, 538-552.

Virtanen, S.M., Kenward, M.G., Erkkola, M., Kautiainen, S., Kronberg-Kippila, C., Hakulinen, T., Ahonen, S., Uusitalo, L., Niinisto, S., Veijola, R.*, et al.* (2006). Age at introduction of new foods and advanced beta cell autoimmunity in young children with HLA-conferred susceptibility to type 1 diabetes. Diabetologia *49*, 1512-1521.

Virtanen, S.M., Nevalainen, J., Kronberg-Kippila, C., Ahonen, S., Tapanainen, H., Uusitalo, L., Takkinen, H.M., Niinisto, S., Ovaskainen, M.L., Kenward, M.G.*, et al.* (2012). Food consumption and advanced beta cell autoimmunity in young children with HLA-conferred susceptibility to type 1 diabetes: a nested case-control design. The American journal of clinical nutrition *95*, 471-478.

Virtanen, S.M., Rasanen, L., Ylonen, K., Aro, A., Clayton, D., Langholz, B., Pitkaniemi, J., Savilahti, E., Lounamaa, R., Tuomilehto, J.*, et al.* (1993). Early introduction of dairy products associated with increased risk of IDDM in Finnish children. The Childhood in Diabetes in Finland Study Group. Diabetes *42*, 1786-1790.

von Hagen, J. (2011). Proteomics Sample Preparation (Wiley).

von Herrath, M. (2009). Diabetes: A virus-gene collaboration. Nature *459*, 518-519.

von Meyenburg, H. (1940). Über "Insulitis" bei Diabetes. Schweiz Med Wochenschr *21*, 554-557.

von Toerne, C., Laimighofer, M., Achenbach, P., Beyerlein, A., de Las Heras Gala, T., Krumsiek, J., Theis, F.J., Ziegler, A.G., and Hauck, S.M. (2017). Peptide serum markers in islet autoantibody-positive children. Diabetologia *60*, 287-295.

Voskoboinik, I., Whisstock, J.C., and Trapani, J.A. (2015). Perforin and granzymes: function, dysfunction and human pathology. Nature reviews Immunology *15*, 388-400.

Wallberg, M., and Cooke, A. (2013). Immune mechanisms in type 1 diabetes. Trends Immunol *34*, 583-591.

Wang, H., Kadlecek, T.A., Au-Yeung, B.B., Goodfellow, H.E., Hsu, L.Y., Freedman, T.S., and Weiss, A. (2010). ZAP-70: an essential kinase in T-cell signaling. Cold Spring Harbor perspectives in biology *2*, a002279.

Wang, Y., Xiao, Y., Zhong, L., Ye, D., Zhang, J., Tu, Y., Bornstein, S.R., Zhou, Z., Lam, K.S., and Xu, A. (2014). Increased neutrophil elastase and proteinase 3 and augmented NETosis are closely associated with beta-cell autoimmunity in patients with type 1 diabetes. Diabetes *63*, 4239-4248.

Wang, Z., Xie, Z., Lu, Q., Chang, C., and Zhou, Z. (2017). Beyond Genetics: What Causes Type 1 Diabetes. Clin Rev Allergy Immunol *52*, 273-286.

Warncke, K., Krasmann, M., Puff, R., Dunstheimer, D., Ziegler, A.G., and Beyerlein, A. (2013). Does diabetes appear in distinct phenotypes in young people? Results of the diabetes mellitus incidence Cohort Registry (DiMelli). PloS one *8*, e74339.

Washburn, M.P., Wolters, D., and Yates, J.R., 3rd (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature biotechnology *19*, 242-247.

Weber, K.S., Raab, J., Haupt, F., Aschemeier, B., Wosch, A., Ried, C., Kordonouri, O., Ziegler, A.G., and Winkler, C. (2015). Evaluating the diet of children at increased risk for type 1 diabetes: first results from the TEENDIAB study. Public Health Nutr *18*, 50-58.

Weiss, S.J. (1989). Tissue destruction by neutrophils. N Engl J Med *320*, 365-376.

Weiss, S.J., Young, J., LoBuglio, A.F., Slivka, A., and Nimeh, N.F. (1981). Role of hydrogen peroxide in neutrophil-mediated destruction of cultured endothelial cells. The Journal of clinical investigation *68*, 714-721.

Wen, L., Ley, R.E., Volchkov, P.Y., Stranges, P.B., Avanesyan, L., Stonebraker, A.C., Hu, C., Wong, F.S., Szot, G.L., Bluestone, J.A.*, et al.* (2008). Innate immunity and intestinal microbiota in the development of Type 1 diabetes. Nature *455*, 1109-1113.

Wenzlau, J.M., and Hutton, J.C. (2013). Novel Diabetes Autoantibodies and Prediction of Type 1 Diabetes. Current diabetes reports *13*, 10.1007/s11892-11013-10405-11899.

Wenzlau, J.M., Juhl, K., Yu, L., Moua, O., Sarkar, S.A., Gottlieb, P., Rewers, M., Eisenbarth, G.S., Jensen, J., Davidson, H.W.*, et al.* (2007). The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. Proceedings of the National Academy of Sciences of the United States of America *104*, 17040-17045.

References

Wherry, E.J., and Kurachi, M. (2015). Molecular and cellular insights into T cell exhaustion. Nature reviews Immunology *15*, 486-499.

Whiting, D.R., Guariguata, L., Weil, C., and Shaw, J. (2011). IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes research and clinical practice *94*, 311-321.

Whitney, A.R., Diehn, M., Popper, S.J., Alizadeh, A.A., Boldrick, J.C., Relman, D.A., and Brown, P.O. (2003). Individuality and variation in gene expression patterns in human blood. Proceedings of the National Academy of Sciences of the United States of America *100*, 1896-1901.

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H.*, et al.* (2014). Mass-spectrometry-based draft of the human proteome. Nature *509*, 582-587.

Willcox, A., Richardson, S.J., Bone, A.J., Foulis, A.K., and Morgan, N.G. (2009). Analysis of islet inflammation in human type 1 diabetes. Clinical and experimental immunology *155*, 173-181.

Winkler, C., Krumsiek, J., Buettner, F., Angermuller, C., Giannopoulou, E.Z., Theis, F.J., Ziegler, A.G., and Bonifacio, E. (2014). Feature ranking of type 1 diabetes susceptibility genes improves prediction of type 1 diabetes. Diabetologia *57*, 2521-2529.

Winkler, C., Krumsiek, J., Lempainen, J., Achenbach, P., Grallert, H., Giannopoulou, E., Bunk, M., Theis, F.J., Bonifacio, E., and Ziegler, A.G. (2012). A strategy for combining minor genetic susceptibility genes to improve prediction of disease in type 1 diabetes. Genes and immunity *13*, 549-555.

Winkler, C., Lauber, C., Adler, K., Grallert, H., Illig, T., Ziegler, A.G., and Bonifacio, E. (2011). An interferon-induced helicase (IFIH1) gene polymorphism associates with different rates of progression from autoimmunity to type 1 diabetes. Diabetes *60*, 685-690.

Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nature methods *6*, 359-362.

Wong, S.L., Demers, M., Martinod, K., Gallant, M., Wang, Y., Goldfine, A.B., Kahn, C.R., and Wagner, D.D. (2015). Diabetes primes neutrophils to undergo NETosis, which impairs wound healing. Nature medicine *21*, 815-819.

Woo, M.M., Patterson, E.K., Clarson, C., Cepinskas, G., Bani-Yaghoub, M., Stanimirovic, D.B., and Fraser, D.D. (2016). Elevated Leukocyte Azurophilic Enzymes in Human Diabetic Ketoacidosis Plasma Degrade Cerebrovascular Endothelial Junctional Proteins. Crit Care Med *44*, e846-853.

Xu, M., Luo, W., Elzi, D.J., Grandori, C., and Galloway, D.A. (2008). NFX1 Interacts with mSin3A/Histone Deacetylase To Repress hTERT Transcription in Keratinocytes. Molecular and cellular biology *28*, 4819-4828.

Yabuhara, A., Yang, F.C., Nakazawa, T., Iwasaki, Y., Mori, T., Koike, K., Kawai, H., and Komiyama, A. (1996). A killing defect of natural killer cells as an underlying immunologic abnormality in childhood systemic lupus erythematosus. The Journal of rheumatology *23*, 171-177.

Yates, J.R., 3rd (1998). Mass spectrometry and the age of the proteome. Journal of mass spectrometry : JMS *33*, 1-19.

Yeung, W.C., Rawlinson, W.D., and Craig, M.E. (2011). Enterovirus infection and type 1 diabetes mellitus: systematic review and meta-analysis of observational molecular studies. Bmj *342*, d35.

Yoo, B.C., Kim, K.H., Woo, S.M., and Myung, J.K. (2017). Clinical multi-omics strategies for the effective cancer management. Journal of proteomics.

Yoon, J.W. (1990). The role of viruses and environmental factors in the induction of diabetes. Current topics in microbiology and immunology *164*, 95-123.

Yoon, J.W., Austin, M., Onodera, T., and Notkins, A.L. (1979). Isolation of a virus from the pancreas of a child with diabetic ketoacidosis. N Engl J Med *300*, 1173-1179.

Young, J.D., Peterson, C.G., Venge, P., and Cohn, Z.A. (1986). Mechanism of membrane damage mediated by human eosinophil cationic protein. Nature *321*, 613-616.

Zeng, Y., Ramya, T.N., Dirksen, A., Dawson, P.E., and Paulson, J.C. (2009). High-efficiency labeling of sialylated glycoproteins on living cells. Nature methods *6*, 207-209.

Zhang, K., Lin, G., Han, Y., Xie, J., and Li, J. (2017a). Circulating unmethylated insulin DNA as a potential non-invasive biomarker of beta cell death in type 1 Diabetes: a review and future prospect. Clinical Epigenetics *9*, 44.

Zhang, Q., Fillmore, T.L., Schepmoes, A.A., Clauss, T.R., Gritsenko, M.A., Mueller, P.W., Rewers, M., Atkinson, M.A., Smith, R.D., and Metz, T.O. (2013a). Serum proteomics reveals systemic dysregulation of innate immunity in type 1 diabetes. The Journal of experimental medicine *210*, 191-203.

Zhang, S.-g., Song, Y.-x., Shu, X.-m., Shen, H.-l., Yang, H.-b., Duo, R.-x., and Wang, G.-c. (2017b). A simple method for removing low-density granulocytes to purify T lymphocytes from peripheral blood mononuclear cells. Journal of Zhejiang University Science B *18*, 605-614.

Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.-C., and Yates, J.R. (2013b). Protein Analysis by Shotgun/Bottom-up Proteomics. Chemical reviews *113*, 2343-2394.

Zhi, W., Sharma, A., Purohit, S., Miller, E., Bode, B., Anderson, S.W., Reed, J.C., Steed, R.D., Steed, L., Hopkins, D.*, et al.* (2011). Discovery and validation of serum protein changes in type 1 diabetes patients using high throuput two dimensional liquid chromatography-mass spectrometry and immunoassays. Molecular & cellular proteomics : MCP *10*, M111.012203.

Ziegler, A.G., and Bonifacio, E. (2012). Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. Diabetologia *55*, 1937-1943.

Ziegler, A.G., Meier-Stiegen, F., Winkler, C., and Bonifacio, E. (2012). Prospective evaluation of risk factors for the development of islet autoimmunity and type 1 diabetes during puberty--TEENDIAB: study design. Pediatric diabetes *13*, 419-424.

Ziegler, A.G., Pflueger, M., Winkler, C., Achenbach, P., Akolkar, B., Krischer, J.P., and Bonifacio, E. (2011). Accelerated progression from islet autoimmunity to diabetes is causing the escalating incidence of type 1 diabetes in young children. Journal of autoimmunity *37*, 3-7.

Ziegler, A.G., Rewers, M., Simell, O., Simell, T., Lempainen, J., Steck, A., Winkler, C., Ilonen, J., Veijola, R., Knip, M.*, et al.* (2013). Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. Jama *309*, 2473-2479.

Ziegler, A.G., Schmid, S., Huber, D., Hummel, M., and Bonifacio, E. (2003). Early infant feeding and risk of developing type 1 diabetes-associated autoantibodies. Jama *290*, 1721-1728.

Zipitis, C.S., and Akobeng, A.K. (2008). Vitamin D supplementation in early childhood and risk of type 1 diabetes: a systematic review and meta-analysis. Archives of disease in childhood *93*, 512-517.

Zoeller, R.A., Grazia, T.J., LaCamera, P., Park, J., Gaposchkin, D.P., and Farber, H.W. (2002). Increasing plasmalogen levels protects human endothelial cells during hypoxia. Am J Physiol Heart Circ Physiol *283*, H671-679.

7 Abbreviations

2D	two-dimensional	MS1	precursor ion spectrum/survey scan
ANCOVA	analysis of covariance	MS2	fragment ion spectrum
APC	antigen presenting cell	MudPIT	multidimensional protein identification technology
CCL	C-C motif chemokine ligand	NETs	neutrophil extracellular traps
CCR	C-C chemokine receptor	NK	natural killer
CID	collision-induced dissociation	NOD	non-obsese diabetic
CTL	cytotoxic T lymphocyte	nPOD	network for pancreatic organ donors with diabetes
CV	coefficient of variation	Orbitrap	orbital trap
DDA	data-dependent acquisition	PAL	periodate oxidation and aniline-catalyzed oxime ligation
DIA	data-independent acquisition	PBMC	peripheral blood mononuclear cell
DiMelli	Diabetes Mellitus Incidence cohort study	PCA	prinicpal component analysis
DIPP	type 1 diabetes prediction and prevention project	pDC	plasmacytoid dendritic cell
DM	diabetes mellitus	PSM	peptide-to-spectrum matching
ESI	electrospray ionization	PTM	post-translational modification
FACS	fluorescence activated cell sorting	QC	quality control
FASP	filter-aided sample preparation	ROS	reactive oxygen species
FDA	food and drug administration	RP	reversed phase
FDR	false discovery rate	RT	room temperature
GADA	glutamic acid decarboxylase 65 autoantibody	SAM	significance analysis of microarrays
GWAS	genome wide association study	SAX/SCX	strong anion/strong cation exchange chromatography
HbA1c	glycated hemoglobin, type 1Ac	SDB-RPS	styrene divinylbenzene reversed phase sulfonate
HILIC	hydrophilic liquid interaction chromatography	SILAC	stable isotope labeling by amino acids in cell culture
HLA	human leukocyte antigen	SLE	systemic lupus erythematosus
HPLC	high performance liquid chromatography	SNPs	single nucleotide polymorphisms
HPM	human proteome map	SOP	standard operating procedure
IA-2A	insulinoma antigen-2 autoantibody	T1D	type 1 diabetes
IAA	insulin autoantibody	T2D	type 2 diabetes

IFN	interferon		TCR	T cell receptor
IL	interleukin		TGF	transfroming growth factor
iRT	indexed retention time		TMD	transmembrane domain
iST	in stop-and-go extraction tip		TMT	tandem mass tag
LADA	latent autoimmune diabetes in adults		TNF	tumor necrosis factor
LDG	low density granulocyte		TOF	time-of-flight
LTQ	linear trap quadrupole		Treg	regulatory T cell
MACS	magnetic activated cell sorting		VNTR	variable number tandem repeat
MALDI	matrix-assisted laser desorption/ionization		WBC	white blood cell
MHC	major histocompatibility complex		WHO	world health oragnization
miR/miRNA	microRNA		XIC	extracted ion chromatogram
MS	mass spectrometry		Znt8A	zinc transporter-8 autoantibody
MS/MS	tandem mass spectrometry			

8 Annex

8.1 Figure Index

8.2 Table Index

Annex

8.3 Publications and Presentations

8.3.1 Peer-reviewed publications

<u>Lepper MF</u>, Ohmayer U, von Toerne C, Maison N, Ziegler AG, Hauck SM (2018). "The proteomic landscape of patient-derived CD4+ T cells in recent-onset type 1 diabetes". J Proteome Res. *17*, 618-634.

Hauck SM, <u>Lepper MF</u>, Hertl M, Sekundo W, Deeg CA (2017). "Proteome dynamics in biobanked horse peripheral blood Derived lymphocytes (PBL) with induced autoimmune uveitis". Proteomics *17*.

Grosche A, Hauser A, <u>Lepper MF</u>, May R, von Toerne C, Merl-Pham J, Hauck SM (2016). "The proteome of native adult Muller glial cells from murine retina". MCP *15*, 462-480.

8.3.2 Poster presentations

Proteomic Forum 2017, Potsdam, Germany: "The proteomic landscape of patient-derived CD4+ T cells in recent-onset type 1 diabetes" (April 2017).

EMBO Workshop Advanced Proteomics, Varna, Italy: "Proteomic profiling of patient-derived CD4+ T cells in recent-onset type 1 diabetes" (August 2016).

The Immunology of Diabetes Society Meeting 2015, Munich, Germany: "Proteomic profiling of patient-derived CD4+ T cells for biomarker discovery in type 1 diabetes" (April 2015).

Proteomic Forum 2015, Berlin, Germany: "Proteomic profiling of patient-derived CD4+ T cells for biomarker discovery in type 1 diabetes" (March 2015).

8.3.3 Oral presentations

Childhood diabetes: features, pathogenesis and intervention. Oktoberfestsymposium 2017, Munich, Germany: "T cell proteomics" (September 2017).

Helmholtz Diabetes Center Seminar; Portfolio Project: "Proteomic signatures of patient-derived PBMCs in type 1 diabetes" (February 2016).

8.4 Danksagung