Norwegian University
of Life Sciences

Master's Thesis 2020 60 ECTS
Faculty of Chemistry, Biotechnology and Food Sciences

Protein Dysregulation in Immune Cells of Multiple Sclerosis Patients

Oda Glomstad Kråbøl
Biotechnology
Faculty of Chemistry, Biotechnology and Food Sciences

Protein Dysregulation in Immune Cells of Multiple Sclerosis Patients

Oda Glomstad Kråbøl

Multiple Sclerosis Research Group
Oslo University Hospital

Norwegian University
of Life Sciences

Main supervisor: Tone Berge

Co-supervisor: Chiara Cappelletti

Internal supervisor: Siv Kjølsrud Bøhn

Oslo University Hospital – University of Oslo

Multiple Sclerosis Research Group

and

The Norwegian University of Life Sciences,

Faculty of Chemistry, Biotechnology and Food Sciences

June 2020

Protein Dysregulation in Immune Cells of Multiple Sclerosis Patients

https://nmbu.brage.unit.no

Acknowledgements

This thesis was carried out at the MS research group in the Neuroscience Research Unit of Oslo University Hospital (OUS) from August 2019 to June 2020. The thesis is a part of a Master's Degree in Biotechnology at the Norwegian University of Life Sciences (NMBU), main affiliation being the Faculty of Chemistry, Biotechnology and Food Sciences (KBM).

Firstly, I would like to thank Professor Tone Berge for being my supervisor, and for arranging this interesting project. Her expertise and enthusiasm have given me great insights into the challenging yet fascinating multiple sclerosis research field. The guidance I have received throughout this time has been exceptional! I am likewise thankful for my co-supervisor Chiara Cappelletti for devoting her time supervising me in the lab and answering all my questions. She has provided me with protocols as well as theoretical and practical explanations for everything I have needed, which has been invaluable.

I would also like to thank the MS research group members for making me feel welcome and included. A special thanks to Anna Eriksson, Ina Brorson and Steffan D. Bos, for sharing their time, knowledge and good advice. Additionally, I would like to thank my supervisor at NMBU, associate professor Siv Kjølsrud Bøhn, for supportive and constructive feedback.

Furthermore, I have appreciated the company of my fellow master students Hannah and Stine. Thank you for sharing lunch breaks, experiences and occasional frustrations. Lastly, I am ever grateful for the support from my family and friends during this past year, and especially for Adrian, whose motivational words and encouragement have been essential for finalizing this work. Thank you for always pushing me and inspiring me to do my very best!

Oda Glomstad Kråbøl

Oslo, June 2020

Abstract

Multiple sclerosis (MS) is a complex, autoimmune and inflammatory disease affecting the central nervous system. The pathogenesis of MS is not completely understood, but T cell activation is believed to be an important part of disease etiology. MS susceptibility is provided through a combination of several single contributions involving environmental exposures and multiple genetic risk variants. To date, more than 200 MS associated risk variants have been identified, many of which are linked to immune cell reactions in T cells. In the current project, we aimed at (i) using available proteomic data from $CD4_+$ and $CD8_+$ T cells of genome-wide genotyped MS patients and healthy controls to identify novel MS-linked proteins or pathways. Secondly, (ii) we aimed at collecting new samples from activated $CD4_+$ T cells from the same groups for proteomics profiling by mass spectrometry.

(i) A proteomic data set from $CD4_+$ and $CD8_+$ T cells was utilized to correlate the genotype at MS susceptibility variants with the expression of proteins encoded by genes located 100kb upstream and downstream of the MS risk variants. Furthermore, pathway analysis was performed on proteins that were differentially expressed between MS patients and healthy controls in both cell types. (ii) Live $CD4_+$ T cells are stored on liquid nitrogen, and cells from 28 female relapsing-remitting MS patients and 28 healthy individuals were thawed and activated *in vitro* with α-CD3 and α-CD28 antibodies. The cells were analyzed by flow cytometry for cell surface expression of the T cell activation marker CD69 and with a LIVE/DEAD stain distinguishing live from dead cells. Cells with >50 % CD69 positive cells after activation and with >70 % viability were sent for liquid chromatography-tandem mass spectrometry analysis to analyze the proteome of the samples.

(i) Two novel protein quantitative trait loci (pQTL) candidates were identified, i.e. rs1800693 - *CD9* and rs137955 - *UQCRFS1P1.* 26 proteins were differentially expressed in both $CD4_+$ and $CD8_+$ T cells between the groups, and pathway analysis did not identify any specific enriched biological pathways. The identified pQTL candidates might have relevance for MS disease, and can inspire for functional studies to seek a broader understanding of the effect of genetic risk variants, as well as the mechanisms behind MS

development. (ii) A total of 20 samples from MS patients and 20 samples from healthy controls fulfilled the inclusion criteria, and were included in the proteomic analysis. Thus, samples from un-activated and cells activated for 24 hours were shipped to the proteomics core facility in Bergen. The samples are lysed an prepared for proteomic profiling, but the results of the analysis is not available yet, due to the Covid-19 situation. Whether the proteomic analysis of activated $CD4_+$ T cells lead to identification of proteins and pathways of importance for MS, remains to be shown.

Sammendrag

Multippel sklerose er en autoimmun, nevro-inflammatorisk sykdom i sentralnervesystemet med komplekst sykdomsbilde. Årsak til utviklingen av denne kroniske sykdommen er ikke fullstendig kartlagt, men aktiverte T-celler antas å ha en betydningsfull rolle. Mottakelighet for MS er knyttet til kombinasjonen av en rekke genetiske risikovarianter samt eksponering for ulike miljøfaktorer. Over 200 MS-assosierte mottakelighets varianter er i dag indentifisert, hvor mange av disse er koblet opp mot immuncelle responser hos T-celler. Dette prosjektet hadde som hensikt (i) å benytte tilgjengelige proteomikk-data fra $CD4_+$ og $CD8_+$ T-celler fra MS pasienter og friske kontroller i kombinasjon med individenes genotype-data, for å identifisere nye MS-assosierte proteiner eller reaksjonsveier. Videre ønsket vi å (ii) samle nye prøver fra aktiverte $CD4_+$ T-celler fra de samme gruppene til proteomikk-analyser ved massespektrometri.

(i) Proteomikk-data fra $CD4_+$ og $CD8_+$ T-celler ble benyttet til korrelasjon av genotypen ved MS mottakelighets-variantene med ekspresjon av proteiner kodet av gener lokalisert i et område på 100kb oppstrøms og nedstrøms for hver MS mottakelighets-variant. Videre ble reaksjonsvei-analyser utført for proteiner som var differensielt uttrykte mellom MS pasienter og friske kontroller i begge celletypene. (ii) Levende $CD4_+$ T celler blir lagret i flytende nitrogen, og celler fra 28 kvinnelige relapserende remitterende MS pasienter og 28 friske kontroller ble tinet og aktivert *in vitro* ved platebundet α-CD3 og løselig α-CD28 antistoff. Cellene ble analysert ved væskestrømsanalyse ved å detektere T-celle aktiverings-markøren CD69 i kombinasjon med en LIVE/DEAD stain for å skille levende fra døde celler. Celler med > 50 % CD69 positive celler etter aktivering, samt > 70 % levelighet, ble sent til analyser ved væskekromatografi/tandem massespektrometri, for å analysere proteomet i prøvene.

(i) To nye protein «quantitative trait loci» (pQTL) kandidater ble identifisert, dvs. rs1800693 - *CD9* og rs137955 - *UQCRFS1P.* 26 proteiner var differensialt uttrykte i både $CD4_+$ og $CD8_+$ T-celler mellom gruppene, men ingen anrikede reaksjonsveier ble identifisert i reaksjonsvei-analysene. De identifiserte pQTL kandidatene kan være relevante for MS og samtidig inspirere til funksjonelle studier som kan belyse effektene

av de genetiske risiko-variantene samt de underliggende mekanismene i MS utvikling. (ii) 20 prøver fra MS-pasienter og 20 prøver fra friske kontroller oppfylte inklusjons-kravene og ble inkludert i proteomikk-analysene. Ikke-aktiverte prøver og prøver aktivert i 24 timer sendt til kjerneanlegget for proteomikk i Bergen. Prøvene er lysert og preparert til analyse, men resultatene er ikke tilgjengelige grunnet Covid-19 situasjonen. Det gjenstår å evaluere om proteomikk-resultatene kan bidra til identifisering av proteiner eller reaksjonsveier med betydning for MS.

Table of contents

1 - Introduction

Multiple sclerosis (MS) is an autoimmune, chronic inflammatory disease of the central nervous system (CNS), causing mild to severe physical and cognitive impairment, loss of vision and fatigue. Inflammation in the CNS is initiated by autoreactive lymphocytes crossing the blood-brain (BBB), leading to demyelination of the myelin sheath insulating the nerve fibers and degeneration of the axons (1). Although the mechanism behind the pathology of MS is still unknown, a combination of genetic susceptibility and environmental factors including early Epstein Barr virus (EPV) infection, smoking, obesity at a young age and vitamin D insufficiency are known to contribute to disease risk (1). The strongest genetic factor contributing to MS risk is the human leukocyte antigen (HLA), but today more than 200 non-HLA genetic variants associated with MS risk have been identified (2-4). Because these genetic variants typically are located in non-coding parts of the genome, their function can be challenging to study. Enrichment of the genetic variants associated with MS is however observed in regulatory regions of DNA in cells of the adaptive immune system, such as the T and B lymphocytes (5). For this reason, it is hypothesized that these genetic variants may play a role in the lack of education and/or regulation of the immune system promoting an autoimmunity phenotype. The impact of the genetic variants for MS phenotype is still not known.

1.1 The immune system

The immune system, divided into the innate and the adaptive system, is characterized as our body's defense machinery, protecting against bacteria, viruses, fungi and other invading pathogens. It is an intricate and regulated network of interacting cells and molecules, specialized to eliminate a wide range of pathogens without damaging the body's own cells. The consequence of dysfunction or disruption of this complex system might be development of allergies or autoimmune diseases (6)

The primary mechanism of innate immunity is protection obtained by the external structures, which prevents infection, spreading of microbes, and isolates the internal environment from external factors (7). Innate immunity is comprised of immune cells derived from myeloid stem cells from the bone marrow, such as monocytes, granulocytes, macrophages, dendritic cells (DCs) and natural killer (NK) cells. Upon infection, these cells

will immediately neutralize a wide range of pathogens, due to their expression of different receptor molecules (8). When a pathogen has been identified, the immune system will become activated and make the endothelial tissue permeable for immune cells and plasma proteins to enter the infection site. This process of plasma proteins and fluids leaking into the connective tissues, as well as increased blood flow and heat, causes inflammation, a state characterized by edema, swelling and redness. If this innate immune response fails to withstand the amount of pathogens, the adaptive immune system has to step in (9). The innate immunity provides a rapid response, and can recognize and eliminate a range of pathogenic molecular patterns. However, pathogens have developed the ability to mutate to avoid recognition, and this ability have driven the evolution of the adaptive immune system (10).

In contrast to the non-specific response of the innate immune system, the adaptive immune system provides a broader and more refined recognition of pathogens, as well as being able to recognize self-peptides. The response is directed towards unique and distinctive structures of a single bacteria or a virus particle (11). This part of the immune system requires an interplay between different specialized immune cells, in particular B and T lymphocytes (10). In addition to the specificity, the development of immunological memory is an important feature, enabling a faster and more powerful defense in the case of secondary pathogen encounters (12). The initial response of the adaptive immune system involves activation of the T and B lymphocytes, both derived from a common lymphoid precursor cell of the bone marrow (13). B and T cells express receptors on their surface, called immunoglobulins and T cell receptors (TCR), that specifically recognize a wide range of molecules derived from microbes and non-infectious substances. Any molecule or substance recognized by lymphocytes or antibodies through their receptors are referred to as an antigen (6). Inactive T and B lymphocytes circulate through lymphoid organs and nonlymphoid tissues to find foreign antigens, and if an interaction occurs, they will initiate the adaptive immune system responses. This response leads to lymphocyte emigration to different sites in the body to perform effector functions (14).

The activated T cells differentiate into three different types of effector cells, cytotoxic T cells, helper T cells and regulatory T cells. The cytotoxic T cells express the transmembrane protein and TCR co-receptor Cluster of differentiation 8 (CD8), while the

T helper cells express CD4, hence they are referred to as $CD8_+$ and $CD4_+$ T cells (13). Activated B cells can differentiate into plasma cells secreting antibodies, a soluble version of their immunoglobulins. Although the adaptive immune system takes longer time to develop, it is able to generate immunological memory (6), as long-lived memory T and B cells are established after the first encounter with a specific antigen. In the case of a secondary encounter of the same antigen, these memory cells are quickly activated, enabling an enhanced and enforced response to defeat the pathogen (10).

1.1.1 T cell differentiation and development

All T cells originate in the bone marrow and differentiate in the thymus, a small and specialized primary lymphoid organ. The differentiation process is crucial for our immune system, as the T cells develop the ability to discriminate between foreign substances and antigens naturally present in healthy individuals, called self-antigens. Without this feature, the body's immune system could register self-antigens as pathogenic and hence initiate a response causing cell and tissue damage (6).

The precursor of T cells ultimately derive from hematopoietic stem cells in the bone marrow, which subsequently migrate through the blood and reside in the thymus for further differentiation (15). The development depend on signals from non-hematopoietic stromal cells, located in specific sites in the thymus. As a result of the signal, the T cells start to proliferate, rearrange genes of the TCR, and eventually express their TCR and co-receptors, such as CD8 and CD4 (16). During the maturation process, the precursors of the $CD8_+$ T cells are exposed to antigens presented on the major histocompatibility complex (MHC) class I molecule, expressed by all nucleated cells. The precursor of the $CD4_+$ T cells are presented to antigens on the MHC class II molecule, only expressed by APCs (8). In humans, the MHC is referred to as the HLA. The ligand specificity of the TCR controls the fate of the T cell, as they further undergo a strict selection process where only cells with functional TCRs that does not react to self-antigens will pass (16).

Positive selection is an immune process making sure that T cell precursors are capable of binding MHC molecules. The T cell precursors are presented peptides on an MHC, and are positively selected as they receive survival signals through binding of their TCR. T cell precursors incapable of binding will undergo apoptosis (17, 18). The positive selection is

a checkpoint for MHC restriction. This term involves that the T cells only react with an antigen if the cooperative cell express the same type of MHC molecule on their surface as the cells in the thymus where the T cell developed into a mature T cell (8). A $CD4_+$ T cell will by this principle only react with an antigen on a MHC class II molecule expressed by a APC (18).

The second step in the selection process is the negative selection eliminating any developing T cells that are self-reactive. The dendritic cells in the thymus play an important role in this process, when T cell precursors that react too strongly towards self-peptides presented on MHC molecules are killed by apoptosis (8). However, this process in imperfect, and some autoreactive T cells can be released into the periphery, resulting in autoimmunity. In healthy individuals, the mechanisms of positive selection keep these escaped cells in check (14). T cells that have passed the two selection steps leave the thymus and enter the circulation system, each equipped with TCRs, giving a unique specificity for recognizing antigens (19).

1.1.2 T cell activation

The TCR complex is a multi-subunit complex consisting of two glycosylated polypeptide chains, the alpha (α) and beta (β) chains, linked together by disulfide bonds, forming the TCR, along with several co-receptors. The two chains are each composed of a variable (V), constant (C) and joining (J) regions, and therefore closely resembling Ig chains (19). Because of a complex and random genetic rearrangement process called somatic recombination, these different regions are able to generate millions of different receptors (20). Some populations of T cells, about 5 %, express γ and δ as the main chains in the receptor. These receptors are constructed in the same manner as the $\alpha\beta$-receptor (8).

The co-receptors of the TCR complex consist of either the CD4 or the CD8 co-receptor, along with the CD3 co-receptor. The CD3 molecule is non-covalently associated with the α and β chains, and consist of four different polypeptide chains termed gamma (γ), delta (δ), epsilon (ε), and zeta (ζ). The complex contains a homodimer of two ζ chainsat the bottom, two ε chains, one γ chain and one δ chain (Figure 1.1)(8, 11). The CD3 co-receptor is important for signaling leading to activation of T cells.

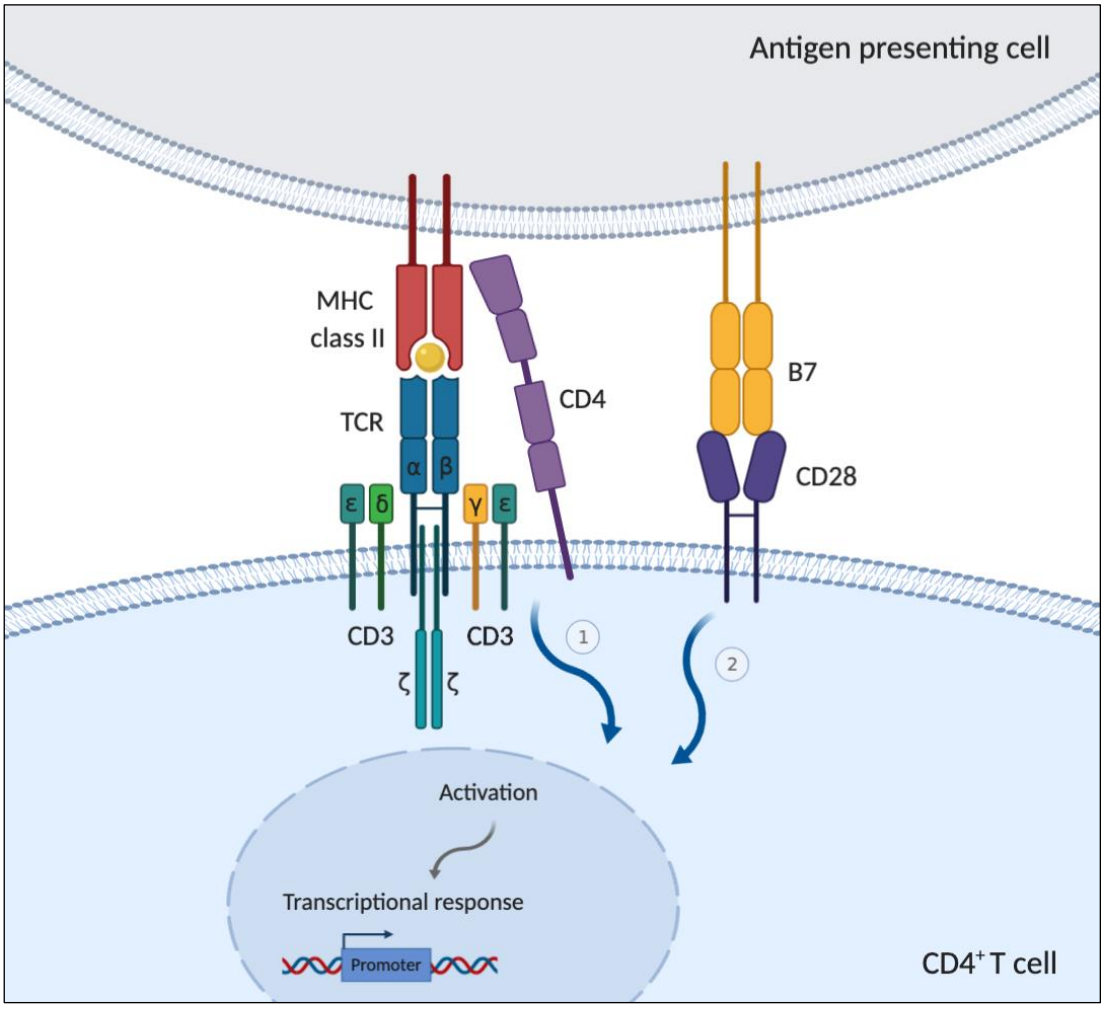

Figure 1.1.: Signals needed to successfully activate a CD4+ T cell. *Two signals are required for the activation of CD4+ T cells, transmitted through the CD3 molecules (1) and the CD28 molecule (2). The signaling complex in T cells involve the T cell receptor (TCR), two ζ-chains, two CD3 molecules and the co-receptor CD4. An antigen presenting cell (APC) presents an antigen in its major histocompability complex (MHC) class 2 and interacts with the TCR complex and the CD4 co-receptor. This leads to signal transmission through the CD3 molecule. If an infection occurs, the B7 receptor is expressed on the surface of an APC and interacts with CD28 on the T cell surface, leading to transmission of the second signal required for T cell activation. The activation generates downstream effects such as a transcriptional response (11). The figure was designed in BioRender.*

In the case where a naive CD4+ T cell encounters an APC expressing peptide-bound MHC-II at its surface, the TCR complex will be assembled. The activation and differentiation of the T cell require at least two signals delivered by the APC. Signal 1 is generated through interaction between the TCR complex and the co-receptor CD4 to a peptide-MHC (11). Signal 2 is generated when the ligand B7 on the same APC is interacting with the

costimulatory receptor CD28 on the naïve $CD4_+$ T cell. Even though the $CD4_+$ T cells receive signal 2 via the B7-CD28 pathway, there are other T cells using different costimulatory molecules. These molecules can deliver either stimulatory (positive) or inhibitory (negative) signals for T cell activation (19). The CD3 co-receptor is responsible for transferring the activation signal across the membrane from the TCR. A successful activation induces a cascade of intracellular signaling that eventually triggers T cell maturation, proliferation and production of immune mediators (11).

The interaction between the TCR and antigen bound to MHC is not the only factor determining the effector functions of the effector T cells, as cytokines produced by APCs are a part of the activation process (21). Activated $CD8_+$ T cells excrete perforin and granzymes, that in turn stimulate virus-infected cells to undergo apoptosis. $CD4_+$ T cells stimulate phagocytes for a more efficient degradation of foreign peptides, by secreting cytokines upon TCR-MHC binding. $CD4_+$ T cells additionally help the maturation process of B cells, by secreting interleukins. The B cells proliferate into memory cells or plasma cells secreting antibodies.

1.2 Multiple sclerosis

MS is a chronic neuroinflammatory disease of the CNS and is the leading cause of non-traumatic neurologic disability in young adults, especially women (22). Worldwide, the disease affects approximately 2.5 million individuals (14), with a higher prevalence in Europe and North America presumably linked to the latitude gradient (23). In Norway, the MS prevalence is one of the highest in the world, as approximately 1 in 490 citizens are affected (24). The symptoms of MS vary greatly between patients, but the majority have symptoms involving motor, sensory, visual and autonomic systems as well as other physical and cognitive deficits (1).

1.2.1 Multiple sclerosis etiology

MS is a multifaceted disease and a definite cause of its development still remains elusive, but a well-established principle is that MS is triggered by environmental exposures and genetic susceptibility. Among the environmental factors contributing to MS risk are EBV

infection, smoking, obesity at a young age and low vitamin D levels due to insufficient sun exposure and/or dietary intake (25).

The prevalence of MS is observed to be higher in family members of affected individuals compared to the general population (26). However, MS is not considered a Mendelian disease (26) and accordingly not caused by inheriting a single mutated gene. In fact, studies of MS etiology points to MS having a complex inheritance pattern and also suggest that epigenetic mechanisms may play a role in the disease development (27).

1.2.2 Multiple sclerosis pathogenesis

Even though the pathogenesis of MS is still not entirely understood, T lymphocytes, both $CD4_+$ and $CD8_+$ T cells, are considered to play central roles in the process (19, 28). Inflammation in the CNS is caused by autoreactive lymphocytes crossing the BBB leading to demyelination of the myelin sheath insulating the nerve fibers and degeneration of the axons (1). This leads to disturbance in the transmission of nerve impulses, which can lead to a wide variety of MS symptoms (14).

The MS pathology is portrayed by demyelinated regions of the white and grey matter of the brain and spinal cord. These areas are referred to as lesions or scars (sclerae), indicating myelin sheath loss and loss of the myelin-producing oligodendrocytes. The nerve fibers and the axons are moderately intact in the early stages, but as the disease evolves, the axon degeneration increase with the disability of the patient. Although there are large individual variations in MS symptoms and disease course, over 85 % are affected by relapsing-remitting MS (RRMS), characterized by an initial incidence of neurological dysfunction, followed by a period of remission. The relapses are often accompanied by inflammation in the CNS and demyelination that can be detected as lesions by magnetic resonance imaging (MRI)(14).

1.2.3 Multiple sclerosis is an autoimmune disorder

During the negative selection of T cells in the thymus, autoreactive T cells are removed by inducing apoptosis. Usually, these cells will be controlled by positive selection, but in some cases, autoreactive B and T cells can be activated in the periphery and lead to

autoimmune diseases. These activated cells can develop into aggressive effector cells infiltrating the CNS, causing inflammation and damage of tissue (14).

Autoreactive cells released into the periphery can be activated through a mechanism called molecular mimicry, involving the presentation of a foreign antigen with shared sequences or sequence similarities with an antigen of the CNS, generating an autoimmune reaction (29). As an example, immune responses towards EBV may promote CNS inflammation, as studies purpose that antibodies against an ion channel expressed in the CNS of MS patients can recognize a fragment of an EBV nuclear antigen (30). A second scenario purposes bystander activation as a contributing factor to the autoreactive cell activation, implicating that the activation occurs due to nonspecific inflammation caused by an infection. The bystander activation can occur through the many inflammatory cytokines and chemokines produced during an infection, and these molecules are considered the main activators of virus-specific $CD8_+$ T cells and autoimmune response inducers (31).

1.3 Genetics

The human genome involves about three billion base pairs (32), generating an estimated amount of 19.000 protein-coding genes (33). Approximately 2 % of the human genome encodes proteins, whereas a large portion of the genome is considered non-coding regions (34). The traits of common genetics are comprised of monogenic and polygenic traits, referring to the phenotype being influenced by a single, or multiple alleles or genes, respectively (35). Most autoimmune diseases are defined by the impact of several genes, giving rise to broad research fields that develop treatments and study causes of disease development.

1.3.1 Genetic variations

Genetic diversity is the degree of genetic differences among individuals of a population, and some genetic variants are common contributors to various diseases. The variations are a consequence of gene mutations (36), occurring at varying mutation rates across the genome (37). The mutations arise mostly in the DNA of somatic cells, while some

mutations affect germline cells and can be passed on from generation to generation. Populations of gene variants are a result of mutated genes inherited through several generations, in humans most commonly occurring as a single nucleotide polymorphism (SNP). A SNP represents a mutation in a single base of the DNA, and they are spread widely across the genome. Most of the SNPs are however located in genome regions that are non-coding (38). Additionally, the SNPs often have many candidate variants in linkage disequilibrium (LD), indicating that they are located in close proximity and inherited together, making the identification of specific causal variants and genes a challenge (39). To detect associations between genetic variants and traits in samples from different populations, genome-wide association studies (GWAS) was developed (40). In studies of complex diseases, GWAS is widely utilized to find SNPs that are more frequently present in disease cases, compared to healthy individuals.

1.3.2 Gene expression and regulation

Regulation mechanisms of mRNA transcription from DNA are sophisticated processes (34) that prevents overproduction of the gene products. To maintain this regulation, cells have several control points during the transcription of mRNA from DNA and the translation of protein products from mRNA, in addition to the regulation of synthesized proteins. The control of these processes play an important role in determining the quantity and the identity of proteins that are present in a cell (41).

The expression of a gene is affected by an intricate collaboration between regulatory elements located proximal to, or distant from the transcribed gene (34). A group of proteins called transcription factors (TFs) retain DNA binding domains that can bind areas of specific genes, i.e. promoter and enhancer regions. The TFs additionally possess domains interacting with RNA polymerase II and other transcription factors, leading to the regulation of mRNA expression (42). By this direct or indirect binding, the regulators can activate or repress transcription. The primary transcripts include sequences referred to as introns which are removed before mature mRNA leave the nucleus. In a process called splicing, the protein-coding regions of the transcript, called exons, are joined together to form the mature mRNA. In addition to the splicing, ends of the mature mRNA is modified to influence their stability and translation (41).

While the sequence of the mRNA determines the amino acid sequence of the synthesized protein, there is no cohesive relationship between the transcript concentration and the concentration of the protein (43). Some genetic loci may correlate with the quantitative expression of genes, and are collectively called expression quantitative trait loci (eQTLs). Similarly to eQTLs, genetic variants may also correlate with the quantitative expression of proteins, and they are termed protein QTLs (pQTLs). Combining pQTL discoveries with SNPs identified by GWAS may contribute to insights in the role of the proteome in disease, in addition to finding SNP-disease relationships (44).

1.3.3 Multiple sclerosis genetics

Studies provide evidence that genetic factors have a notable role in MS development, as increased heritability is seen within families. In addition, decreased risk seems to be proportional with the level of relatedness between individuals (45). The first identified genetic MS risk locus was the HLA region, and today the HLA-DRB1*15:01 allele provides the strongest genetic association to MS (2, 46-48), accounting for 20,2 % of the heritability. Consequently, carriers of this allele are about three times more at risk to develop MS, compared to non-carriers (3).

The genetic variants postulating an MS risk, most commonly exist as SNPs. The relationship between SNPs and disease has been assessed through GWAS, mapping a large amount of SNPs in the genome of MS patients and healthy controls, to find possible MS associated SNPs. The first non-HLA genetic variants identified with small effects on the risk, included SNPs in the *IL7RA* and *IL2RA* genes (45, 49), followed by identification of several new associations including regions of *TYK2*, *STAT3*, *CD58* and *TNFRSF1A*, through GWAS combined with meta-analysis (45, 50). At this time, more than 200 non-HLA SNPs associated with MS risk have been identified through GWAS and meta-analysis (2-4). This information contributes to the genetic architecture of MS, in addition to indicating the immune system as important factors of disease (45).

While studies of MS genetics have focused on the identification of genetic variants contributing to MS risk, studies of gene activity in patient-derived samples are aiming to identify genes of relevance for MS development or genes affected by MS disease. Alterations in histone post-translational modifications and DNA methylation have been

detected in MS patients. The histone post-translational modifications regulate gene expression by changing the structure of chromatin into a permissive or repressive state (27), while DNA methylations regulate gene expression by inhibiting the binding of TFs to DNA (51). Several studies have assessed genome-wide DNA methylation changes in peripheral blood mononuclear cells (PBMCs) and sorted $CD4_+$ and $CD8_+$ T cells from MS patients and healthy controls. These studies point to global differences in DNA methylation between MS patients and healthy controls (52-54). Increased DNA methylation have been observed in the SLFN12 gene of $CD4_+$ and $CD8_+$ T cells from MS patients, as well as in the HLA-DRB1 loci (52, 55).

1.4 Liquid chromatography-tandem mass spectrometry analysis

In analytical chemistry, liquid chromatography mass spectrometry (LC-MS) is a technique combining the liquid chromatography (LC) feature of separating solution components and the mass spectrometry (MS) feature of analyzing the mass of these components. A much applied workflow for proteomic studies is based upon the digestion of proteins within a sample into peptides, followed by LC separation and tandem MS (MS/MS) for peptide measurements and protein identification by database searching (56).

In LC, a fluid termed the mobile phase carries the solution of interest through a column, in which the components of the solution has various interactions with the column materials, termed the stationary phase. Components of the solution with low levels of interaction with the stationary phase, will eluate prior to the components with high levels of interaction with the stationary phase. The components will reach a detector, and peaks representing these separated components will be formed on the basis of time spent in the column. In LC, one peak does not represent one component (57), and mass spectrometry is by this reason used to analyze the molecular weight of the different components by separating ions in the gas phase. This technique measures the ratio of mass and charge in the separated ions. In analysis of peptides, the peptide solution eluting from the LC is often ionized by electrospray ionization (ESI), involving the sample being sprayed out of a capillary into a field of high voltage (58). To catch the separated ions, an ion trap is attached, whereas the ion trap termed Orbitrap is normally applied for complex samples where the peptides is unknown (59, 60).

Where the traditional LC-MS would struggle due to overlapping peptide mass and background, operating with LC/MS-MS will enable detection of multiple components simultaneously, which is highly applicable in proteomic studies. For identification, the mass spectrometry data are scanned across amino acid sequence databases and ultimately find the peptide sequence most likely to give rise to the tandem mass spectrum data (61).

In proteomics based on mass spectrometry, quantification is a central part, and can determine the absolute amount of each protein in a mixture. The quantification is mainly performed by either isotope-based or label-free methods. When signals signal of a given peptide is compared to the same peptide in different experiments, the quantification method is termed label-free (61).

2 - Aims of the study

The overall objective of the current thesis is to unravel proteomic T cell dysregulation in MS by LC/MS-MS, and to correlate protein expression with genotype at MS risk SNPs.

In a published study (54), proteomic profiles of un-activated $CD4_+$ and $CD8_+$ T cells from RRMS patients and healthy controls were analyzed, revealing differentially expressed proteins between MS patients and healthy controls. This data, in combination with genotype-data of the same MS patients and healthy controls were used with the aim to identify novel proteins or pathways associated with MS. The same dataset was utilized to perform pathway analysis on proteins differentially expressed in both cell types.

While the available dataset analyzed the proteome of un-activated T cells, we aimed to study the $CD4_+$ T cell response to activation in RRMS patients and healthy controls. Samples of *in vitro* activated $CD4_+$ T cells from MS patients and healthy controls were collected and sent for proteomic profiling by LC/MS-MS, to potentially identify proteins or pathways of relevance for MS disease.

3 - Methods

The following chapter include the methods and protocols applied in this study. A complete list of equipment, materials and software are presented in appendix B.

3.1 Sample collections and cellular techniques

In the laboratory experiments of this thesis, $CD4_+$ T lymphocytes from MS patients and healthy controls were studied. Prior to the start of this master project, blood was drawn from untreated, female Norwegian RRMS patients and female, age-matched Norwegian healthy controls. After blood withdrawal and $CD4_+$ T cell isolation, the samples of live $CD4_+$ T cells were stored in liquid nitrogen. The inclusion of patients and cell isolation was performed by members of our lab, between 2011 and 2016.

All work with mammalian cells was performed under sterile conditions to prevent environmental contamination. The work station and the equipment were disinfected with 70 % ethanol and the cells were cultured at 37°C and 5 % CO_2, otherwise stored on liquid nitrogen.

3.1.1 MS patients and heathy controls included in the study

Patients were recruited by clinicians in the MS research group from the MS out-patient clinic at Oslo University Hospital, Oslo, Norway and the healthy controls among the employees at the hospital. All patients and healthy controls were self-declared of Nordic ancestry. All MS patients met the updated McDonald criteria for MS (62), did not have an ongoing infection, and had not experienced a relapse or received steroids in the three months prior to admission. The patients were diagnosed less than a year prior to study inclusion and the healthy controls did not report MS disease in near relatives. The regional Committee for medical and Health Research Ethics South East, Norway, approved the study and all study participants received oral and written information. Written informed consent was obtained from all study participants (54).

3.1.2 Isolation of peripheral blood mononuclear cells from blood with LymphoPrep

The density gradient LymphoPrep (STEMCELL Technologies) is an agent commonly used to isolate mononuclear cells from peripheral blood, cord blood and bone marrow. The differences in cell density are exploited to separate eosinophils, neutrophils (granulocytes) and erythrocytes from PBMCs. Due to the higher density of granulocytes and erythrocytes at the osmotic pressure of the LymphoPrep, these cells would precipitate through the layer of LymphoPrep during centrifugation. After centrifugation, the PBMCs form a layer on top of the remaining LymphoPrep and can be collected (Figure 3.1).

Eight EDTA tubes each containing 8 mL blood were collected from each patient. The method was performed by other members of the MS research group for all samples included in the study. In collaboration with PhD student Chiara Cappelletti, the work performed in this thesis involved contributing to the isolation of cells used as control samples and for activation procedure establishment.

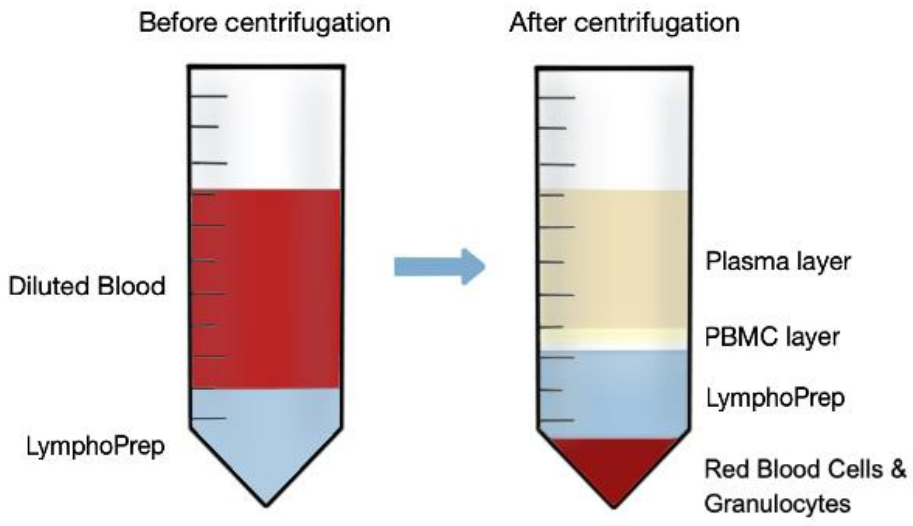

Figure 3.1.: Illustration of layers formed after centrifugation during PBMC isolation with LymphoPrep *A layer of LymphoPrep is syringed at the bottom of a tube of diluted blood, and centrifuged. During centrifugation, the red blood cells and the granulocytes will pass through the LymphoPrep layer, while the PBMCs form a collectable layer on top of the remaining LymphoPrep.*

Isolation procedure

Isolation buffer was prepared one day in advance;

- PBS with 2% fetal calf serum (FCS) or fetal bovine serum (FBS) and 1 mM EDTA
- 48,5 mL PBS, 1 mL FCS and 500 µL EDTA (100 mM) was mixed.

The LymphoPrep was tempered to room temperature (RT) before starting the procedure.

1. The LAF-bench was covered with paper towels.
2. 1 mL of 100 mM EDTA was added to a 75 mL cell culture flask.
3. Blood was transferred from EDTA tubes to the flask.
4. To wash out the remaining blood from the flask, 10 mL cold RPMI-1640 medium was added to the EDTA tubes and transferred to the flask. The same 10 mL RPMI-1640 medium was used to wash all the tubes. The flasks were swirled.
5. The blood was diluted by adding RPMI-1640 medium to a total volume of 100 mL (for 8 tubes of blood). The content was mixed by pipetting.
6. 25 mL of blood/RPMI-1640 solution was gently pipetted to a 50 mL tube (4 tubes in total).
7. The syringe needle (10 mL syringe + 2,1x80mm needle) was inserted with 10 mL LymphoPrep through the blood and to the very bottom of the tube. Pressure was gently applied to the syringe to place the LymphoPrep at the bottom, distinct from the blood. This procedure was repeated for all the 4 tubes.
8. The tubes were centrifuged at 2000 rpm for 20 min at RT (if blood had been in RT for more than 2h, centrifugation time was increased to 30 min). Brake was adjusted to 0 and acceleration to 1.
9. The mononuclear cells formed a band on the mediums interface (See Fig. X). A Pasteur pipette was used to transfer the cells in the interface into two separate 50 mL tubes. The upper layer was avoided when transferring the mononuclear cells.
10. The volume was adjusted to 50 mL by adding PBS to the tubes, reducing the density of the sample.
11. The tubes were centrifuged at 1800 rpm at RT for 10 min. The brake and acceleration were set to 9.
12. The supernatant was removed from both tubes.

13. The cells were resuspended in one of the tubes in 10 mL PBS and transferred to the other tube.
14. 10 mL PBS was added to the first tube to wash out remaining PBMCs and transferred to the rest to have a final volume of 20 mL.
15. The tubes were centrifuged at 1000 rpm at RT for 10 min to wash away any blood plates and to form a pellet.
16. The supernatant was carefully removed.
17. 20 mL PBS was added and the mixed by resuspending. 50 µL of the cell suspension was collected to count cells (as in section 3.1.4).
18. The tubes were centrifuged at 1100 rpm at RT for 10 min to remove all contaminating blood plates.
19. The supernatant was carefully removed without disturbing the pellet.

3.1.3 Isolation of $CD4_+$ T cells from PBMCs with EasySep Human $CD4_+$ T cell Isolation Kit

The EasySep Human $CD4_+$ T cell Isolation Kit (STEMCELL Technologies) can be used to easily and rapidly isolate purified, untouched and viable $CD4_+$ T cells from fresh or previously frozen human PBMCs. The kit is column-free and targets non-$CD4_+$ T cells to be removed by antibodies recognizing specific cell surface markers. The unwanted cells are marked with antibodies and magnetic particles and separated using an EasySep magnet, while the desired $CD4_+$ T cells are poured off into a separate tube.

Procedure

The isolation cocktail and the Rapid Spheres was kept on ice at all times.

1. The cell pellet was resuspended in isolation buffer (PBS with 2% FCS and 1mM EDTA) to get a final concentration of 5 x 10_7 cells/mL in a volume of 1 – 8,5 mL.
2. The sample was transferred to a 14 mL flow tube.
3. 50 µL Isolation Cocktail was added for each mL cell suspension and mixed well by pipetting.

4. The tube was incubated at RT for 5 min.
5. The RapidSpheres was vortexed for 30 seconds until particles appeared evenly dispersed.
6. 50 µL RapidSpheres was added for each mL of sample and mixed well by pipetting.
7. PBS with 2% FCS and 1 mM EDTA was added to a total volume of 5 mL (for samples less than 4 mL) or 10 mL (for samples greater than 4 mL). The cells were mixed well by pipetting up and down 2-3 times.
8. The tube without lid was placed in the magnet and incubated at RT for 3 min.
9. The tube was left in the magnet and inverted for 2-3 seconds in one continuous motion and the supernatant containing $CD4_+$ T cells was poured off in a 15 mL tube. The cells were kept on ice.
10. The $CD4_+$ T cells was counted as described in section 3.1.4.

3.1.4 Counting cells

Evaluating viability and the total number of cells in cell suspensions was performed using TC20™ Automated Cell Counter provided by Bio-Rad. By adding Trypan blue solution to the cell suspension, the cell counter discriminates viable from non-viable cells. Due to their intact plasma membrane, the Trypan blue will not stain live cells and they will appear white. In contrast, the defective plasma membrane of dead cells will enable Trypan blue to protrude through the plasma membrane and the cells will be stained blue.

Procedure

1. 10 µL Trypan blue was added to 10 µL cell suspension.
2. 10 µL of the mix was added to a TC20™ Automated Cell Counter chamber slide and inserted into the cell counter.
3. It was pressed "count cells" on the TC20™ Automated Cell Counter to calculate the number of living cells, total number of cells, as well as an estimate of the viability percentage.

3.1.5 Freezing of living cells

To freeze the $CD4_+$ T cells, the protective agent Dimethyl sulfoxide (DMSO) was used. DMSO is commonly used in freezing procedures of cells in culture. The presence of DMSO prevents cell death by protecting the cells against extracellular and intracellular development of ice crystals by lowering the freezing point of the medium. This way, the cells can be stored in -80°C without damage.

Procedure

1. Live cells were batched for freezing with DMSO according to standard procedure. $2x10_6$ cells was frozen down per tube.
2. The cells were spun down for 10 min at 1400 rpm in a 15 mL tube at slow acceleration (5).
3. The freezing solution was made (0,5 mL solution per tube to be frozen, always 10% more than needed):
    i. 20% DMSO
    ii. 80% Serum
    iii. Prepared by adding DMSO to serum on ice. The solution was kept on ice until use.
4. The supernatant was removed from the centrifuged cells and the pellet was resuspended in cold RPMI (0.5ml RPMI per tube to be frozen).
5. Cryotubes was placed on ice.
6. A drop of freezing solution was added one at the time, to the tube of RPMI containing the cells. The cells were kept on ice at all times. The tube was swirled during the procedure.
7. The solution was mixed, and 1 ml of the solution was transferred to chilled cryotubes on ice, before they were transferred to a freezing box at 4°C.
8. The box was placed in -80°C for minimum 24 hours before it was transferred to liquid nitrogen.

3.1.6 Thawing CD4+ T cells

When thawing CD4+ T cells from liquid nitrogen, the cells are slowly reheated on a water bath at 37 °C to maximize the amount of viable and undamaged cells and to maintain their functionality. The freshly thawed T cells are centrifuged to remove residual medium.

Thawing procedure

1. The cells were thawed in a 37°C water bath, rocking gently until some of the ice was melted.
2. The cells were transferred to a 15 mL tube with 2 mL RPMI-1640 medium at RT. 7 mL RPMI-1640 at RT was added. Some medium was used to wash the cryotube.
3. The cells were centrifuged for 7 min at 1500 rpm at RT.
4. All the supernatant was removed.
5. The cell pellet was gently resuspended in 1 mL warm (37°C) X-VIVO medium at RT and transferred to a 12-well plate. The remaining cells were washed out (1000 µL pipette) in 1 mL X-VIVO medium at RT and added to the plate.

3.1.7 Activation of CD4+ T cells

The CD4+ T cells were stimulated *in vitro* by culturing freshly thawed cells in plates coated with α-CD3 antibody and medium with soluble α-CD28 antibody. The cells were left unstimulated or stimulated on two different levels dependent on the amount of α-CD3 and α-CD28 antibody present in the well, i.e. high, low and non-activated cells (Table 3.1).

Table 3.1: *Dose of stimulating α-CD3 and α-CD28 antibodies present in the culturing medium during the CD4+ T cell activation.*

Antibody dose	α-CD3 concentration (µg/mL)	α-CD28 concentration (µg/mL)
High	5,0	2
Low	0,5	0,2
No	0	0

$CD4_+$ T cell activation procedure

Coating cell culture wells with α-CD3 antibodies

1. The desired volume of α-CD3 was prepared with a final concentration of 5µg/mL in PBS.
2. The appropriate volume was added to each well (35µL per well in a 96-well).
3. The plates were incubated at 37°C, 5 % CO_2 for 2 hours or at 4°C overnight.
4. PBS with antibody was carefully removed and the plate was gently washed with PBS before the cells were added.

Preparation of α-CD28 antibody solution

1. The desired volume of α-CD28 was prepared with a concentration of 4 µg/mL in X-VIVO medium. The final concentration when mixing the cells (see above) was 2 µg/mL.
2. The appropriate volume was added to each well.

$CD4_+$ T cell activation

1. Thawed $CD4_+$ T cells was transferred to a 15 mL tube.
2. The cells were centrifuged at 1200 rpm for 8 min at RT.
3. The supernatant was removed.
4. The cells were resuspended in X-VIVO medium to a concentration of 2 x 10^6 cells/mL.
5. The same volume of cells as the volume of α-CD28 already present in the well was added to a final concentration of 1 x 10^6 cells/mL and 2 µg/mL α-CD28.
6. The plate was incubated at 37°C, 5 % CO_2 for 24 hours.

3.2 Flow Cytometry

Flow cytometry is an instrument in which a cell in suspension pass through a light source, quantitatively and simultaneously analyzing multiple physical and chemical characteristics of the individual cell. The technology can determine size and granularity

of the cell (63), and is commonly used in immunological laboratories due to its high sensitivity and rapid results(64).

The principle of flow cytometry is based on scattering of light and fluorescence emission which occurs when light from a laser beam collides with moving particles. The scattering of lights can be directly associated with the morphological and structural properties of a cell. The forward scattered (FSC) light is measured along the same axis as the laser beam, and is proportional to the area of the cell-surface or the cell size, while the side scattered (SSC) light is measured at approximately 90° to the laser beam, and is proportional to the internal complexity or the granularity of the cell. Furthermore, the fluorescence emission descending from a fluorescence probe or an antibody labelled with a fluorochrome, is proportional to the amount of the probe or antibody bound to the particular cell (63). This fluorescent labelling enables distinction of specific cell types based on their cell-surface molecules.

Flow cytometry was applied by other group-members in the immunophenotyping of peripheral blood cells and for validation of the purity of isolated $CD4_+$ T cells. In this thesis, flow cytometry was applied for quantification and distinguishing between activated, non-activated, live and dead cells. The analysis of cells during the initial experiments was performed on the Attune Acoustic focusing cytometer coupled with an NxT auto sampler, whereas the samples to be sent for proteomic analysis was analyzed on FACSCanto™ II Flow Cytometer.

3.2.1 Staining cells for flow cytometry

When staining cells for analysis using flow cytometry, cell surface antigens located on the plasma membrane are commonly used to identify and characterize the cell type (Adan, 2017)(63). Although antibodies specifically bind to particular antigens, they might have interactions that are not associated with the antigen-binding domain, but still give rise to a background signal. To compensate for this background signal, a fluorescently conjugated isotype control is normally included for each sample. The isotype control antibody is labelled with the same fluorochrome as the antibody used to stain the main samples but is generated against non-specific and irrelevant antigens that are not present in the sample.

In this thesis, $CD4_+$ T cells were stained with CD69 antibody conjugated with the fluorochrome fluorescein isothiocyanate (FITC), along with a LIVE/DEAD Far Red cell dye. For each sample stained with α-CD69-FITC antibody, a parallel sample with an IgG-FITC isotype control was included. Following $CD4_+$ T cell activation, CD69 is one of the earliest cell surface antigens expressed, hence the activation state can easily be evaluated by analyzing the amount or the percentage of cells displaying CD69. The LIVE/DEAD cell dye is used to evaluate the viability of mammalian cells, where the reactive dye can permeate the membranes of necrotic cells and react with free amines in the cell interior and on the cell surface. This results in an intense fluorescent staining. In the viable cells, only the cell-surface amines are available to react with the dye, resulting in a relatively weak staining compared to necrotic cells.

In addition to the isotype control, a set of controls prepared by the same procedure were included to determine the settings on the flow cytometer. The samples and the controls were stained in parallel, the controls consisting of one unstained and two single stained samples. All samples were prepared in 96-well plates, with eight samples each. In each plate, one internal healthy control was included. To avoid excitation of the fluorescent antibodies, the procedure was performed protected from light. An overview of staining antibodies and final concentration is presented in table 3.2.

***Table 3.2**: Overview of antibodies used for staining of cells for flow cytometry, and their final concentration.*

Antibody	***Final concentration***
α-CD96-APC	100 µL/1 x 10_6 cells
α-IgG-APC	100 µL/1 x 10_6 cells
α-CD69-PE	100 µL/1 x 10_6 cells
α-CD69-FITC	100 µL/1 x 10_6 cells
α-IgG-FITC	100 µL/1 x 10_6 cells

Before starting the procedure, half of the cells for the LIVE/DEAD single stained control were killed by heating at 60 °C for 20 min.

Procedure

LIVE/DEAD Dye preparation

1. One vial of the fluorescent reactive dye (Component A) and the vial of anhydrous DMSO (Component B) was brought to room temperature before removing the caps.
2. 50 µL of DMSO was added to the vial of reactive dye, and mixed well until visual confirmation that all of the dye had dissolved.
3. The solution of reactive dye was used as soon as possible (see below), ideally within a few hours of reconstitution.
4. The unused portions could be used for up to 2 weeks if stored at –20°C, protected from light and moisture. Otherwise unused portions could be aliquoted and stored at -80°C, avoiding freeze-thaw cycles.

Cell staining

1. 100.000 cells were added to a v-bottom well of a 96-well plate.
2. The plate was centrifuged at 300 x g for 5 min before the supernatant was discarded.
3. The cells were washed with 150 µL PBS.
4. The plate was centrifuged at 600 x g for 30 s, and the supernatant was discarded.
5. The appropriate antibodies were diluted 1:10 in PBS. 100 µL was used per $1x10^6$ cells.
6. 10 µL antibody was added to each well.
7. The plate was incubated on ice for 30 min protected from light.
8. The plate was centrifuged at 600 x g for 30 s, before the supernatant was discarded.
9. The cells were washed with 150 µL PBS twice.
10. 1 µL of LIVE/DEAD reconstructed dye was added to 1 mL PBS for 1×10^6 cells.
11. The diluted LIVE/DEAD reconstructed dye was added to each well to a final concentration of 1×10^6 cells/mL.
12. The plate was incubated on ice for 30 min protected from light.
13. The plate was centrifuged at 600 x g for 30 s, before the supernatant was discarded.

14. The cells were washed with 150 µL PBS twice.
15. The cells were fixed by adding 20 µL PBS with 1% PFA.
16. The plate was incubated at RT for 15 min protected from light.
17. The plate was centrifuged at 600 x g for 30 s, before the supernatant was discarded.
18. The cells were resuspended in 300 µL PBS with 1% BSA.
19. Stained cells were stored at 4°C protected from light.

3.2.2 Data analysis

Following the flow cytometry analysis, a principle part of the analysis is gating populations or cells with characteristics of interest. By creating histograms and plots based on different parameters, the cell properties can be distinguished. Gates can be applied to these plots to positively select or exclude cell populations.

The strategy of gating in this thesis is based on evaluating the viability and the activation of the $CD4_+$ T cells. The cells are stained with a fluorescently labelled CD69 antibody, as well as a LIVE/DEAD dye, enabling the visualization of both live and activated cell present in the samples. The data from the initial experiments was analyzed using the Attune® Cytometric Software, whilst data obtained during the quality control was analyzed using the software FCS Express 6 Flow Cytometry. The gating strategy was identical using both analysis software's.

Gating strategy

1. The $CD4_+$ T cell population was identified and gated in a SSC/FCS density plot.
2. Live $CD4_+$ T cells was identified and gated in a single parameter histogram displaying LIVE/DEAD signal. The dead population gave rise to a more intense signal than the live population in the LIVE/DEAD channel, leading to two easily distinguished peaks in the histogram.
3. Among the live $CD4_+$ T cells, the CD69 positive cells was identified in a single parameter histogram displaying α-CD69 signal. The placement of the gate was determined by correction using the isotype control.

3.3 Cell lysis for Liquid chromatography tandem mass spectrometry analysis (LC-MS/MS).

A cell lysis is performed prior to the LC/MS-MS proteomic analysis to access the $CD4_+$ T cells proteome. The term lysis is referring to breaking down the cell membrane components, still keeping the integrity of the DNA and proteins within the cell.

In this thesis, activated $CD4_+$ T cells were lysed by the Proteomic Core Facilities at the University of Bergen (PROBE), using the following procedure.

Cell lysis procedure

The samples were distributed in seven 96 well plates, with 200.000 cells in each well.
The following was added to the RIPA buffer;

- 130 µL H_2O
- 10 µL NP40
- 10 µL 10 % SDS
- 100 µL 5 % Sodium deoxycholate (SDC)
- 100 µL 0,5M TrisHCl (pH 7.6)
- 150 µL 1M NaCl
- 500 µL 2x cOmplete protease inhibitor

1. Each 96 well plate was centrifuged for 1 min at 200 x G
2. Each well was added 40 µL ice cold RIPA buffer and left on ice for 15 min
3. Each plate was treated by ultrasonic bath for 4 x 30 seconds in ice cold water (without ice) and left on ice for 15 min
4. The plate was centrifuged for 1 min at 200 x G
5. Each sample was mixed by pipetting and transferred to LO-BIND tubes.
6. The tubes were centrifuged for 10 min at 13 000 rpm and 4 °C.
7. 5 µL of each sample was used for BCA (BCA standard 0-1-2-4-6-8-10)

3.4 Statistical analyses

The statistical analyses in this thesis is performed to evaluate whether the protein abundances provide any significant difference when comparing protein abundance in T

cells from MS patients compared to healthy controls. The statistical analysis was performed in Microsoft Excel (Microsoft) and GraphPad Prism 8.

Methods in descriptive statistics are used to describe a group of data, and include calculation of sample mean ($\bar{x}$) and standard deviation (SD) for the sample size *n*. The SD is a measure of variability or how data varies around the sample mean ($\bar{x}$) for the sample size *n*.

3.4.1 F-test for equality of two variances

In statistics, an f-test is defined as any test that apply an f-distribution. The f-test is used to compare if two variances (s_1, s_2) provide the two-sided probability that the variances in two groups are not significantly different. As variances are always positive, the result of the test is always a positive number.

The significance level is the probability of rejecting the null hypothesis, when the null hypothesis is true. In this test, the null hypothesis states that the variances are equal. As the data of protein abundances are distributed in both positive and negative directions, a two tailed f-test was performed, meaning that the region of rejection is on both sides of the sampling distribution. The variance of the groups and the f-test was calculated using the equations 3.1 and 3.2. A p-value < 0,05 was considered significant.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \tag{3.1}$$

$$F = \frac{s_1^2}{s_2^2} \tag{3.2}$$

3.4.2 Student's t-test and Mann Whitney U test

In statistics, a Student's t-test is a method to evaluate significance of differences between groups, generally conducted in a one-tailed or a two-tailed manner. By comparing the means of two groups, the test provides the probability that the group difference could have appeared by chance. For these analyses the assumption is that the data is normally distributed. In a Mann Whitney U test, however, the assumption states that the

observations are not normally distributed and the test is non-parametric. This test is used to compare differences between two independent groups.

In this thesis, a two-tailed Students t-test is performed to evaluate the possible difference in protein abundance between MS cases and controls. In the case of a significant f-test (section 3.4.1), a Student's t-test with unequal variances was used. If the f-test did not return a significant p-value, a Student's t-test with equal variances was used. The Student's t-test was additionally used to compare the differences between individuals carrying the MS risk variant of a SNP and individuals carrying the non-risk variant of the same SNP. In this genotype-dependent expression analysis, an unpaired Student's t-test with equal variances was conducted if the data was normally distributed. For data that was not normally distributed, a Mann Whitney U test was performed.

3.4.3 Fischer's exact test

The Fischer's exact test is used to determine if there is an association between two categorical variables. A significant Fischer's exact test calculates the probability that the given distribution of observations can occur by chance.

In this thesis, the Fisher's exact test is essentially preformed to evaluate enrichment of proteins expressed by MS susceptibility genes among proteins differentially expressed between MS patients and healthy controls.

3.4.4 Correlation of protein abundance with MS risk genotype

Correlation is a statistical measure indicating how strong pairs of variables are related and if they are related. In this study, the correlation between the genotype at MS risk SNPs and protein abundance was assessed. All genes present in a 100 kB area upstream and downstream from the each of the 200 MS associated SNPs were extracted, using the "biomart" function in Ensembl Genome Browser. Proteins that was expressed in $CD4_+$ and $CD8_+$ T cells was sorted separately and according to the proteomics dataset of $CD4_+$ and $CD8_+$ T cell expression from the published paper (54). An f-test and a Student's t-test were performed as described in section 3.4.1 and 3.4.2, to find proteins differentially expressed between MS patients and healthy controls. The proteins conferring a differential protein expression was excluded in the further correlation analysis. Genotype data from all

homozygous and heterozygous carriers of the risk allele of the SNPs was pooled, and compared to heterozygous carriers of the non-risk allele. The samples in each of the two groups were divided independently of the individual being a MS case or a healthy control. The correlation was assessed through an unpaired Student's t-test with equal variances (for normally distributed data) or a non-parametric Mann U Whitney test, as described in section 3.4.2.

3.5 Pathway analysis

Pathway analysis is generally utilized for identification of genes or proteins that are related within specific pathways, and the analysis is valuable for studying differential expression of a gene or a protein. Databases of known pathways and interaction networks, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome databases can be explored to find protein or genes involved in pathways based on specific organisms or cells.

The STRING online database of known and predicted protein-protein interactions, is a tool based upon predictions of genomic context, high-throughput laboratory experiments, co-expression, text mining and previous knowledge in databases. Searches in this tool provide network maps and statistics that can be exported and used in further analysis.

In this thesis the STRING database combined with the KEGG database was used to find pathways enriched among protein differentially expressed in $CD4_+$ and $CD8_+$ T cells combined. All STRING searches were performed with medium confidence level (0.40), and with experiments, databases, co-expression, neighborhood and gene fusion as interaction sources.

4 - Results

The overall purpose of this study was to unravel potential proteomic differences in T cells from RRMS cases and healthy controls, and to identify potential pQTLs. In a previous study (54), proteomic profiles of freshly isolated $CD4_+$ and $CD8_+$ T cells from RRMS patients and healthy controls were analyzed to get insights into immune processes in MS. A number of differentially expressed proteins were detected, as well as several pathways involved in T cell activation. Additionally, three pQTLs for proteins encoded by genes most proximal to MS risk SNPs were identified. In the current thesis, the aim was to prepare and analyze proteomic profiles of activated $CD4_+$ T cells from MS cases and healthy controls, to study the activation effect on immune-cell processes and the impact on MS disease.

Additionally, extended analyses of the existing proteomics data set was performed to seek functional links between proteins and MS disease. In this thesis, the search region surrounding each MS risk SNP was expanded, not only including the most proximal genes to identify additional pQTLs. The results and data from the proteomic study of $CD4_+$ and $CD8_+$ T cells were utilized (54).

4.1 Analysis of available proteomics dataset

Proteomic profiles of $CD4_+$ and $CD8_+$ T cells from MS cases and healthy controls were analyzed in a previous study conducted by other members of our lab (54). The data sets of protein abundances from this study were available for further analysis in this thesis.

4.1.1 Differentially expressed proteins in both $CD4_+$ and $CD8_+$ T cells

The study detected a total of 228 proteins in $CD4_+$ T cells and 195 proteins in $CD8_+$ T cells as differentially expressed between MS patients and healthy controls. The analysis was performed on the cell types separately, whilst this study aimed to study the 26 proteins that were differentially expressed in both $CD4_+$ and $CD8_+$ T cells.

4.1.1.1 Pathway analysis of differentially expressed proteins in $CD4_+$ and $CD8_+$ T cells

The previous study performed an ingenuity pathway analysis (IPA), on proteins showing differential expression between MS cases and healthy controls in $CD4_+$ and $CD8_+$ T cell

data sets, separately. The study identified 14 biological processes in $CD4_+$ T cells affected by MS disease, while no pathways were significant for $CD8_+$ T cells. In the current study, the STRING online database was utilized to examine if the 26 proteins that were differentially expressed in both cell types were connected to any specific biological pathways or mechanisms (section 3.5). A list of these proteins with corresponding p-values and fold change is presented in appendix C, table S6.

The IDs of the genes encoding the 26 differentially expressed proteins were imported into the STRING Multiple protein tool (confidence level=0,4) with experiments, databases, co-expression, neighborhood and gene fusion as interaction sources (Fig. 4.1). The network displayed 10 out of 26 nodes interacting, where one node is represented by one protein. The network contained seven node edges, representing the end node of each collection of interacted nodes, whereas the expected number of edges calculated by STRING was at three. This indicated that the network had more node edges than to be expected if the proteins were selected randomly. In addition, the significant ($p<0.05$) enrichment p-value of 0.044, indicated that the proteins were somewhat biologically connected as a group. The functional enrichments in the network based on gene ontology (GO), indicated common cellular components between the proteins, while the citrate cycle was noted as an enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway.

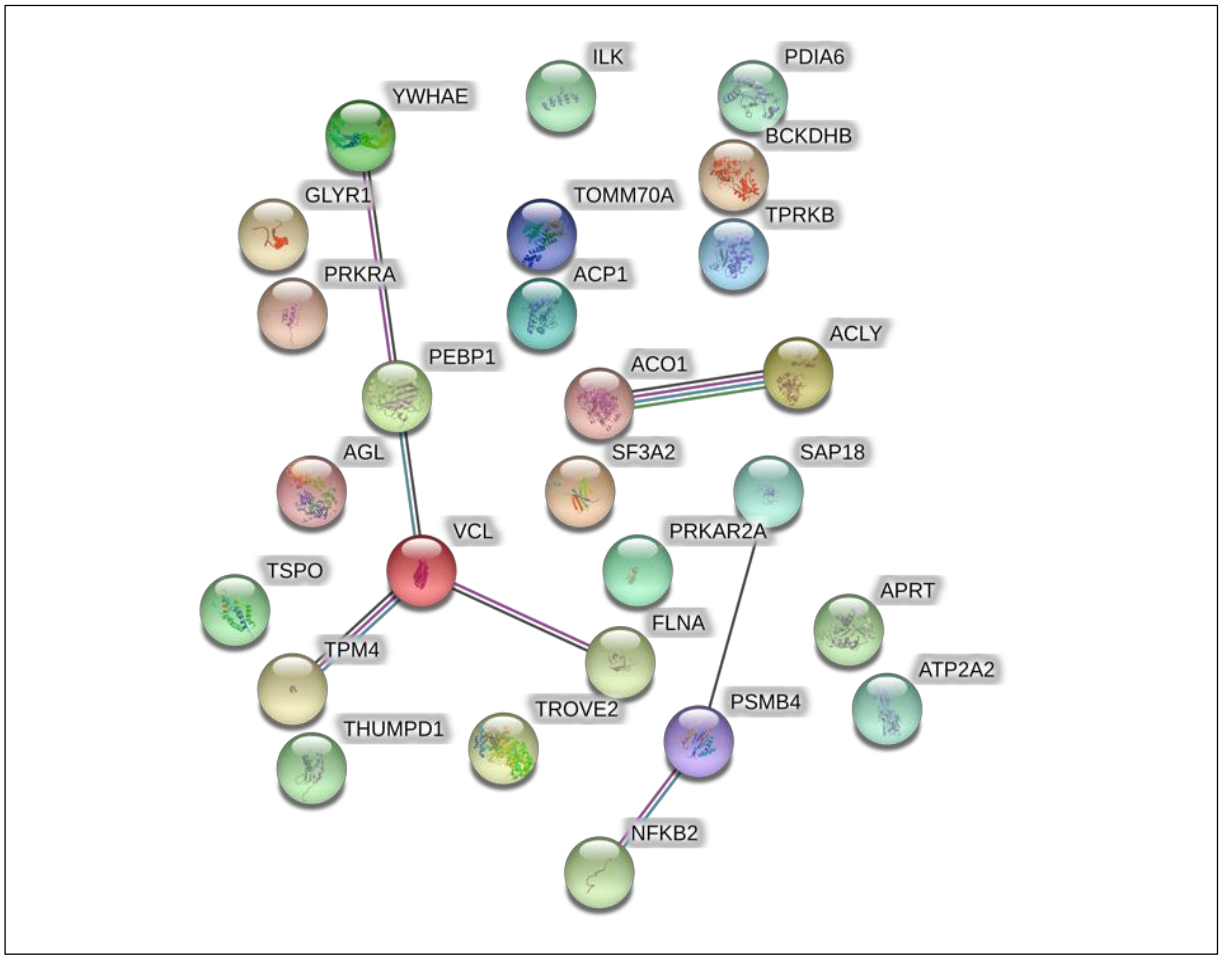

Figure 4.1: STRING network of proteins differentially expressed in both CD4+ and CD8+ T cells. *26 proteins differentially expressed between MS patients and healthy controls in both CD4+ and CD8+ T cells (54) were imported into STRING Multiple protein tool. The confidence level was set to 0,4 and experiments, databases, co-expression, neighborhood and gene-fusion as interaction sources. Each node represents a protein, with the predicted protein structure in 3D illustrated in each node. Known interactions are illustrated by lines representing interactions experimentally determined (pink) and interactions from curated databases (light blue). Predicted interactions are illustrated by lines representing interactions based on gene neighborhood (green). Other interactions can be visualized as lines representing interaction based on co-expression (grey).*

As a control measure, a random selection of 26 proteins that were not differentially expressed between MS cases and healthy controls in neither CD4+ or CD8+ T cells in our data sets was imported into the STRING database and analyzed by the same interaction sources and confidence level (Fig. 4.2). In this network, 16 out of 26 nodes were interacting, while there were 12 node edges. The expected number of edges was 7,

indicating that there were more node edges than to be expected by random protein selection. As expected, the enrichment p-value was not significant (p=0.067), indicating that the proteins were a random selection of proteins that were not very well connected. The functional enrichments in the network based on gene ontology indicated common biological processes, molecular functions and cellular components, however at low counts. Essentially, there were no big differences in the number of nodes comparing the network of differentially expressed proteins with the network of randomly selected proteins.

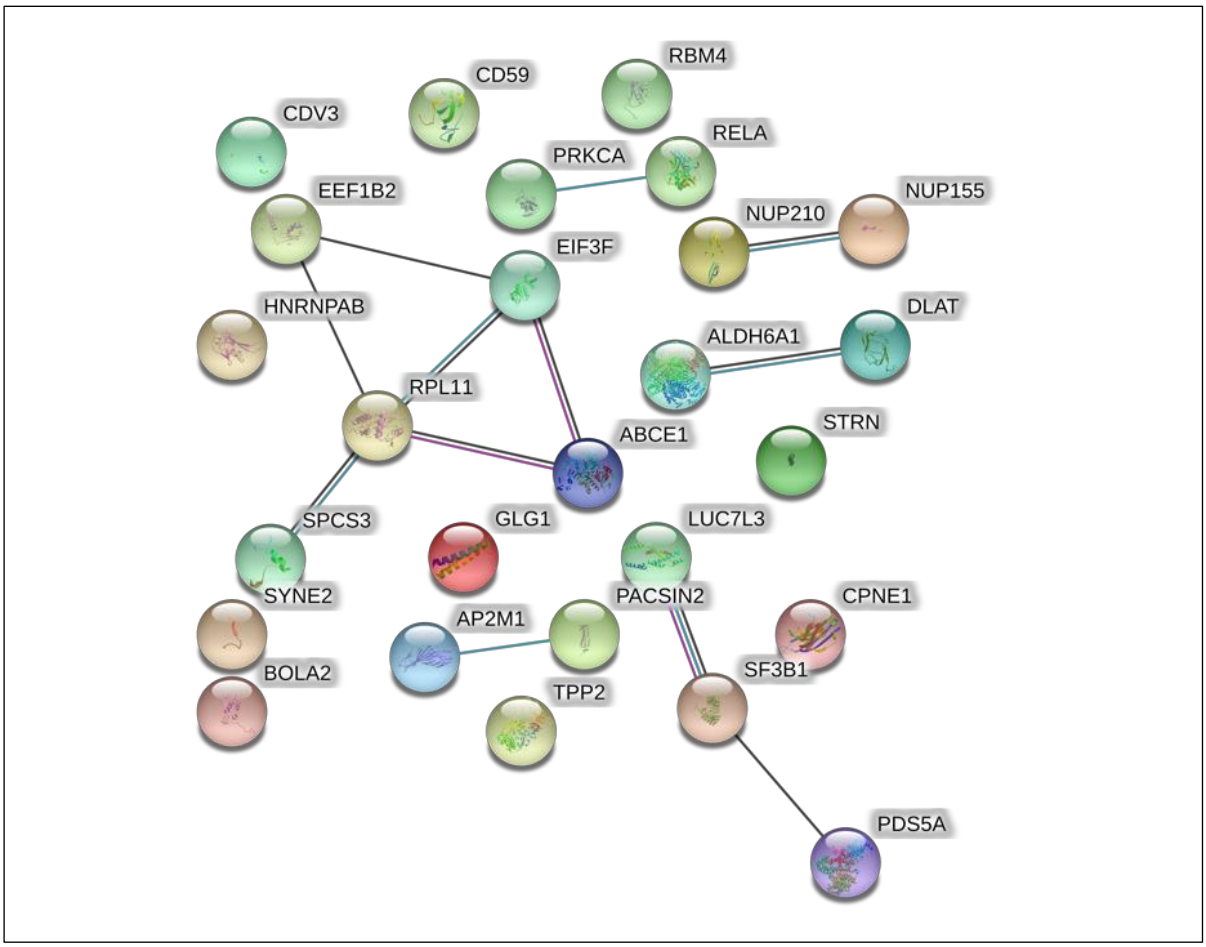

Figure 4.2: STRING network of proteins not differentially expressed in both CD4+ and CD8+ T cells. *26 randomly selected proteins not differentially expressed between MS patients and healthy controls in both CD4+ and CD8+ T cells (54) were imported into STRING Multiple protein tool. The confidence level was set at 0,4 and experiments, databases, co-expression, neighborhood and gene-fusion as interaction sources. Each node represents a protein, with the predicted protein structure in 3D illustrated in each node. Known interactions are illustrated by lines representing interactions experimentally determined (pink) and interactions from curated databases (light blue). Predicted interactions are illustrated by lines representing interactions based on gene neighborhood (green). Other interactions can be visualized as lines representing interaction based on co-expression (grey).*

4.1.1.2 Statistical testing of differentially expressed proteins in CD4+ and CD8+ T cells

To extract more information on the presence of the differentially expressed proteins, a two-sided Fischer's Exact test was performed (section 3.4.3).

The first test was conducted with the aim to determine if there were an enrichment of differentially expressed proteins (between MS patients and healthy controls) than expected by chance (Table 4.1). It was tested if there was a significant difference in the distribution of differentially expressed proteins among proteins expressed by both CD4+ and CD8+ T cells, than expected by chance. The test did not result in a significant ($p<0.05$) p-value ($p=0.051$), pointing in the direction that among proteins expressed in both CD4+ and CD8+ T cells, the distribution of proteins differentially expressed between MS cases and healthy controls, were no different than the distribution expected by random chance.

Table 4.1: No significant difference in the distribution of differentially expressed proteins among proteins expressed in both CD4+ and CD8+ T cells. *The table presents the data used in the Fisher's exact test, with the number of differentially and non-differentially expressed proteins (between MS cases and healthy controls) in both CD4+ and CD8+ T cells, along with the corresponding p-value.*

		CD8+ T cells		*Total*
		Not DE (p>0.05)	DE (p<0.05)	
CD4+ T cells	Not DE (p>0.05)	1543	131	1674
	DE (p<0.05)	194	26	220
Total		1737	157	1894

Fischer Exact test p-value
0,051

4.1.2 Analysis of proteins encoded by genes located in expanded region surrounding MS risk SNPs

In order to identify proteins with a possible impact on MS disease, the published study (54) selectively analyzed the abundance of proteins expressed from the genes most proximal to the 200 non-HLA associated MS risk SNPs, i.e. MS susceptibility genes. For the intergenic MS risk SNPs, the abundance of proteins expressed from the most proximal gene both upstream and downstream of the SNPs were analyzed. This search for proteins

encoded by genes close to MS risk SNPs was expanded in the current project, i.e. analyzing genes in the 100kb region upstream and downstream from the MS-associated SNPs (Fig. 4.3).

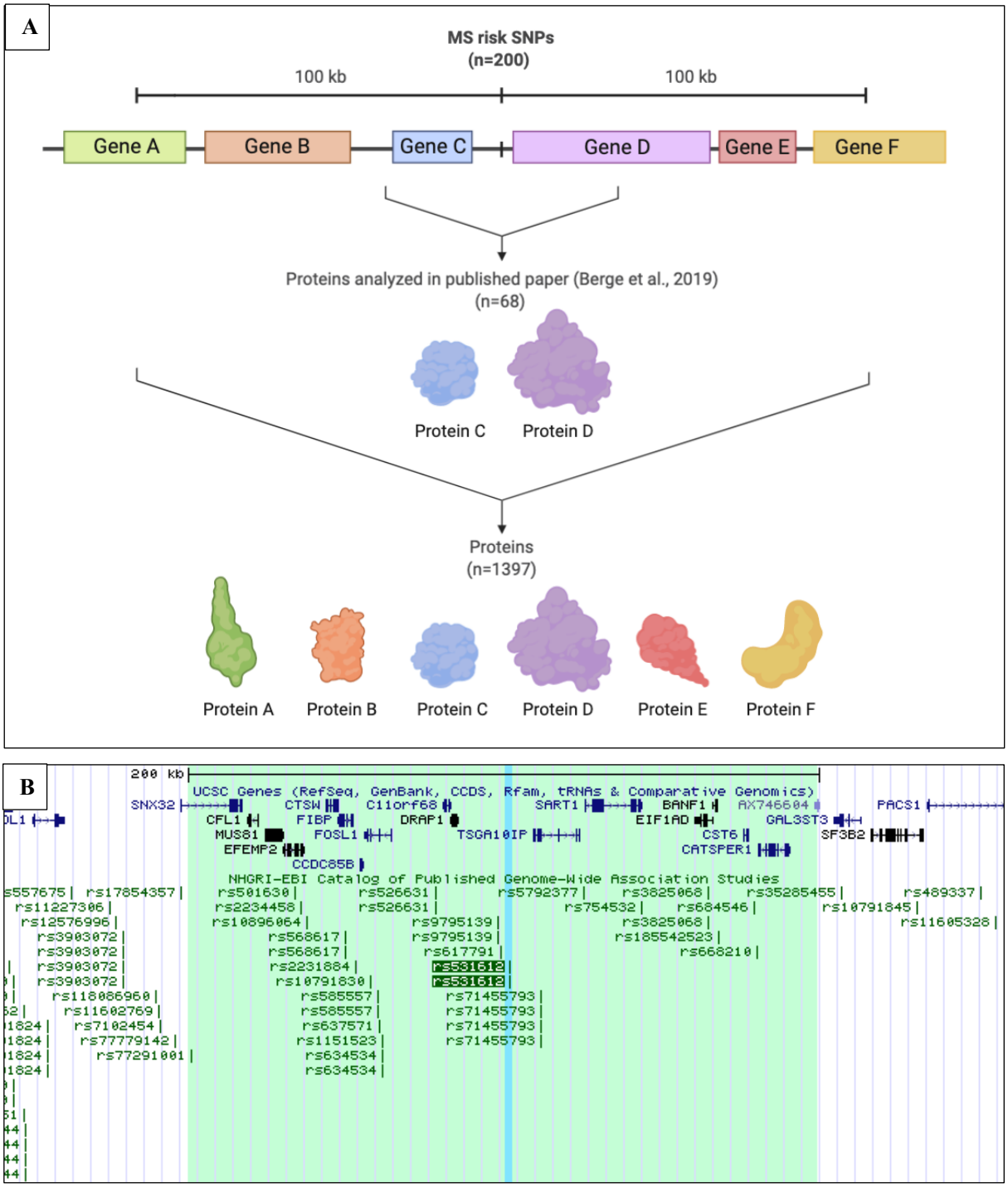

Figure 4.3: Illustration of search region surrounding an MS risk SNP, e.g. rs531612. *(A) The figure is an illustration of the area enclosing one of the 200 non-HLA MS risk SNPs, where the given area comprise gene A-F. In a published paper (54), the expression of proteins encoded by the most proximal genes to the MS risk SNP was analyzed, represented by protein C and D. In the current thesis, the search region was expanded to 100kb upstream and downstream from each SNP, resulting*

in the expression analysis of the proteins represented by protein A-F. (B) Presentation of the MS risk SNP rs531612 (dark green and blue line), with an illustration of the expanded search region (light green; 100kb upstream and downstream of rs531612). Proteins encoded by these genes were analyzed in our data sets. Genes are indicated in dark blue and black lines, and the SNPs located nearby in dark green. Figure (A) was designed in BioRender, figure B was retrieved from (65).

A list of 1397 proteins were extracted from the expanded region, of which 83 individual proteins were expressed in $CD4_+$ T cells, and 101 individual proteins were expressed in $CD8_+$ T cells, according to the available dataset. Of the 83 proteins expressed in $CD4_+$ T cells, 13 were differentially expressed in samples from MS patients and healthy controls (Appendix C, Table S7), while six of the 101 proteins were differentially expressed in $CD8_+$ T cells (Fig. 4.4 and appendix C, Table S8).

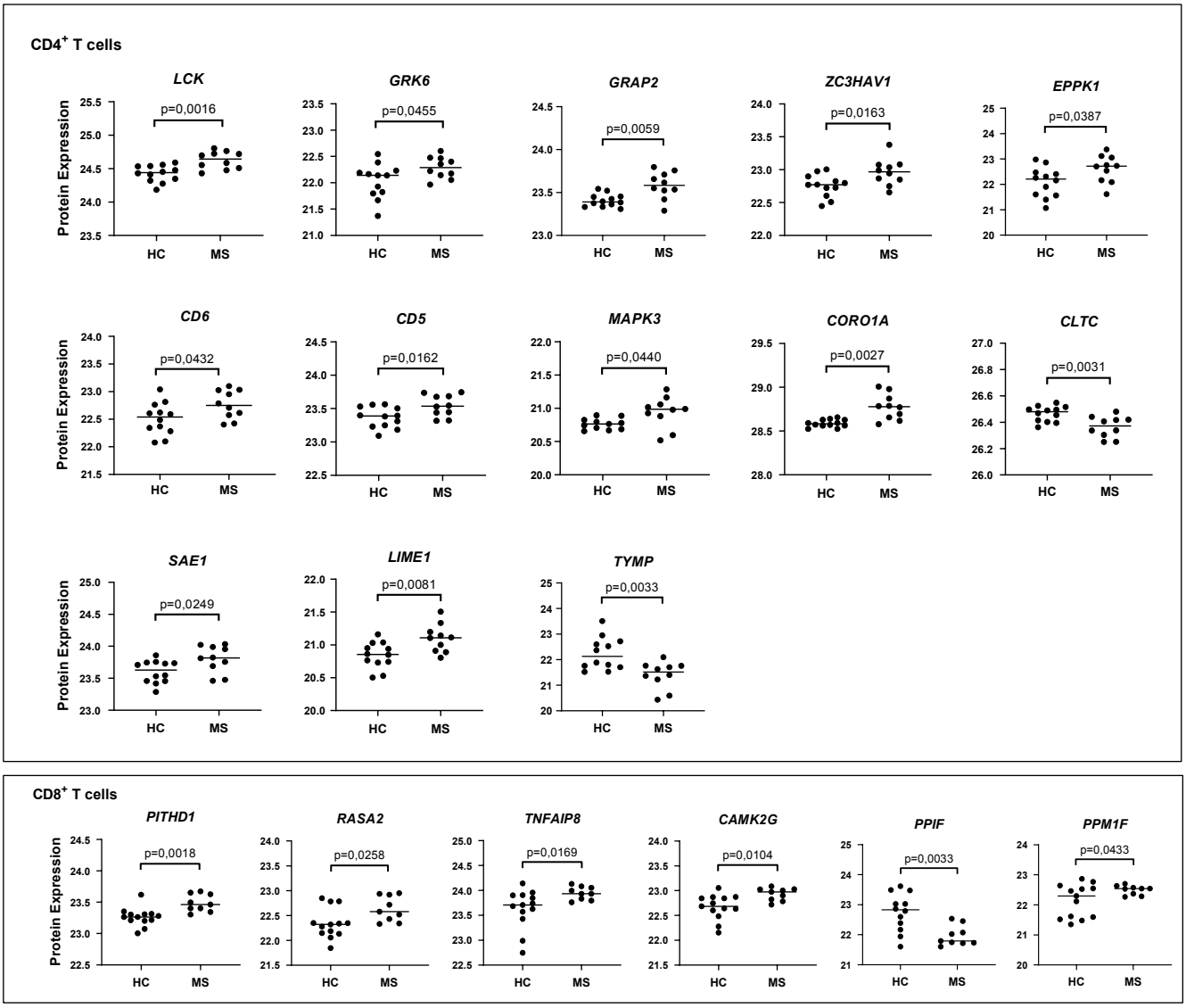

Figure 4.4. Genes encoding proteins differentially expressed in CD4+ and CD8+ T cells from MS patients and healthy controls. *The plots present log2-transformed protein abundances of proteins encoded by the indicated genes present in a 100kb region upstream and downstream from MS risk SNPs in CD4+ and CD8+ T cells from MS patients and healthy controls (HC). A two-tailed Student's t-test with unequal variances was used to compare the protein abundances of the groups if the f-test was significant ($p<0.05$). A Student's t-test with equal variances was used otherwise. The horizontal lines indicates the median within the groups.*

4.1.2.1 Correlation of MS risk genotype and protein expression

To detect a potential functional link between the MS risk SNPs and disease, an evaluation of the correlation between the MS risk genotypes and the expression of proteins encoded by genes located in the expanded region surrounding each SNP was conducted. Any SNP

with a genotype significantly correlating with the expression of proteins encoded by genes in the expanded region could act as a potential pQTL.

Most of the proteins that were extracted from the extended region did not display a significant difference in abundances when comparing MS cases to healthy controls, i.e. 70 and 95 proteins were not differentially expressed in $CD4_+$ and $CD8_+$ T cells, respectively. The genotype data of the SNPs proximal to these non-differentially expressed proteins were collected (from MS patients and healthy controls). Individuals that were homozygous and heterozygous carriers of the risk allele at each SNP were merged and compared to samples from individuals homozygous for the non-risk allele for each SNP(section 3.4.4).

In $CD4_+$ T cells, the previous study identified a significant correlation between genotype of two MS risk SNPs and the expression of the protein encoded by *LEF1* and *STAT3* (54)(Appendix C Fig. S1). No additional pQTL candidates in $CD4_+$ T cells were discovered among the list of proteins encoded by genes in the expanded region. In $CD8_+$ T cells, there was a known correlation between the genotype of an MS risk SNP and the expression of the protein encoded by *RUNX3*, identified in the published study (54)(Appendix C, Fig. S2). In our expanded protein list, the genotype of rs1800693 and rs137955 were significantly correlating with the expression of proteins encoded by *CD9* and *UQCRFS1P1,* respectively, in $CD8_+$ T cells (Fig. 4.5).

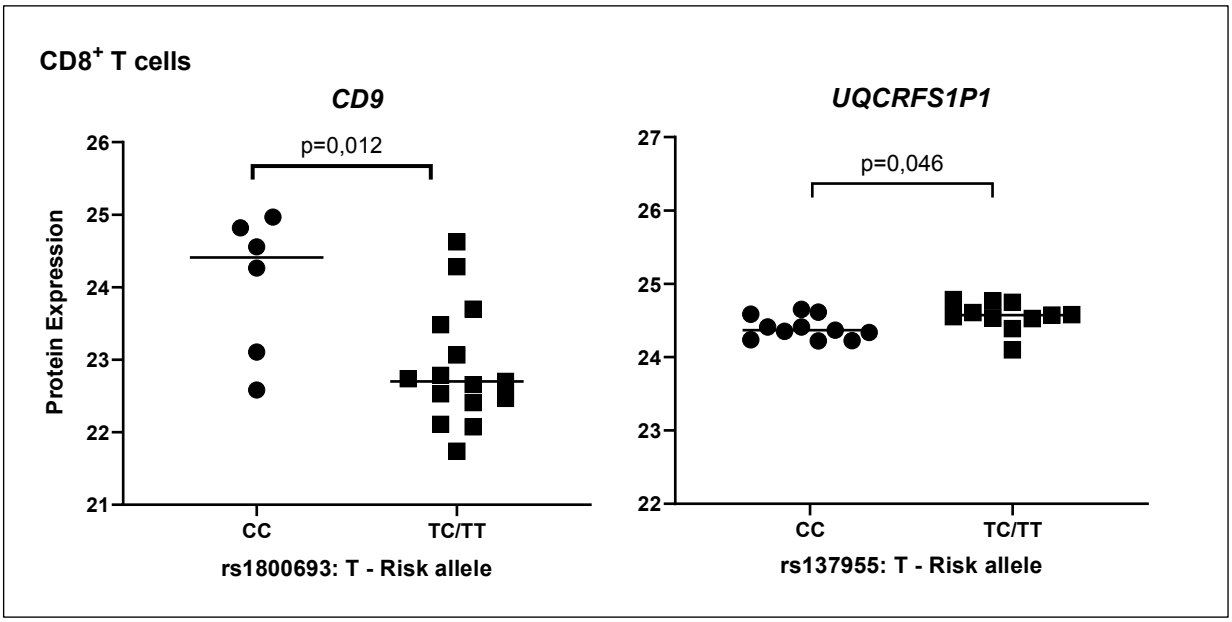

***Figure 4.5**: **Genotype-dependent expression of proteins in CD8+ T cells encoded from genes located in a 100kb upstream and downstream region from MS risk SNPs.** The scatterplot present log2-transformed protein intensities of CD8+ T cell proteins expressed from genes in the selected region as a function of the MS risk SNP genotype in samples from MS cases and healthy controls. The protein intensities are sorted according to groups, i.e. carriers or non-carrier of the risk allele. An unpaired students t-test was performed, with the p-value illustrated in each plot. The horizontal lines in each group represent the median within the group*.

4.1.2.2 Statistical testing of MS risk genes

It was examined if there were an enrichment of differentially expressed proteins encoded by MS risk genes, and additionally if there were an enrichment of differentially expressed proteins encoded by genes located in the expanded region.

To assess this, Fischer's Exact tests were conducted on the basis of proteins expressed in CD4+ T cells and CD8+ T cells, separately. For proteins expressed by CD4+ T cells, the test calculated the probability that the distribution of differentially expressed proteins and the distribution of proteins encoded by MS risk genes were different than expected by chance (Table 4.2). The test did not result in a significant p-value (p=0.072), leading towards no enrichment of proteins encoded by MS risk genes among differentially expressed proteins in CD4+ T cells. In CD8+ T cells, an identical test was performed (Table 4.2). This test did not return a significant p-value (p=0.368), indicating that among the

differentially expressed proteins in CD8+ T cells, there was no enrichment of proteins encoded by MS risk genes.

Table 4.2: No enrichment of differentially expressed (DE) proteins between MS patients and healthy controls among proteins encoded by MS risk genes, in CD4+ or CD8+ T cells *The table presents data used in Fisher's exact test, with the distribution of the number of differentially expressed proteins and the number of proteins encoded by MS risk genes, among proteins expressed by CD4+ (top) and CD8+ (bottom) T cells, with the corresponding p-value.*

		MS risk gene		*Total*
		No	Yes	
CD4+ T cells	Not DE (p>0.05)	1772	23	1795
	DE (p<0.05)	223	7	230
Total		1995	30	2025

Fischer's Exact test p-value
0,072

		MS risk gene		*Total*
		No	Yes	
CD8+ T cells	Not DE (p>0.05)	2022	36	2058
	DE (p<0.05)	194	1	195
Total		2216	37	2253

Fischer's Exact test p-value
0,368

As comparison, Fischer's Exact tests were conducted on the basis of the proteins of the extended search, separately in proteins expressed in CD4+ and CD8+ T cells. In proteins expressed in CD4+ T cells, the test calculated the probability that the distribution of differentially expressed proteins and the distribution of proteins encoded by genes located in the extended region, were different than expected by chance (Table 4.7). The test did not result in a significant p-value (p=0.215), indicating that among proteins differentially expressed in CD4+ T cells, there was no enrichment of proteins encoded by genes of the expanded region.

An identical test was performed in CD8+ T cells (Table 4.3) and did not result in a significant p-value (p=0.466). This indicated no enrichment of proteins encoded by genes in the expanded region among proteins differentially expressed in CD8+ T cells.

Table 4.3: No enrichment of proteins differentially expressed (DE) between MS patients and healthy controls among proteins encoded by genes located in the extended region surrounding MS risk SNPs, in CD4+ or CD8+ T cells. *The table presents data used in Fisher's exact test, with the distribution of the number of differentially expressed proteins and the number of proteins encoded by genes present the extended region (i.e. 100kb upstream and downstream from MS risk SNPs), among proteins expressed by CD4+ (top) and CD8+ (bottom) T cells, and the corresponding p-value.*

		Extended region		*Total*
		No	Yes	
CD4+ T cells	Not DE (p>0.05)	1725	70	1795
	DE (p<0.05)	217	13	230
Total		1942	83	2025

Fischer's Exact test p-value
0.215

		Extended region		*Total*
		No	Yes	
CD8+ T cells	Not DE (p>0.05)	1963	95	2058
	DE (p<0.05)	189	6	195
Total		2152	101	2253

Fischer's Exact test p-value
0,466

By comparing the tests performed on proteins expressed in CD4+ T cells and tests performed on proteins expressed in CD8+ T cells, the p-values were lower for proteins expressed by MS risk genes, compared to proteins encoded by genes present in the extended region surrounding the MS risk SNPs.

4.2 Sample preparation of activated CD4+ T cells from MS patients and healthy controls

This part of the project aimed at collecting samples from activated CD4+ T cells to be analyzed with LC/MS-MS at PROBE. CD4+ T cells from RRMS patients and healthy controls had previously been isolated by other members of our lab, and live cells had been frozen on liquid nitrogen after cell purification. These cells were thawed and left unstimulated or stimulated *in vitro* with low or high doses of activating antibodies. A PhD student in our lab, had previously tested the thawing procedure and performed several tests of the activation procedure on freshly isolated CD4+ T cells. In this thesis, the thawing procedure was performed in collaboration with this student. The activation procedure was additionally tested at the start of the current project, with the purpose to contribute to the selection of staining reagents to apply in the quality control by flow cytometry. After activation, the cells were centrifuged and stored at -80 °C until all samples were collected and thereupon sent to PROBE for sample preparation and proteomic profiling by LC/MS-MS (Fig. 4.6).

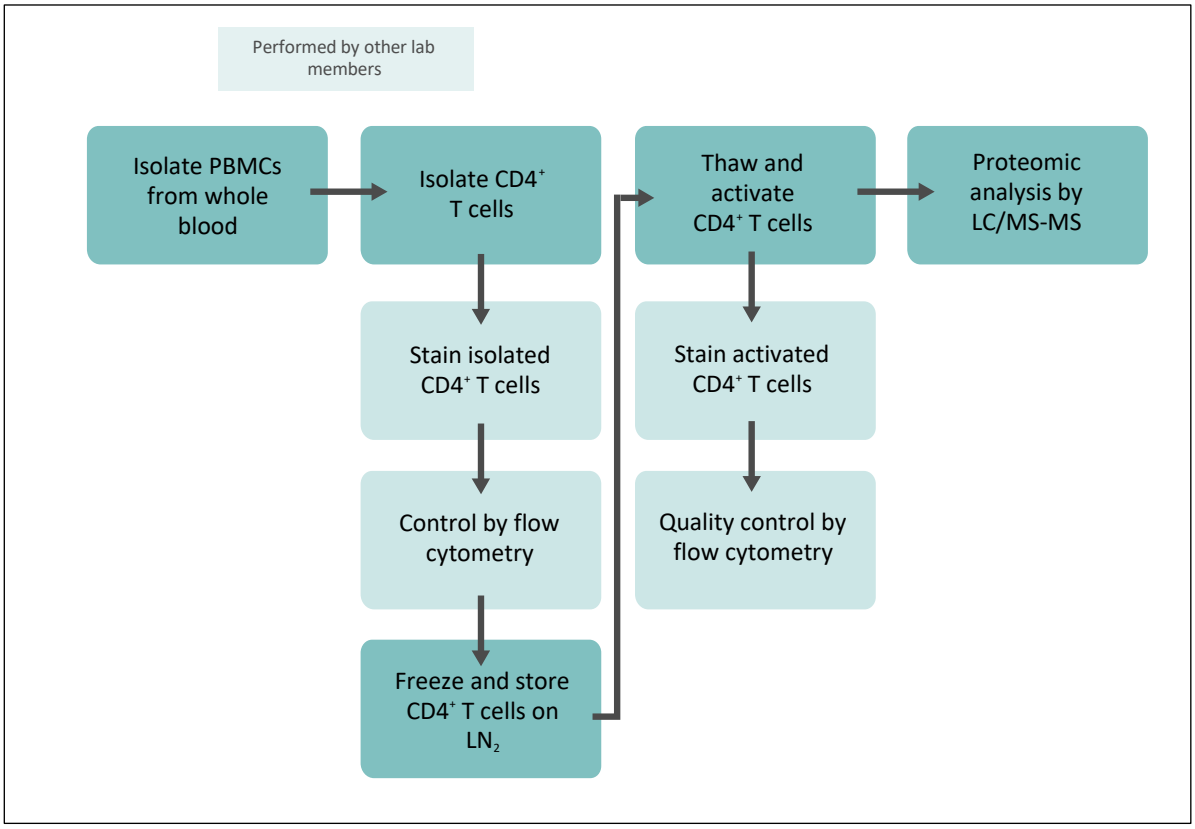

Figure 4.6: Overview of section 4.2; Sample preparation of activated CD4+ T cells from MS patients and healthy controls. *PBMCs were isolated from blood, followed by CD4+ T cell isolation by magnetic separation. Flow cytometry was performed to evaluate the purity of the cells, before the live CD4+ T cells were frozen down and stored on liquid nitrogen (LN_2). CD4+ T cells were thawed and activated in vitro and parts of each sample were collected for quality control analysis, by analyzing cell viability and T cell activation by flow cytometry. Samples passing the quality control of cell viability and cell activation were shipped for proteomic analysis by LC/MS-MS. Light green boxes indicate control steps. Procedures performed previously by other members of the lab are indicated in the area on the left side of the figure. Whereas procedures performed in this thesis is shown on the right side of the figure. The figure was designed using Biorender.*

4.2.1 Establishment of staining protocol for the quality control

T cell activation is routinely being performed in the lab. The activation and thawing procedures were previously tested by a PhD student in our group, and tested three times in the current project. The two initial activation tests were performed on the same healthy donor, whilst the last test on a different healthy donor. The purpose of these tests was to establish the protocol to use in staining of the activated CD4+ T cells, during the quality control.

All activation tests were performed using the same procedure (section 3.1.7 and 3.2), but with different staining reagents as specified below (Table 4.4). The $CD4_+$ T cells were thawed and left unstimulated or stimulated *in vitro* for 24 hours with two different doses of α-CD3 and α-CD28 antibodies, i.e. high and low dose of antibodies. The cells were harvested and stained with a fluorescently conjugated α-CD69 antibody and a LIVE/DEAD stain, accompanied in parallel by fluorescently conjugated isotype controls.

Table 4.4: Overview of staining reagents applied in initial tests of $CD4_+$ T cell activation. *The table presents the three activation tests with corresponding staining reagents applied in for samples and isotype controls.*

	Activation test 1	**Activation test 2**	**Activation test 3**
Staining	α-CD69-APC LIVE/DEAD Red	α-CD69-PE LIVE/DEAD Far Red	α-CD69-FITC LIVE/DEAD Far Red
Staining isotype control	α-IgG-APC LIVE/DEAD Red	N/A	α-IgG-FITC LIVE/DEAD Far Red

The initial activation test displayed that the percentage of CD69 expressing cells increased with the dose of α-CD3 and α-CD28 antibodies present in the culturing medium, where 80 % of the cells stimulated with high dose of antibodies were CD69 positive (Fig. 4.7). During the flow cytometry analysis, compensation was required to cancel out the overlapping signals between the α-CD69-APC antibody and the LIVE/DEAD Red staining. Because there were other staining options available for the use of two staining combinations simultaneously, the activation test was repeated with different staining reagents.

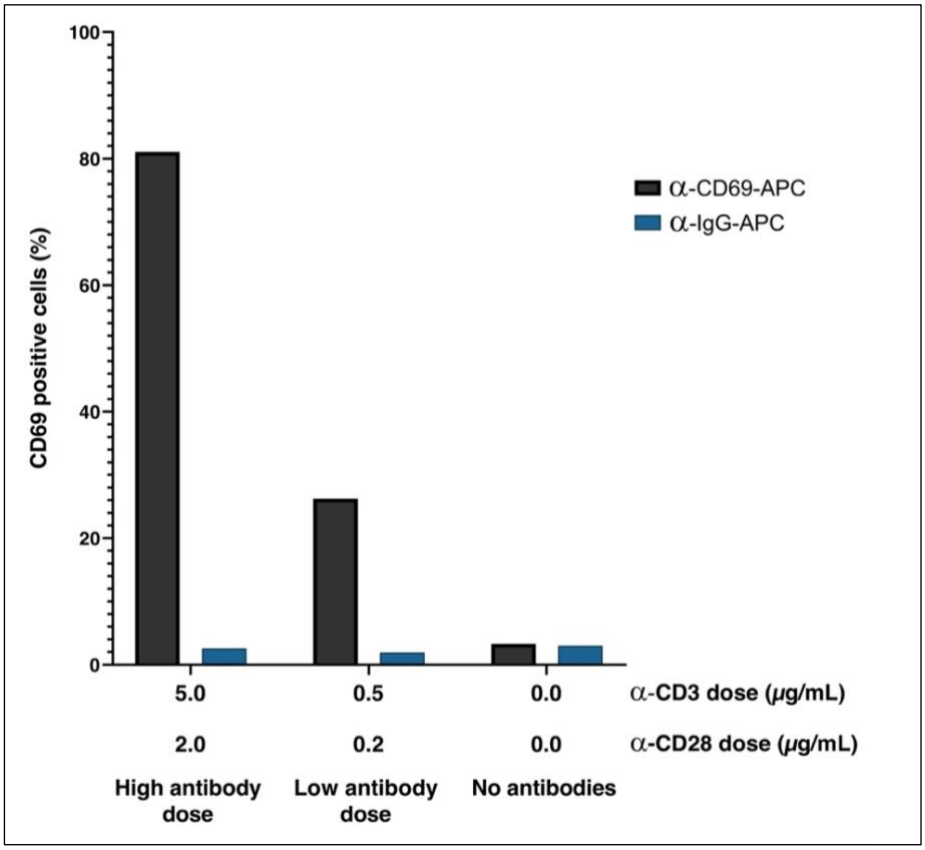

Figure 4.7: The percentage of CD69 positive CD4+ T cells increases with the dose of stimulating antibodies. *CD4+ T cells were left unstimulated or stimulated with the indicated amount of α-CD3 and α-CD28 antibodies for 24 hours. The cells were harvested, stained with α-CD69-APC and LIVE/DEAD Red, with corresponding isotype controls stained with α-IgG-APC and LIVE/DEAD Red, and analyzed by flow cytometry. The graph display the percentage of CD69 positive cells in live CD4+ T cells stimulated as depicted.*

The next activation test was performed to see if the results of the previous test were reproducible, in addition to testing a different staining combination to avoid the extensive need for compensation. As observed in the previous test, the percentage of CD69 positive cells increased with the higher dose of α-CD3 and α-CD28 added to the medium (Fig. 4.8). In cells stimulated with high dose of antibodies, 66 % were CD69 positive. Even though this test and the previous test evaluated cells from the same donor, some degree of variation in the percentage of CD69 positive cells were observed by comparing to the previous test. The compensation procedure in this test was not as extensive as the previous, though compensation was still necessary. Despite this observation, staining strategies without compensation requirements were explored.

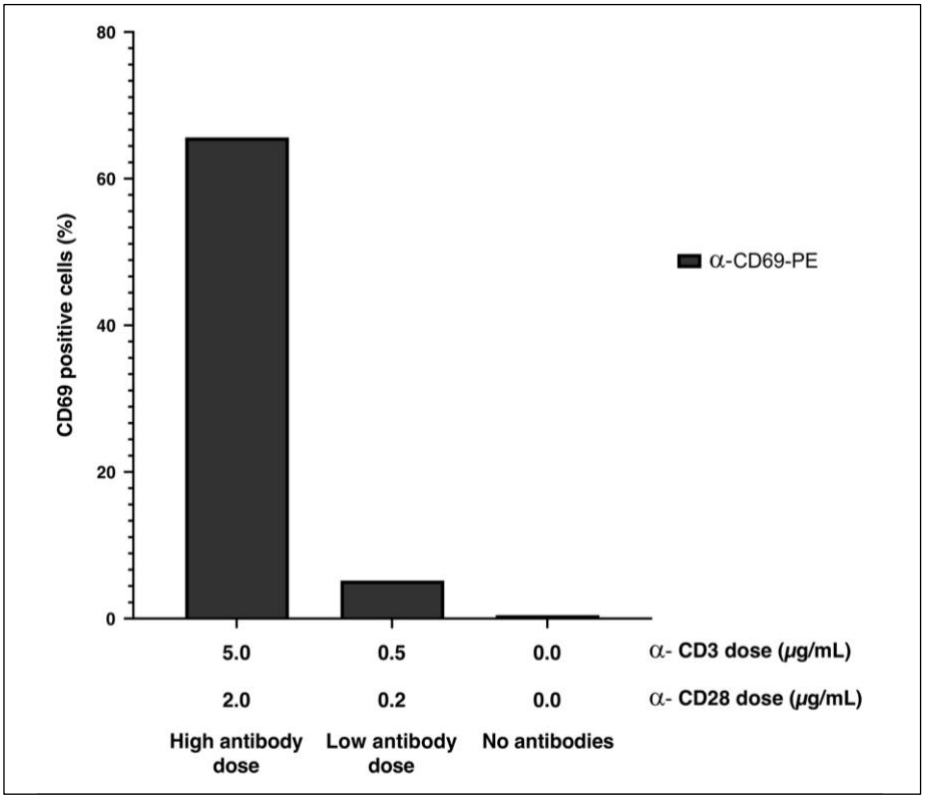

Figure 4.8: The percentage of CD69 positive CD4+ T cells increased with the dose of stimulating antibodies. *CD4+ T cells were left unstimulated or stimulated with the indicated amount of α-CD3 and α-CD28 for 24 hours. The cells were harvested, stained with α-CD69-PE antibody and LIVE/DEAD Far Red and analyzed by flow cytometry. Half of the cells for the LIVE/DEAD Far Red single stained control was heated to induce cell death prior to staining. The graph displays the percentage of CD69 positive cells in live CD4+ T cells stimulated as depicted. The histograms show (B) CD69 expression in single stained control cells after compensation and (C) LIVE/DEAD Far Red single stained control signal after compensation.*

The third activation test was performed on a different healthy donor than the previous activation tests (Fig. 4.9). As expected, the percentage of CD69 positive cells increased with the dose of α-CD3 and α-CD28 in the medium (Fig 4.9A). However, the percentage of CD69 positive cells in the sample with high antibody dose were observed at 43%, which was considerably low compared to the other activation tests.

It was not necessary to compensate when combining α-CD69-FITC with LIVE/DEAD Far Red due to the distance of the detection channels in the emission spectrum. Figure 4.9 present the single α-CD69-FITC stained control (B) and the LIVE/DEAD Far Red single

stained control (C). As expected, no overflow was observed between the LIVE/DEAD Far Red channel (RL1) and the α-CD69-FITC channel (BL1).

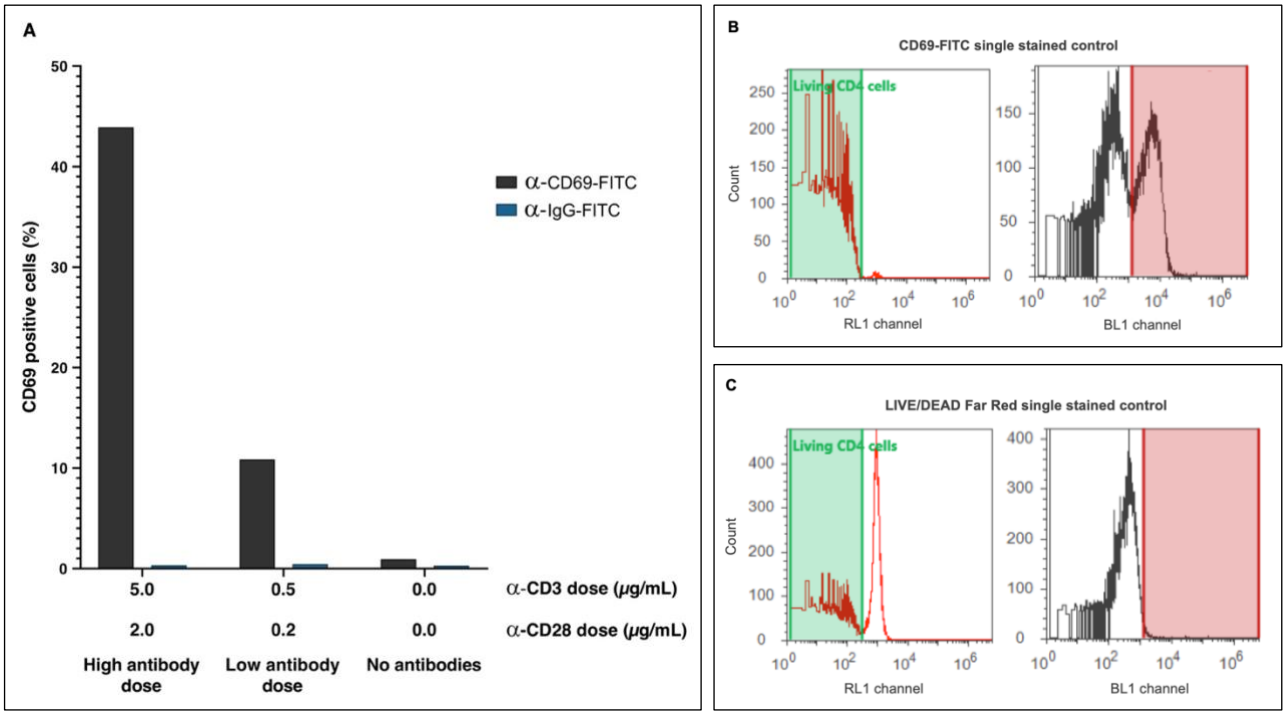

Figure 4.9: Increasing percentage of CD69 positive CD4+T cells with higher dose of stimulating antibodies and no overlap between the α-CD69-FITC and the LIVE/DEAD Far Red channels. *CD4+ T cells were left unstimulated or stimulated in vitro for 24 hours with α-CD3 and α-CD28 in the culturing medium. The cells were harvested, stained with α-CD69-FITC and LIVE/DEAD Far Red stain and analyzed by flow cytometry. Half of the cells for the LIVE/DEAD Far Red single stained control was heated to induce cell death prior to staining. (A) The graph displays the percentage of CD69 positive cells in live CD4+ T cells stimulated as indicated in the figure. The histograms show signal in the RL1 channel (detecting LIVE/DEAD Far Red signal) and the BL1 channel (detecting α-CD69-FITC signal) with (B) expression of CD69 in single stained control cells and (C) LIVE/DEAD Far Red single stained control.*

With the observations described above taken into account, it was confirmed that the pairing of α-CD69-FITC antibody and LIVE/DEAD Far Red stain were reasonable choice of stainings. Consequently, this staining pair was applied when performing the quality control by flow cytometry (section 3.2).

4.2.1.1 Evaluation of washing procedure

Upon collection of cells to be shipped to LC/MS-MS analysis, it was preferable to ship an many cells as possible to achieve sufficient concentration of proteins upon cell lysis. During activation of the cells, antibodies were added to the media. To test if remains of these antibodies affect the results from mass spectrometry, a test was conducted to analyze cells that were harvested by centrifugation without washing and cells that were washed prior to collection.

The samples in this test were prepared as the previous tests, by thawing a tube of isolated $CD4_+$ T cells from a healthy donor, before leaving the cells unstimulated or performing *in vitro* stimulation by culturing for 24 hours in medium containing α-CD3 and α-CD28 (section 3.1.6 and 3.1.7). The cells were harvested directly or washed by spinning down the cells before gently removing the medium. An overview of samples prepared and analyzed by LC/MS-MS analysis are presented in table 4.5.

Table 4.5: Samples analyzed in evaluation of the washing procedure in the $CD4_+$ T cells activation. *The table display samples of $CD4_+$ T cells that were analyzed, along with the dose of stimulating antibodies added to the culturing medium, description of how the cells were treated after stimulation, and the number of protein detected in the liquid chromatography tandem mass spectrometry(LC/MS-MS) analysis.*

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**	**Sample 5**
Antibody dose	High	High	Low	No	High
Treatment	Harvested directly	Washed	Washed	Washed	Medium only
Number of proteins detected	4823	4959	4426	4924	353

Among the proteins detected by LC/MS-MS analysis in the samples, none of the antibodies used for stimulation were detectable. Essentially, approximately the same number of proteins were detected and quantified in samples harvested directly compared to

samples that were washed. In cells harvested directly, 4823 proteins were detected, while 4959 proteins were detected in washed cells. Of these, 4741 proteins were detected both in cells harvested directly and in washed cells, while 196 and 57 proteins were unique for cells harvested directly and washed cells, respectively (Fig. 4.10). As expected, the majority of the protein identities were detected both in cells harvested directly and in washed cells.

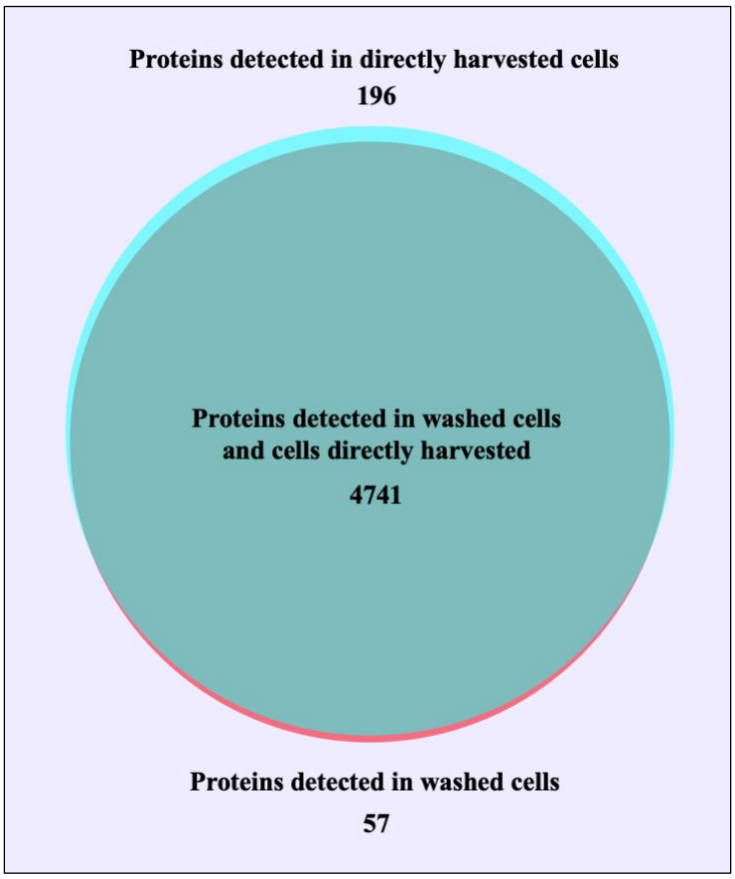

Figure 4.10: Venn diagram comparing the proteome of cells harvested directly and washed cells. *The diagram display the number of proteins detected only in CD4+ T cells harvested directly (blue), the number of proteins detected only in washed cells (red) and the number of proteins detected in both directly harvested and washed cells (green). The diagram was created in Biovenn (66).*

When comparing the proteome of CD4+ T cells stimulated with high antibody dose to the proteome of unstimulated cells, the majority of the protein identities (4809) were detected in both stimulated and unstimulated cells (Fig. 4.11). A minimal difference in the number of detected proteins in stimulated cells compared to unstimulated cells was observed, as 4959 proteins were detected in stimulated cells and 4924 proteins were detected in unstimulated cells. Of these, 128 proteins were detected only in stimulated cells, while 86 proteins were only detected in unstimulated cells.

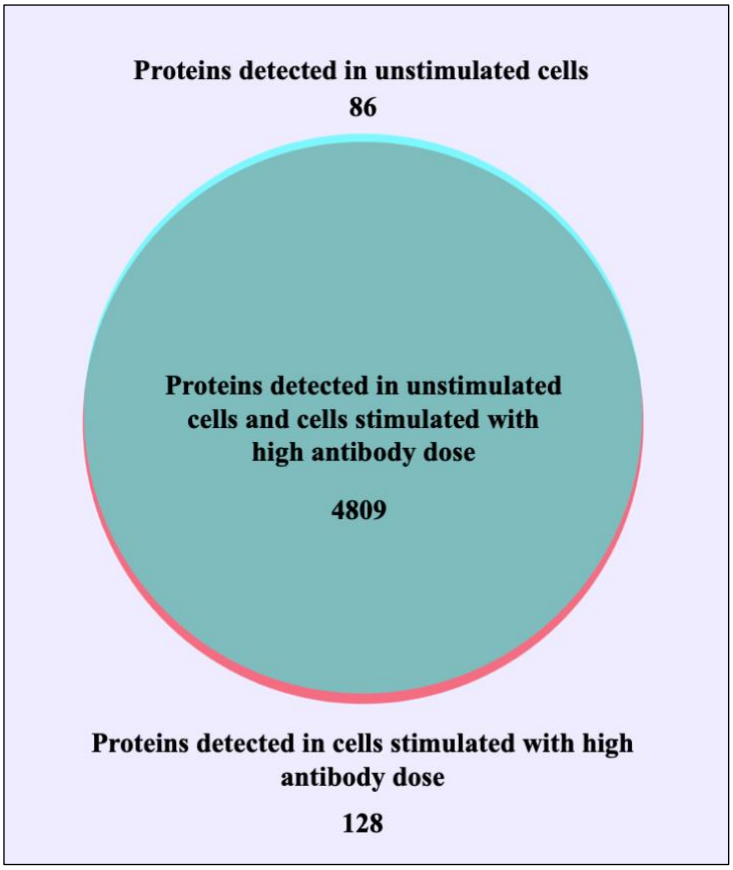

Figure 4.11: Venn diagram comparing the proteome of cells detected in stimulated and unstimulated CD4+ T cells. *The diagram display the proteins only detected in unstimulated CD4+ T cells (blue), proteins only detected in CD4+ T cells stimulated with high antibody dose (red) and the proteins detected in both stimulated and unstimulated CD4+ T cells (green). The diagram was created in Biovenn (66).*

In conclusion, proteins in the medium were not found to interfere with the detection of proteins by LC/MS-MS in the CD4+ T cells, and washing the cells did not result in changes in the number of detected proteins. Directly harvesting the cells without washing was therefore considered sufficient in further analysis.

4.2.2 Sample collection

As the optimization of the washing procedure in the stimulation of CD4+ T cells and the choice of staining to use in the quality control were complete, the collection of samples to bring into the LC/MS-MS proteomic analysis was gathered. The aim of the quality control was to investigate and exclude samples with insufficient cell activation and low cell viability, ensuring good quality of the samples to be analyzed by LC/MS-MS.

Initially, CD4+ T cells from 20 RRMS patients and 20 healthy controls were thawed (section 3.1.6), left unstimulated or stimulated *in vitro* for 24 hours in medium containing high or low doses of α-CD3 and α-CD28 (section 3.1.7). The cells were harvested, before portions of each sample was stained and analyzed by flow cytometry (section 3.2). All

samples were stained with α-CD69-FITC and LIVE/DEAD Far Red stain, accompanied in parallel by isotype controls stained with α-IgG-FITC and LIVE/DEAD Far Red stain.

Presented in table 4.6 are the percentage of live and activated cells in the internal healthy controls stimulated with high dose of antibodies. Only cells from the same donor was used as internal controls in the first batch of thawed samples, while cell from two different donors was used in the second batch. A small variation in the percentage of live cells and CD69 positive cells was observed in cells from the same donor. As expected, a slightly bigger variation was observed in samples with cells from different donors. The percentage of CD69 positive cells in batch 2 was observed to be low.

Table 4.6: *The table presents internal control (HC) samples stimulated with high antibody dose, with corresponding percentage of live and CD69 positive cells detected by flow cytometry. The cells in samples in Batch 1 was from the same donor, while cells in samples of Batch 2 were from different donors.*

	Control	*Live cells (%)*	*CD69 positive cells (%)*
Batch 1	**HC100**	85	62
	HC101	83	60
	HC102	89	74
	HC103	85	70
	HC104	81	73
Batch 2	**HC105**	83	42
	HC106	78	55

The following standards of viability and activation were set for the included samples to fulfill. Among samples unstimulated and stimulated with high antibody dose, a threshold of 70 % live cells were set. In addition, samples stimulated with high antibody dose with <50 % activated cells were excluded. Because 14 of the initially thawed samples did not fulfill one or more of the inclusion standards that were set, a new batch of $CD4_+$ T cells from eight RRMS patients and eight healthy controls were thawed (Table 4.7). Among samples stimulated with low antibody dose, 39 of the 56 thawed samples displayed a low percentage of activated cells (<10 %). Consequently, all samples that were stimulated

with low antibody dose were included and lysed, but not analyzed by LC/MS-MS at this time.

Table 4.10: *The table displays all thawed and analyzed CD4+ T cell samples from MS patients (left) and healthy controls (right) stimulated with high antibody dose, with the percentage of live and activated cells obtained from flow cytometry analysis. Batch 1 indicates the initial 40 samples that were thawed, whereas Batch 2 represents the following 16 samples that were thawed. (N/A- not applicable)*

	Patient	*Live cells (%)*	*Activated cells (%)*	*Control*	*Live cells (%)*	*Activated cells (%)*
Batch 1	**MS1**	95	45	**HC1**	89	52
	MS2	92	55	**HC2**	88	60
	MS4	69	60	**HC3**	30	38
	MS5	88	34	**HC4**	89	54
	MS6	89	50	**HC5**	93	48
	MS7	82	50	**HC6**	79	55
	MS9	75	43	**HC8**	90	75
	MS10	68	41	**HC12**	86	58
	MS12	88	62	**HC13**	87	67
	MS15	85	63	**HC14**	91	57
	MS16	72	55	**HC15**	88	53
	MS20	75	60	**HC16**	83	64
	MS21	93	48	**HC19**	65	53
	MS22	92	49	**HC20**	83	69
	MS23	86	59	**HC21**	83	67
	MS24	94	52	**HC22**	75	67
	MS25	97	63	**HC23**	11	33
	MS26	92	51	**HC25**	63	55
	MS27	91	63	**HC27**	85	71
	MS28	79	47	**HC28**	N/A	N/A
Batch 2	**MS3**	61	63	**HC7**	87	50
	MS8	68	65	**HC9**	92	63
	MS11	75	70	**HC10**	23	55
	MS13	65	50	**HC11**	90	53
	MS14	67	65	**HC17**	61	61
	MS17	69	64	**HC18**	68	58
	MS18	84	75	**HC24**	66	66
	MS19	67	57	**HC26**	55	63

An example of one of the thawed samples is HC20, a sample from a healthy control stimulated with high dose of antibodies. The results from the flow cytometry quality

control analysis (Fig. 4.12) displayed 83 % live CD4+ T cells, which was above the set threshold of 70% live cells. In addition, the 68 % of the live cells expressed CD69, which is above the activation threshold of 50 %. On the basis of these observations, the HC20 sample was included in the list of samples to be analyzed by LC/MS-MS.

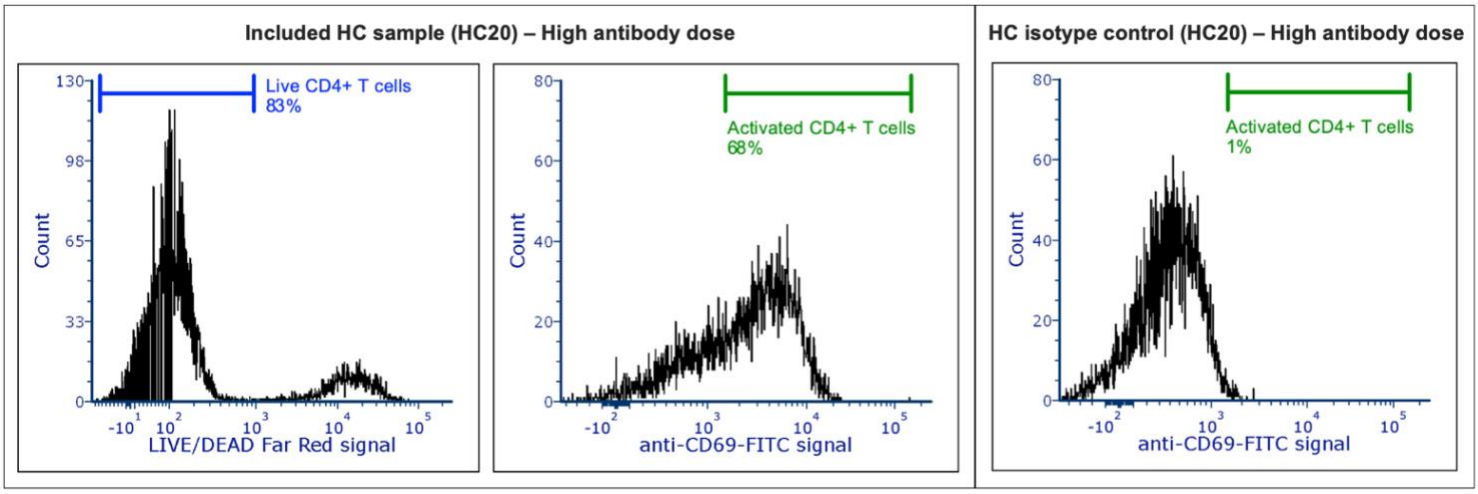

Figure 4.12: Healthy control sample fulfilling the inclusion criteria. *CD4+ T cells were stimulated in vitro for 24 hours with a high dose of α-CD3 and α-CD28 in the culturing medium. The cells were harvested and stained with α-CD69-FITC and LIVE/DEAD Far Red accompanied by a parallel isotype control sample stained with α-IgG-FITC and analyzed by flow cytometry. The histograms display the LIVE/DEAD Far Red signal (left) and the α-CD69-FITC signal (middle) in cells from the healthy control sample (HC20), as well as the α-CD69-FITC signal in the corresponding isotype control (right). The markers indicates the live CD4+ T cells (blue) and the live and activated CD4+ T cells (green).*

A different example of a thawed sample is the MS patient, MS5, where the cells had been stimulated with high dose of antibodies. Results from the flow cytometry quality control analysis (Fig. 4.13), displayed a sufficient percentage of live CD4+ T cells at 88 %. However, the sample displayed only 34% activated cells, and was therefore excluded for further analysis.

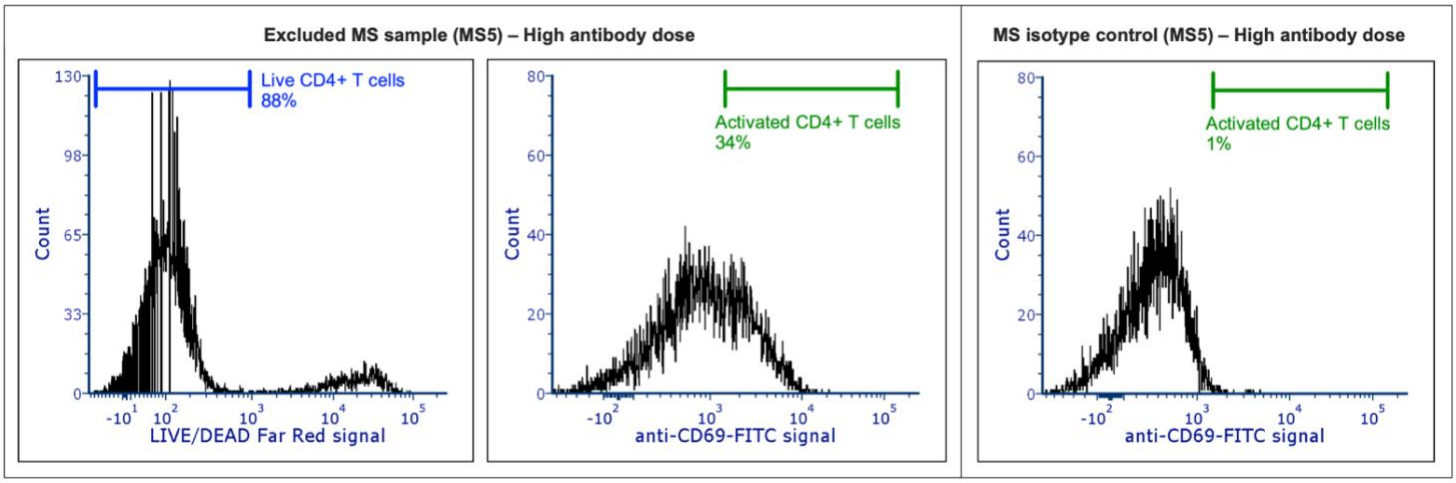

Figure 4.13: MS patient sample excluded from LC/MS-MS analysis due to insufficient activation. *$CD4_+$ T cells were stimulated in vitro for 24 hours with a high dose of α-CD3 and α-CD28 in the culturing medium. The cells were harvested and stained with α-CD69-FITC and LIVE/DEAD Far Red accompanied by a parallel isotype control sample stained with α-IgG-FITC and analyzed by flow cytometry. The histograms display the LIVE/DEAD Far Red signal (left) and the α-CD69-FITC signal (middle) in cells from MS patient (MS5), as well as the α-CD69-FITC signal in cells of the corresponding isotype control (right). The markers indicates the live $CD4_+$ T cells (blue) and the live and activated $CD4_+$ T cells (green).*

Samples from 20 MS patients and 20 healthy controls fulfilled the criteria (Appendix C, Table S9), and were shipped to Bergen for LC/MS-MS analysis. The summarized data of the key information about the included MS patients and healthy controls is presented in table 4.14. Data for the individual MS patients and a complete list of thawed samples is presented in the appendix C, Table S10 and S11, respectively. All samples that passed the standards of the quality control were included in the list of samples to be analyzed by LC/MS-MS. All thawed and analyzed samples were lysed and the protein concentration was measured (Appendix C, Table S12). The original plan was to analyze the proteomics results within the first weeks of March 2020, but due to the Covid-19 situation, the LC/MS-MS analysis was put on hold. On this note, it was not possible to analyze the LC/MS-MS results within the time schedule of the current thesis.

	Age	EDSS	Years since disease onset	Age at disease onset	Heterozygous at HLA-DR*B1:1501	Homozygous for non-risk variant
MS patients mean/median (range)*	43 (29 - 72)	1.5* (0 - 5.5)	15 (4 - 47)	28 (14 - 38)	15	5
Healthy controls mean (range)	44 (28 - 57)	N/A	N/A	N/A	8	12

Table 4.14. *The table presents the summarized statistics of the included MS patients and healthy controls with mean (range) of age, years since MS disease onset and age at MS disease onset, median* (range) of EDSS, number of individuals heterozygous for the HLA-DRB1*1501 loci and individuals homozygous for the non-risk variant. (EDSS – expanded disability status scale, N/A – not applicable)*

4.2.3 Analysis of quality control data

The flow cytometry data retrieved from the quality control of activated $CD4_+$ T cells, was utilized in further analysis. The purpose of these analyses was to evaluate the effect of storage time on the viability and the activation of the cells, as well as to find potential activation differences in cells from MS patients and healthy controls.

4.2.3.1 Effect of storage months on cell viability and activation

The $CD4_+$ T cells stored on liquid nitrogen had been collected between 2011 and 2016. To analyze whether the storage time on liquid nitrogen had an impact on cell viability or cell activation, a linear regression test was performed. Furthermore, it was analyzed whether there was any difference in cell activation between MS patients and healthy controls.

A linear regression test was performed by assessing the relationship between the number of storage months and the detected percentage of live $CD4_+$ T cells (Fig. 4.14A). The analysis showed no correlation between the number of months stored on liquid nitrogen and the cell viability in the $CD4_+$ T cells that were left unstimulated or stimulated with high or low antibody dose, with p=0.192, 0.115 and 0.103, respectively. This result

implied that there was no relationship between CD4+ T cell viability and the number of storage months in liquid nitrogen.

An identical test was implemented to assess the relationship between the number of storage months on liquid nitrogen and the percentage of CD69 positive cells (Fig. 4.14B). The linear regression test showed no correlation between the number of storage months and the percentage of CD69 expressing CD4+ T cells that were left unstimulated or stimulated with high or low antibody dose, with p=0.617, 0.114 and 0.784, respectively. Based on this result, it was implied that the number of months stored on liquid nitrogen did not have an impact on the CD4+ T cell activation.

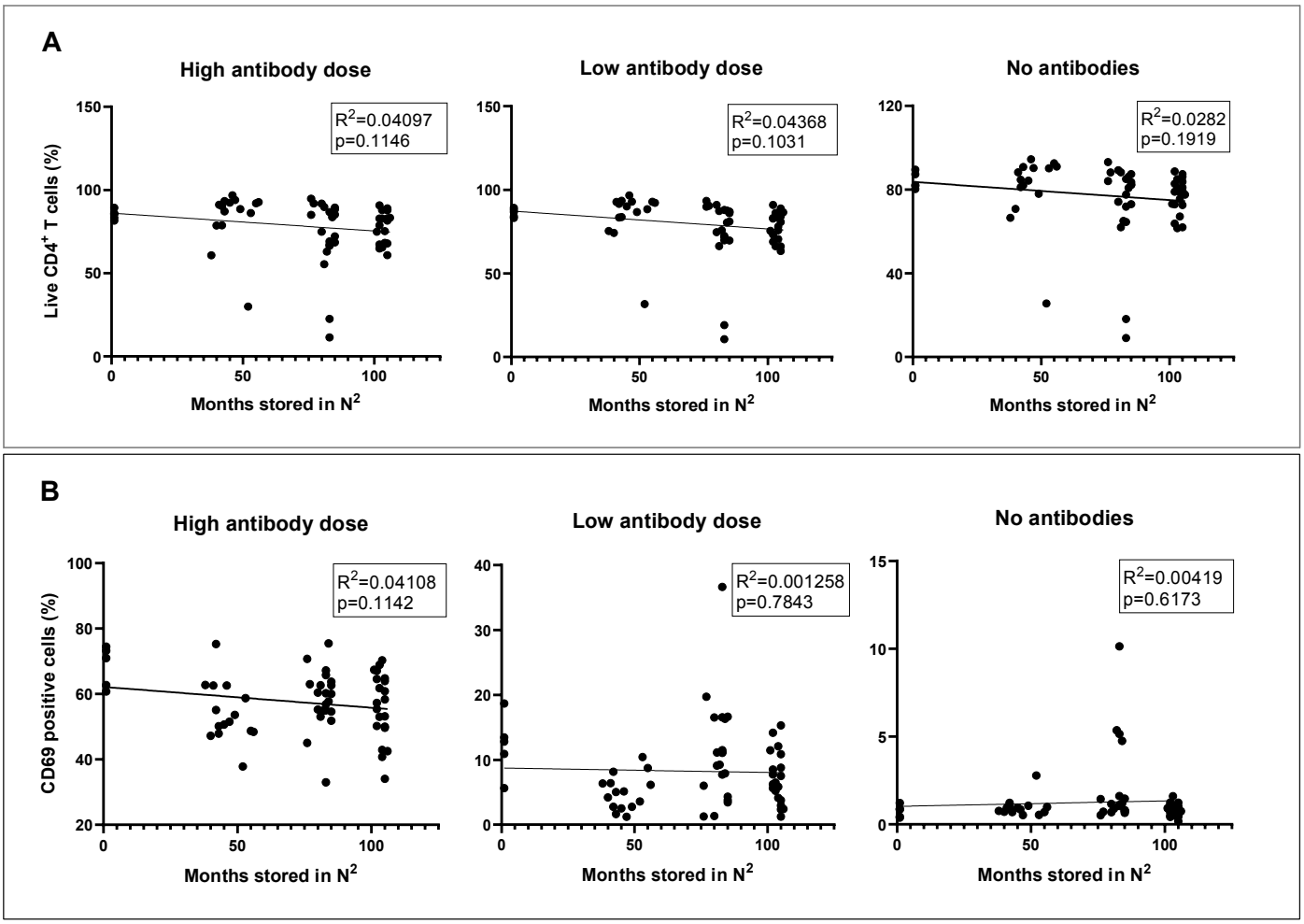

Figure 4.14: No correlation between number of storage months on LN_2 and CD4+ T cell viability and activation. *CD4+ T cells were left unstimulated or activated with high or low dose of antibodies for 24 hours prior to measurement of live cells and cell surface expression of CD69 by flow cytometry.*

The graph display results of linear regression tests performed to evaluate the relationship between the number cell storage months and (A) the cell viability and (B) the cell activation, with R_2 and p-values presented in each graph.

4.2.3.2 Differences in activation between CD4+ T cells from MS patients and healthy controls

The available data from the quality control was used to evaluate whether the level of activation was different in CD4+ T cells from MS patients compared to healthy controls.

The activation differences was assessed by comparing the percentage of CD69 expression between CD4+ T cells from MS patients and healthy controls. A non-parametric Mann Whitney U test was conducted in each group, i.e. unstimulated cells, cells stimulated with high and low dose of antibodies (Fig. 4.15). No significant differences was found between cells from MS patients and healthy controls with high antibody dose or in cells with low antibody dose, while there was a significant difference (p=0,024) between patients and controls in the unstimulated cells.

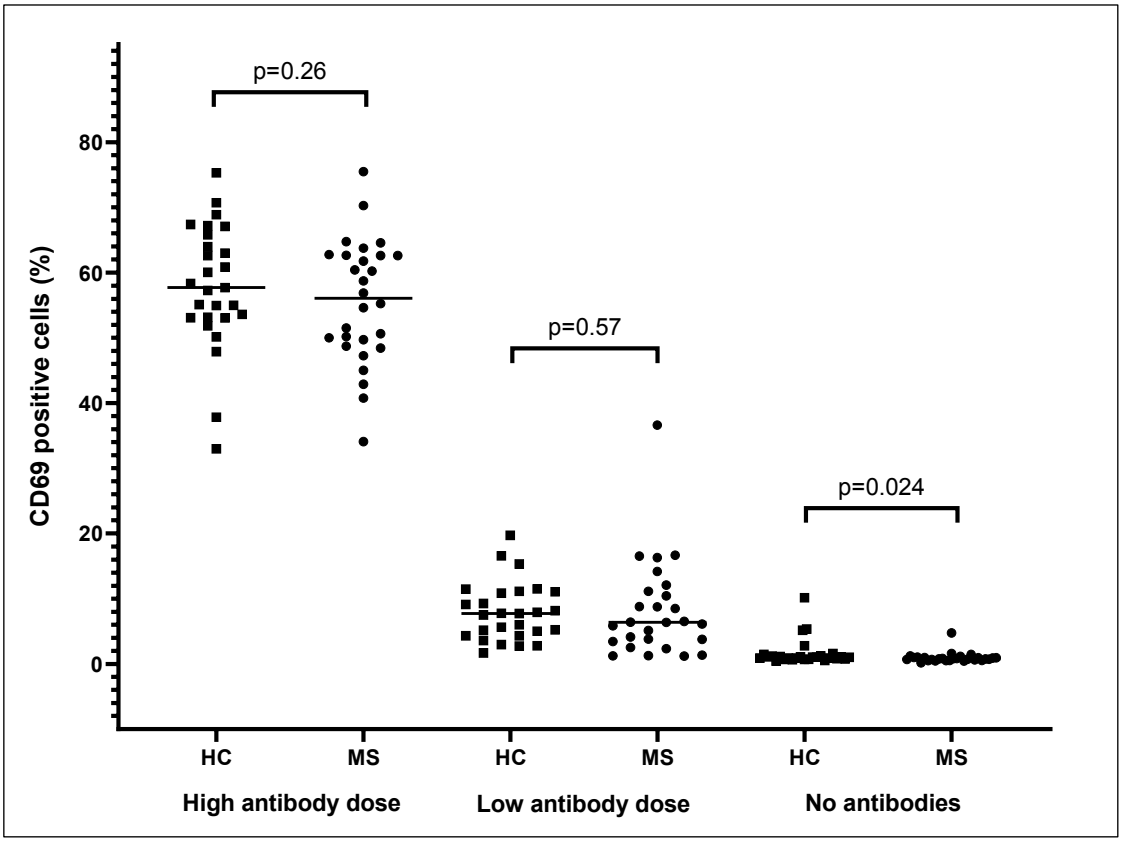

Figure 4.15: Difference in cell surface expression of CD69 between healthy control samples and MS patients in unstimulated CD4+ T cells. *CD4+ T cells were left unstimulated or stimulated in vitro with high or low dose of antibodies for 24 hours before measurement of cell surface expression of CD69. The graph display results from non-parametric Mann Whitney U tests evaluating differences in CD69 expression between MS patients and healthy controls (HC) in unstimulated cells and in cells stimulated with high or low dose of antibodies. The horizontal lines represent the median within each group.*

5 - Discussion

The overall objective of the thesis was to unravel proteomic differences in T cells from MS patients and healthy controls, and to investigate the relationship between protein expression and genotype at MS susceptibility variants.

$CD4_+$ T cells from MS patients and healthy controls were activated *in vitro,* with the aim to study the effect of $CD4_+$ T cell activation at the protein level, to find possible MS associated pathways or proteins that were not found in proteome profiling of freshly isolated cells. A possible correlation between genotype at known MS risk SNP and the expression of proteins encoded by genes located in an expanded region surrounding the MS risk SNPs was investigated was investigated. These pQTL candidates could contribute to identification of potential links between the MS risk variants and disease at the protein level.

5.1 Why study activated T cells?

MS is assumed to be triggered by self-reactive T cell activation in the periphery, followed by infiltration of the CNS through BBB. Re-activation of the self-reactive T cells by APCs located in the CNS, results in recruitment of additional T cells and macrophages leading to the development of inflammatory lesions (21). For this reason, and because the T cells are considered important regulators of adaptive immunity, this study aims to examine T cells.

The strongest genetic risk factor for MS, HLA-DRB1*15:01 or the MHC class II molecule (2, 47, 48), indicates that $CD4_+$ T cells are involved in MS pathogenesis, because they are activated by peptide presentation on MHC class II. $CD8_+$ T cells are linked to MS due to their activation through peptide presentation by MHC class I, which is associated with MS through the HLA-A*0301 allele(67, 68). Moreover, $CD4_+$ and $CD8_+$ T cells and their products are observed in CNS lesions of MS patients (11, 69). This points out both $CD4_+$ and $CD8_+$ T cells as interesting cell types in MS research, as well as for this study.

Former MS studies have typically examined gene expression profiles of unstimulated cells, providing information on untreated cells that reflects the activation status at the

time of blood withdrawal. Independently of the cell activation however, the association between expression levels of mRNA and protein copy numbers are not completely correlating (70). Studies of the dynamic response to CD4+ T cell activation at the RNA level, have interestingly unmasked new pathways associated with MS disease (71), while quantitative proteomic studies have shown differences in the proteome of untreated T cells between MS patients and healthy controls (54). Considering the results from Hellberg and colleagues (71), we hypothesized that studies of T cell activation at the protein level could reveal new and interesting insights in the understanding of MS disease.

Based on quantitative proteomic studies of immune cells, it is expected that the proteomic profiles of activated cells are different from profiles of unstimulated cells (72). In our ongoing study, the proteomic analyses are disease-specific, providing insight in the proteins involved in MS compared to healthy individuals. The results of the proteomic studies for which samples were prepared in this project, were however not finalized due to the Covid-19 situation.

5.2 CD4+ T cell activation

In vivo, T cells are activated via interaction of its TCR and accompanied by co-receptor with an MHC-bound peptide on an APC. This results in a series of downstream effects, triggering T cell maturation, proliferation and immune mediator production, leading to an immune response (11). *In vitro,* traditional T cell activation involves stimulation with antibodies that specifically bind to components of the TCR complex, such as the CD3 molecule. A co-stimulatory signal is crucial for sufficient activation, and is mediated through α-CD28 antibodies specifically recognizing the CD28 co-receptor (73).

Magnetic beads mimicking APCs can be used for *in vitro* T cell stimulation, as the bead surface binds α-CD3 and α-CD8, simultaneously (74). As the size of the beads is similar to the size of APCs, the stimulation conditions are close to the physiological conditions of *in vivo* activation (75). Stimulation by plate-bound α-CD3 antibodies in combination with soluble α-CD28 antibodies is broadly applied in T cell activation studies, as well as in our lab. Previously, members of our lab have used magnetic beads for T cell activation. However, these studies have showed high levels of cell death, inspiring a study of

systematic comparison of T cell activation by plate-bound α-CD3 with soluble α-CD28 and CD3/CD28-bound magnetic beads (Raja, Fitsum and Berge, Unpublished). By evaluating these T cell activation methods based on cell viability and cell activation, the results revealed lower cell death levels and a comparable activation response by using plate-bound α-CD3 and soluble α-CD28.

At the initiation of this project, the $CD4_+$ T cell activation procedure by plate-bound α-CD3 and soluble α-CD28 was already well established. Although the results of the initial testing in this project indicated activation variations, we knew that the variation was small, based on the former procedure establishment. As these tests were performed at the initiation of this project, the variation could be a result of inconsistencies or mistakes during the staining procedure. During the collection of samples to be analyzed by LC-MS-MS, the activation protocol was performed in collaboration with an experienced PhD student in our lab, ensuring that the cells were treated as identical as possible for each sample plate.

For the proteomic profiling by LC/MS-MS, we wanted to include samples stimulated with a high dose of α-CD3 and α-CD28 as well as unstimulated cells. Additionally, we included samples with cells stimulated with low dose of antibodies. Since T cells in our body can exist in different activation states, we wanted to evaluate the differences in the proteome of cells stimulated with high and low antibody dose. Formerly, our lab had shown that freshly isolated $CD4_+$ T cells stimulated with low dose of antibodies displayed adequate activations. Our samples had been isolated between 2011 and 2016, and stored on LN_2 until thawing. Unfortunately, our samples showed low response to stimulation with low antibody doses. However, based on the quality control data, we did not see a significant relationship between the number of storage months and the viability or the activation of the cells.

We observed that unstimulated cells had significant differences in CD69 expression between MS patients and healthy controls, of which the $CD4_+$ T cells from the controls seemed to display a higher CD69 expression than $CD4_+$ T cells from MS patients. After removal of an outlier displaying 10% CD69 expression, the p-value was still significant.

5.3 pQTL candidates

MS is caused by multiple genetic and environmental contributions. Thus, it was interesting to not only assess proteomic differences between MS patients and healthy controls, but also the proteomic differences between individuals that are carriers of risk-alleles and non-risk alleles at the MS risk SNPs. An MS risk SNP is likely to affect genes located in close proximity, but also genes present at more remote distances in the genomic region. The published study correlated the protein expression from genes proximal to the MS risk SNPs with the genotype at these SNPs. In this study, proteins encoded by genes present within an expanded region surrounding the SNPs were assessed. Two new pQTL candidates were identified in $CD8_+$ T cells, i.e. rs1800693 correlating with the protein expression of *CD9* and rs137955 correlating with the protein expression of *UQCRFS1P1.*

While eQTLs have continuously been studied for some time, recent studies implement pQTL analysis as well (54, 76, 77). These analyses can enable explorations of proteins involved in MS pathogenesis. Studies have uncovered that classical eQTLs for gene expression levels are extremely widespread throughout the genome (78). These eQTLs could explain the mechanisms behind the correlation of genotype and protein expression. By searches in the Genotype-Tissue Expression (GTEx) project portal (79), no indications were found of rs1800693 or rs137955 acting as eQTLs for *CD9* or *UQCRFS1P1,* respectively. However, rs1800693 is reported to act as an eQTL in the atrial appendage of the heart and tibial arteries for the proximal *TNFRSF1A* gene, a previously identified MS susceptibility gene (79, 80). This SNP is furthermore reported to have a direct impact on the *TNFRSF1A* gene, and it is located within a splice acceptor site in an intron of *TNFRSF1A* (80). Whether the *CD9* gene has an influence on MS disease remains to be shown, but due to the direct impact of rs1800693 on *TNFRSF1A*, it may be less likely that the *CD9* gene has an influence on MS disease. The position of the *TNFRSF1A* gene relative to *CD9* is presented in figure 5.1.

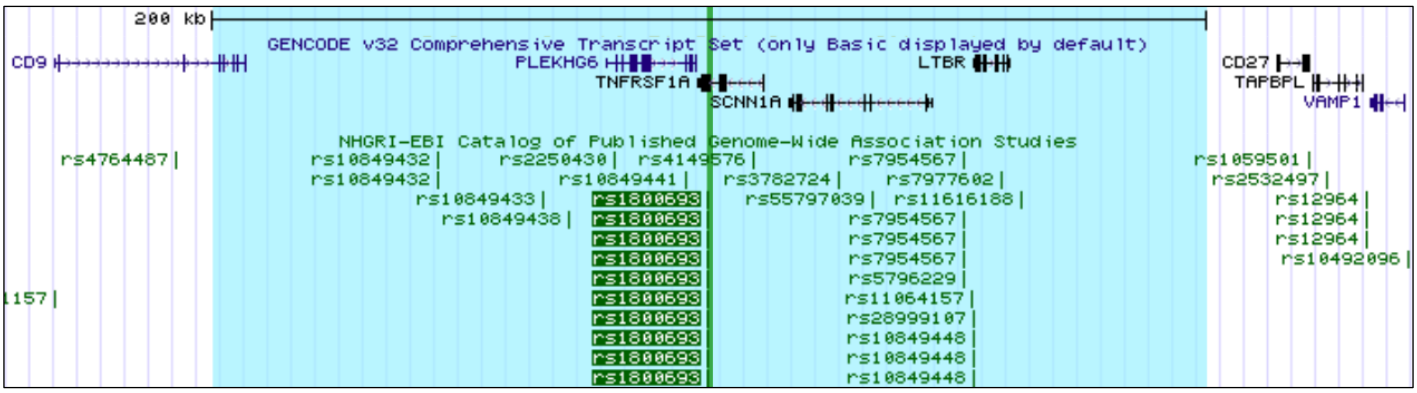

Figure 5.1: Illustration of expanded region surrounding the MS risk SNP rs180069. *The figure presents the region surrounding rs1800693, where the blue area indicates the expanded region(100kb upstream and downstream the SNP) and the green line represent the SNP position. Genes present in the region are indicated in dark blue. (65)*

To investigate if the SNPs with genotypes correlating with protein expression were located in regulatory regions of the genome, their genomic locations were explored in the UCSC Genome Browser. Rs1800693 is located in a region of low levels of histone H3 acetylation, indicating a relatively closed chromatin structure leading to low transcription levels (65). The SNP is not located within DNase hypersensitive sites, but by chromatin immunoprecipitation sequencing (ChIP-Seq), it is shown that the region is bound by RNA polymerase 2 (*POLR2A*). This might suggest that the SNP is positioned in a region that is transcribed and can affect gene expression (65). Similarly, rs137955 is not located in DNase hypersensitivity sites, but it is located within two putative transcription factor (TF) ChIP-seq clusters, i.e. regions of the genome highly bound by TFs. TFs bind to DNA and interact with RNA polymerases to regulate gene expression, hence rs137955 could potentially be involved in this regulation (65). Whether the SNP can affect transcription or the mRNA quantity and lead to alterations of the protein abundance, is not known.

SNPs are highly abundant and distributed throughout the human genome (81) and do not change significantly from generation to generation, as they are often inherited in haplotype blocks (82). A SNP which functions as a pQTL, might not be the causative SNP, if other nearby SNPs are in high LD. The location of each SNP was therefore examined using the LDproxy tool by LDlink (83), to find possible recombination blocks of SNPs inherited together. Rs35558398 and rs767455, two neighboring SNPs of rs1800693,

displayed an R_2 value above 0.8, indicating high LD. Examining rs137955, 12 neighboring SNPs were found to display an R_2 value above 0.8. Whether these neighboring SNPs in high LD, or the SNP whose genotype is correlating with protein expression in our data, have a functional role in regulation of the genes, requires further investigations and studies.

The *CD9* gene encodes a cell surface protein that is a member of the tetraspanin protein family, and is expressed by the major leukocyte subsets, as well as at a high level in endothelial cells (84). Studies suggest that *CD9* is involved in T cell activation (85) and furthermore, studies of $CD4_+$ T cells obtained from brain and spinal cord of experimental autoimmune encephalomyelitis (EAE) mice showed decreased expression of *CD9* compared to the *CD9* expression in naïve $CD4_+$ T cells (86). In our data, the *CD9* protein abundance was observed to be decreased in individuals carrying the risk allele compared to non-risk carriers. Whether these finding can be translated to the EAE-study, or if other mechanisms are involved, remains to be shown.

The *UQCRFS1P1* gene encodes ubiquinol-cytochrome c reductase Rieske iron-sulfur polypeptide 1 pseduogene 1, a supposed pseudogene. In theory, pseudogenes are remains of former genes that has lost their biological function, and would not encode protein products. However, proteomic studies in humans have identified peptides encoded by 107 pseudogenes (87), and while some translated pseudogenes have functions at the protein level, not all pseudogenes do (88). Our findings suggest that the *UQCRFS1P1* gene encodes proteins, although this could be due to accidental translations. To our knowledge, no other studies have identified the biological function of the protein.

Even though several experiments must be implemented to validate these findings, we present relevant candidates for further pQTL studies related to MS.

5.4 Pathway analysis

Members of our lab had selectively and separately analyzed differentially expressed proteins in $CD4_+$ and $CD8_+$ T cells to find cell-specific pathways involved in MS (54). As the previous analysis was conducted on the T cell subtypes separately, the objective for

the analysis of this thesis was to examine enriched pathways in proteins differentially expressed in both $CD4_+$ and $CD8_+$ T cells.

The published study identified 14 biological pathways enriched in $CD4_+$ T cells from MS patients, including T cell activation pathways such as signaling through CD28, CTLA4 and TCR (54). Other studies on gene expression have indicated that transcription of genes involved in translational regulation, oxidative phosphorylation, immune synapse and antigen presentation pathways are dysregulated in MS patients, all connected to T cell pathways (89).

To examine the 26 proteins differentially expressed in both $CD4_+$ and $CD8_+$ T cells from MS patients and healthy controls, searches were made in the KEGG pathway database. Among the 26 proteins, five proteins were involved in metabolic pathways, while three proteins were involved in focal adhesion pathways. These pathways are common cellular mechanisms, and do not link the proteins to pathways specifically involved in MS disease. Some of the 26 proteins were linked to biologically relevant pathways like the EBV infection pathway and the leukocyte transendothelial migration pathway.

A literature search was conducted to evaluate if any of the 26 proteins could be linked to MS disease, inflammation or degeneration in the CNS. The integrin-linked kinase is encoded by the *ILK* gene. Interestingly, studies in mice indicate that deletion of *ILK* reduces the proliferation and differentiation of oligodendrocyte precursor cells (OPCs) during CNS development (90). Oligodendrocytes are myelin generating cells, and myelin oligodendrocyte glycoprotein are thought to be targets for the autoreactive T cells that have crossed the BBB during MS disease. Moreover, one of the 26 proteins was the translocator protein encoded by *TSPO*. This protein is considered a marker of brain and spinal cord inflammation in neurodegenerative disorders (91). In many neuropathic conditions, *TSPO* is showing an increased expression. In MS patients, the expression is however shown to decrease (92). In our data, the protein encoded by *TSPO* is downregulated in $CD4_+$ and $CD8_+$ T cells from MS patients. Lastly, the protein encoded by *NFKB2* was among the 26 differentially expressed proteins. Nuclear factor (NF)-kB pathways are changed in MS, leading to increased levels of NF-kB signal activation in cells and excessive expression of effector molecules that relies on NF-kB for transcription (93,

94). In our data, the protein encoded by the NFKB2 gene is shown to be upregulated in both cell types from MS patients compared to healthy controls.

5.5 Future perspectives

Proteome profiling is a suitable tool to evaluate inflammatory mechanisms, as inflammatory responses often involve alterations of the protein expression (95). These analyses can contribute to identification of new proteins or biological pathways that are of relevance for disease development.

For further analysis, the results of the LC/MS-MS analysis would need to be explored and could be mapped to the proteomic profiles of freshly purified $CD4_+$ and $CD8_+$ T cells from the published paper (54). By comparing protein profiles of stimulated cells to unstimulated cells, new insights in the impact of T cell activation could be established. Identifications of proteins differentially expressed, upregulated or downregulated between MS patients and healthy controls would be made in the data of activated cells. These findings could be compared to the previously identified differentially expressed proteins, to potentially find new candidates for future studies of their role in the immune system and in the development of MS.

The proteomic study of freshly isolated T cells from MS patients and healthy controls, $CD4_+$ and $CD8_+$ T cells were examined. These T cells can be categorized into several additional subtypes. Further analysis could include investigation of the distinct roles of the detected proteins in these subtypes of $CD4_+$ and $CD8_+$ T cells. Our studies have analyzed T cells isolated from whole blood, while the CNS is central in MS disease. Hence, it would be interesting to look at the T cells specifically located in the CNS, to find characteristics defining CNS T cells compared to circulating T cells.

Proteomic profiling can identify proteins or genes with known or common biological functions. The proteins or genes can have functions known to have an impact on other autoimmune or inflammatory diseases. Hence, functions of these proteins or genes in other diseases can be translated to functions or mechanisms of MS disease. Some proteins can have known functions and mechanisms in other cell types. Functional studies can

accordingly be implemented in immune cell subsets, such as T cells, to examine the effect of these proteins on the T cell activation or other immune cell reactions. In addition, proteins of unknown functions can be detected. A general mapping of the function of these proteins can be initiated, to find potential links to MS disease. It could be interesting to look at the sub-cellular localization of these newly identified proteins, and one could gain further insights in how the proteins are localized before and upon activation, and their involvement in immune cell responses.

In combination with genotyping, results from the proteomic study can contribute to identification of additional SNPs whose genotype is correlating with protein expression. The mechanism behind the correlation could be assessed through examination of the location of the SNPs and if their neighboring SNPs are in high LD. One could additionally find potential transcriptional regulations or epigenetic modification of the region that might explain the relationship between genotype and protein expression.

6 - Concluding remarks

The aim of this thesis was to gain insights in the proteomic differences between MS patients and healthy controls, and investigate association between genotype at MS risk gene variants and protein expression in proteomic samples from $CD4_+$ and $CD8_+$ T cells. The analysis of the available proteomics dataset did not result in findings of MS related biological pathways, and there was no significant enrichment of proteins differentially expressed in both $CD4_+$ and $CD8_+$ T cells between MS patients and healthy controls. When specifically analyzing proteins expressed from genes located 100kb upstream and downstream from each of the 200 non-HLA MS risk SNPs, two additional pQTLs candidates were found. $CD4_+$ T cell samples from 20 MS patients and 20 healthy controls were activated and prepared for proteomic profiling by LC/MS-MS. These analyses were interrupted by the Covid-19 situation in Norway, and the results were not available.

The pQTL candidates would require confirmatory studies to be considered reliable pQTLs. However, these findings might contribute to, and inspire to introduce studies involving *CD9* and *UQCRFS1P1*, to further understand patterns and processes leading to MS disease, as well as insights in identified MS risk gene variants. The awaiting results of the LC/MS-MS analysis, might contribute to identification of T cell specific proteins or pathways involved in MS pathogenesis.

Bibliography

1. Compston A, Coles A. Multiple sclerosis. Lancet (London, England). 2008;372(9648):1502-17.
2. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9.
3. International Multiple Sclerosis Genetics C. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science (New York, NY). 2019;365(6460).
4. Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kemppinen A, Cotsapas C, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nature genetics. 2013;45(11):1353-60.
5. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science (New York, NY). 2012;337(6099):1190-5.
6. Abbas AK, Lichtman AH, Pillai S. Basic Immunology - Functions and Disorders of the Immune system. Fifth ed. St. Louis, Missouri ELSEVIER; 2016. 335 p.
7. Riera Romo M, Pérez-Martínez D, Castillo Ferrer C. Innate immunity in vertebrates: an overview. Immunology. 2016;148(2):125-39.
8. Lea T. Immunologi og immunologiske teknikker: Fagbokforlaget Vigmostad & Bjørke AS; 2006. 400 p.
9. Iwasaki A, Medzhitov R. Control of adaptive immunity by the innate immune system. Nat Immunol. 2015;16(4):343-53.
10. Bonilla FA, Oettgen HC. Adaptive immunity. The Journal of allergy and clinical immunology. 2010;125(2 Suppl 2):S33-40.
11. Legroux L, Arbour N. Multiple Sclerosis and T Lymphocytes: An Entangled Story. J Neuroimmune Pharmacol. 2015;10(4):528-46.
12. Netea MG, Schlitzer A, Placek K, Joosten LAB, Schultze JL. Innate and Adaptive Immune Memory: an Evolutionary Continuum in the Host's Response to Pathogens. Cell host & microbe. 2019;25(1):13-26.
13. Wherry EJ, Masopust D. Chapter 5 - Adaptive Immunity: Neutralizing, Eliminating, and Remembering for the Next Time. In: Katze MG, Korth MJ, Law GL, Nathanson N, editors. Viral Pathogenesis (Third Edition). Boston: Academic Press; 2016. p. 57-69.
14. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. Nature reviews Immunology. 2015;15(9):545-58.
15. Bhandoola A, Sambandam A. From stem cell to T cell: one route or many? Nature reviews Immunology. 2006;6(2):117-26.
16. Starr TK, Jameson SC, Hogquist KA. Positive and negative selection of T cells. Annual review of immunology. 2003;21:139-76.
17. Hagel K. Positive and Negative Selection of T Cells. 2018. Available from: https://immunobites.com/2018/08/20/positive-and-negative-selection-of-t-cells/.
18. Takaba H, Takayanagi H. The Mechanisms of T Cell Selection in the Thymus. Trends in immunology. 2017;38(11):805-16.
19. Chitnis T. The role of CD4 T cells in the pathogenesis of multiple sclerosis. International review of neurobiology. 2007;79:43-72.

20. Siu G, Clark SP, Yoshikai Y, Malissen M, Yanagi Y, Strauss E, et al. The human T cell antigen receptor is encoded by variable, diversity, and joining gene segments that rearrange to generate a complete V gene. Cell. 1984;37(2):393-401.
21. Kaskow BJ, Baecher-Allan C. Effector T Cells in Multiple Sclerosis. Cold Spring Harbor perspectives in medicine. 2018;8(4):a029025.
22. Browne P, Chandraratna D, Angood C, Tremlett H, Baker C, Taylor BV, et al. Atlas of Multiple Sclerosis 2013: A growing global problem with widespread inequity. Neurology. 2014;83(11):1022-4.
23. Simpson S, Jr., Blizzard L, Otahal P, Van der Mei I, Taylor B. Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis. Journal of neurology, neurosurgery, and psychiatry. 2011;82(10):1132-41.
24. Berg-Hansen P, Moen SM, Harbo HF, Celius EG. High prevalence and no latitude gradient of multiple sclerosis in Norway. Multiple sclerosis (Houndmills, Basingstoke, England). 2014;20(13):1780-2.
25. Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. Nature reviews Neurology. 2017;13(1):25-36.
26. Patsopoulos NA. Genetics of Multiple Sclerosis: An Overview and New Directions. Cold Spring Harbor perspectives in medicine. 2018;8(7).
27. Zheleznyakova GY, Piket E, Marabita F, Pahlevan Kakhki M, Ewing E, Ruhrmann S, et al. Epigenetic research in multiple sclerosis: progress, challenges, and opportunities. Physiological genomics. 2017;49(9):447-61.
28. Zhang J, Weiner HL, Hafler DA. Autoreactive T Cells in Multiple Sclerosis. International Reviews of Immunology. 1992;9(3):183-201.
29. Libbey JE, McCoy LL, Fujinami RS. Molecular mimicry in multiple sclerosis. International review of neurobiology. 2007;79:127-47.
30. Tengvall K, Huang J, Hellström C, Kammer P, Biström M, Ayoglu B, et al. Molecular mimicry between Anoctamin 2 and Epstein-Barr virus nuclear antigen 1 associates with multiple sclerosis risk. Proc Natl Acad Sci U S A. 2019;116(34):16955-60.
31. Sospedra M, Martin R. Immunology of multiple sclerosis. Annual review of immunology. 2005;23:683-747.
32. NIH. Human Genome Project Bethesda, MD, USA: National Human Genome Research Institute; 2020 [updated 24.02.2020. Available from: https://www.genome.gov/human-genome-project/Completion-FAQ.
33. Piovesan A, Antonaros F, Vitale L, Strippoli P, Pelleri MC, Caracausi M. Human protein-coding genes and gene feature statistics in 2019. BMC Research Notes. 2019;12(1):315.
34. Elkon R, Agami R. Characterization of noncoding regulatory DNA in the human genome. Nature biotechnology. 2017;35(8):732-46.
35. Turnpenny PD, Allard S. Emery's Elements of Medical Genetics. 2017. In: Emery's Elements of Medical Genetics [Internet]. Amsterdam, Netherlands: Elsevier Limited. 15.
36. Osada N. Genetic diversity in humans and non-human primates and its evolutionary consequences. Genes & Genetic Systems. 2015;90(3):133-45.
37. Supek F, Lehner B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. Nature. 2015;521(7550):81-4.
38. Alvaro R-I. Genetic variation within transcriptional regulatory elements and its implications for human disease. Biological Chemistry. 2014;395(12):1453-60.

39. Liu B, Montgomery SB. Identifying causal variants and genes using functional genomics in specialized cell types and contexts. Human Genetics. 2020;139(1):95-102.
40. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101(1):5-22.
41. Nature. Gene Expression: Nature Education; 2014 [Available from: https://www.nature.com/scitable/topicpage/gene-expression-14121669/.
42. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018;172(4):650-65.
43. Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell. 2016;165(3):535-50.
44. Yao C, Chen G, Song C, Keefe J, Mendelson M, Huan T, et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nature Communications. 2018;9(1):3268.
45. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. Lancet (London, England). 2018;391(10130):1622-36.
46. Patsopoulos NA, Esposito F, Reischl J, Lehr S, Bauer D, Heubach J, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Annals of neurology. 2011;70(6):897-912.
47. Barcellos LF, Oksenberg JR, Begovich AB, Martin ER, Schmidt S, Vittinghoff E, et al. HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course. Am J Hum Genet. 2003;72(3):710-6.
48. Sawcer S, Ban M, Maranian M, Yeo TW, Compston A, Kirby A, et al. A high-density screen for linkage in multiple sclerosis. Am J Hum Genet. 2005;77(3):454-67.
49. Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, et al. Risk alleles for multiple sclerosis identified by a genomewide study. The New England journal of medicine. 2007;357(9):851-62.
50. Sawcer S, Franklin RJ, Ban M. Multiple sclerosis genetics. The Lancet Neurology. 2014;13(7):700-9.
51. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013;38(1):23-38.
52. Maltby VE, Graves MC, Lea RA, Benton MC, Sanders KA, Tajouri L, et al. Genome-wide DNA methylation profiling of CD8+ T cells shows a distinct epigenetic signature to CD4+ T cells in multiple sclerosis patients. Clinical epigenetics. 2015;7:118.
53. Graves MC, Benton M, Lea RA, Boyle M, Tajouri L, Macartney-Coxson D, et al. Methylation differences at the HLA-DRB1 locus in CD4+ T-Cells are associated with multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England). 2014;20(8):1033-41.
54. Berge T, Eriksson A, Brorson IS, Høgestøl EA, Berg-Hansen P, Døskeland A, et al. Quantitative proteomic analyses of CD4(+) and CD8(+) T cells reveal differentially expressed proteins in multiple sclerosis patients and healthy controls. Clin Proteomics. 2019;16:19-.
55. Rhead B, Brorson IS, Berge T, Adams C, Quach H, Moen SM, et al. Increased DNA methylation of SLFN12 in CD4+ and CD8+ T cells from multiple sclerosis patients. PLoS One. 2018;13(10):e0206511-e.
56. Li X, Wang W, Chen J. Recent progress in mass spectrometry proteomics for biomedical research. Science China Life Sciences. 2017;60(10):1093-113.
57. Blum F. High performance liquid chromatography. British journal of hospital medicine (London, England : 2005). 2014;75(2):C18-21.

58. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422(6928):198-207.
59. Perry RH, Cooks RG, Noll RJ. Orbitrap mass spectrometry: instrumentation, ion motion and applications. Mass spectrometry reviews. 2008;27(6):661-99.
60. Scigelova M, Makarov A. Orbitrap mass analyzer--overview and applications in proteomics. Proteomics. 2006;6 Suppl 2:16-21.
61. Cox J, Mann M. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. Annual review of biochemistry. 2011;80(1):273-99.
62. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Annals of neurology. 2011;69(2):292-302.
63. Adan A, Alizada G, Kiraz Y, Baran Y, Nalbant A. Flow cytometry: basic principles and applications. Critical reviews in biotechnology. 2017;37(2):163-76.
64. Kanegane H, Hoshino A, Okano T, Yasumi T, Wada T, Takada H, et al. Flow cytometry-based diagnosis of primary immunodeficiency diseases. Allergology International. 2018;67(1):43-54.
65. The UCSC Genome browser. [cited 02.05.2020]. Available from: https://genome.ucsc.edu/.
66. Hulsen T, de Vlieg J, Alkema W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC genomics. 2008;9:488.
67. Harbo HF, Lie BA, Sawcer S, Celius EG, Dai KZ, Oturai A, et al. Genes in the HLA class I region may contribute to the HLA class II-associated genetic susceptibility to multiple sclerosis. Tissue antigens. 2004;63(3):237-47.
68. Fogdell-Hahn A, Ligers A, Grønning M, Hillert J, Olerup O. Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. Tissue antigens. 2000;55(2):140-8.
69. Wu GF, Alvarez E. The immunopathophysiology of multiple sclerosis. Neurol Clin. 2011;29(2):257-78.
70. Payne SH. The utility of protein and mRNA correlation. Trends in biochemical sciences. 2015;40(1):1-3.
71. Hellberg S, Eklund D, Gawel DR, Köpsén M, Zhang H, Nestor CE, et al. Dynamic Response Genes in CD4+ T Cells Reveal a Network of Interactive Proteins that Classifies Disease Activity in Multiple Sclerosis. Cell reports. 2016;16(11):2928-39.
72. Rieckmann JC, Geiger R, Hornburg D, Wolf T, Kveler K, Jarrossay D, et al. Social network architecture of human immune cells unveiled by quantitative proteomics. Nat Immunol. 2017;18(5):583-93.
73. Frauwirth KA, Thompson CB. Activation and inhibition of lymphocytes by costimulation. J Clin Invest. 2002;109(3):295-9.
74. Trickett A, Kwan YL. T cell stimulation and expansion using anti-CD3/CD28 beads. Journal of immunological methods. 2003;275(1-2):251-5.
75. Scientific TF. T cell activation with Dynabeads Waltham, MA,USA: Thermo Fisher Scientific; [Available from: https://www.thermofisher.com/no/en/home/life-science/cell-analysis/cell-isolation-and-expansion/cell-expansion.html.
76. Lin X, Deng FY, Mo XB, Wu LF, Lei SF. Functional relevance for multiple sclerosis-associated genetic variants. Immunogenetics. 2015;67(1):7-14.

77. Partha R, Bronson PG, Sangurdekar D. Discovering Context-dependent Whole Blood Gene Expression Quantitative Trait Loci (eQTL) In Relapsing Remitting Multiple Sclerosis (P2.2-051). Neurology. 2019;92(15 Supplement):P2.-051.
78. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013;501(7468):506-11.
79. The GTEx portal. [cited 02.05.2020]. Available from: https://gtexportal.org/.
80. Ottoboni L, Frohlich IY, Lee M, Healy BC, Keenan BT, Xia Z, et al. Clinical relevance and functional consequences of the TNFRSF1A multiple sclerosis locus. Neurology. 2013;81(22):1891-9.
81. Takeuchi F, Yanai K, Morii T, Ishinaga Y, Taniguchi-Yanai K, Nagano S, et al. Linkage disequilibrium grouping of single nucleotide polymorphisms (SNPs) reflecting haplotype phylogeny for efficient selection of tag SNPs. Genetics. 2005;170(1):291-304.
82. Edwards D, Forster JW, Chagné D, Batley J. What Are SNPs? In: Oraguzie NC, Rikkerink EHA, Gardiner SE, De Silva HN, editors. Association Mapping in Plants. New York, NY: Springer New York; 2007. p. 41-52.
83. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics (Oxford, England). 2015;31(21):3555-7.
84. Reyes R, Cardeñes B, Machado-Pineda Y, Cabañas C. Tetraspanin CD9: A Key Regulator of Cell Adhesion in the Immune System. Frontiers in immunology. 2018;9:863-.
85. Kobayashi H, Hosono O, Iwata S, Kawasaki H, Kuwana M, Tanaka H, et al. The tetraspanin CD9 is preferentially expressed on the human CD4(+)CD45RA+ naive T cell population and is involved in T cell activation. Clinical and experimental immunology. 2004;137(1):101-8.
86. Lombardo SD, Mazzon E, Basile MS, Campo G, Corsico F, Presti M, et al. Modulation of Tetraspanin 32 (TSPAN32) Expression in T Cell-Mediated Immune Responses and in Multiple Sclerosis. Int J Mol Sci. 2019;20(18):4323.
87. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. Nature. 2014;509(7502):575-81.
88. Xu J, Zhang J. Are Human Translated Pseudogenes Functional? Molecular Biology and Evolution. 2015;33(3):755-60.
89. Gandhi KS, McKay FC, Cox M, Riveros C, Armstrong N, Heard RN, et al. The multiple sclerosis whole blood mRNA transcriptome and genetic associations indicate dysregulation of specific T cell pathways in pathogenesis. Human molecular genetics. 2010;19(11):2134-43.
90. Hussain R, Macklin WB. Integrin-Linked Kinase (ILK) Deletion Disrupts Oligodendrocyte Development by Altering Cell Cycle. J Neurosci. 2017;37(2):397-412.
91. Nutma E, Stephenson JA, Gorter RP, de Bruin J, Boucherie DM, Donat CK, et al. A quantitative neuropathological assessment of translocator protein expression in multiple sclerosis. Brain : a journal of neurology. 2019;142(11):3440-55.
92. Harberts E, Datta D, Chen S, Wohler JE, Oh U, Jacobson S. Translocator protein 18 kDa (TSPO) expression in multiple sclerosis patients. J Neuroimmune Pharmacol. 2013;8(1):51-7.
93. Leibowitz SM, Yan J. NF-κB Pathways in the Pathogenesis of Multiple Sclerosis and the Therapeutic Implications. Front Mol Neurosci. 2016;9:84-.
94. Li Q, Verma IM. NF-kappaB regulation in the immune system. Nature reviews Immunology. 2002;2(10):725-34.

95. Assadsangabi A, Evans CA, Corfe BM, Lobo A. Application of Proteomics to Inflammatory Bowel Disease Research: Current Status and Future Perspectives. Gastroenterol Res Pract. 2019;2019:1426954-.

Appendix A: Abbreviations

APC - Antigen presenting cell

BBB - Blood-brain barrier

BCA – Bicinchoninic acid assay

CD - Cluster of differentiation

CNS - Central nervous system

DC - Dendritic cell

DE - Differentially expressed

DMSO - Dimethyl sulfoxide

EAE - Experimental autoimmune encephalitis

EBV - Epstein Barr virus

EDTA - Ethylenediaminetetraacetic acid

EDSS – Expanded disability status scale

eQTL - Expression quantitative trait loci

ESI - Electrospray ionization

FBS - Fetal bovine serum

FCS - Fetal calf serum

FITC - Fluorochrome fluorescein isothiocyanate

FSC - Forward scattered

GTEx - Genotype-tissue expression

GWAS - Genome wide association studies

HC - Healthy control

HLA - Human leukocyte antigen

Ig - Immunoglobulin

IMSGC - The International Multiple Sclerosis Genetics Consortium

IPA - Ingenuity pathway analysis

Kb - Kilo base

KEGG - Kyoto Encyclopedia of Genes and Genomes

LAF - Laminar flow

LC - Liquid chromatography

LC/MS-MS - Liquid chromatography tandem mass spectrometry

LN_2 - Liquid nitrogen

M - Molar

MBP - Basic myelin protein

MHC - Major histocompatibility complex

MRI - Magnetic resonance imaging

MS - Multiple Sclerosis

MS - Mass spectrometry

NK - Natural killer

OCB – Oligoclonal bands

OPC - Oligodendrocyte precursor cell

PAMP - Pathogen-associated molecular pattern

PBMC - Peripheral blood mononuclear cells

PBS - Phosphate-buffered saline

PBS - Phosphate buffered saline

pQTL - Protein quantitative trait loci

PROBE - The Proteomics unit at the University of Bergen

PRR - Pattern recognition receptor

RIPA - Radioimmunoprecipitation assay buffer

RRMS - Relapsing-remitting multiple sclerosis

RT - Room temperature

SD - Standard deviation

SDC - Sodium deoxycholate

SDS - Sodium dodecyl sulfate

SNP - Single nucleotide polymorphism

SSC - Side scattered

TCR - T cell receptor

TF - Transcription factor

$\bar{x}$ - Sample mean

Appendix B: Materials

Reagents, kits, instruments and software

Table S1: Reagents with corresponding suppliers and catalog number.

Reagents	Supplier	Cat. no.
Bovine Serum Albumin (BSA)	Bio-Rad Laboratories, Hercules, CA, USA	500-0206
Dimethyl Sulfoxide (DMSO)	Sigma-Aldrich®, Darmstadt, Germany	D2650
Dulbecco's phosphate buffered saline (PBS)	Sigma-Aldrich®, Darmstadt, Germany	D8537
Tris-EDTA buffer solution	Sigma-Aldrich®, Darmstadt, Germany	T9285
Fetal Calf Serum (FCS)	GE Healthcare Life Sciences, Pittsburg, PA, USA	SH30072.03
LIVE/DEAD™ Fixable Far Red Dead Stain Kit	Invitrogen by Thermo Fisher Scientific, Carlsbad, California, USA	L10120
LIVE/DEAD™ Fixable Red Dead Stain Kit	Invitrogen by Thermo Fisher Scientific, Carlsbad, California, USA	L23102
Lymphoprep™	STEMCELL™ Technologies, Vancouver, Canada	07801/07811
Paraformaldehyde (PFA)	Sigma-Aldrich®, Darmstadt, Germany	158127
RPMI Medium 1640	Gibco™ by Thermo Fisher Scientific, Waltham, MA, USA	21875-034
Trypan Blue	Bio-Rad Laboratories, Hercules, CA, USA	1450013
X-VIVO medium	Sartorius Stedim Biotech, Aubagne, France	BE02-061Q

Table S2: Commercial kits and assays with corresponding suppliers and catalog number.

Kit	Supplier	Cat. no
EasySep™ Human CD4+ T cell Isolation Kit	STEMCELL™ Technologies, Vancouver, Canada	17952

Table S3: Antibodies with corresponding suppliers and catalog number

Antibodies	**Supplier**	**Cat. no**
Anti-human CD69 APC	ImmunoTools, Friesoythe, Germany	21620694X2
Anti-human CD69 FITC	ImmunoTools, Friesoythe, Germany	21459693
Anti-human CD69 PE	ImmunoTools, Friesoythe, Germany	21620694X2
CD3 Mouse anti-Human	eBioscience™ by Thermo Fisher Scientific, San Diego, CA, USA	15276737
Mouse IgG1 control APC	ImmunoTools, Friesoythe, Germany	21275516X2
Mouse IgG1 control FITC	ImmunoTools, Friesoythe, Germany	21815013
Purified NA/LE Mouse Anti-Human CD28	BD Biosciences, Franklin Lakes, NJ, USA	555725

Table S4: Instruments with corresponding supplier

Instruments	**Supplier**
Attune Acoustic Focusing Cytometer	Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA
Attune NxT Auto Sampler	Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA
Centrifuge 5810 R	Eppendorf, Hamburg, Germany
FACSCanto™ II Flow Cytometer	BD Biosciences, San Jose, CA, USA
HERAcell™ 150i CO2 incubator	Thermo Fisher Scientific, Waltham, MA, USA
TC20™ Automated Cell Counter	Bio-Rad Laboratories, Hercules, CA, USA

Table S5: Software with corresponding supplier

Software	**Supplier**
Attune® Cytometric Software 2.1	Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA
FCS Express 6 Flow Cytometry	De Novo Software, Glendale, CA, USA
GraphPad Prism 8	GraphPad Software, Inc., La Jolla, CA, USA
Microsoft® Excel	Microsoft, Redmond, WA, USA

Appendix C: Supplementary tables and figures

Table S6: *Overview of the 26 proteins differentially expressed in both CD4+ and CD8+ T cells between MS patients and healthy controls, with corresponding gene name, Student's t-test p-value and fold change (MS/HC) in CD4+ and CD8+ T cells.*

Gene name	Protein name	CD4+ T cells		CD8+ T cells	
		Student's t-test	Fold change MS/HC	Student's t-test	Fold change MS/HC
ACLY	ATP-citrate synthase	0,023	-0,062	0,030	-0,132
ACO1	Cytoplasmic aconitate hydratase	0,008	0,170	0,022	0,213
ACP1	Low molecular weight phosphotyrosine protein phosphatase	0,005	0,224	0,008	0,221
AGL	Glycogen debranching enzyme;4-alpha-glucanotransferase;Amylo-alpha-1,6-glucosidase	0,002	0,322	0,010	0,463
APRT	Adenine phosphoribosyltransferase	0,009	0,192	0,019	0,292
ATP2A2	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	0,003	-0,340	0,010	-0,505
BCKDHB	2-oxoisovalerate dehydrogenase subunit beta, mitochondrial	0,030	0,243	0,004	0,259
FLNA	Filamin-A	0,024	-0,234	0,022	-0,324
GLYR1	Putative oxidoreductase GLYR1	0,002	0,230	0,038	0,100
ILK	Integrin-linked protein kinase	0,042	-0,543	0,046	-1,173
NFKB2	Nuclear factor NF-kappa-B p100 subunit;Nuclear factor NF-kappa-B p52 subunit	0,028	0,194	0,016	0,322
PDIA6	Protein disulfide-isomerase A6	0,004	-0,304	0,045	-0,366
PEBP1	Phosphatidylethanolamine-binding protein 1;Hippocampal cholinergic neurostimulating peptide	0,048	0,149	0,043	0,256
PRKAR2A	cAMP-dependent protein kinase type II-alpha regulatory subunit	0,009	0,131	0,023	0,126
PRKRA	Interferon-inducible double-stranded RNA-dependent protein kinase activator A	0,019	0,187	0,035	0,268
PSMB4	Proteasome subunit beta type-4	0,040	0,146	0,030	0,139
SAP18	Histone deacetylase complex subunit SAP18	0,004	0,377	0,048	0,238
SF3A2	Splicing factor 3A subunit 2	0,011	0,302	0,041	0,111

THUMPD1	THUMP domain-containing protein 1	0,028	0,137	0,047	0,134
TOMM70A	Mitochondrial import receptor subunit TOM70	0,008	0,217	0,005	0,235
TPM4	Tropomyosin alpha-4 chain	0,016	-0,396	0,016	-0,598
TPRKB	EKC/KEOPS complex subunit TPRKB	0,018	0,335	0,013	0,208
TROVE2	60 kDa SS-A/Ro ribonucleoprotein	0,024	0,196	0,002	0,177
TSPO	Translocator protein	0,005	0,538	0,044	0,295
VCL	Vinculin	0,020	-0,576	0,014	-1,186
YWHAE	14-3-3 protein epsilon	0,010	-0,190	0,030	-0,246

Table S7: *Non-HLA associated MS risk SNPs with proximal (100kb upstream and downstream) genes and protein products in CD4+ T cells, with corresponding chromosome numbe, f-test (marked * if significant) and Student's t-test results of differences in protein expression between MS patients and healthy controls. Significant Student's t-test p-values(p<0,05) and corresponding gene names are outlined. (n=87)*

Chr.	SNP ID	Proximal gene	Protein name	f-test	Sign. f-test (*)	Student's t-test
1	chr1:32738415	**LCK**	Tyrosine-protein kinase Lck	0,850		**0,002**
5	rs67111717	**GRK6**	G protein-coupled receptor kinase 6	0,169		**0,046**
7	rs10271373	**ZC3HAV1**	Zinc finger CCCH-type antiviral protein 1	0,582		**0,016**
8	rs3923387	**EPPK1**	Epiplakin	0,787		**0,039**
11	rs4939490	**CD6**	T-cell differentiation antigen CD6	0,739		**0,043**
11	rs4939490	**CD5**	T-cell surface glycoprotein CD5	0,888		**0,016**
16	rs3809627	**MAPK3**	Mitogen-activated protein kinase 3	0,003	*	**0,044**
16	rs3809627	**CORO1A**	Coronin-1A	0,000	*	**0,003**
17	rs2150879	**CLTC**	Clathrin heavy chain 1	0,323		**0,003**
19	rs11083862	**SAE1**	SUMO-activating enzyme subunit 1;SUMO-activating enzyme subunit 1, N-terminally processed	0,569		**0,025**
20	rs6742	**LIME1**	Lck-interacting transmembrane adapter 1	0,829		**0,008**
22	rs137955	**GRAP2**	GRB2-related adapter protein 2	0,024	*	**0,006**
22	rs140522	**TYMP**	Thymidine phosphorylase	0,587		**0,003**
1	rs67934705	PITHD1	PITH domain-containing protein 1	0,541		0,437
1	rs67934705	LYPLA2	Acyl-protein thioesterase 2	0,037	*	0,403

1	rs67934705	HMGCL	Hydroxymethylglutaryl-CoA lyase, mitochondrial	0,134		0,221
1	rs67934705	SRSF10	Serine/arginine-rich splicing factor 10	0,005	*	0,775
1	chr1:32738415	EIF3I	Eukaryotic translation initiation factor 3 subunit I	0,472		0,231
1	chr1:32738415	HDAC1	Histone deacetylase 1	0,003	*	0,099
1	chr1:154983036	PBXIP1	Pre-B-cell leukemia transcription factor-interacting protein 1	0,535		0,994
1	chr1:154983036	FLAD1	FAD synthase;Molybdenum cofactor biosynthesis protein-like region;FAD synthase region	0,604		0,431
1	rs3737798	COPA	Coatomer subunit alpha;Xenin;Proxenin	0,258		0,084
1	rs6427540	CD48	CD48 antigen	0,851		0,268
1	rs983494	CD48	CD48 antigen	0,851		0,268
2	rs13414105	LBH	Protein LBH	0,679		0,245
2	rs35540610	SP110	Sp110 nuclear body protein	0,022	*	0,137
3	rs6789653	RASA2	Ras GTPase-activating protein 2	0,821		0,453
4	rs2726479	PPA2	Inorganic pyrophosphatase 2, mitochondrial	0,031	*	0,336
4	rs9992763	LEF1	Lymphoid enhancer-binding factor 1	0,304		0,301
5	rs34681760	NSUN2	tRNA (cytosine(34)-C(5))-methyltransferase	0,068		0,778
5	rs32658	TNFAIP8	Tumor necrosis factor alpha-induced protein 8	0,144		0,546
5	rs32658	HSD17B4	Peroxisomal multifunctional enzyme type 2;(3R)-hydroxyacyl-CoA dehydrogenase;Enoyl-CoA hydratase 2	0,995		0,096
5	rs244656	SKP1	S-phase kinase-associated protein 1	0,170		0,245
5	rs244656	PPP2CA	Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	0,354		0,272
5	rs2546890	UBLCP1	Ubiquitin-like domain-containing CTD phosphatase 1	0,901		0,153
5	rs67111717	LMAN2	Vesicular integral-membrane protein VIP36	0,162		0,854
5	rs67111717	RGS14	Regulator of G-protein signaling 14	0,354		0,658
6	rs802730	THEMIS	Protein THEMIS	0,133		0,053
7	rs55858457	EIF3B	Eukaryotic translation initiation factor 3 subunit B	0,293		0,937
7	rs60600003	ELMO1	Engulfment and cell motility protein 1	0,712		0,381

7	chr7:50328339	IKZF1	DNA-binding protein Ikaros	0,322		0,427
8	rs3923387	PUF60	Poly(U)-binding-splicing factor PUF60	0,847		0,271
8	rs3923387	PLEC	Plectin	0,323		0,190
9	rs7855251	ANP32B	Acidic leucine-rich nuclear phosphoprotein 32 family member B	0,454		0,464
9	rs7855251	NANS	Sialic acid synthase	0,613		0,532
10	rs17741873	CAMK2G	Calcium/calmodulin-dependent protein kinase type II subunit gamma	0,233		0,416
11	rs35218683	PSMD13	26S proteasome non-ATPase regulatory subunit 13	0,555		0,489
11	rs61884005	COPB1	Coatomer subunit beta	0,166		0,460
11	rs2269434	PSMC3	26S protease regulatory subunit 6A	0,322		0,484
11	rs11231749	FKBP2	Peptidyl-prolyl cis-trans isomerase FKBP2	0,985		0,943
11	rs11231749	TRMT112	Multifunctional methyltransferase subunit TRM112-like protein	0,084		0,164
11	rs11231749	PRDX5	Peroxiredoxin-5, mitochondrial	0,177		0,762
11	rs531612	CFL1	Cofilin-1	0,379		0,419
11	rs531612	C11orf68	UPF0696 protein C11orf68	0,650		0,850
11	rs531612	DRAP1	Dr1-associated corepressor	0,406		0,210
11	rs531612	SART1	U4/U6.U5 tri-snRNP-associated protein 1	0,827		0,101
11	rs531612	BANF1	Barrier-to-autointegration factor;Barrier-to-autointegration factor, N-terminally processed	0,244		0,952
11	rs34026809	ARCN1	Coatomer subunit delta	0,392		0,824
11	rs12365699	DDX6	Probable ATP-dependent RNA helicase DDX6	0,304		0,167
11	rs6589706	DDX6	Probable ATP-dependent RNA helicase DDX6	0,304		0,167
11	rs4262739	ETS1	Protein C-ets-1	0,291		0,524
12	rs701006	AGAP2	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 2	0,738		0,749
14	rs11852059	GNG2	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-2	0,884		0,250
14	rs12434551	ACTN1	Alpha-actinin-1	0,025	*	0,417
15	rs6496663	IQGAP1	Ras GTPase-activating-like protein IQGAP1	0,599		0,556
16	rs3809627	ALDOA	Fructose-bisphosphate aldolase A	0,977		0,101
16	rs3809627	PPP4C	Serine/threonine-protein phosphatase 4 catalytic subunit	0,909		0,361

17	rs883871	PSMD3	26S proteasome non-ATPase regulatory subunit 3	0,415		0,483
17	rs1026916	STAT5A	Signal transducer and activator of transcription 5A	0,762		0,461
17	rs1026916	STAT3	Signal transducer and activator of transcription 3	0,049	*	0,461
17	rs7222450	FMNL1	Formin-like protein 1	0,448		0,275
17	rs7222450	NPEPPS	Puromycin-sensitive aminopeptidase	0,164		0,383
17	rs7222450	KPNB1	Importin subunit beta-1	0,274		0,841
17	rs9900529	GRB2	Growth factor receptor-bound protein 2	0,376		0,714
19	rs34536443	RAVER1	Ribonucleoprotein PTB-binding 1	0,034	*	0,283
19	rs34536443	ICAM3	Intercellular adhesion molecule 3	0,745		0,675
19	rs34536443	CDC37	Hsp90 co-chaperone Cdc37;Hsp90 co-chaperone Cdc37, N-terminally processed	0,723		0,952
19	rs28834106	CDC37	Hsp90 co-chaperone Cdc37;Hsp90 co-chaperone Cdc37, N-terminally processed	0,723		0,952
19	rs28834106	KRI1	Protein KRI1 homolog	0,772		0,608
19	rs58166386	EPS15L1	Epidermal growth factor receptor substrate 15-like 1	0,733		0,109
19	rs58166386	CHERP	Calcium homeostasis endoplasmic reticulum protein	0,098		0,116
20	rs2585447	PFDN4	Prefoldin subunit 4	0,943		0,215
20	rs2248137	PFDN4	Prefoldin subunit 4	0,943		0,215
21	rs9808753	GART	Trifunctional purine biosynthetic protein adenosine-3;Phosphoribosylamine--glycine ligase;Phosphoribosylformylglycinamidine cyclo-ligase;Phosphoribosylglycinamide formyltransferase	0,667		0,432
22	rs9610458	MAPK1	Mitogen-activated protein kinase 1	0,714		0,829
22	rs9610458	PPM1F	Protein phosphatase 1F	0,263		0,357
22	rs140522	SBF1	Myotubularin-related protein 5	0,821		0,108

Table S8: *Non-HLA associated MS risk SNPs with proximal (100kb upstream and downstream) genes and protein products in CD8+ T cells, with corresponding chromosome number f-test (marked * if significant) and Student's t-test results of differences in protein expression between MS patients and healthy controls. Significant Student's t-test p-values (p<0,05) and the corresponding genes are outlined. (n=107)*

Chr.	SNP ID	Proximal gene	Protein name	f-test	Sign. f-test (*)	Student's t-test
1	rs67934705	**PITHD1**	PITH domain-containing protein 1	0,905		**0,002**
3	rs6789653	**RASA2**	Ras GTPase-activating protein 2	0,642		**0,026**
5	rs32658	**TNFAIP8**	Tumor necrosis factor alpha-induced protein 8	0,004	*	**0,017**
10	rs17741873	**CAMK2G**	Calcium/calmodulin-dependent protein kinase type II subunit gamma;Calcium/calmodulin-dependent protein kinase type II subunit beta	0,063		**0,010**
10	rs1250551	**PPIF**	Peptidyl-prolyl cis-trans isomerase F, mitochondrial	0,080		**0,003**
22	rs9610458	**PPM1F**	Protein phosphatase 1F	0,001	*	**0,043**
1	rs2986736	ACOT7	Cytosolic acyl coenzyme A thioester hydrolase	0,007	*	0,118
1	rs1801133	AGTRAP	Type-1 angiotensin II receptor-associated protein	0,197		0,327
1	rs6427540	CD48	CD48 antigen	0,557		0,808
1	rs983494	CD48	CD48 antigen	0,557		0,808
1	rs3737798	COPA	Coatomer subunit alpha;Xenin;Proxenin	0,259		0,980
1	chr1:32738415	EIF3I	Eukaryotic translation initiation factor 3 subunit I	0,843		0,522
1	chr1:32738415	HDAC1	Histone deacetylase 1	0,357		0,448
1	rs67934705	HMGCL	Hydroxymethylglutaryl-CoA lyase, mitochondrial	0,304		0,155
1	rs72922276	JAK1	Tyrosine-protein kinase JAK1	0,536		0,373
1	chr1:32738415	LCK	Tyrosine-protein kinase Lck	0,466		0,161
1	rs67934705	LYPLA2	Acyl-protein thioesterase 2	0,055		0,139
1	rs483180	PHGDH	D-3-phosphoglycerate dehydrogenase	0,526		0,469
1	rs6672420	RUNX3	Runt-related transcription factor 3	0,584		0,143
1	rs67934705	SRSF10	Serine/arginine-rich splicing factor 10	0,081		0,133
2	rs12622670	PLEK	Pleckstrin	0,792		0,227

2	rs6738544	STAT4	Signal transducer and activator of transcription 4	0,029	*	0,811
3	rs11919880	GLB1	Beta-galactosidase	0,191		0,092
4	rs6533052	CISD2	CDGSH iron-sulfur domain-containing protein 2	0,214		0,162
4	rs9992763	LEF1	Lymphoid enhancer-binding factor 1	0,134		0,297
4	rs2726479	PPA2	Inorganic pyrophosphatase 2, mitochondrial	0,431		0,064
5	rs67111717	DBN1	Drebrin	0,694		0,909
5	rs67111717	GRK6	G protein-coupled receptor kinase 6	0,018	*	0,146
5	rs32658	HSD17B4	Peroxisomal multifunctional enzyme type 2;(3R)-hydroxyacyl-CoA dehydrogenase;Enoyl-CoA hydratase 2	0,196		0,135
5	rs67111717	LMAN2	Vesicular integral-membrane protein VIP36	0,736		0,632
5	rs34681760	NSUN2	tRNA (cytosine(34)-C(5))-methyltransferase	0,843		0,098
5	rs244656	PPP2CA	Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform;Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform	0,744		0,575
5	rs67111717	RGS14	Regulator of G-protein signaling 14	0,205		0,311
5	rs244656	SKP1	S-phase kinase-associated protein 1	0,361		0,926
5	rs2546890	UBLCP1	Ubiquitin-like domain-containing CTD phosphatase 1	0,037	*	0,063
6	rs802730	THEMIS	Protein THEMIS	0,625		0,302
7	rs55858457	EIF3B	Eukaryotic translation initiation factor 3 subunit B	0,347		0,932
7	rs60600003	ELMO1	Engulfment and cell motility protein 1	0,081		0,706
7	chr7:50328339	IKZF1	DNA-binding protein Ikaros	0,214		0,899
7	rs4728142	TNPO3	Transportin-3	0,191		0,806
7	rs10271373	ZC3HAV1	Zinc finger CCCH-type antiviral protein 1	0,884		0,456
8	rs3923387	PLEC	Plectin	0,115		0,115
8	rs3923387	PUF60	Poly(U)-binding-splicing factor PUF60	0,910		0,353
9	rs7855251	ANP32B	Acidic leucine-rich nuclear phosphoprotein 32 family member B	0,783		0,341
9	rs7855251	NANS	Sialic acid synthase	0,412		0,872
10	rs12722559	RBM17	Splicing factor 45	0,286		0,421

10	rs11256593	RBM17	Splicing factor 45	0,286		0,421
11	rs34026809	ARCN1	Coatomer subunit delta	0,891		0,777
11	rs531612	BANF1	Barrier-to-autointegration factor;Barrier-to-autointegration factor, N-terminally processed	0,504		0,604
11	rs4939490	CD5	T-cell surface glycoprotein CD5	0,168		0,487
11	rs4939490	CD6	T-cell differentiation antigen CD6	0,131		0,333
11	rs531612	CFL1	Cofilin-1	0,452		0,549
11	rs61884005	COPB1	Coatomer subunit beta	0,731		0,638
11	rs531612	CTSW	Cathepsin W	0,512		0,864
11	rs12365699	DDX6	Probable ATP-dependent RNA helicase DDX6	0,813		0,069
11	rs6589706	DDX6	Probable ATP-dependent RNA helicase DDX6	0,813		0,069
11	rs531612	DRAP1	Dr1-associated corepressor	0,725		0,526
11	rs4262739	ETS1	Protein C-ets-1	0,365		0,327
11	rs11231749	FKBP2	Peptidyl-prolyl cis-trans isomerase FKBP2	0,994		0,709
11	rs35218683	IFITM1	Interferon-induced transmembrane protein 1;Interferon-induced transmembrane protein 2;Interferon-induced transmembrane protein 3	0,334		0,623
11	rs35218683	IFITM2	Interferon-induced transmembrane protein 1;Interferon-induced transmembrane protein 2;Interferon-induced transmembrane protein 3	0,334		0,623
11	rs35218683	IFITM3	Interferon-induced transmembrane protein 1;Interferon-induced transmembrane protein 2;Interferon-induced transmembrane protein 3	0,334		0,623
11	rs11231749	PRDX5	Peroxiredoxin-5, mitochondrial	0,062		0,450
11	rs2269434	PSMC3	26S protease regulatory subunit 6A	0,371		0,836
11	rs35218683	PSMD13	26S proteasome non-ATPase regulatory subunit 13	0,646		0,820
11	rs531612	SART1	U4/U6.U5 tri-snRNP-associated protein 1	0,071		0,566
11	rs11231749	TRMT112	Multifunctional methyltransferase subunit TRM112-like protein	0,152		0,461
11	rs6589939	UBASH3B	Ubiquitin-associated and SH3 domain-containing protein B	0,419		0,431

12	rs701006	AGAP2	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 2	0,429		0,772
12	rs1800693	CD9	CD9 antigen	0,805		0,086
12	rs12832171	CD9	CD9 antigen	0,805		0,086
12	rs701006	TSFM	Elongation factor Ts, mitochondrial	0,997		0,195
13	rs9591325	ST13P4	Hsc70-interacting protein;Putative protein FAM10A5;Putative protein FAM10A4	0,215		0,724
13	chr13:100026952	TM9SF2	Transmembrane 9 superfamily member 2	0,863		0,137
14	rs12434551	ACTN1	Alpha-actinin-1	0,176		0,464
14	rs11852059	GNG2	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-2	0,897		0,521
15	rs6496663	IQGAP1	Ras GTPase-activating-like protein IQGAP1	0,073		0,983
16	rs3809627	ALDOA	Fructose-bisphosphate aldolase A	0,013	*	0,088
16	rs3809627	CORO1A	Coronin-1A	0,098		0,365
16	rs3809627	PPP4C	Serine/threonine-protein phosphatase 4 catalytic subunit	0,515		0,393
17	rs2150879	CLTC	Clathrin heavy chain 1	0,417		0,173
17	rs7222450	FMNL1	Formin-like protein 1	0,286		0,680
17	rs9900529	GRB2	Growth factor receptor-bound protein 2	0,027	*	0,628
17	rs11079784	KPNB1	Importin subunit beta-1	0,071		0,444
17	rs11079784	NPEPPS	Puromycin-sensitive aminopeptidase	0,274		0,716
17	rs883871	PSMD3	26S proteasome non-ATPase regulatory subunit 3	0,163		0,342
17	rs1026916	STAT3	Signal transducer and activator of transcription 3	0,810		0,210
17	rs1026916	STAT5A	Signal transducer and activator of transcription 5A	0,354		0,633
18	rs2469434	CD226	CD226 antigen	0,633		0,297
19	rs34536443	CDC37	Hsp90 co-chaperone Cdc37;Hsp90 co-chaperone Cdc37, N-terminally processed	0,934		0,137
19	rs28834106	CDC37	Hsp90 co-chaperone Cdc37;Hsp90 co-chaperone Cdc37, N-terminally processed	0,934		0,137
19	rs58166386	CHERP	Calcium homeostasis endoplasmic reticulum protein	0,311		0,344
19	rs58166386	EPS15L1	Epidermal growth factor receptor substrate 15-like 1	0,402		0,098
19	rs34536443	ICAM3	Intercellular adhesion molecule 3	0,806		0,767

19	rs4808760	IFI30	Gamma-interferon-inducible lysosomal thiol reductase	0,246		0,319
19	rs34536443	RAVER1	Ribonucleoprotein PTB-binding 1	0,340		0,367
19	rs11083862	SAE1	SUMO-activating enzyme subunit 1;SUMO-activating enzyme subunit 1, N-terminally processed	0,796		0,075
20	rs6742	LIME1	Lck-interacting transmembrane adapter 1	0,227		0,838
20	rs2585447	PFDN4	Prefoldin subunit 4	0,708		0,922
20	rs2248137	PFDN4	Prefoldin subunit 4	0,708		0,922
21	rs9808753	GART	Trifunctional purine biosynthetic protein adenosine-3;Phosphoribosylamine--glycine ligase;Phosphoribosylformylglycinamidine cyclo-ligase;Phosphoribosylglycinamide formyltransferase	0,045	*	0,471
22	rs140522	ARSA	Arylsulfatase A;Arylsulfatase A component B;Arylsulfatase A component C	0,643		0,372
22	rs137955	GRAP2	GRB2-related adapter protein 2	0,137		0,409
22	rs9610458	MAPK1	Mitogen-activated protein kinase 1	0,110		0,595
22	rs140522	SBF1	Myotubularin-related protein 5	0,585		0,188
22	rs140522	TYMP	Thymidine phosphorylase	0,115		0,075
22	rs137955	UQCRFS1P1	Cytochrome b-c1 complex subunit Rieske, mitochondrial;Cytochrome b-c1 complex subunit 11;Putative cytochrome b-c1 complex subunit Rieske-like protein 1	0,870		0,951

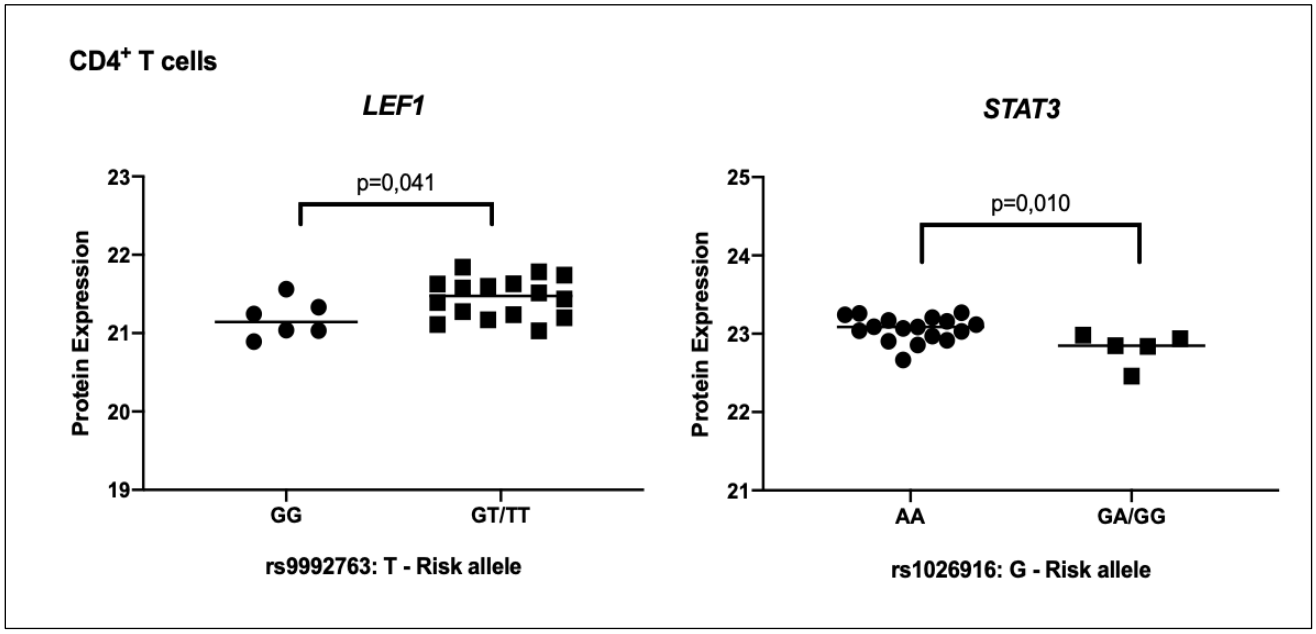

Figure S1: Genotype dependent expression of proteins encoded by LEF1 and STAT3

Genotype-dependent expression of proteins in CD4+ T cells encoded from genes located in a 100kb upstream and downstream region from MS risk SNPs. The scatterplot present log2-transformed protein intensities of CD4+ T cell proteins expressed from genes in the selected region as a function of the MS risk SNP genotype in samples from MS cases and healthy controls. The protein intensities are sorted according to groups, i.e. carriers or non-carrier of the risk allele. An unpaired students t-test was performed, with the p-value illustrated in each plot. The horizontal lines in each group represent the median within the group.

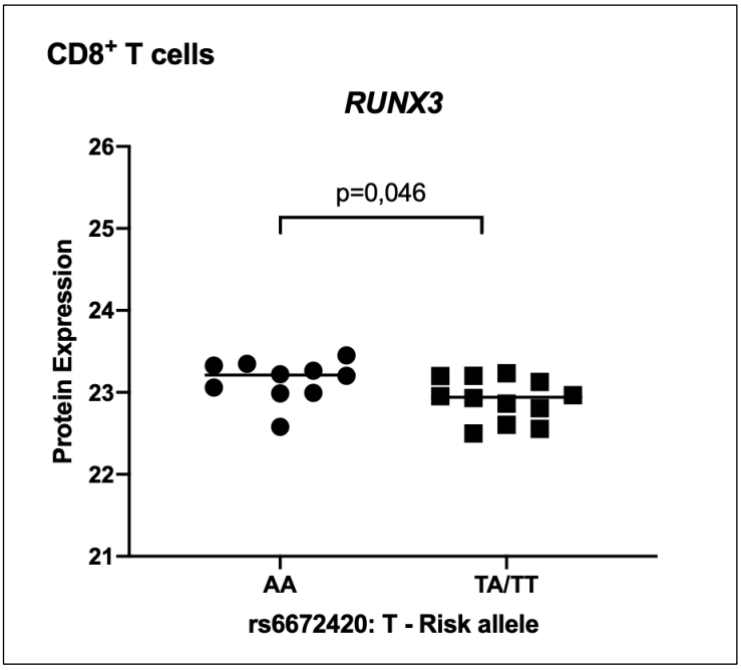

Figure S2: Genotype-dependent expression of protein encoded by RUNX3. *Genotype-dependent expression of proteins in CD8+ T cells encoded from genes located in a 100kb upstream and downstream region from MS risk SNPs. The scatterplot present log2-transformed protein intensities of CD8+ T cell proteins expressed from genes in the selected region as a function of the MS risk SNP genotype in samples from MS cases and healthy controls. The protein intensities are sorted according to groups, i.e. carriers or non-carrier of the risk allele. An unpaired students t-test was performed, with the p-value illustrated in each plot. The horizontal lines in each group represent the median within the group.*

Table S9: *Included CD4+ T cell samples from MS patients and healthy controls, with sample ID, percentage of live and activated CD4+ T cells after in vitro stimulation in samples stimulated with high dose and no antibodies, and the number of storage months on liquid nitrogen (LN_2).*

Sample	High antibody dose		No antibodies		Months on LN_2
	Live cells (%)	Activated cells (%)	Live cells (%)	Activated cells (%)	
MS2	91,8	55,3	89,3	0,7	80
MS4	69,1	60,2	71,9	1,6	83
MS6	89,1	50,0	87,4	0,8	105
MS9	75,3	42,9	67,2	0,5	104

MS11	75,4	70,3	78,8	1,0	104
MS12	88,0	61,8	85,0	0,9	103
MS15	85,3	62,6	82,3	0,7	85
MS16	72,2	54,6	82,4	0,8	85
MS17	68,6	63,7	73,0	1,5	85
MS18	83,7	75,5	80,8	4,8	84
MS19	67,2	56,8	64,6	1,1	83
MS20	75,0	60,4	74,2	1,2	80
MS21	92,8	48,4	91,1	1,0	56
MS22	92,0	48,7	92,5	0,7	55
MS23	86,2	58,7	90,2	0,5	53
MS24	94,0	51,5	90,3	0,5	47
MS25	96,8	62,6	94,5	0,9	46
MS26	92,4	50,6	84,3	1,0	45
MS27	91,2	62,6	88,2	1,0	41
MS28	78,7	47,2	70,8	0,7	40
HC1	89,5	51,8	83,5	0,8	85
HC2	88,3	60,0	87,4	0,7	85
HC4	88,6	53,6	78,0	1,1	49
HC6	78,8	55,1	81,1	1,1	42
HC7	87,2	50,1	82,0	0,7	43
HC8	90,3	75,3	84,7	1,2	42
HC9	91,9	63,0	88,2	0,7	77
HC11	89,8	53,1	88,2	0,9	81
HC12	85,7	57,7	86,1	1,2	84
HC13	86,8	67,2	85,1	1,1	83
HC14	90,9	57,3	88,7	0,7	102
HC15	88,2	53,2	81,0	1,0	105
HC16	82,9	64,0	86,2	0,5	105
HC18	67,8	58,3	62,0	1,1	105
HC20	82,9	68,9	80,9	0,9	103
HC21	82,9	67,1	82,8	1,1	102
HC22	74,9	67,4	73,2	0,9	101
HC24	66,5	65,8	77,6	1,1	83
HC25	63,0	54,9	65,0	5,4	82
HC27	85,1	70,7	84,0	1,4	76

Table S10: *Key information of MS patients whose CD4+ T cells were analyzed, including age category(*), years since disease onset, Expanded disability status scale (EDSS), oligoclonal bands (OCB) and age at disease onset.*

** Age category 1=25-29; 2=30-34; 3=35-39; 4=40-44; 5=45-49; 6=50-54; 7=55-59; 8=60-64; 9=65-69;10=70-74.*

Patient	Age category	Years since disease onset	EDSS	OCB	Age at MS Onset
MS1	8	37	2,0	N/A	23
MS2	6	20	0,0	Yes	31
MS3	5	19	3,0	Yes	27
MS4	6	26	2,0	Yes	27
MS5	5	20	0,5	Yes	26
MS6	10	47	5,5	No	25
MS7	2	10	2,0	Yes	23
MS8	7	10	2,0	Yes	45
MS9	3	17	3,5	Yes	22
MS10	6	14	5	Yes	38
MS11	5	18	1,5	Yes	30
MS12	4	15	1,5	Yes	24
MS13	4	15	2,0	Yes	27
MS14	6	12	2,5	Yes	42
MS15	4	14	2,0	Yes	26
MS16	5	11	0,0	Yes	38
MS17	3	9	4	Yes	28
MS18	1	8	1,5	Yes	21
MS19	5	10	1,5	Yes	35
MS20	5	12	2,5	Yes	34
MS21	4	17	1,5	No	25
MS23	4	11	2,5	Yes	31
MS24	2	5	1,5	Yes	24
MS25	2	20	3,5	Yes	14

MS26	2	5	1,0	Yes	29
MS27	7	23	1,5	Yes	33
MS28	2	4	1,0	Yes	28
Summary					
Mean/median* (range)	44.9 (29 - 72)	15,8 (5 - 47)	2.0* (0.0 - 5.5)	N/A	28.8 (14 - 45)

Table S11: *Samples of CD4+ T cells from MS patients and healthy controls (HC) that were thawed and stimulated in vitro, with corresponding percentage of live and activated cells detected in Flow cytometry quality control, and the number of storage months on liquid nitrogen (LN_2)*

Sample	**High antibody dose**		**Low antibody dose**		**No antibodies**		**Months on LN_2**
	Live cells (%)	**Activated cells (%)**	**Live cells (%)**	**Activated cells (%)**	**Live cells (%)**	**Activated cells (%)**	
MS1	94,9	45,0	93,4	1,3	93,2	0,5	76
MS2	91,8	55,3	91,0	1,3	89,3	0,7	80
MS3	60,8	62,7	75,4	6,4	66,5	0,8	38
MS4	69,1	60,2	72,4	36,6	71,9	1,6	83
MS5	87,9	34,1	87,9	3,8	83,8	0,8	105
MS6	89,1	50,0	88,9	2,4	87,4	0,8	105
MS7	81,6	49,7	80,8	1,2	79,3	0,2	105
MS8	67,9	64,8	66,2	8,8	72,6	1,2	105
MS9	75,3	42,9	70,5	4,1	67,2	0,5	104
MS10	68,3	40,8	78,1	5,9	75,6	0,6	104
MS11	75,4	70,3	75,8	12,1	78,8	1,0	104
MS12	88,0	61,8	86,2	6,5	85,0	0,9	103
MS13	64,9	50,2	69,0	14,2	73,0	0,5	102
MS14	67,3	64,6	73,2	8,5	63,8	0,4	102
MS15	85,3	62,6	87,0	3,8	82,3	0,7	85

MS16	72,2	54,6	81,0	3,4	82,4	0,8	85
MS17	68,6	63,7	69,7	16,7	73,0	1,5	85
MS18	83,7	75,5	80,5	16,3	80,8	4,8	84
MS19	67,2	56,8	71,8	11,2	64,6	1,1	83
MS20	75,0	60,4	74,6	16,5	74,2	1,2	80
MS21	92,8	48,4	92,2	6,1	91,1	1,0	56
MS22	92,0	48,7	92,9	8,8	92,5	0,7	55
MS23	86,2	58,7	88,5	10,4	90,2	0,5	53
MS24	94,0	51,5	93,0	1,2	90,3	0,5	47
MS25	96,8	62,6	96,7	5,1	94,5	0,9	46
MS26	92,4	50,6	90,1	2,5	84,3	1,0	45
MS27	91,2	62,6	92,8	6,4	88,2	1,0	41
MS28	78,7	47,2	74,2	4,2	70,8	0,7	40
HC1	89,5	51,8	86,9	4,3	83,5	0,8	85
HC2	88,3	60,0	86,0	4,3	87,4	0,7	85
HC3	30,0	37,8	31,7	3,6	25,6	2,8	52
HC4	88,6	53,6	86,7	2,8	78,0	1,1	49
HC5	93,4	47,9	93,5	1,7	90,8	0,8	43
HC6	78,8	55,1	83,5	2,8	81,1	1,1	42
HC7	87,2	50,1	83,8	5,0	82,0	0,7	43
HC8	90,3	75,3	91,6	8,2	84,7	1,2	42
HC9	91,9	63,0	90,4	19,7	88,2	0,7	77
HC10	22,6	55,0	19,0	11,5	18,2	5,2	83
HC11	89,8	53,1	87,4	9,1	88,2	0,9	81
HC12	85,7	57,7	87,7	7,9	86,1	1,2	84
HC13	86,8	67,2	87,9	16,6	85,1	1,1	83
HC14	90,9	57,3	91,0	5,6	88,7	0,7	102
HC15	88,2	53,2	86,9	3,0	81,0	1,0	105
HC16	82,9	64,0	84,1	15,3	86,2	0,5	105
HC17	60,9	60,9	66,2	10,9	72,7	0,4	105
HC18	67,8	58,3	63,4	7,5	62,0	1,1	105
HC19	65,4	53,0	66,3	5,2	61,6	1,6	103
HC20	82,9	68,9	83,5	5,2	80,9	0,9	103

HC21	82,9	67,1	83,1	7,8	82,8	1,1	102
HC22	74,9	67,4	75,5	11,5	73,2	0,9	101
HC23	11,5	33,0	10,6	7,7	9,1	10,1	83
HC24	66,5	65,8	69,8	11,1	77,6	1,1	83
HC25	63,0	54,9	75,9	9,3	65,0	5,4	82
HC26	55,4	62,6	66,4	11,1	62,0	0,9	81
HC27	85,1	70,7	89,9	6,0	84,0	1,4	76

Table S12: *Samples of CD4+ T cells from MS patients and healthy controls (HC), with corresponding protein concentrations of samples unstimulated and stimulated with high and low antibody doses. Samples below the marked line in addition to all samples stimulated with low antibody dose were excluded from further analysis. The remaining samples above the line were included.*

Patient	**Protein Concentration (µg/µl)**			**Control**	**Protein Concentration (µg/µl)**		
	High antibody dose	**Low antibody dose**	**No antibodies**		**High antibody dose**	**Low antibody dose**	**No antibodies**
MS2	0,300	0,242	0,276	**HC1**	0,446	0,477	0,367
MS4	0,799	1,012	0,772	**HC2**	0,763	0,717	0,352
MS6	0,270	0,507	0,273	**HC4**	0,361	0,422	0,236
MS9	0,708	0,641	0,662	**HC6**	0,748	0,684	0,291
MS11	0,818	0,478	0,413	**HC7**	0,369	0,416	0,390
MS12	0,602	0,285	0,297	**HC8**	0,571	0,608	0,315
MS15	0,339	0,336	0,266	**HC9**	0,715	0,617	0,989
MS16	0,339	0,369	0,331	**HC11**	0,313	0,393	0,466
MS17	0,384	1,054	0,407	**HC12**	0,318	0,598	0,638
MS18	0,656	0,384	0,540	**HC13**	0,461	0,316	0,336
MS19	0,484	0,650	0,791	**HC14**	0,318	0,580	0,276
MS20	0,641	0,297	0,410	**HC15**	0,301	0,286	0,339
MS21	0,461	0,297	0,273	**HC16**	0,425	0,658	0,413

MS22	0,355	0,248	0,318	**HC18**	0,635	0,310	0,260
MS23	0,316	0,431	0,428	**HC20**	0,300	0,279	1,003
MS24	0,653	0,366	0,366	**HC21**	0,939	0,349	0,303
MS25	0,313	0,289	0,351	**HC22**	0,361	0,355	0,897
MS26	0,638	0,369	0,325	**HC24**	0,650	0,336	0,313
MS27	0,399	0,419	0,263	**HC25**	0,668	0,321	0,668
MS28	0,345	0,307	0,656	**HC27**	0,364	0,239	0,696
MS1	0,245	0,461	0,525	**HC3**	0,300	0,735	0,215
MS3	0,446	0,721	0,685	**HC5**	0,513	0,556	0,550
MS5	0,416	0,422	0,898	**HC10**	0,283	0,325	0,425
MS7	0,279	0,325	0,373	**HC17**	0,848	0,638	1,119
MS8	0,856	0,793	0,592	**HC19**	0,306	0,452	0,611
MS10	0,662	0,688	0,546	**HC23**	0,334	0,310	0,797
MS13	0,788	0,499	0,679	**HC26**	0,378	0,744	0,726
MS14	0,271	0,541	0,629	**HC28**	0,382	0,556	0,370

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway