Genetic and molecular studies of multiple sclerosis associated genes

Pankaj Kumar Keshari

*Dissertation for the degree of Philosophiae Doctor*

Department of Neurology
Oslo University Hospital
and
Institute of Clinical Medicine
Faculty of Medicine
University of Oslo
Norway
© Pankaj Kumar Keshari, 2017

Series of dissertations submitted to the Faculty of Medicine, University of Oslo

ISBN 978-82-8377-008-7

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.
# Table of Contents

## Contents

AKNOWLEDGEMENTS ................................................................................................. 1

ABBREVIATIONS ........................................................................................................... 4

1 INTRODUCTION ........................................................................................................... 7

1.1 Multiple sclerosis .................................................................................................. 7

1.1.1 Epidemiology ...................................................................................................... 7

1.1.2 Genetic epidemiology of MS ............................................................................. 7

1.1.3 Clinical course .................................................................................................... 8

1.1.4 Hallmarks of MS pathology ............................................................................. 9

1.1.5 Treatment .......................................................................................................... 10

1.2 The immune system ............................................................................................. 10

1.2.1 T lymphocytes .................................................................................................. 11

1.2.2 Different subsets of T cells .............................................................................. 13

1.2.3 B lymphocytes .................................................................................................. 14

1.2.4 Immunopathogenesis of MS .......................................................................... 14

1.3 Mapping of genetic risk factors .......................................................................... 15

1.3.1 Genetic variations and their importance for human health ......................... 16

1.3.2 Haplotypes and linkage disequilibrium ......................................................... 17

1.3.3 Association of HLA in MS ............................................................................ 18

1.3.4 Complex diseases and genome-wide association studies .......................... 19

1.3.5 Discovery of non-HLA associations with MS through GWAS .................... 20

1.3.6 Functional follow-up of the GWAS signal .................................................... 21

1.4 Environmental factors in MS .............................................................................. 22

1.4.1 Sun exposure and vitamin D ......................................................................... 22

1.4.2 Epstein–Barr virus .......................................................................................... 22

1.4.3 Smoking ............................................................................................................ 23

1.4.4 Obesity ............................................................................................................ 23

1.4.5 Other environmental factors .......................................................................... 23

1.5 Gene-environment interaction ............................................................................ 24

1.6 The 16p13.13 region and MS susceptibility ...................................................... 24

1.7 Allele-specific expression and its biological mechanism ............................... 26

2 STUDY OBJECTIVE .................................................................................................. 28

3 SUMMARY OF THE PAPERS IN THE THESIS .................................................... 29
# METHODOLOGICAL CONSIDERATIONS

## 4.1 Study population

## 4.2 Collection of human peripheral T cells

## 4.3 DNA and RNA extraction and quality analysis

## 4.4 SNP genotyping

## 4.5 Gene expression analysis by quantitative real-time PCR

## 4.6 Assessment of allele-specific expression

## 4.7 Transfection of Jurkat T cells

## 4.8 Activation of T cells for *in vitro* experiments

## 4.9 Measurement of T cell stimulation

## 4.10 Statistical analysis

# GENERAL DISCUSSION

## 5.1 *Cis*-regulatory roles for MS-susceptibility SNPs in *IQGAP1* and *IKZF3*

## 5.2 The complex role of the 16p13.13 chromosomal region in MS susceptibility

## 5.3 Regulatory role of MS-associated intronic *CLEC16A* SNPs

## 5.4 Co-expression of 16p13.13 locus genes *CIITA, DEXI, CLEC16A*, and *SOCS1* in peripheral T cells

## 5.5 Functional analysis of DEXI

# FUTURE PERSPECTIVES

# CONCLUSIONS

# REFERENCES

APPENDIX, PAPERS I-III

**3.1 Paper I** ........................................................................................................ 29

**3.2 Paper II** ........................................................................................................ 30

**3.3 Paper III** ........................................................................................................ 30

**4 METHODOLOGICAL CONSIDERATIONS** .......................................................... 32

**4.1 Study population** .......................................................................................... 32

**4.2 Collection of human peripheral T cells** .......................................................... 33

**4.3 DNA and RNA extraction and quality analysis** ............................................... 33

**4.4 SNP genotyping** .......................................................................................... 34

**4.5 Gene expression analysis by quantitative real-time PCR** ............................... 36

**4.6 Assessment of allele-specific expression** ....................................................... 38

**4.7 Transfection of Jurkat T cells** ....................................................................... 38

**4.8 Activation of T cells for *in vitro* experiments** .............................................. 39

**4.9 Measurement of T cell stimulation** ............................................................... 40

**4.10 Statistical analysis** ...................................................................................... 40

**5 GENERAL DISCUSSION** .................................................................................. 43

**5.1 *Cis*-regulatory roles for MS-susceptibility SNPs in *IQGAP1* and *IKZF3*** ........ 44

**5.2 The complex role of the 16p13.13 chromosomal region in MS susceptibility*** ... 48

**5.3 Regulatory role of MS-associated intronic *CLEC16A* SNPs** ......................... 49

**5.4 Co-expression of 16p13.13 locus genes *CIITA, DEXI, CLEC16A*, and *SOCS1* in peripheral T cells** ................................................................. 51

**5.5 Functional analysis of DEXI** ........................................................................ 52

**6 FUTURE PERSPECTIVES** ................................................................................ 54

**7 CONCLUSIONS** ............................................................................................. 55

**8 REFERENCES** ................................................................................................... 56

**APPENDIX, PAPERS I-III** .................................................................................. 68
ACKNOWLEDGEMENTS

The work presented in this thesis was carried out in the research group of Professor Hanne F. Harbo, Department of Neurology, Oslo University Hospital, and Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway from August 2013 – December 2016. The work was performed at Neuroscience Research Unit, Domus Medica 4. I am grateful to South-Eastern Norway Regional Health Authority for providing me a three years Doctoral fellowship.

I would sincerely like to thank many people who have been a part of my PhD journey, and without whom it would not have been possible to complete this Doctoral thesis.

First of all, I would like to express my heartiest gratitude to my supervisors, especially my main supervisor Dr. Tone Berge, who has been an excellent guide throughout my PhD education. Her enthusiasm and great knowledge has inspired me enormously and helped me all the time during my research and writing of this thesis. She was always polite and keeping the door of her office open for all scientific discussions. She has been an important part in reviewing my manuscripts critically and developing my scientific writing. Thank you for your encouragement and motivation during my research.

I wish to express my gratitude to my co-supervisor Dr. Steffan D. Bos-Haugen for his constant support and being instrumental in my scientific development as a researcher. Thank you for being an excellent advisor and for being available whenever I needed during the journey of my PhD study. You have been very nice in teaching and explaining the basic genetics using colourful drawing. I appreciate all the input and supervision I had from you.

I wish to express my heartfelt gratitude to my co-supervisor and group leader Professor Hanne F. Harbo for providing me an excellent opportunity to carry out my PhD in her research facilities. Her enthusiasm, positive spirit and great knowledge in multiple sclerosis have inspired and motivated me enormously. She has always been very supportive and encouraging. It was really a great pleasure and good experience to work in a friendly research environment. I truly could not have imagined having better advisors and mentors for my PhD education.
Besides my supervisors, I would like to thank my colleagues in the multiple sclerosis research group Elisabeth, Gro, Piotr, Pål, Marte, Sigrid, Stine Marit, Mona, Christian, Ingvild, Anna, Ina, Einar, Anja, Ingrid and Fernanda for creating an inspiring environment, both scientifically and socially. In addition, I would also like to thank the other research groups at Domus Medica 4, for providing a good working and social environment.

I would also thank to the co-authors on the papers for their significant contributions.

I would like to thank Department of Neurology, Oslo University Hospital represented by Professor Espen Dietrichs and Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway for allowing me to work for three years on Multiple sclerosis.

A heartily thanks to my fellow colleague and friend Zafar Iqbal with whom I shared several scientific and non-scientific discussions. You have been always inspiring and motivating me saying that one day I will be there and no need to worry. I heartily thank for all your support and input including fruitful social discussions. Additionally, I would like to thank my all dear friends with whom I shared my all frustrations and concerns. Thank you for your joyful moments, inspiration and motivation. Thank you for everything.

I would like to heartily thank my family for their constant support and motivation. I express my deepest and heartiest gratitude to my dear Father Mr. Shyamlal Keshari, Mother Mrs. Shanti Keshari, Elder brother Mr. Shambhunath Kesharwani, and Sister in-law Mrs. Vimala Kesharwani who have always supported me financially, emotionally and encouraged me at each and every step of my life. Truly, without your blessings and precious supports in hard times of my life, it would not be possible to achieve this rewarding goal.

I wish to express my gratitude to my uncle Mr. Laxmandas Gupta and relatives including Mr. Hari Prasad Gupta for their well wishes and motivation throughout my life and PhD study. Thank you for your blessings.

Last, but not least, I thank my beloved wife Juhi Kesharwani for her patience and constant support along the way, especially during writing of my doctoral thesis. You have been very kind and understanding throughout my PhD study and in life in general. You are an important person who filled my life with joy and happiness. Thank you for your love, care and all your support in my life. It would have been very difficult to achieve this goal without your cooperation. I love you. Finally, I hope your positive support throughout my life.
I am blessed with one angel who is my daughter Pankhudi Keshari. She was about to be born when I got this exciting opportunity to pursue my PhD education. I feel so fortunate to have you in the journey of my life. Every evening you are waiting for me that makes me a special and important person. Your smiley face and hugs made me tired less, and gave me inspiration. Thanks for being such a perfect baby when daddy had to focus on his PhD work and thesis writing. I wish you all the best for every thing you do in your life.

Finally, I thank my nephews Mukesh, Nikesh and nieces Ranjana, Kanchan and Arti for their love, support and making me happy always that counts a lot in the journey of my life. I wish you all the best.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASE</td>
<td>allele-specific expression</td>
<td>EBV</td>
<td>Epstein-Barr virus</td>
</tr>
<tr>
<td>AI</td>
<td>allelic imbalance</td>
<td>EDSS</td>
<td>expanded disability status scale</td>
</tr>
<tr>
<td>APC</td>
<td>antigen presenting cell</td>
<td>eQTL</td>
<td>expression quantitative trait locus</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
<td>FRET</td>
<td>fluorescence resonance energy transfer</td>
</tr>
<tr>
<td>CD</td>
<td>clusters of differentiation</td>
<td>GWAS</td>
<td>genome-wide association study</td>
</tr>
<tr>
<td>CDCV</td>
<td>common disease–common variant</td>
<td>HLA</td>
<td>human leucocyte antigen</td>
</tr>
<tr>
<td>cDNA</td>
<td>complimentary DNA</td>
<td>IKZF3</td>
<td>IKAROS family zinc finger 3</td>
</tr>
<tr>
<td>CLEC16A</td>
<td>C-type lectin-like domain family 16A</td>
<td>IL</td>
<td>interleukin</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
<td>IM</td>
<td>infectious mononucleosis</td>
</tr>
<tr>
<td>CSF</td>
<td>cerebrospinal fluid</td>
<td>IMSGC</td>
<td>International Multiple Sclerosis Genetics Consortium</td>
</tr>
<tr>
<td>C&lt;sub&gt;T&lt;/sub&gt;</td>
<td>threshold cycle</td>
<td>INF-γ</td>
<td>interferon gamma</td>
</tr>
<tr>
<td>CVID</td>
<td>common variable immunodeficiency</td>
<td>IO</td>
<td>ionomycin</td>
</tr>
<tr>
<td>CIITA</td>
<td>major histocompatibility complex class II transactivator</td>
<td>IQGAP1</td>
<td>IQ Motif containing GTPase activating protein 1</td>
</tr>
<tr>
<td>DEXI</td>
<td>dexamethasone-induced transcript</td>
<td>ITAM</td>
<td>immunoreceptor tyrosine-based activation motif</td>
</tr>
<tr>
<td>DCs</td>
<td>dendritic cells</td>
<td>LD</td>
<td>linkage disequilibrium</td>
</tr>
<tr>
<td>DHSs</td>
<td>DNase I hypersensitive sites</td>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
<td>MRI</td>
<td>magnetic resonance imaging</td>
</tr>
<tr>
<td>EAE</td>
<td>experimental autoimmune encephalomyelitis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>multiple sclerosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NK cells</td>
<td>natural killer cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCB</td>
<td>oligoclonal bands</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>odds ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PBMC</td>
<td>peripheral blood mononuclear cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKC</td>
<td>protein kinase C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMA</td>
<td>phorbol 12-myristate-13-acetate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PML</td>
<td>progressive multifocal leukoencephalopathy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPMS</td>
<td>primary progressive MS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>qPCR</td>
<td>quantitative real-time PCR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA</td>
<td>rheumatoid arthritis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIN</td>
<td>RNA integrity number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RRMS</td>
<td>relapsing remitting MS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>siRNA</td>
<td>small interfering RNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLE</td>
<td>systemic lupus erythematosus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNPs</td>
<td>single nucleotide polymorphisms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOCS1</td>
<td>suppressor of cytokine signalling 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPMS</td>
<td>secondary progressive multiple sclerosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1D</td>
<td>type 1 diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBP</td>
<td>TATA-binding protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCR</td>
<td>T-cell receptor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TF</td>
<td>transcription factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGFβ</td>
<td>transforming growth factor beta</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Th</td>
<td>T helper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNF</td>
<td>tumour necrosis factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tregs</td>
<td>regulatory T cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


3. **Keshari PK**, Leikfoss IS, Page C, Celius EG, Harbo HF, Bos SD, Berge T. *DEXI* expression is down-modulated in Jurkat T cells upon cell activation. Unpublished manuscript
INTRODUCTION

1 INTRODUCTION

1.1 Multiple sclerosis

Since 1868, when the first detailed clinical description of multiple sclerosis (MS) was published by J.M. Charcot, our knowledge of the aetiology, epidemiology, pathogenesis, and treatment of the disease has increased. MS is a chronic, demyelinating inflammatory disease of the central nervous system (CNS), often resulting in severe neurological disability in young adults. The cause of MS is largely unknown, but evidence from genetic and epidemiological studies suggests that the onset of MS is triggered by an interplay between genetic and environmental factors (Compston et al., 2008).

1.1.1 Epidemiology

MS is estimated to affect more than 2.5 million individuals worldwide (Browne et al., 2014). It is twice as common in women as in men (Compston et al., 2008), and is more frequent in Northern Europe, Canada, Northern United States, Southern Australia, and New Zealand than in countries close to the equator. There is a correlation between latitude and the incidence and prevalence of MS, and the spread of MS is affected by migration (Simpson et al., 2011). In Norway, the prevalence of MS is among the highest in the world and a crude prevalence estimate of 203 cases per 100,000 inhabitants has recently been reported (Berg-Hansen et al., 2014). The incidence of MS is low during childhood, but increases significantly in young adulthood. The mean age at onset in Norwegian patients is 32 years, with a peak incidence in the age range 20–40 years (Berg-Hansen et al., 2013).

1.1.2 Genetic epidemiology of MS

The cause of MS is largely unknown but the best-supported hypothesis is that MS is triggered by genetic and environmental factors and their interaction (Figure 1). Epidemiological studies have indicated that the risk of developing MS is higher (2–5%) in first-, second-, and third-degree relatives of MS patients compared with the general population (0.2%) (Carton et al., 1997; Robertson et al., 1996; Sadovnick, 1993). Additionally, a number of studies have
reported a higher risk of MS in monozygotic twins when the one sibling in a twin has the disease (20–30%) than in dizygotic twins (3–5%) (Willer et al., 2003). Further, adoption studies have highlighted the importance of genetics, as there is an increased risk of the development of MS in genetically related family members compared with adopted family members (Ebers et al., 1995).

Figure 1. A schematic overview of the complex aetiology of MS
There is no single cause of MS; however, the susceptibility to MS is a result of genetic, environmental and complex interplay between genetic and environmental factors. The figure was designed using Illustrator (Adobe).

1.1.3 Clinical course
MS is characterized by episodes of neurological dysfunction. The first episode of clinical neurological symptoms of MS is defined as clinically isolated syndrome. Based on the clinical presentation, the disease has two clinical courses: relapsing remitting MS (RRMS: 80–85%) and primary progressive MS (PPMS) (Figure 2). Disability in MS patients can be measured using Kurtzke’s Expanded Disability Status Scale (EDSS) (score 0–10) (Kurtzke, 1983). RRMS patients may experience attacks of neurological dysfunction and full or partial recovery (Noseworthy et al., 2000). Over time, the majority of these patients enter a progressive phase (secondary progressive MS – SPMS) (Noseworthy et al., 2000). Generally, 15–20% of MS patients exhibit a primary progressive disease course with steady decline in
neurological function (Compston et al., 2008). The diagnosis of MS is based on spatially and temporally disseminated symptoms and signs of multifocal inflammatory demyelination in the CNS, using clinical assessment, magnetic resonance imaging (MRI), and usually also analyses of cerebrospinal fluid (CSF) and sometimes visual evoked potentials. The symptoms of MS are variable but, in addition to fatigue and cognitive changes, cerebellar, brainstem, sensory, motor, bowel, and bladder symptoms are common. The diagnostic criteria for MS have been revised a number of times, most recently in 2010 by an international panel (Polman et al., 2011). Although there is no single diagnostic test for MS, MRI scans showing white matter lesions in the brain have proved useful for both diagnosis and follow-up. Advances in MRI in recent years have led to improvements in our understanding of MS disease pathology. The presence of oligoclonal bands (OCBs) in CSF has been observed in 90–95% of European MS patients (Lechner-Scott et al., 2012).

Figure 2. Schematic presentation of the clinical courses of MS
The majority of MS patients experience episodes of neurological dysfunction (relapsing remitting MS). Over time, the majority of patients will enter a stage of secondary progressive MS, when their disability progresses without remission. In primary progressive MS, disability slowly progresses from disease onset without recovery. The figure was designed using Illustrator (Adobe).

1.1.4 Hallmarks of MS pathology
In recent decades, major progress has been made in understanding the inflammatory process and pathogenic mechanisms involved in MS. The pathological hallmarks of MS are multifocal inflammation, axonal damage, neurodegeneration, demyelination, remyelination, and glial scar formation, which occur either focally or diffusely throughout the white matter in the brain (Lassmann et al., 2012). Additionally, a number of researchers have described damage to grey matter in patients with MS (van Munster et al., 2015; Nygaard et al., 2015). These pathological features are present in different disease subtypes and vary over time. It has been
suggested that axonal loss is the main pathological substrate of disease progression (Kornek et al., 2000). Axonal loss occurring independently of inflammation is the main determinant of the irreversible impairment characterizing progressive disease activity (Correale et al., 2016). Inflammation is frequently present in RRMS but infrequent in PPMS and SPMS (Frischer et al., 2009). Inflammatory lesions in MS patients consist of perivascular and parenchymal infiltrates of lymphocytes and activated macrophages (Prineas et al., 1978).

1.1.5 Treatment

Currently, no drugs can effectively cure MS. However, the existing treatments aim to slow down the inflammation and reduce the rate of progression of the disease. Available treatment options for MS disease include anti-inflammatory and immunosuppressive drugs that target parts of the immune system. Several disease-modifying medications for RRMS are approved for clinical use (Wingerchuk et al., 2014). The first-line disease-modifying agents for RRMS include interferon-beta (IFN-β), glatiramer acetate, teriflunomide, and dimethyl fumarate (Dhib-Jalbut, 2002), which reduce the risk of relapses by ~30%. Natalizumab (anti-α4-integrin antibody), fingolimod and alemtuzumab are regarded as second-line treatments for RRMS. Natalizumab is administered intravenously and has been shown to result in a 68% reduction in relapse rate (Hutchinson, 2007). There is an increased risk of progressive multifocal leukoencephalopathy (PML) caused by the John Cunningham virus in MS patients who have received natalizumab (McGuigan et al., 2016). Fingolimod was the first oral MS drug and may reduce the relapse rate by 54%, but it may have side-effects such as heart rhythm block and macular oedema (D'Amico et al., 2016). Alemtuzumab, a lymphocyte-depleting monoclonal antibody, which was introduced in Norway in 2013, is given intravenously and decreases the relapse rate by 69% but may have autoimmune side-effects (Cohen et al., 2012). To date, no biological markers have been developed that can predict the drug response in MS patients.

1.2 The immune system

All multicellular organisms have an immune system that defends against a variety of infections. The immune system comprises of complex networks of interacting cells, tissues, and organs. The first-line defence includes physical barriers, such as skin or an epithelial
The human immune system can be divided into innate and adaptive components with distinct roles and functions. The innate immune system is non-specific, is present at birth and lasts throughout life. This system is ‘trained’ to recognize and neutralize a variety of pathogens following infection. The cells of the innate immune system are of myeloid lineage, developed from multipotent haematopoietic stem cells present in the bone marrow, and consist of macrophages, granulocytes, monocytes, mast cells, dendritic cells (DCs), and natural killer (NK) cells (Murphy et al., 2012). These cells specialize in identifying and eliminating pathogens through pattern recognition receptors specific to microbes (Janeway et al., 2002). DCs are also known as professional antigen presenting cells (APCs), which mainly participate in activating T lymphocytes. DCs form a link between the innate and adaptive immune response (Said et al., 2015).

Unlike the innate immune system, the adaptive immune system is more specific and can provide long-lasting protection (Alberts et al., 2002). The cells of the adaptive immune system are derived from the lymphoid progenitor of the multipotent haematopoietic stem cells present in the bone marrow. The major types of adaptive lymphocytes include B cells and T cells that recognize antigens through highly specific antigen receptors present on their cell surfaces (Murphy et al., 2012). The hallmark of the adaptive immune system is a clonal expansion of lymphocytes that occurs in response to a specific infection. During clonal expansion, both effector and memory cells are acquired, and these cells recognize a specific antigen and contribute to an effective immune response (Polonsky et al., 2016). Whereas B cells play a major role in the humoral immunity of the adaptive immune system by producing antibodies, T cells are involved in cell-mediated responses.

1.2.1 T lymphocytes

T lymphocytes are white blood cells that circulate around the body and constantly scan for infectious agents. T lymphocytes can be distinguished by the presence of T-cell receptors (TCRs). T-cell development takes place in the thymus, which is a bilobed organ located above the heart. Precursors of T cells arise in the bone marrow and migrate to the thymus where precursor T cells proliferate and differentiate into mature T cells. In the thymus, T cells mature and develop their T-cell surface markers and undergo positive and negative selection to ensure tolerance towards self-antigens (Koch et al., 2011). During positive selection that occurs in thymic cortex, only those T cells whose TCRs have the capacity to recognize self
major histocompatibility complex (MHC) molecules expressed by cortical epithelial cells will survive. The process for negative selection occurs in the thymic medulla, where APCs (macrophages and dendritic cells) derived from bone marrow present self peptides to the T cells (Murphy et al., 2012). T cells that react too strongly to self peptides or to self MHC will die due to apoptosis and thereby be eliminated (Klein et al., 2014). Only those cells that are capable of passing both positive and negative selection will leave the thymus and enter the blood, where they will circulate through secondary lymphoid organs. T cells can recognize specific antigens through their TCRs when the antigens are presented by professional APCs via MHC molecules (Figure 3).

**Figure 3. T cell and antigen interaction**

Activation of a T cell by antigen-bound MHC molecule on an antigen presenting cell (APC). Antigen is processed by the APC and presented by its MHC molecule to the TCR on the T-cell surface. The figure was designed using Illustrator (Adobe).

The majority of T cells contain a heterodimeric TCR composed of two transmembrane glycoproteins of one alpha (α) and one beta (β) chain linked by disulphide bonds, whereas in a minority of T cells the TCR consists of gamma (γ) and delta (δ) chains (Murphy et al., 2012). Each α-chain and β-chain contains two extracellular domains: the variable (V) region that binds to the antigen presented by the MHC, and a constant (C) region that is in close proximity to the cell membrane followed by a transmembrane region. A TCR is non-covalently associated with CD3 to form the TCR–CD3 complex (Blumberg et al., 1990). The intracellular part of the CD3 contains a single immunoreceptor tyrosine-based activation motif (ITAM), which is important for intracellular signalling. T lymphocytes are subdivided into two main classes: CD4+ T cells and CD8+ T cells. Via their TCRs, CD4+ T cells bind to
peptides presented by the MHC class II (MHC-II) on APCs, whereas CD8+ T cells recognize peptides displayed by MHC class I (MHC-I) that is expressed on all nucleated cells. In addition to antigen-specific signals during the activation of T cells, a co-stimulatory signal is also required for an effective immune response (Murphy et al., 2012).

1.2.2 Different subsets of T cells

CD4+ T cells or helper T cells are typically classified based on the cytokines they produce. Naive CD4+ T cells have the capacity to differentiate into different T-helper cell (Th cell) subsets such as Th1, Th2, Th9, Th17, and Th22 or induced regulatory T cells (Tregs) (Bouchery et al., 2014; Zhu et al., 2010) as depicted in Figure 4.

Figure 4. Schematic diagram of Th cell lineages

The activation of naive CD4+ T cells is directed through the contact between an APC and a naive Th cell upon the binding of the TCR to the antigen-loaded MHC-II. Depending on a suitable microenvironment and the cytokine milieu, the naive T cell will have the capacity to differentiate into Th subsets. Each Th subset produces different types of cytokines. The figure was designed using Illustrator (Adobe).
Each Th subset produces specific cytokines. For example, Th1 releases interferon-gamma (IFN-γ) and tumour necrosis factor alpha (TNFα); Th2 releases interleukin-4 (IL-4), IL-5, and IL-13; Th9 produces IL-9; and Treg secretes TGF-β and IL-10, and are involved in cell-mediated immunity (Raphael et al., 2015). When Th cells are activated, they can activate B cells to produce antibodies to kill microbes and stimulate cytotoxic CD8⁺ T cells (Luckheeram et al., 2012; Hamaoka et al., 1973). Activated CD8⁺ T cells secrete cytokines and cytotoxic granules that can kill infected cells and malignant cells (Murphy et al., 2012).

### 1.2.3 B lymphocytes

B lymphocytes contribute to the adaptive immune system by producing antibodies. B cells are also known as APCs and express B-cell receptors on their cell surfaces, which bind to specific antigens and secrete antibodies (Murphy et al., 2012). B cells develop in bone marrow and immature B cells then migrate from the bone marrow to the spleen, where they finally develop into mature B cells (Cerutti et al., 2013; Loder et al., 1999). There are several types of B cells, including plasmablasts (short-lived), plasma cells (long-lived), memory B cells (dominant), follicular B cells, and regulatory B cells (immunosuppressive) (Rosser et al., 2015; Murphy et al., 2012).

### 1.2.4 Immunopathogenesis of MS

Despite considerable scientific efforts, the etiology of MS is unknown. However, the disease is believed to be caused by immune dysregulation (Ascherio et al., 2007). MS is thought to be mediated by autoreactive T cells and B cells in the CNS, leading to damage of the myelin sheets and axonal loss. T cells have been identified as central players, but also dysregulation of other immune cells, such as B cells and NK cells, has been found (Harp et al., 2010). NK cells are large granular cytotoxic and cytokine-producing lymphocytes that have important roles both in the innate and adaptive immune system (Poggi et al., 2014). There has been renewed interest in the NK cells due to their potential involvement in the immunoregulation of the CNS inflammation in MS patients (Poggi et al., 2014). CD4⁺ T helper cells play an important role in MS pathogenesis due to their ability to recognize myelin antigens and induce experimental autoimmune encephalomyelitis (EAE) (Vandenbark et al., 1985; Ben-Nun et al., 1981). EAE is a commonly used experimental animal model for MS, and EAE
studies have suggested the involvement of both Th1 and Th17 cells in the development of disease (Domingues et al., 2010). CD8$^+$ T cells are also implicated in MS pathogenesis (Saxena et al., 2011). Additionally, an oligoclonal expansion of CD8$^+$ T cells and higher myelin basic protein reactive CD8$^+$ T cells have been demonstrated in the CSF of MS patients (Zang et al., 2004; Jacobsen et al., 2002). Further, studies have shown increased amounts of Th17 cells, which secrete inflammatory cytokines IL-17A, IL-17F, IL-21, IL-9, IL-22 and TNFα, in peripheral blood and the CSF of RRMS patients during relapses (Mehling et al., 2010; Brucklacher-Waldert et al., 2009; Durelli et al., 2009; Matusevicius et al., 1999). A number of studies have indicated the role of regulatory T cells in MS (Viglietta et al., 2004; Hafler et al., 1997). Recently, impaired Treg function has been found in blood from RRMS patients (Schneider et al., 2013). Another subset of T lymphocytes involved in MS pathogenesis is γδ T cells, which are found in increased levels in CNS lesions and the CSF of MS patients (Stinissen et al., 1995; Shimonkevitz et al., 1993). IL-17, IL-21, and IL-22 producing γδ T cells activated by IL-1β and IL-23 have been found in increased levels in the brains of mice in EAE, suggesting that γδ T cells have an important role in mediating MS pathogenesis (Sutton et al., 2009). Based on B-cell immunomodulatory therapies, there has been an increase in the knowledge of the role of B cells in MS (Hauser et al., 2008), but it is not yet known which B-cell subsets that have a pathogenic role in MS.

1.3 Mapping of genetic risk factors

Sequencing of the human genome and mapping of common genetic variation has led to progress and new knowledge in the field of genetics, and has provided a new tool with which to study complex genetics (The 1000 Genomes Project Consortium et al., 2015; International Human Genome Sequencing Consortium, 2004; The International HapMap Consortium, 2003). Such studies have provided the foundation for identifying the risk factors that contribute to complex diseases. Linkage studies aim to determine the co-segregation of a disease phenotype with a genetic marker. Such studies are most appropriate for the identification of rare disorders with Mendelian inheritance, and they are less relevant for complex diseases in which combinations of several risk variants act together. A genetic association study is typically conducted by comparing allele frequencies of patients and healthy controls or family trios. Trios can be defined as normal parents (mother and a father) and an affected child. With large samples sizes, genetic association studies are powerful for
identifying small genetic effects in complex disorders (Risch, 2000).

1.3.1 Genetic variations and their importance for human health

The human genome consists of ~3 billion base pairs (bps) arranged within 23 chromosome pairs. Approximately 98% of the human genome encompasses non-coding deoxyribonucleic acid (DNA) (i.e. DNA that does not directly encode proteins). Non-coding DNA contains regulatory regions that may affect the expression of the coding regions (Hrdlickova et al., 2014). Approximately 2% of human genome contains coding DNA that can be translated into proteins. There are ~20,000–25,000 protein-coding genes (International Human Genome Sequencing, 2004). Genetic variation is defined as variation in the DNA sequence among individuals of the same species. Genetic studies are important for identifying and understanding how genetic variation contributes to disease mechanisms (Lander et al., 2001). Genetic variation includes single nucleotide polymorphisms (SNPs), variation at only one nucleotide (Figure 5), insertion or deletions of shorter or longer fragments, copy number variation, and larger re-arrangements of DNA segments (Gonzaga-Jauregui et al., 2012). Usually, SNPs are classified into three main classes on the basis of their frequency in the general population: < 1%, 1–5%, and > 5% variants are categorized as rare, low-frequency, and common SNPs, respectively (Bodmer et al., 2008). To date, ~84 million SNPs and ~60,000 structural variations in the human genome have been catalogued (The 1000 Genomes Project Consortium et al., 2015; Sudmant et al., 2015). SNPs are the most common type of genetic variation in the DNA sequence. In the human genome, the frequency of SNPs is approximately 1 in 1,000 bps (Brookes, 1999). Depending on the SNPs location in the coding region of a gene, they may cause an amino acid change, in which case they are known as non-synonymous SNPs. They may not change the identity of the amino acids, in which case they are known as synonymous SNPs (Koberle et al., 2016). SNPs can also be located in regulatory gene regions and may therefore influence the promoter activity, gene splicing, or messenger RNA (mRNA) stability, thus potentially affecting the gene activity that can contribute to disease or increase the risk of disease. In general, a SNP can be used as a biological marker to locate the genes or specific parts of the DNA contributing to disease development.
Figure 5. Single nucleotide polymorphism (SNP)
The position of a single nucleotide in DNA that differ between two individuals. Nucleotides contain the bases shown as adenine (A), cytosine (C), guanine (G), and thymine (T). The figure was designed using Illustrator (Adobe).

1.3.2 Haplotypes and linkage disequilibrium

A haplotype is defined as the set of SNP alleles from a region of one chromosome and is inherited together from one parent (Figure 6). When genetic variations occur on the same haplotype, they will be inherited together more frequently than would be expected by chance (Lohhoff, 2010). This non-random association of alleles at two genetic loci is referred to as linkage disequilibrium (LD). During human evolution, a new mutation may arise on an ancestral haplotype, which is then carried over to following generations. Recombination’s occurring between the SNPs on a haplotype may then give rise to further branching and over generations give rise to a complex population-specific haplotype tree. The strength of LD for two SNPs is expressed by D’ and r², with values ranging from zero to one, where one is perfect LD, and close to one means very strong LD. The measures between D’ and r² have different interpretations. D’ with a value of one implies that two alleles exist on a shared haplotype and were co-inherited in all meiosis (Wall et al., 2003). The r² is the square of the statistical correlation coefficient between two loci, meaning that alleles with different allele frequencies can never reach an r² of one (VanLiere et al., 2008). Therefore, the D’ and r² measures may be very different for a set of SNPs. For example, when the minor allele frequency for one of the SNPs is very low, while for the other SNP the minor allele is frequent, the D’ value may be high and the r² value may be low.
Figure 6. Schematic presentation of haplotypes

Combination of alleles at adjacent loci on a chromosome is shown as haplotype 1-3. The SNPs are indicated in bold that are inherited together from one parent. The figure was designed using Illustrator (Adobe).

1.3.3 Association of HLA in MS

The MHC is a group of genes in a large genomic region present in all vertebrates and encodes immune molecules that can present peptides to other immune cells as described in section 1.2. In humans, the MHC is referred to as the human leukocyte antigen (HLA). The HLA genes reside on the short arm of chromosome 6 and have been shown to be associated with over 100 human diseases, including several autoimmune disorders (Holoshitz, 2013). This genomic region is divided into three classes: class I, class II and class III gene regions (Figure 7). HLA class I and class II contain the HLA loci that are associated with MS. The HLA class III region is located between class I and class II and includes many genes of importance for the complement system. HLA class I includes the HLA-A, HLA-C, and HLA-B loci, and HLA class II includes the HLA-DR, HLA-DQ, and HLA-DP loci.

In 1972, the HLA gene region was identified as associated with MS (Jersild et al., 1972). The HLA-DRB1*15:01 locus has been proven to confer the primary genetic association in MS (Oksenberg et al., 2004), accounting for 14–50% of the genetic MS risk (Hafler et al., 2005; Ebers et al., 1996). An odds ratio (OR) of 3.1 has been reported for the HLA-DRB1*15:01 allele (International Multiple Sclerosis Genetics Consortium et al., 2011). The HLA-DRB1*15:01 allele is 15–20% more frequent in Northern Europe than in Central Europe or USA (www.allelefrequencies.net). However, a reduced frequency of the DRB1*15-DQB1*06 haplotype was found in the Norwegian Sami population compared with the general Norwegian population. This probably contributes to the low prevalence of MS in the Sami population (Harbo et al., 2007). The association between the HLA-DRB1*15:01 allele and
MS has been demonstrated in almost all studied MS patients in Europe (Ballerini et al., 2004). In addition, several studies have shown the protective effect of the HLA-A*02:01, HLA-DRB1*07, and HLA-DRB1*01 alleles in MS (Zhang et al., 2011; Brynedal et al., 2007; Dyment et al., 2005; Harbo et al., 2004). Furthermore, it has been shown that the involvement of HLA-DRB1*15:01 in combination with HLA-DRB1*14 (HLA-DRB1*15/-HLA-DRB1*14 heterozygotes) is associated with a reduced risk of MS, indicating epistatic effects within the HLA region (Barcellos et al., 2006). Recently, HLA-DRB1*03:01 (OR = 1.26) and HLA-DRB1*13:03 (OR = 2.40) have also been reported as risk alleles for MS (Sawcer et al., 2014; Patsopoulos et al., 2013; International Multiple Sclerosis Genetics Consortium et al., 2011). In Sardinia, the HLA-DRB1*03 allele is significantly associated with the risk of MS, whereas a lower effect has been found for the HLA-DRB1*15:01 allele (Marrosu et al., 2001). In addition, HLA-C has been shown to influence the risk of MS independently of other HLA alleles (Yeo et al., 2007). It has also been reported that HLA-DRB1*15:01 allele is associated with the early onset of MS (International Multiple Sclerosis Genetics Consortium et al., 2011). Collectively, these data confirm the association between HLA genes and MS.

Figure 7. The HLA complex
The HLA complex is located on the short arm of chromosome 6 (6p21.3). The HLA genomic region is divided into three classes; class I, class II and class III gene regions, where HLA class I and class II are shown to be associated with MS. The figure was received from Professor B. A. Lie (Department of Medical Genetics, Oslo University Hospital, Ullevål, Norway), and modified and re-created using Illustrator (Adobe).

1.3.4 Complex diseases and genome-wide association studies
Unlike Mendelian diseases, complex diseases in humans are caused by a combination of
multiple risk SNPs interacting with environmental factors (Risch et al., 1996). Examples of common complex diseases includes most inflammatory and immune-mediated, psychiatric, and cardiovascular disorders. Two main models have been suggested for the heritability of complex diseases. The first model is the common disease–common variant (CDCV) hypothesis, which suggests the involvement of several common variants with low penetrance in disease susceptibility. The second model is the common disease–rare variant or disease heterogeneity hypothesis, which suggests the involvement of multiple rare genetic variants with high penetrance in disease susceptibility (Gibson, 2012). Another proposed model is a combination of common and rare genetic variants (Gibson, 2012). The genome-wide association study (GWAS) approach was designed to identify common genetic variants, which fits with the CDCV hypothesis (Andersson et al., 2009).

The International Human Genome Project, the 1000 Genomes Project, and the HapMap Project are the hallmarks of genetic research that have provided a foundation for GWASs (The 1000 Genomes Project Consortium et al., 2015; International Human Genome Sequencing Consortium, 2004; The International HapMap Consortium, 2003). GWASs have focused on genetic associations between SNPs and complex diseases. Series of GWASs have reported the associations of thousands of SNPs with different complex diseases. A large sample size is crucial in order to achieve sufficiently robust power to identify common variants that typically have small effect sizes.

1.3.5 Discovery of non-HLA associations with MS through GWAS

The use of GWASs has led to the identification of more than 110 MS susceptibility genes outside the HLA locus. Genetic variants located in the interleukin-2 receptor alpha (IL2Ra) and interleukin-7 receptor alpha (IL7Ra) genes were the first risk factors outside the HLA region that were identified in MS through a moderately powered GWAS (International Multiple Sclerosis Genetics Consortium et al., 2007). Since then, a series of GWASs and follow-up studies have been conducted. The International Multiple Sclerosis Genetics Consortium (IMSGC) and the Wellcome Trust Case Control Consortium collaborated in a study involving samples from 9,772 MS patients and 17,376 healthy controls. This GWAS identified 52 non-HLA MS-associated loci (International Multiple Sclerosis Genetics Consortium et al., 2013; International Multiple Sclerosis Genetics Consortium et al., 2011). The majority of these non-HLA MS-associated loci were found in or near genes, which have
been described as having a role in the immune system, confirming that MS is an immune-mediated disease (International Multiple Sclerosis Genetics Consortium et al., 2011). In a later study, the ImmunoChip custom genotyping array was used in the analysis of samples from 14,498 MS cases and 24,091 healthy controls, and the results led to a doubling of the non-HLA associations in MS to 110 (OR = 1.1–1.3). Interestingly, one-third of the MS susceptibility SNPs significantly overlap with genetic hits in other autoimmune diseases (Gourraud et al., 2012; International Multiple Sclerosis Genetics Consortium et al., 2011). In an ongoing follow-up, the IMMSGC anticipates that 200 non-HLA MS-associated loci will be identified (Principal Investigator at IMMSGC, personal communication 2016). Only 5% of identified MS susceptibility SNPs from the GWAS analysis are estimated to be causal SNPs (Farh et al., 2015), thus indicating the need for fine mapping of associated regions. In order to translate the GWAS results into biological insights how MS risk is mediated, functional characterization of the most likely causal variants in these regions is needed.

GWAS analyses have identified genetic regions associated with disease rather than the true causal genetic variants. The genetic variants identified through large-scale GWASs may not be the causal variant; however, the causal variant(s) could be in high LD with the identified SNP. Imputation of additional variants or sequencing the LD region associated with disease may identify plausible candidates for the causal variant.

1.3.6 Functional follow-up of the GWAS signal

GWAS studies need to be followed up by functional studies. For example, functional follow-up of the GWAS signal in the TNFRSF1A gene encoding tumour necrosis factor (TNF) receptor superfamily 1A has indicated an alternative splicing induced by the MS risk variant at this locus (Gregory et al., 2012). Likewise, the expression of one of the first non-HLA genes associated with MS, C-type lectin-like domain family 16A (CLEC16A) (International Multiple Sclerosis Genetics Consortium et al., 2007), has been shown to be expressed at higher level in peripheral blood mononuclear cells (PBMCs) and white matter brain tissue from MS patients compared with controls (van Luijn et al., 2015). The gene expression level of CLEC16A in specific cell types is correlated with the genotype of MS risk SNPs in CLEC16A (Table 3, in Section 5.3) (Soleimanpour et al., 2014; Leikfoss et al., 2013; Mero et al., 2011). Recently, Couturier and colleagues have shown that the MS-associated TYK2 variant rs34536443 influences T-lymphocyte polarization (Couturier et al., 2011).
In the pathogenesis of MS, the involvement of multiple signalling pathways has been uncovered in pathway and network analysis (Cotsapas et al., 2011) and probably implicates the role of several distinct cell types (International Multiple Sclerosis Genetics Consortium et al., 2011). The majority of MS-associated SNPs have been shown to affect the regulatory region of genes (Farh et al., 2015; Maurano et al., 2012), thus indicating tissue-specific changes to gene expression (Raj et al., 2014). Understanding the functional mechanisms underlying the associations between genetic variants and MS may provide entry points for targeting disease-specific biological mechanisms and the development of therapeutics.

1.4 Environmental factors in MS

1.4.1 Sun exposure and vitamin D

Vitamin D plays a vital role in maintaining the level of serum calcium and promoting bone mineral density as well as in modulating the immune system (Hart et al., 2011). Humans obtain vitamin D from two sources: diet or dietary supplements such as fatty fish, cod liver oil, fortified food, and UV light. Low serum levels of vitamin D have been associated with the risk of several neurological and autoimmune diseases, including MS (Smolders et al., 2011). Growing evidence indicates that vitamin D has a role in the regulation of the immune system, including the proliferation of T cells (Bhalla et al., 1984) and shifting the T-cell repertoire from a pro-inflammatory towards anti-inflammatory state (Smolders et al., 2008; Boonstra et al., 2001). Through GWAS and other genetic studies, SNPs in \textit{CYP27B1} and \textit{CYP24A1} have been found associated with the risk of MS (International Multiple Sclerosis Genetics Consortium et al., 2013; International Multiple Sclerosis Genetics Consortium et al., 2011; Ramagopalan et al., 2011). Expression of the MS-associated HLA-DRB1*15:01 as well as \textit{IL2RA} and \textit{TAGAP}, have been shown to be regulated by vitamin D \textit{in vitro} (Berge et al., 2016; Ramagopalan et al., 2009).

1.4.2 Epstein–Barr virus

Epstein–Barr virus (EBV) infection has been regarded as a risk factor for MS development but the mechanism behind this association is not yet clear (Tselis, 2011). EBV may cause infectious mononucleosis (IM). The relative risk of MS in individuals with a history of IM has been reported as 2.3 compared with individuals with no IM infection (Thacker et al., 2006).
The risk of developing MS among EBV-negative individuals is 15-fold lower than in EBV-positive individuals (Ascherio, 2013). Molecular mimicry is suggested as a mechanism behind this risk factor (Lang et al., 2002; Wucherpfennig et al., 1995), which means that peptides from pathogens share structural similarities with self-antigens. These structural similarities may misguide the adaptive immune system into becoming reactive against self-antigens.

### 1.4.3 Smoking

Smoking is yet another established risk factor for MS. A number of studies have shown an association between smoking and an increased risk of MS (Salzer et al., 2013; Sundstrom et al., 2008; Simpson et al., 1966). The Nurses’ Health Study revealed that the risk of women developing MS was 70% higher for heavy smokers (for more than 25 years) than for women who had never smoked (Hernan et al., 2001). However, an association study conducted in Scandinavia has shown that the use of tobacco snuff seems protective against MS and therefore nicotine might not be a triggering factor for MS (Hedstrom et al., 2009). The biological mechanism behind these observations are not yet clear, but the irritation of the lungs due to smoking has been suggested (Hedstrom et al., 2011).

### 1.4.4 Obesity

Obesity during childhood and adolescence has been reported to increase MS susceptibility. The risk has also been reported for obese women, based on data from the Nurses’ Health Study (Munger et al., 2009) and this finding was further replicated in studies conducted in Sweden and Denmark (Munger et al., 2013; Hedstrom et al., 2012). Low levels of serum vitamin D or higher proinflammatory activity in fatty tissue might be one of the possible mechanisms contributing to the risk of obese individuals developing MS (Munger et al., 2009).

### 1.4.5 Other environmental factors

It is still a challenge to identify the role of other environmental factors in the aetiology of MS (McDonald et al., 2016). A recent study has indicated that high sodium chloride intake is associated with risk of MS (Farez et al., 2015). A study published in 2016 found that salt triggers a pro-inflammatory response and worsens EAE through modulation of pathogenic
Th17 cells (McDonald et al., 2016). Recently, a growing body of literature has suggested the role of the gut microbiome in MS pathology (Chen et al., 2016; Jangi et al., 2016), and this might support the evidence for the role of the environment in MS.

1.5 Gene-environment interaction

It has been broadly assumed that there is a complex interaction between genes and environmental factors in MS susceptibility (Hedstrom et al., 2015; Gourraud et al., 2012). An interaction between smoking and two MS-associated HLA alleles has been reported. The carriage of the risk-allele HLA-DRB1*15:01 and absence of the HLA A*02 allele showed an OR of 13.5 in people with MS who were smokers compared with non-smokers without these genetic risk factors (Hedstrom et al., 2011). Recently, interaction between passive smoking and HLA genes (Hedstrom, Bomfim, et al., 2014), EBV infection and HLA genes (Sundstrom et al., 2009), and adolescent obesity and HLA genes (Hedstrom, Lima Bomfim, et al., 2014) have been reported to confer increased risk of MS development. An interaction of non-HLA gene Nacetyltransferase-I with smoking has also been suggested (Briggs, Acuna, et al., 2014).

1.6 The 16p13.13 region and MS susceptibility

The 16p13.13 chromosomal region of the genome has been convincingly shown to be associated with several autoimmune diseases, including MS. The first MS GWAS, which was published in 2007 identified an MS-associated SNP in the CLEC16A gene (International Multiple Sclerosis Genetics Consortium et al., 2007). The finding has since been confirmed and this genetic region has been fine mapped in several follow-up studies (Mero et al., 2011; International Multiple Sclerosis Genetics Consortium, 2009; Zoledziewska et al., 2009; Hoppenbrouwers et al., 2009), including a large MS GWAS published in 2011 (International Multiple Sclerosis Genetics Consortium et al., 2011). In addition to MS-associated SNPs, the CLEC16A region harbours several other SNPs with an association with autoimmune diseases such as type-1 diabetes (T1D), Crohn’s disease, rheumatoid arthritis (RA), and systemic lupus erythematosus (SLE), as well as the common variable immunodeficiency (CVID) disorder (Li et al., 2015; Berge et al., 2013), thus indicating that this gene region has a role in immune regulation. These autoimmune disease associated SNPs reside within introns of CLEC16A and typically display strong LD, and therefore it is a challenge to identify the primary genetic
CLEC16A is important for endosomal trafficking and autophagy, and regulates antigen presentation and HLA-II expression in APCs (van Luijn et al., 2015; Soleimanpour et al., 2014; Kim et al., 2012; Kim et al., 2010), thereby linking this gene to the regulation of strongest genetic factor in MS. In addition to **CLEC16A**, the 16p13.13 gene region contains other genes, such as **CIITA** (major histocompatibility complex class II transactivator), **DEXI** (dexamethasone-induced transcript), and **SOCS1** (suppressor of cytokine signalling 19) (Figure 8b). **CIITA** and **SOCS1** are immunoregulatory genes that respectively encode an essential transcription factor (TF) important for the activation of MHC-II gene expression (Kern et al., 1995) and a suppressor of cytokine signalling (Fenner et al., 2006). Importantly, independent SNPs in these genes have recently been highlighted as potential susceptibility candidates for MS and other autoimmune diseases (Gyllenberg et al., 2014; Fenner et al., 2006; Swanberg et al., 2005). The dexamethasone-induced gene (**DEXI**), which encodes a protein with unknown function, is located between **CIITA** and **CLEC16A** (Davison et al., 2012; Edgar et al., 2001), and intronic SNPs in **CLEC16A** have been found to be expression quantitative trait loci (eQTLs) for **DEXI** in thymic tissues, monocytes, and lymphoblastoid cell lines (Tomlinson et al., 2014; Leikfoss et al., 2013; Davison et al., 2012). In addition, by chromatin capture assay, Davison and colleagues showed that intron sequences within **CLEC16A**, which contains the MS-associated SNPs rs12708716, is in physical proximity to the **DEXI** promoter, and this in turn provides a plausible mechanism for how intronic **CLEC16A** sequences can regulate **DEXI** expression from a distance (Davison et al., 2012). This might explain how elements in the intronic region of **CLEC16A** could regulate **DEXI** expression (Davison et al., 2012). The ImmunoChip study has revealed two MS-associated SNPs located in the intronic region of the **CLEC16A** gene: SNP rs12927355 (primary hit) in strong LD with the earlier identified rs12708716 ($r^2 = 0.82, D' = 1.00$), and SNP rs4780346 (secondary hit) located in the **CLEC16A-SOCS1** intergenic region (International Multiple Sclerosis Genetics Consortium et al., 2013).
Figure 8. CLEC16A and the 16p13.13 genetic region
(a) the CLEC16A gene containing several SNPs associated with autoimmunity; (b) the chromosome 16p13.13 genetic region encompassing CIITA, DEXI, CLEC16A, and SOCS1 (abbreviations: MS = multiple sclerosis, T1D = type 1 diabetes, PBC = primary biliary cirrhosis, PAI = primary adrenal insufficiency, AA = alopecia areata, CVID = common variable immunodeficiency, CD = Crohn’s disease, JIA = juvenile idiopathic arthritis, RA = rheumatoid arthritis) (modified from (Berge et al., 2013)).

1.7 Allele-specific expression and its biological mechanism

Humans have two copies of each gene, which are in general expressed simultaneously and are normally expected to be expressed at equal levels. However, when the ratio of gene transcripts is unequal between two chromosomes, this is referred to as allelic imbalance (AI) (Figure 9). Allele-specific expression (ASE) can be the result of sequence variation between the chromosomes in an element regulating the gene expression. The biological mechanism through which cis-acting genetic variations regulate ASE includes transcription factor binding sites, differential epigenetic configuration, and mRNA processing. AI measurements are relatively robust for detecting and quantifying variations in allelic expression in heterozygous carriers. The advantage of this type of analysis is that relative expression levels of two alleles are measured within the same biological sample, thus avoiding the possibility of variations introduced by the environment or differences in the physiological background of the sample.
Figure 9. Allele-specific expression

This figure shows the expression of two alleles as either equal (allelic balance) or unequal (allelic imbalance, AI). AI is a situation where a copy of a gene is expressed comparatively less than another, which can be measured by allele-specific expression (ASE) analysis. ASE can be used to detect AI in the transcription of heterozygous individuals. The colours green and orange indicate two copies of a gene. The figure was designed using Illustrator (Adobe).


**STUDY OBJECTIVE**

2 STUDY OBJECTIVE

The aim of the study on which this thesis is based was to characterize the molecular functions of selected MS risk loci in immune cells in order to understand their role in MS disease risk. Our specific objectives were:

- To analyse allele-specific expression of selected MS-susceptibility SNP-containing genes in human peripheral blood

- To analyse whether MS-associated SNPs in intronic regions of CLEC16A act as eQTLs for CLEC16A itself and its nearby genes (i.e. CIITA, DEXI, and SOCS1) in T cells from MS patients and controls

- To study the expression and function of DEXI in T cells.
3 SUMMARY OF THE PAPERS IN THE THESIS

3.1 Paper I

In the first paper, we studied the functional role of selected genetic MS risk variants identified through GWAS analysis. To date, GWASs have revealed more than 110 SNPs associated with susceptibility to MS. However, the functional contribution of MS-associated SNPs to disease development is largely unknown. Measures of relative expression levels of alleles in heterozygous samples for the SNP of interest (i.e. allele-specific expression (ASE)), is a powerful approach for the identification of cis-acting regulatory variants. We selected three genes – CD69 (Cluster of differentiation 69), IKZF3 (IKAROS family zinc finger 3), and IQGAP1 (IQ Motif containing GTPase activating protein 1) – with an MS-associated SNP in their coding region or in strong LD with a coding SNP, and then performed ASE analyses of whole blood samples from individuals who were heterozygous for the studied SNPs. Among the 92 MS patients that were heterozygous for at least one of the three SNPs studied, 58 patients were heterozygous for rs11052877 (CD69), 30 patients were heterozygous for rs907091 (IKZF3) and 61 were heterozygous for rs11609 (IQGAP1). In these samples, we observed consistent AI for rs907091 in IKZF3 and rs11609 in IQGAP1, which are in strong LD with the MS-associated SNPs rs12946510 and rs8042861, respectively. Similarly, AI was observed also in samples from healthy controls (n = 8), indicating that the observed AI is independent of disease status. The MS risk alleles at IKZF3 and IQGAP1 were expressed at higher levels than the protective alleles. Furthermore, individuals who were homozygous for the MS risk allele at IQGAP1 had a significantly higher total expression of IQGAP1 than individuals who were homozygous for the protective allele. Our data indicate a possible regulatory role for MS-associated IKZF3 and IQGAP1 gene variants. The study highlighted the usefulness of ASE measurements for identifying disease-associated SNPs or SNPs in LD with cis-acting regulatory properties. This study may provide a functional mechanism behind the MS-association of SNPs near IQGAP1 and IKZF3.
3.2 Paper II

CLEC16A has been shown to be associated with several autoimmune diseases, including MS and T1D. Recently, our research group has reported that expression of two of its neighbouring genes, DEXI and SOCS1, is reduced in the thymus of individuals carrying MS risk alleles in the CLEC16A intron 19 (Leikfoss et al., 2013). However, this correlation has not been seen in whole blood (Leikfoss et al., 2013), which indicates a cell-specific mechanism for the association between CLEC16A genotype and gene expression. In Paper II we therefore performed analyses of peripheral T cells (i.e. CD4+ and CD8+ T cells) that were isolated from treatment-naïve, female RRMS patients (n = 33) and healthy controls (n = 29). Gene expression levels were analysed using real-time quantitative polymerase chain reaction (qPCR) and we assayed DEXI and SOCS1 as well as CLEC16A and CIITA, all of which are present in the 16p13.13 region. We observed no significant differences in gene expression between MS cases and healthy controls for any of the three genes. When samples were sorted according to the two CLEC16A MS risk SNPs rs12927355 (primary ImmunoChip signal) and rs4780346 (secondary ImmunoChip signal) ($r^2 = 0.18$, D’ = 1.00), we observed a significantly higher expression of SOCS1 and CLEC16A in CD4+ T cells from MS patients homozygous for the risk allele as compared to the protective allele for rs12927355. This effect was not observed in samples from CD8+ T cells. However, when analysed for rs4780346, no association was found between genotype and gene expression for any of the studied genes, neither in CD4+ T nor in CD8+ T cells. Interestingly, a high correlation of gene expression of CIITA, DEXI, CLEC16A, and SOCS1 was observed in CD4+ T cells when a pairwise linear regression analysis was applied. Our results indicate a possible regulatory role for the MS-associated CLEC16A SNP rs12927355 in peripheral CD4+ T cells.

3.3 Paper III

The DEXI gene is located between CIITA and CLEC16A in the 16p13.13 chromosomal region and has been identified as a novel autoimmune susceptibility gene. Autoimmune disease associated intronic CLEC16A SNPs have been demonstrated to act as eQTLs for DEXI in thymic tissue samples, monocytes, and lymphoblastoid cell lines (Tomlinson et al., 2014; Leikfoss et al., 2013; Davison et al., 2012). DEXI encodes a protein with unknown function. However, DEXI has been reported to be upregulated in the human lung adenocarcinoma A549 cell line in response to dexamethasone (Edgar et al., 2001), which is a glucocorticoid drug.
commonly used as an anti-inflammatory and immunosuppressive drug.

In Paper III we aimed both to analyse the expression of the DEXI gene using real-time qPCR in different human immune cells and upon T cell stimulation, and to explore whether it affects the T cell activation. Our data show that DEXI is expressed in several human immune cells purified from healthy individuals, specifically, CD19+ B cells, CD4+ and CD8+ T cells, CD56+ NK cells, and CD14highCD16− monocytes, as well as in human immune cell lines, specifically Jurkat (T leukaemia cell line), Raji (Burkitt’s lymphoma B cell line), and THP-1 (acute monocytic leukaemia cell line), with the highest expression in monocytes and the lowest in B cells, both in the primary cells and in the cell lines. Using Jurkat T cells as a model system, we found that DEXI expression was reduced upon T cell activation with phorbol-12-myristate-13-acetate (PMA) and ionomycin. However, by using small interfering RNA (siRNA) to knock down DEXI in Jurkat cells, we found that DEXI had no impact on T cell activation when measured by flow cytometry of CD69 cell-surface expression marker. Further functional studies of DEXI in T cells are warranted to determine whether the association between DEXI and autoimmune diseases is mediated through an effect in T cells or in other cells.
METHODOLOGICAL CONSIDERATIONS

4 METHODOLOGICAL CONSIDERATIONS

4.1 Study population

For Paper I, the data from the Norwegian MS patients and samples from whole blood used for the isolation of RNA (ribonucleic acid) and DNA (n = 140) were accessed from the Norwegian MS Registry and Biobank located at Haukeland University Hospital in Bergen, Norway. Whole blood samples from healthy controls (n = 46) were recruited among hospital employees in Oslo, Norway. An overview of the samples used in the study are summarized in Table 1 in Paper I. The patient group included in the study was representative of the Norwegian MS patient population in terms of different disease stages and different treatments. This heterogeneity among the sampled patients may have affected the gene expression levels in blood, leading to variations in the results. Treatment data for the patients at the time of blood drawing was not available; and therefore this covariate may have influenced the gene expression. However, the advantage of using these blood samples was that DNA as well as blood for RNA isolation were available.

For Paper II, peripheral CD4\(^+\) and CD8\(^+\) T cells from 33 female Norwegian RRMS patients and 29 age-matched and sex-matched healthy controls were recruited from the MS clinic at the Department of Neurology, Oslo University Hospital, Ullevål. At the time of sample collection, all controls were healthy with no known autoimmune diseases, while all patients included in the study had recently been diagnosed with MS according to the latest revision of the McDonald criteria (Polman et al., 2011). The included patients had never received any immunomodulatory drugs or recent steroid treatment and had not had any relapsing attacks in the three months prior to enrolment in the study. The patients had no other autoimmune diseases. The patient group was homogenous with respect to gender and disease course and were in the age range 21–63 years (mean 39.5 years). Healthy controls were carefully age-matched with the patient group (mean age 39.6 years).
4.2 Collection of human peripheral T cells

PBMCs from whole blood were separated using density gradient centrifugation, from which peripheral T lymphocytes, CD4+ and CD8+ T cells were subfractionated using the autoMACS® Pro Separator and magnetic beads. The CD8+ T cells were purified by positive selection using anti-CD8+ antibodies coupled to magnetic beads. The binding of anti-CD8 magnetic particles will not activate CD8+ T cells. In the case of CD4+ T cells separation, we routinely used a negative selection procedure, whereby monoclonal antibodies (conjugated with the biotin) were used against CD8, CD14, CD16, CD19, CD36, CD56, CD123, γδ TCR, and Glycophorin A. Further, purity of cells was assessed routinely by flow cytometry, resulting > 95% for both CD4+ and CD8+ T cells. During blood sampling, the patients and controls were healthy also from infections.

4.3 DNA and RNA extraction and quality analysis

There are several methods available to isolate DNA and RNA from whole blood and cells. For our study, we paid close attention to sample collection, preparation, and storage in order to extract good quality DNA and RNA before the analytical pipeline. Compromised quality of input may introduce adverse effect when genetic and gene expression experiments are performed. Isolation of DNA from whole blood was performed using DNA isolation columns (QIAmp DNA Blood Mini Kit). The concentration of DNA samples used for sensitivity analyses of the ASE assays were measured by Qubit fluorometry (Invitrogen, Carlsbad, CA, USA). Qubit fluorometry is a highly sensitive technique that enables accurate measurements of DNA, RNA, and protein concentrations. In order to verify the sensitivity of the ASE analysis, we used homozygous samples for either allele mixed in known ratios. In order to prepare the right mixtures, an accurate measure of input DNA for either allele was a prerequisite, since small differences in allelic input should be detectable.

Blood for purification of RNA can be collected in various pre-made collection systems including Tempus and PAXgene tubes. Nikula and colleagues compared the performance of the Tempus (Applied Biosystems) and PAXgene (PreAnalytiX, Qiagen BD, Valencia, CA) blood RNA systems for whole blood collection, and found that the quality of total RNA was high in both systems (Nikula et al., 2013). However, they observed 443 transcripts that were differentially expressed between RNA samples preserved in Tempus and PAXgene tubes.
Since it was important to use only one procedure throughout the study, we used Tempus tubes for the entire collection, and RNA from whole blood was isolated by the Tempus Spin RNA Isolation Kit protocol (Applied Biosystems) followed by DNase treatment during extraction. DNA and RNA concentrations were analysed by a nanodrop 2000c spectrophotometer (Thermo Fisher Scientific Inc., Madison, WI, USA). Nanodrop is a UV-based spectrophotometer and is used to assess the quantity and purity of DNA and RNA. For the RNA samples, an RNA integrity number (RIN) above 7.0 in a random set of RNA samples was confirmed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), since high-quality RNA is an important requirement in gene expression analyses. RIN is a commonly used index of RNA quality. The Bioanalyzer technique provides quality control of DNA and RNA on a single platform. Since the RNA quality of selected samples was good, we assumed that all samples included in our study were good quality and could be used for gene expression analysis.

4.4 SNP genotyping

There are several methods available for SNP genotyping. For Paper I, we used TaqMan assays, which have been designed to be specific, sensitive, and cost and time effective. The technique is based on two allele-specific probes and requires DNA as an input for genotyping. When deciding upon the method for SNP genotyping in a large-scale study, Affymetrix SNP GeneChip and Illumina GoldenGate BeadChips assays can be an ideal choice (Shen et al., 2009). For Paper I, we only analysed a few SNPs and therefore the TaqMan platform for genotyping was a more attractive method. Genotyping was performed in-house with TaqMan technology, using Life Technologies “ViiA7 instrument” (Life Technologies, Foster City, CA, USA). For Paper II, the genotype data were already available using a chip-based method (Human Omni Express BeadChip (Illumina, San Diego, CA, USA)) as described by Bos et al. (Bos et al., 2015).

We used commercially available primers and probes from Applied Biosystems. Each TaqMan assay is designed with two probes with different fluorophores that detect the specific alleles of the desired polymorphism, and two sequence-specific primers that amplify the specific DNA sequence (Figure 10a). TaqMan utilizes the fluorescence resonance energy transfer (FRET) mechanism that occurs when the reporter and quencher are in close proximity. The probe in the TaqMan assay is fluorescently labelled by a reporter dye at the 5’ end of the probe and
non-fluorescent quencher dye on the 3’ end of the probe. During PCR reaction, each probe is
designed to anneal specifically to its complimentary target sequence, and DNA polymerase
extends the primer in the 5’ – 3’ direction followed by cleaving the dye from the quencher
from the sequence-specific probe (Figure 10a).

**Figure 10. Schematic illustration of the TaqMan probe principle for genotyping**
(a) region-specific primers and sequence-specific probes that contains a reporter dye (different colours
for the different target alleles) and a quencher; the 5’ nuclease activity of Taq DNA polymerase
cleaves probes and separates the reporter dye from the quencher; upon the completion of the PCR
cycles, the fluorescence for either of the reporter dyes will be identified; the amount of dye will be
proportional to the amount of amplified product, which in turn depends on the input sample; (b) for
samples that are homozygous for either allele, a sequence-specific probe containing a specific dye will
be active, whereas in heterozygous samples roughly equal amounts of dye are expected; the
genotyping result is illustrated in an allelic discrimination plot (Life Technologies’ ViiA7 instrument).
Figure 10a was designed using Illustrator (Adobe), figure 10b is a screen capture from the ViiA7
software (Life Technologies’).

When the reporter and quencher are separated, FRET is no longer effective, and this leads to a
fluorescent signal proportional to the amount of amplified target in the sample (Whitcombe et
In Figure 10, genotyping results are depicted in an allelic discrimination plot. In our study, all genotyping experiments were run in duplicate and included positive controls with known genotypes that had already been genotyped in-house as well as a negative control with water. In addition, genotyping of the samples included for Paper II had been done using the Human Omni Express BeadChip (Bos et al., 2015).

4.5 Gene expression analysis by quantitative real-time PCR

The polymerase chain reaction (PCR) technique was invented by Mullis in 1983 and aims to amplify specific regions of DNA sequences. In order to check whether the target sequences were amplified successfully, it was necessary to run an agarose gel. Further improvement of the PCR technique was accomplished by Higuchi and colleagues in 1992, who introduced the quantitative real-time PCR (qPCR) technique, whereby a specific DNA sequence can be amplified, monitored, and measured at each step of the PCR cycle. The advantage of qPCR is that the PCR reaction can be monitored as it progresses, enabling the quantity of amplified DNA to be measured accurately at each cycle. This technique has been increasingly used for quantitative gene expression measures. For our study, we used qPCR to monitor the expression of IKZF3 and IQGAP1 (Paper I), CLEC16A, CIITA, DEXI, and SOCS1 (Paper II), and DEXI (Paper III). The principle of real-time qPCR is illustrated in Figure 11a. Unlike genotyping, real-time qPCR is based on one probe and requires complimentary DNA (cDNA) as input. The threshold cycle (C_T) value is defined as the number of PCR cycles needed for the fluorescent signal to exceed the threshold (Figure 11b) (Walker, 2002). The C_T value is inversely related to the amount of cDNA in a reaction. Higher cDNA input at the beginning of the qPCR reaction will result in a lower C_T value indicating higher gene expression, whereas a higher C_T value will indicate lower gene expression.

There are two ways to analyse gene expression data by real-time qPCR: absolute and relative quantification (Livak et al., 2001). We used the standard curve method for relative quantification, which identifies the gene expression changes of the target gene relative to a reference gene, generating the cDNA from RNA with known concentrations. We serially diluted the cDNA with known quantity and generated a standard curve by plotting the log quantity against the C_T values to assess the quantity of target samples and reaction efficiency. The quantity of an unknown sample can be measured by interpolating the unknown values from the standard curve. Further, the normalization of target gene expression was performed
relative to a validated reference gene that is line with the MIQE (minimum information for publication of quantitative real-time PCR experiments) guidelines (Bustin et al., 2009). A reference or housekeeping gene can be described as a stably expressed gene that is equally expressed in different cell types and unaffected by any treatments. Another extensively used quantification approach is the comparative $C_T$ or delta delta $C_T$ method, which involves comparison of $C_T$ values between target and control genes, using a reference gene as a normalizer (Livak et al., 2001). This method assumes equal PCR efficiencies across all assays; if the efficiency varies, it will lead to unreliable results.

Figure 11. Schematic illustration of the principle behind real-time quantitative PCR
(a) sequence-specific primers and a probe that contains a reporter dye and a quencher; the 5’ nuclease activity of Taq DNA polymerase cleaves probes during cDNA extension and separates the reporter dye from the quencher; upon completion of the PCR cycles, the fluorescence for the reporter dye will be identified; the amount of dye will be proportional to the amount of amplified product, which in turn depends on the input sample; (b) amplification plot showing threshold level and threshold cycle ($C_T$); the $C_T$ value is defined as number of PCR cycles needed for the fluorescent signal to exceed the threshold and is inversely related to the amount of cDNA. The figure was designed using Illustrator (Adobe).

The qPCR method is a sensitive method for quantifying genes and requires proper handling of samples. For accurate quantification, selection of the reference gene for the normalization of gene expression in target samples is an important step. We tested several reference genes and presented the results with 18s rRNA encoding 18s ribosomal RNA or $TBP$ encoding the TATA-binding protein (TBP) that displayed lowest variance in our samples. Bas et al.
concluded that 18s rRNA was the most stable reference gene when analysing the expression of different genes in T-lymphocytes at different activation stages (Bas et al., 2004), and the results of our analysis support their finding. In order to check for contamination such as residual DNA, we included –RT (all components in the reaction except the reverse transcriptase enzyme) and NTC (all components in the reaction except template, i.e. cDNA) in our analysis.

4.6 Assessment of allele-specific expression

Several functional characterization approaches can be taken to gain a better understanding of the molecular mechanisms behind associations of SNPs in MS. One approach is studying the ASE of selected MS susceptibility genes. In our case, we aimed to measure the differential expression levels of the gene depending on different alleles in its coding region. Studying AI of MS-associated genes may shed new light on how pathogenic processes are initiated in MS. For this purpose, we selected heterozygous carriers of MS-associated SNPs in coding region of the gene. The presence of a coding allele creates a unique opportunity to investigate the relative abundance of both alleles in samples that are heterozygous for the SNP. Genomic heterozygous DNA, in which the alleles are present in 1:1 ratio, is used to calibrate the measurement, as there may be intrinsic differences in fluorophore intensity. A variety of methods have been developed to measure ASE in subsets of human genes, including primer extension and capillary electrophoresis (Matyas et al., 2002), sequencing (Fukuda et al., 2006), pyrosequencing (Wang et al., 2007), microarray technology (Lo et al., 2003), and other methods (Jordheim et al., 2008). For ASE assays, we used the TaqMan technology, due to its sensitivity and specificity, and because it can directly detect differences in allelic output when assessed in heterozygous carriers, as shown in Figure 10a. If there is AI, the expression of the two alleles will differ and this can be detected by comparing the fluorescence from different allele-specific probes.

4.7 Transfection of Jurkat T cells

Transfection is a technique for introducing foreign DNA and RNA into mammalian cells via a temporary opening in the cell membrane, which can be achieved through physical or chemical methods. To study the function of DEXI in T cells, we introduced commercially available
DEXI-specific siRNA or an expression plasmid using BTX electroporation (Genetronix, San Jose, CA) to knock down and overexpress DEXI. The advantage of this technique is that it is cheaper than other available transfection methods, results in better transfection efficiency of T cell lines, and the method was well established in our laboratory. The technique is frequently employed in biological research, whereby an electric field is applied to the cells to increase the permeability of the cell membrane, thus allowing the introduction of foreign genetic material. However, the method can lead to cell death and imbalance in cell homeostasis when an electric field is applied to the cells (Rubinsky, 2007). Therefore, transfection using BTX electroporation is not well suited for primary cells because they are difficult to transfect (Chicaybam et al., 2013).

Transfection with electroporation may affect the phenotype of cells. Accordingly, in our experimental setup, we routinely included electroporated cells with control siRNA or an empty plasmid. We used SMARTpool siRNA for DEXI (Paper III). SMARTpool siRNA is a mixture of four siRNAs that are specific to the same target gene. One disadvantage of gene knockdown using siRNA is an off-target effect (Boutros et al., 2008). In order to control for this effect, it is advised that competent and different types of siRNA should be used. We harvested cells at different time points after transfection (e.g. 24 hours (h), 48h, and 72h) to assess the knockdown and overexpression of DEXI. Our results demonstrate a 60–80% knockdown of DEXI at mRNA level 24h and 48h after transfection. The knockdown effect lasted up to 72 hours after transfection. In our experimental setting, we included a green fluorescent protein expressing plasmid to measure the transfection efficiency by flow cytometry. The transfection efficiency was routinely 30–60%.

### 4.8 Activation of T cells for in vitro experiments

We investigated the functional impact of DEXI in human T cells using Jurkat TAg cell line as a model system. Jurkat TAg is an immortalized cell line derived from human leukemic T cells (Gillis et al., 1980). The cell line is extensively used to study T-cell receptor and cytokine signalling pathways (Abraham et al., 2004; Pawelec et al., 1982). Jurkat TAg is easy to transfect and has routinely been used in our laboratory for functional studies. In contrast to primary cells, which are generally harder to obtain in sufficient quantities, Jurkat cells can be easily handled and expanded in cell culture. The results obtained from Jurkat cells should be interpreted with care because these cells are immortalized and have the capacity to expand
indefinitely, and therefore they may not act like normal cells. Importantly, any results observed from these cells should be confirmed using primary cells.

Activation of T cells occurs through the engagement of peptide complexed with MHC presented to TCRs and a co-stimulatory molecule, as shown in Figure 3 (Boesteanu et al., 2009). Activation of T cells may result in intracellular downstream signalling, leading to regulation of the expression of cell-surface activation markers and secretion of cytokines (Klausner et al., 1991). In order to study T cell activation, we stimulated Jurkat cells with phorbol 12-myristate-13-acetate (PMA) and ionomycin. PMA is a small organic compound that can diffuse through the plasma membrane into the cytoplasm and can directly activate protein kinase C (PKC), thus avoiding the need for surface receptor stimulation, since PKC transduces the signal on T cells after encountering with antigen. Additionally, ionomycin, a calcium ionophore, is used to trigger intracellular calcium release and it synergizes with PMA in the activation of PKC. A suitable combination of PMA and ionomycin can provide specific and strong signalling on T cells (Chatila et al., 1989). Since activation with PMA and ionomycin omits the need for surface receptor stimulation, we were unable to study the impact of DEXI on the activation of T cells upstream of PKC.

4.9 Measurement of T cell stimulation

Activation of T cells results in the release of cytokines, such as IL-2 and interferon gamma (INF-γ) and change in cell-surface expression markers, such as an increase in the early T-cell activation marker CD69. We measured the cell-surface expression of CD69 by flow cytometry, which was the only marker we analysed in our study. We cannot exclude the possibility that DEXI has an impact on very early signalling events (e.g. immediate events such as protein phosphorylation or calcium-level release) or on later events, such as cytokine secretion.

4.10 Statistical analysis

For Paper I, in ASE measurements we used statistical package for the social sciences (SPSS) (IBM SPSS Statistics v21.0, Chicago, IL, USA) for boxplot analyses to check for outliers among the five replicates of each sample. The outliers were defined as measurements with
extreme values, which were excluded from further analysis. Due to careful handling of studied samples, the frequency of outliers per investigated gene was very low (for details of the excluded measurements, see Paper I, additional file 3: Table S2). For ASE measurements, a two-tailed, unpaired Student’s t-test was used to compare gene expression between two groups, i.e. cDNA measurements per sample against gDNA as reference group as well as all cDNA measurements as a single group against the gDNA reference group. The cDNA and gDNA measurements were logarithmically transformed and followed a normal distribution around zero.

For all three papers, a Mann–Whitney U test was performed to compare gene expression levels or protein expression between MS patients versus controls and for the gene expression in relation to genotypes. For ASE reported in Paper I, gene expression and correlation with genotype was done in a single sample, whereas for Papers II and III, samples from different individuals with different genotypes were compared.

Unlike the independent-samples t-test, the Mann–Whitney U test allows comparison of the differences between two independent groups when the values are not normally distributed, whereas a Student’s t-test analysis allows a hypothesis to be tested, such as whether the mean value of two groups will be statistically different for normally distributed variables. In addition, the Mann–Whitney U test makes no assumptions on outliers or measurement errors.

In order to identify correlation between the gene expression values of CIITA, DEXI, CLEC16A, and SOCS1, the coefficient of determination ($r^2$) was determined by pairwise linear regression analysis. The coefficient of determination represents how well the regression line fits the data points examined. In cases when the regression line falls exactly on every data point, the $r^2$ value is one (i.e. all variations of one variable are predicted by the other variables). The more often the data points fall outside the regression line, the lower the $r^2$ value becomes with a minimum of zero.

A Student’s paired t-test was used to analyse the DEXI expression upon T-cell activation. A paired t-test matched each variable to its counterpart in a different condition or state, thereby increasing the power of the test to identify differences between conditions or states. P-values of $< 0.05$ were considered statistically significant for all analyses that were performed using GraphPad Prism 6 (GraphPad Software, Inc., San Diego, CA, USA). In the gene and protein expression analysis reported in Papers I, II, and III, the relatively small sample size was a
drawback. In all statistical analyses, a large number of samples will provide greater power to make firm conclusions and smaller differences between the groups will be more easily detected. A post-hoc power analysis was performed for the study reported in Paper I. To achieve 80% power to detect a significant difference of the observed magnitude and variance, a sample size of at least 43 individuals in each group would be required (R version 3.3.2, The R Foundation). Power analysis allows the sample size to be decided and the statistical test required for designing an experiment in order to make statistical judgements.
5 GENERAL DISCUSSION

Due to international efforts in recent decades, there has been progress in our understanding of the genetic components of MS. The primary and strongest genetic risk factor for MS in most populations has been identified within the MHC locus (OR = 3.1) (Jersild et al., 1972). However, with the advent of GWAS and fine-mapping studies of immune-related loci, more than 110 non-MHC loci have been identified, each exerting a small effect on the risk of MS development (International Multiple Sclerosis Genetics Consortium et al., 2013; International Multiple Sclerosis Genetics Consortium et al., 2011). The next challenge faced by MS researchers is to place these genetic associations into a functional context and to uncover the genetic architecture behind MS susceptibility.

The main aim of this thesis was to explore the functional implications of selected genetic variants underlying MS susceptibility. Understanding the functional mechanisms driving the MS association of these variants might shed new light on how pathogenic processes are initiated and could provide knowledge for developing new strategies to prevent or treat MS. The majority of disease-associated SNPs (93%) lie within non-coding regions of the genome, and thus they may contribute to disease by regulating, for example, either promoter or enhancer activity (Farh et al., 2015) leading to altered gene expression. An enrichment of genetic variants associated with MS has been observed in the regulatory regions of DNA. These variants are found in DNase hypersensitive sites (i.e. regions with an open chromatin structure) in disease-relevant immune cells, especially T and B cells (Maurano et al., 2012; Disanto et al., 2012). Thus, it seems logical that MS risk SNPs may affect gene expression in immune cells. Moreover, many of the MS risk SNPs are associated to other autoimmune diseases (Kemppinen et al., 2011) and therefore knowledge acquired from MS-specific studies could be of relevance for studying autoimmune diseases in general.

The aim of the studies reported in Papers I and II was to investigate whether MS-associated SNPs identified through large-scale GWAS analysis could act as eQTLs for the genes IKZF3, IQGAP1 and CD69, and the 16p13.13 locus genes CIITA, DEXI, CLEC16A and SOCS1.
5.1 Cis-regulatory roles for MS-susceptibility SNPs in IQGAP1 and IKZF3

We aimed to understand the molecular mechanisms behind the associations between an SNP and increased risk of MS, and found a possible regulatory role for the MS-associated IKZF3 and the MS-associated IQGAP1 SNP in whole blood. There are several functional approaches to characterize SNPs. AI has been established as a ubiquitous phenomenon that may underlie disease risk exerted by disease-associated SNPs. AI may occur when the SNP-containing sequence is within a cis-regulatory region of a gene. A prerequisite for studying the putative effects on allelic output for disease-associated SNPs is that either the SNP itself or a proxy SNP on the same haplotype as the susceptibility SNP is transcribed. In order to prioritize the SNPs studied for Paper I, we performed in-silico analysis of the 110 non-MHC MS-associated risk variants (International Multiple Sclerosis Genetics Consortium et al., 2013; International Multiple Sclerosis Genetics Consortium et al., 2011) utilizing the online bioinformatics resources SNPer (Riva et al., 2002) and SNAP (Johnson et al., 2008) to identify transcribed MS risk SNPs or transcribed proxies. Of the 110 non-MHC MS-associated SNPs, 63 were located within a gene region and 46 in intergenic regions, and one SNP in non-coding RNA, which has also been indicated by other researchers (Briggs, Leung, et al., 2014). Among the risk SNPs within a gene, three SNPs were in exons, three in 3’ untranslated regions (UTR), one in 5’ UTR and 11 in promoters (Table 1).

Table 1. Annotation of 110 non-MHC MS-associated SNPs

<table>
<thead>
<tr>
<th>Location of SNPs</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter region</td>
<td>BCL10, RGS1, PLEK, CD86, TCF7, IL12B, TAGAP, FAM164A, MIR1208, C16orf75, and CYP24A1</td>
</tr>
<tr>
<td>Exonic region</td>
<td>TYK2, SLC44A2, and IFI30</td>
</tr>
<tr>
<td>5’ UTR</td>
<td>CD86</td>
</tr>
<tr>
<td>3’ UTR</td>
<td>EVI5, CD69, and C14orf181</td>
</tr>
<tr>
<td>Non-coding RNA</td>
<td>PVT1</td>
</tr>
<tr>
<td>Introgenic</td>
<td>MMEL1, PLEKHG5, DDAH1, SLC30A7, CD58, PHGDH, FCRL1, SLAMF7, CENPO C1orf106, FLJ16341, MERTK, STAT4, SP140, FOXP1, CBLB, IQCB1, TET2, ANKRD55, NDFIP1, LOC285626, RGS14, PXT1, BACH2, AHI1, JAZF1, ELM01, ZNF767, MLANA, IL2RA, ZMIZ1, AGBL2, CXCR5, TNFRSF1A, TSFM, PITPNM2, ZFP36L1, GALT, TRAF3, IQGAP1, CLEC16A, CDH3, WWOX, STAT3, VMP1, MALTI, TNFSF14, EPS15L1, DKK1, CD40, SLC9A8, SLC2A4RG, ZBTB46, MAPK1, and TTYMP</td>
</tr>
<tr>
<td>Introgenic</td>
<td>BCL10, VCAM1, RGS1, ZFP36L2, PLEK, SATB1, CMC1, CCR4, CD86, IL12A, MANBA, ILR7, PTGER4, TCF7, JARID2, PTPRK, IL22RA2, OLG3, TNFAIP3, TAGAP, CARD11, SKAP2, IKZF1, FAM164A, POU5F1B, MIR1208, ZNF438, C10orf55, HHEX, CD6, PRDX5, TREH, CXCR5, LTBR, MIR548AN, JDP2, MAF, CTS8, SOX8, CLEC16A, RMI2, MAPK3, IRF8, NPEPPS, CYP24A1, and GRB7</td>
</tr>
</tbody>
</table>
Based on these results, we identified one transcribed SNP, rs11052877 in the 3’ UTR of CD69. Thereafter, we searched for proxy SNPs in the LD region of MS-associated SNPs to identify proxy markers in coding regions. We chose not to include proxy SNPs that affect the encoded amino acid within a gene. We found seven SNPs in strong LD with the MS-associated SNPs ($r^2 \geq 0.8$) (rs12479056, rs10802190, rs28445040, rs11609, rs907091, and rs13943) at the PUS10, CD58, SP140, IQGAP1, IKZF3, and MAPK1 loci respectively (Table 2). Based on the expression level in whole blood confirmed by PCR in our laboratory and on its biological relevance, we prioritized three SNPs for further analysis; rs11052877, rs11609, and rs907091, at the CD69, IQGAP1, and IKZF3 loci respectively (in bold in Table 2).

Table 2. Details of transcribed MS-associated SNPs and proxy markers

<table>
<thead>
<tr>
<th>Gene</th>
<th>MS-associated SNP</th>
<th>Proxy</th>
<th>Alleles (MAF)</th>
<th>$r^2$ ($D'$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD69</td>
<td>rs11052877</td>
<td>N/A</td>
<td>A/G (0.34)</td>
<td>N/A</td>
</tr>
<tr>
<td>IKZF3</td>
<td>rs12946510</td>
<td>rs907091</td>
<td>C/T (0.48)</td>
<td>0.80 (0.90)</td>
</tr>
<tr>
<td>IQGAP1</td>
<td>rs8042861</td>
<td>rs11609</td>
<td>G/C (0.36)</td>
<td>0.86 (0.93)</td>
</tr>
<tr>
<td>CD58</td>
<td>rs6677309</td>
<td>rs10802190</td>
<td>A/T (0.18)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>PUS10</td>
<td>rs12479056</td>
<td>rs842639</td>
<td>A/G (0.32)</td>
<td>0.97 (1.00)</td>
</tr>
<tr>
<td>MAPK1</td>
<td>rs2283792</td>
<td>rs13943</td>
<td>C/G (0.46)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>SP140</td>
<td>rs9989735</td>
<td>rs28445040</td>
<td>C/T (0.20)</td>
<td>0.84 (0.94)</td>
</tr>
</tbody>
</table>

Notes: 1Gene in bold were studied by ASE (Paper I), 2MS-associated SNPs (International Multiple Sclerosis Genetics Consortium et al., 2013; International Multiple Sclerosis Genetics Consortium et al., 2011), 3Transcribed SNPs in LD with non-transcribed MS-associated SNPs, 4Minor allele frequency (MAF) of the MS-associated SNPs or proxy SNPs (Johnson et al., 2008), A/B – major/minor allele, 5LD between MS-associated SNP and proxy SNP, N/A – not applicable

Our data showed that the risk genotype at rs11609 results in higher allelic expression and overall increase in the gene expression of IQGAP1 in whole blood from MS patients and from healthy controls. However, a minority of the samples deviated from the others in the ASE experiment. Rs11609 is a proxy SNP for the MS-associated SNP rs8042861 in IQGAP1 and the lack of AI could be due to lack of full LD ($r^2 = 0.86$, $D' = 0.93$). In order to investigate this, we analysed whether the lack of AI for rs11609 in the inconsistent samples could be due to lack of heterozygosity for MS-associated SNP. However, all samples except one were also heterozygous for rs8042861, thus eliminating the possibility that the absence of AI in the inconsistent samples may have arisen from a different genotype of the MS risk SNP. Since MS-associated SNPs from large-scale GWASs act as representatives for all SNPs located in
the same haplotype, there is a reasonable probability that other SNPs or more than one SNP in high LD with MS-associated SNPs might drive the risk of developing MS through higher expression of genes such as \textit{IQGAP1}. There are 79 proxy SNPs, of which five are in the coding regions of \textit{IQGAP1} or neighbouring genes, in high LD ($r^2 < 0.70$; $D' = 0.93–1.00$) with the MS-associated SNP rs8042861. A total of 15 SNPs are located within transcription factor binding sites (genome-euro.ucsc.edu) and hence might be the causal SNP contributing to the higher expression of the risk allele.

Although a genotype-dependent overall \textit{IQGAP1} gene expression difference was observed at the mRNA level, we could not detect this difference at the protein level in lysates from PBMCs from healthy controls. We did detect a trend towards higher \textit{IQGAP1} protein expression, but this trend was not statistically significant. The lack of correlation between mRNA and protein levels could be due to the small sample size, as discussed in statistical analysis section 4.10.

We observed consistent AI in heterozygous individuals for rs907091 in \textit{IKZF3}, but differences in overall \textit{IKZF3} gene expression in homozygous individuals were not observed. The limited sample size in our study may have lacked the power to detect subtle differences in whole blood. Furthermore, the blood samples were from an MS patient population at different stages of the disease and who were receiving different treatments, and therefore heterogeneity among sampled patients could also be the reason for the lack of differential genotype-dependent gene expression. Since rs907091 is a proxy SNP in high LD ($r^2 = 0.80$, $D' = 0.90$) with the MS-associated SNP rs12946510, we checked whether consistent AI observed for \textit{IKZF3} could arise due to the genotype of the MS-associated SNP. We analysed the samples for double heterozygosity and found five samples that did not display double heterozygosity between the proxy SNP and the MS-associated SNP. The consistent AI observed for rs907091 may not be attributed to MS-associated SNP but to another regulatory SNP. We cannot exclude the possibility that other SNPs in LD with the MS-associated SNP could have been the driver of the observed AI, as there are 77 SNPs in strong LD ($r^2 = 0.70$, $D' = 0.92–1.00$) with the MS-associated rs12946510 SNP in \textit{IKZF3}, of which 13 are found in transcription factor binding sites. Interestingly, rs907091 lies in the binding sites of the SPI1 and GATA2 transcription factors (genome-euro.ucsc.edu) and might therefore be the casual SNP regulating the expression of \textit{IKZF3}. In addition to rs907091, four other proxy SNPs are located in the coding region of \textit{IKZF3} or neighbouring genes, and might drive the AI of
IKZF3 expression. Alternatively, there might be more than one regulatory SNP in LD with MS-associated SNP driving AI for IKZF3.

We did not observe consistent AI for the CD69 gene, which indicates that the SNP associated with MS might not have cis-regulatory properties for CD69 in whole blood. Recent bioinformatics analyses have indicated that the genotype of the MS-associated rs11052877 SNP does not affect any transcription factor DNA binding motifs that is in line with our results, showing no cis-regulatory effect for this SNP in whole blood (Briggs, Leung, et al., 2014). However, blood consists of different types of immune cells, and the ratio between the different immune cells differs from individual to individual. We cannot exclude the possibility that there could be a possible cis-regulatory mechanism for this SNP in more specific immune cell subtypes or other cells relevant to MS, and therefore further investigation is warranted.

Our findings imply that changes in IQGAP1 and IKZF3 gene expression may affect susceptibility to MS and possibly be a molecular mechanism behind the association of two MS-associated SNPs. Although the AI is marked by the SNPs we tested in our experimental setting, the underlying functional variant may be a risk variant in full LD. Further research on the regulatory properties within these gene regions and their possible roles in disease aetiology is warranted. IQGAP1 encodes the IQ Motif Containing GTPase Activating Protein 1, which is implicated in several cytoskeletal and cellular signalling pathways, leukocyte chemotaxis, NK-cell cytotoxicity, and B-cell dysfunctions (Bamidele et al., 2015; Neel et al., 2011; Chu et al., 2008). Further, IQGAP1 has been recognized by autoantibodies in patients with autoimmune bullous skin diseases (Presslauer et al., 2003). In our study, we observed a higher expression of IQGAP that may result in abnormal leukocyte and NK cell activity, leading to an increased risk of MS. IKZF3 encodes AIOLOS, an Ikaros family of zinc finger 3 protein, which is expressed at higher level in mature peripheral B cells and is involved in the regulation of lymphocyte development (Cortes et al., 1999) and chromatin remodelling and histone deacetylation (Koipally et al., 1999; Morgan et al., 1997). B cell function was affected in IKZF3 deficient mice (John et al., 2011), which implies that the higher expression observed for IKZF3 may result in aberrant B-cell response, leading to an increased risk of MS. Our finding that the changes in gene expression of IKZF3 and IQGAP1 in MS samples may increase the risk of MS susceptibility is a finding also of interest for other autoimmune disorders, as several SNPs in IKZF3 are associated with several autoimmune diseases, such as
SLE, RA, ankylosing spondylitis, and asthma (Zhang et al., 2014; Qiu et al., 2013; Kurreeman et al., 2012; Marinho et al., 2012).

5.2 The complex role of the 16p13.13 chromosomal region in MS susceptibility

The 16p13 chromosomal region has frequently been found associated with other autoimmune diseases in addition to MS. However, the mechanism behind the association is largely unknown. Understanding the consequences of genetic variation and its influence on cellular phenotypes such as gene expression may lead to a better understanding of MS and complex diseases. Large-scale genetic analyses have discovered several SNPs in intronic regions of CLEC16A associated with MS onset (International Multiple Sclerosis Genetics Consortium et al., 2013; International Multiple Sclerosis Genetics Consortium et al., 2011). Intronic CLEC16A SNPs may contribute to disease susceptibility by acting as eQTLs for CLEC16A itself or neighbouring genes through disrupting the binding sites of promoters or enhancers, thus influencing the gene expression. The intronic rs6498169 SNP in CLEC16A was among the first genetic variants highlighted for MS (p = 3.83 × 10^{-6}) in the first MS GWAS study (International Multiple Sclerosis Genetics Consortium et al., 2007) and has since been reported to be associated also with other autoimmune diseases (Skinningsrud et al., 2010; Martinez et al., 2010). The CLEC16A gene has been replicated in several studies as an MS susceptibility gene (Mero et al., 2011; Zoledziewska et al., 2009; Perera et al., 2009; International Multiple Sclerosis Genetics Consortium, 2009). Intronic CLEC16A SNPs have been demonstrated as associated with MS and other autoimmune diseases, such as T1D (Hakonarson et al., 2007), Crohn’s disease (Marquez et al., 2009), Addison’s disease (Skinningsrud et al., 2008), primary biliary cirrhosis (Hirschfield et al., 2012), juvenile idiopathic arthritis (Skinningsrud et al., 2010), alopecia areata (Jagielska et al., 2012), and RA (Martinez et al., 2010). Many of the reported SNPs are in strong LD (reviewed in (Berger et al., 2013)), which makes it difficult to identify the causal SNPs. In order to explain the role of the different MS-associated SNPs across the 16p13 locus encompassing the CIITA-DEXI-CLEC16A-SOCS1 gene complex, 149 SNPs were genotyped in a combined American and British sample set (Zuvich et al., 2011). Further, a detailed LD pattern and logistic regression analysis was performed. The data revealed that this region probably contains three independent MS disease loci (Zuvich et al., 2011). However, an intronic CLEC16A rs7184083
SNP, which is in LD with MS-associated SNP rs12708716 \((r^2 = 0.31, D' = 1.00)\), displayed the most significant p-value in this region. Interestingly, as demonstrated by chromatin capture assay, intronic sequences of CLEC16A containing the MS-associated rs12708716 SNP are in physical proximity to the DEXI promoter (Davison et al., 2012), and this CLEC16A genetic region might affect DEXI expression via DNA looping.

### 5.3 Regulatory role of MS-associated intronic CLEC16A SNPs

A number of studies have indicated that the MS risk alleles in CLEC16A correlate with changes in the gene expression of CLEC16A and of neighbouring genes as summarized in Table 3. In the studies reported in Papers II and III, our aim was to elucidate the functional role of MS-associated intronic CLEC16A SNPs rs12927355, rs4780346, and rs6498169 in the regulation of CIITA, DEXI, CLEC16A, and SOCS1 gene expression in T cells. The results of a large-scale consortium-based analysis identified that rs12927355, located in intron 19 of CLEC16A as the primary MS-associated SNP, whereas rs4780346, located in the intergenic CLEC16A-SOCS1 region was identified as a secondary MS-associated SNP (International Multiple Sclerosis Genetics Consortium et al., 2013).

**Table 3.** Overview of MS-associated CLEC16 SNPs and their effects on the expression of 16p13.13 locus genes such as CIITA, DEXI, CLEC16A, and SOCS1

<table>
<thead>
<tr>
<th>SNP</th>
<th>Risk allele</th>
<th>Genomic location</th>
<th>Cell/tissue</th>
<th>eQTL target genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12708716</td>
<td>A</td>
<td>Intron 19</td>
<td>Thymic tissue(^a) Whole blood(^a) (\beta)-cells(^d) Monocytes(^e)</td>
<td>SOCS1 ↓ No effect CLEC16A ↑ DEXI ↓</td>
</tr>
<tr>
<td>rs12927355</td>
<td>G</td>
<td>Intron 19</td>
<td>CD4(^b) T cells(^b) CD8(^b) T cells(^b)</td>
<td>CLEC16A ↑ SOCS1 ↑ No effect</td>
</tr>
<tr>
<td>rs4780346</td>
<td>A</td>
<td>Intergenic (CLEC16A-SOCS1)</td>
<td>CD4(^b) T cells(^b) CD8(^b) T cells(^b)</td>
<td>No effect No effect</td>
</tr>
<tr>
<td>rs6498169</td>
<td>G</td>
<td>Intron 22</td>
<td>Thymic tissue(^a) Whole blood(^a) CD4(^c) T cells(^c) CD8(^c) T cells(^c)</td>
<td>DEXI ↓ SOCS1 ↓ No effect DEXI ↑</td>
</tr>
<tr>
<td>rs17806056</td>
<td>T</td>
<td>Intron 19</td>
<td>Whole blood(^f)</td>
<td>CLEC16A ↓</td>
</tr>
<tr>
<td>rs7200786</td>
<td>A</td>
<td>Intron 19</td>
<td>Whole blood(^g)</td>
<td>No effect</td>
</tr>
</tbody>
</table>

**Notes:**
- \(^a\)(Leikfoss et al., 2013), \(^b\)Paper II (Leikfoss et al., 2015), \(^c\)Paper III: Manuscript (Keshari et al., 2016), \(^d\)(Soleimanpour et al., 2014), \(^e\)(Davison et al., 2012), \(^f\)(Li et al., 2015), and \(^g\)(van Luijn et al., 2015).
In three out of four studies (Table 3), no eQTL effects of MS-associated *CLEC16A* SNPs were observed in whole blood samples, but genotype-dependent effects were observed in more homogenous samples such as immune cell subsets, indicating cell-specific eQTL effects. In the study reported in Paper II, we observed a significantly higher expression of *CLEC16A* and *SOCS1* in CD4$^+$ T cells in samples that were homozygous for the risk allele of *CLEC16A* rs12927355, but not in CD8$^+$ T cells (Leikfoss et al., 2015). The lack of association between *CLEC16A* rs12927355 genotype and gene expression in CD8$^+$ T cells may be due to the lack of sufficient samples or to a cell-specific mechanism. No significant effect was seen for *DEXI* expression in either T cell subtypes. In the study reported in Paper III, we observed a significantly higher *DEXI* expression in samples for SNP rs6498169 risk allele, both in CD4$^+$ and CD8$^+$ T cells. SNPs in intron 19 of *CLEC16A* are enriched in transcription-factor binding sites that may regulate the expression of *CLEC16A*, *DEXI*, and *SOCS1*. The MS-associated rs12927355 is found in an active transcribed region very close to TF binding sites and with enrichment in enhancer mark H3K27Ac. Further, rs4780346 SNP is found in DNase I hypersensitive sites (DHSs) and TF binding site (Karolchik et al., 2014), which means that genetic variants may have an influence on the expression of target genes.

The rs6498169 of *CLEC16A* is located in DHSs in NK and Th2 cells (Maurano et al., 2012). Our research group has previously demonstrated reduced *DEXI* and *SOCS1* expression in thymic tissue samples for the genotype of intronic *CLEC16A* rs6498169 SNP but not for *CIITA* and *CLEC16A* expression (Table 3). However, in Paper III we show that the rs6498169 risk allele correlated with significantly higher *DEXI* expression, both in CD4$^+$ and CD8$^+$ T cells, indicating cell-specific gene regulatory mechanisms. We do not find any correlation between *DEXI* expression and the genotype of the primary ImmunoChip SNPs rs12927355 in CD4$^+$ or in CD8$^+$ T cells. However, we did see a trend of higher risk allele in the same direction as observed for rs6498169. The lack of correlation between *DEXI* expression and the genotype could have been due to incomplete LD between the studied SNPs. Importantly, rs6498169 has been reported to confer susceptibility to other autoimmune diseases, such as thyroid disease, rheumatoid disease, and juvenile idiopathic arthritis (Muhali et al., 2014; Skinningsrud et al., 2010), implying the functional role of this SNP in autoimmunity. Dysregulation of 16p13.13 locus genes *SOCS1*, *DEXI*, and *CLEC16A* can have direct impact on the immune system and that may trigger MS development. The function of *DEXI* is unknown, and we and other researchers have shown altered *DEXI* expression with MS-
associated intronic CLEC16A SNPs (Tomlinson et al., 2014; Davison et al., 2012), indicating the potential role of DEXI in MS pathogenesis.

5.4 Co-expression of 16p13.13 locus genes CIITA, DEXI, CLEC16A, and SOCS1 in peripheral T cells

Tissue-dependent regulation could be the result of different distribution of chromatin accessibility in different cell types (Thurman et al., 2012). Since ~30% of all genes show tissue-dependent regulation of gene expression (Fu et al., 2012; Dimas et al., 2009), we have in Paper II elucidated the co-expression of 16p13.13 locus genes CIITA, DEXI, CLEC16A and SOCS1 in peripheral T cells (CD4⁺ and CD8⁺ T cells). We have shown that all four genes are co-expressed in CD4⁺ T cells. However, CIITA and DEXI were not co-expressed in CD8⁺ T cells, indicating a cell-type specific regulation of genes. Highly significant correlations between gene expression of DEXI, CLEC16A, and SOCS1 have been observed in lymphoblastoid cell lines (Zuvich et al., 2011). Interestingly, Zuvich and colleagues have speculated that the intergenic region between CLEC16A and SOCS1 harbouring an MS-associated SNP regulates the expression of multiple genes in this region, thereby affecting MS risk.

The co-expression of genes on the same chromosomal region may not necessarily be regulated by a common molecular mechanism. However, their co-expression might be needed in cellular signalling pathways and for general cellular functions. CIITA, which encodes a transcriptional regulator of MHC-II, is important for the MHC-II gene expression (van Luijn et al., 2015; Kern et al., 1995). Interestingly, a recent study showed the correlation between gene expression of CLEC16A and CIITA in immature monocyte-derived dendritic cells (van Luijn et al., 2015), which is also seen in CD4⁺ T cells. In addition, CLEC16A was indicated as a direct regulator of antigen presentation by the HLA-II (van Luijn et al., 2015). The functional role of immune regulatory genes CIITA and SOCS1 in inflammation and autoimmunity has already been established (Yoshimura et al., 2012; Friese et al., 2005), and the molecular roles of DEXI and CLEC16A in immune regulation might elucidate the mechanisms behind MS pathogenesis and other autoimmune diseases.
5.5 Functional analysis of DEXI

In our study, we used the Jurkat cell line as a model system to reveal the function of DEXI in T cells. *DEXI* is expressed in several human primary immune cells and cell lines. In Paper III, we showed that DEXI is expressed in unstimulated Jurkat T cells and down-regulated after stimulation with PMA and ionomycin. This suggests that DEXI might play a role in naive, unstimulated cells or that proper DEXI down-modulation is mandatory for proper T-cell activation and function. However, we found that DEXI did not have any influence on the capacity of Jurkat T cells to become activated, when measured by cell-surface expression of CD69 after PMA and ionomycin stimulation. Evidence from earlier research indicates that T-cell activation with PMA and ionomycin induces the activation of PKC, which in turn mediates an important role in several signalling events downstream of TCR and the CD28 costimulatory receptor (Isakov et al., 2012). We cannot rule out the possibility that DEXI might have an effect downstream of TCR stimulation. In addition, DEXI might have influence on other stages in T-cell activation, affecting very early activation and tyrosine phosphorylation of signalling molecules or calcium release or later activation stages, and this warrants further investigation. Thus, the regulation of *DEXI* expression on T-cell activation should possibly be replicated in primary T cells.

To unravel whether the DEXI protein function could be suggested based on *in silico* analyses, we used freely available online tools to analyse its protein sequence. DEXI is a protein of 95 amino acids that are well conserved, with orthologs readily detectable in multiple species (Figure 12a). There is 95% sequence identity between the human and mouse orthologs and 84.5% between DEXI proteins in humans and zebrafish (Larkin et al., 2007). In general, DEXI is predicted to be a well-structured protein except for its first seven amino acids (Ward et al., 2004; Jones, 1999) (Figure 12b) consisting of three α-helices, in which the second helix could be membrane spanning (Jones, 2007; Ward et al., 2004). We found that two additional algorithms (MEMSAT and MEMSAT3) indicated the presence of one or two transmembrane regions (Figure 12c) ([http://bioinf.cs.ucl.ac.uk/psipred](http://bioinf.cs.ucl.ac.uk/psipred)). A possible biological function for DEXI cannot be inferred from these *in silico* analyses, but the data indicated that DEXI might be associated with membranous structures in the cells. However, preliminary results from
Figure 12. Bioinformatics analysis and subcellular localization of DEXI

(a) multiple sequence alignment tools predict either one or two trans-membrane domains for the DEXI protein; this is underlined in black and red and is conserved across many species; amino acids that differ from the human DEXI sequence are highlighted in green; (b) from N terminal, structural disordered domains at amino acid 1–7 and predicted secondary structures of DEXI; α-helices (red boxes) and structural domains (grey boxes); (c) structural domains of DEXI, either with one (upper) or two transmembrane (lower) regions; (d) confocal microscopy images of Jurkat T cells transfected with plasmid expressing myc-DDK-tagged DEXI and stained with Hoechst (blue in the merged right image) and anti-DDK (red in the merged right image); scale = 10 µm; the images are representative of two independent experiments.

Confocal microscopy after anti-DDK immunostaining of myc-DDK-DEXI transfected Jurkat cells, showed an uneven distribution of DEXI in the cytoplasm and in the perinuclear region (Figure 12d). However, further experiments are warranted to confirm this finding. Interestingly, immunofluorescent staining of the human cell line U-251 MG derived from a malignant glioblastoma has indicated that the subcellular localization of DEXI is mainly in nucleus and additional localization has been suggested in cytoplasm (http://www.proteinatlas.org).
6 FUTURE PERSPECTIVES

We analysed whether genotypes of selected MS-associated SNPs or proxies thereof had any impact on the gene expression of nearby genes. Our data indicate a possible cis-regulatory role for the MS-associated variants in *IKZF3*, *IQGAP1*, *CLEC16A*, *SOCS1* and *DEXI*. We suggest that such cis-acting mechanisms may contribute to the MS pathology. The exact molecular mechanisms of how higher *IKZF3*, *IQGAP1*, *CLEC16A*, *SOCS1* and *DEXI* expression may contribute to increased MS risk remains unclear. It would be of great interest to dissect the exact molecular mechanisms behind higher expressions of *IKZF3* and *IQGAP1* affecting the disease risk. Since we did not observe the cis-regulatory mechanism for *CD69* SNP in whole blood, further investigation of specific immune cells is warranted. Furthermore, the regulatory effect of the *IKZF3* and *IQGAP* variants and the mechanisms behind this in sub-phenotypes of MS would add further knowledge. These SNPs reside in large LD blocks that could harbour one or more functional variants, and additional investigations to identify the functional variations are warranted. Whether *CLEC16A* genotype affects expression of 16p13.13 genes in other immune cells or in subtypes of the CD4+ and CD8+ T cell lineages remains to be studied.

The functional properties of these SNPs and the impact on the protein expression in specific cell types may identify which cells drive the association of these SNPs and MS. Tracing the specific mechanism behind changes in gene expression of SNPs may guide the identification of specific pathogenic processes involved in MS. Understanding the role of cis-acting regulatory variants may provide new insights into signalling pathways and networks underlying complex diseases.

The function of DEXI is unknown and we are in the starting phase of identifying its role. Further studies of DEXI and its effect on MS-disease specifically in T cells, B cells, and other immune cells relevant for MS are warranted. Identification of signalling pathways regulating DEXI expression and interaction partners of DEXI in immune cells may provide potential therapeutic targets for the manipulation of the immune response. In addition, identification of subcellular localization of DEXI would add further knowledge that could be used in the development of hypotheses for functional studies. Furthermore, manipulating the genetic
content of cells implicated in MS pathology may help to reveal the exact signalling pathways involved in MS. In relation with DEXI, further research on cytokine secretion, cell proliferation, cell differentiation, apoptosis, and receptor internalization in activated T cells would add further knowledge that could set the stage for new approaches for treating MS and enhance our understanding of why genes affect MS susceptibility or phenotype.

7 CONCLUSIONS

Currently, there is no cure available for MS and current knowledge of MS aetiology is limited. MS research could therefore help to unravel the mechanisms of disease pathogenesis in general, with broad implications for developing strategies to prevent and treat disease. Because MS is a health challenge for affected patients and for society in Norway and worldwide, it is important to increase the research efforts. Through collaborations worldwide, the MS research community has identified more than 110 MS-associated genetic variants located in coding and non-coding gene regions of the genome (International Multiple Sclerosis Genetics Consortium et al., 2013). Based on the results of genetic and molecular analyses, this thesis highlights the functional role of a selection of these MS-associated SNPs and their putative contribution to MS development in whole blood and T cells. Further, genetic research and molecular studies may contribute to unravel mechanisms of MS pathogenesis.
REFERENCES

8 REFERENCES


Hedstrom, A. K., T. Olsson, & L. Alfredsson. (2012). High body mass index before age 20 is associated with increased risk for multiple sclerosis in both men and women. Mult Scler, 18(9), 1334-1336.


Stinissen, P., C. Vandevyver, R. Medaer, L. Vandegaer, J. Nies, et al. (1995). Increased frequency of gamma delta T cells in cerebrospinal fluid and peripheral blood of


APPENDIX, PAPERS I-III