From DEPARTMENT OF CLINICAL NEUROSCIENCE
Karolinska Institutet, Stockholm, Sweden

EPIGENETICS IN HEALTH AND DISEASE: ON SMOKING, MULTIPLE SCLEROSIS, AND RHEUMATOID ARTHRITIS DISEASE STATES

Mikael Ringh

Karolinska Institutet

Stockholm 2019

Published by Karolinska Institutet.
Printed by Eprint AB 2019

ISBN 978-91-7831-581-9

Epigenetics in Health and Disease: On Smoking, Multiple Sclerosis, and Rheumatoid Arthritis Disease States

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Mikael Ringh

AKADEMISK AVHANDLING

som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Lilla Salen, Astrid Lindgrens Barnsjukhus Q-huset (QA31:01), Karolinska Universitetssjukhuset, Solna

Friday 25th October 2019, at 9.00

Principal Supervisor:
Professor Tomas Ekström
Karolinska Institutet
Department of Clinical Neuroscience

Co-supervisor(s):
Assistant Professor Malin Almgren
Karolinska Institutet
Department of Clinical Neuroscience

Professor Anca Catrina
Karolinska Institutet
Department of Medicine

Dr. Heidi Vähämaa
Karolinska Institutet
Department of Medicine

Professor Yun Liu
Fudan University
Department of Biochemistry and Molecular Biology

Opponent:
Associate Professor Colm Nestor
Linköping University
Department of Clinical and Experimental Medicine

Examination Board:
Professor Ann-Kristin Östlund Farrants
Stockholm University, Wenner-Gren Institute
Department of Molecular Biosciences

Associate Professor Åsa Johansson
Uppsala University
Department of Immunology

Professor Anders Lindén
Karolinska Institutet
Institute of Environmental Medicine

ABSTRACT

The main aim of this thesis was to investigate the genome-wide DNA methylation in tissues associated with the immunopathogenesis of two autoimmune diseases, Rheumatoid Arthritis (RA) and Multiple Sclerosis (MS), as well as smoking-associated methylome patterns in the lung.

In Study I, we investigated the methylome of two sets of monozygotic twin pairs, representing two phases of anti citrulline protein antibody (ACPA)-positive RA disease development. The first set included five MZ twin pairs discordant for APCA at risk for developing RA, and the second set included seven pairs discordant for ACPA-positive RA. A differentially methylated region associated with a protocadherin (*PCDH*) gene proposes a temporal epigenetic connection in the progression from ACPA-positivity to clinical RA.

In Study II, we investigated the impact of tobacco smoking on bronchoalveolar lavage (BAL), which mainly consists of tissue-resident alveolar macrophages. We combined methylome and transcriptome data from BAL cells from healthy individuals and provide novel smoking-associated signatures converging to genes involved in migration, signaling and inflammatory response of immune cells. Of note, many of the sites in the smoking-associated methylome signature map to enhancer regions. Our findings propose that the epigenetic landscape of BAL cells is modified in smokers, and that it may involve active demethylation resulting in induced immune-related activities in the lung.

Despite strong evidence that cigarette smoking is a risk factor for MS, little research has focused on smoking-associated changes in the lung of MS patients. In Study III, we show that smoking in MS patients resulted in subtle alterations in their methylome related to neuronal processes. Additionally, non-smoking MS patients displayed very discrete changes in transcriptional and translational processes and enhanced cellular motility, supporting findings on lung involvement in the pathogenesis of MS-like disease in animal studies.

In conclusion, we demonstrate a non-genetically linked temporal epigenetic connection between ACPA-positivity and clinical RA that may be of interest in future studies. We reveal smoking-associated changes in the epigenetic landscape of BAL cells, and increased activity of immune-related processes in the lung, possibly involving active demethylation. We also present new insights into the impact of cigarette smoking on pulmonary inflammation in the immunopathogenesis of MS.

LIST OF SCIENTIFIC PAPERS

I. Gomez-Cabrero D*, Almgren M*, Sjöholm LK*, Hensvold AH*, **Ringh MV**, Tryggvadottir R, Kere J, Scheynius A, Acevedo N, Reinius L, Taub MA, Montano C, Aryee MJ, Feinberg J, Feinberg AP, Tegnér J, Klareskog L, Catrina AI, Ekström TJ.
High-specificity bioinformatics framework for epigenomic profiling of discordant twins reveals specific and shared markers for ACPA and ACPA-positive rheumatoid arthritis.
Genome Med. 2016 Nov 22;8(1):124.

II. **Ringh MV**, Hagemann-Jensen M, Needhamsen M, Kular L, Breeze CE, Sjöholm LK, Slavec L, Kullberg S, Wahlström J, Grunewald J, Brynedal B, Liu Y, Almgren M, Jagodic M, Öckinger J, Ekström TJ.
Tobacco smoking induces changes in true DNA methylation, hydroxymethylation and gene expression in bronchoalveolar lavage cells.
EBioMedicine. 2019 Aug;46:290-304.

III. **Ringh MV***, Hagemann-Jensen M*, Needhamsen M, Kullberg S, Wahlström J, Grunewald J, Brynedal B, Jagodic M, Ekström TJ, Öckinger J, Kular L.
Methylome and transcriptome signature of bronchoalveolar cells from Multiple Sclerosis patients.
Manuscript.

* These authors contributed equally.

CONTENTS

LIST OF ABBREVIATIONS

5caC	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ACPA	Anti-citrullinated protein antibodies
AM	Alveolar macrophages
BAL	Bronchoalveolar lavage
BS	Bisulfite
CGI	CpG Island
CNS	Central nervous system
DMP	Differentially methylated position
DMR	Differentially methylated region
DNMT	DNA methyltransferases
DVP	Differentially variable position
IPA	Ingenuity Pathway Analysis
MS	Multiple sclerosis
oxBS	Oxidative bisulfite
PTM	Post-translational modifications
RNA-seq	RNA sequencing
RA	Rheumatoid arthritis
SLE	Systemic lupus erythematous
SNP	Single nucleotide polymorphism
TDG	Thymine DNA glycosylase
TET	Ten eleven translocation
TF	Transcription factor
TSS	Transcription start site
VMR	Variably methylated region

1 INTRODUCTION

1.1 EPIGENETICS

With a few exceptions, all cells of one individual carry the same genetic code. When cells divide and grow, they express different genes leading to differentiated cell types. The term epigenetics refers to mechanisms that can regulate gene expression without involving changes to the DNA sequence. Epigenetic alteration is a natural occurrence and essential when cells regulate biological phenomena such as cell differentiation, embryogenesis, genomic imprinting, and X-chromosome inactivation. Cellular differentiation and cancer are well-described examples where epigenetic changes occur (1, 2).

The main molecular mechanisms of epigenetics include DNA methylation, histone modifications, and non-coding RNAs. According to the early definitions, epigenetic changes were described as heritable through cell divisions (3), but according to that definition, this would strictly viewed only include DNA methylation. More recent definitions have progressively relaxed the heritability criteria, since the mitotic and meiotic stability of some histone modifications and non-coding RNAs remain unclear, yet still are epigenetic marks that influence how the genome is interpreted (4-6).

Epigenetic mechanisms direct gene programs and are affected by external and internal influences. In the event of disease, epigenetic deregulation can cause cells to behave and respond abnormally. In this context, environmental factors such as pollution, silica dust and smoking, can alter gene activity and thereby contribute to pathogenesis (7, 8). The studies included in this thesis focus on DNA methylation in health and disease states, which is further described in upcoming sections.

1.2 HISTONE MODIFICATIONS

Chromatin is a complex of DNA and nuclear proteins called histones (9, 10). Chromatin exists as open euchromatin accessible for transcription, and as highly condensed heterochromatin that is inaccessible and rarely transcribed. Briefly, DNA is wrapped tightly around eight histones, and the structure of chromatin and accessibility of DNA for gene transcription is largely controlled by these histone proteins (Figure 1). Histone-DNA interactions are influenced by post-translational modifications (PTM), including some that promote euchromatin by disturbing these interactions, and others that strengthen the interactions and promote heterochromatin. Histone modifications can thereby change the structure of chromatin and regulate gene activity (11).

1.3 DNA METHYLATION AND THE DEMETHYLATION PATHWAY

In mammals, DNA methylation refers to the covalent addition of a methyl group to the 5-position of cytosine (5mC) by DNA methyltransferases (DNMTs), preferentially in the context of a CpG dinucleotide. DNA methylation is among the most and best studied epigenetic

mechanisms and depending on the genomic location, it can correlate with both gene activation and repression (2).

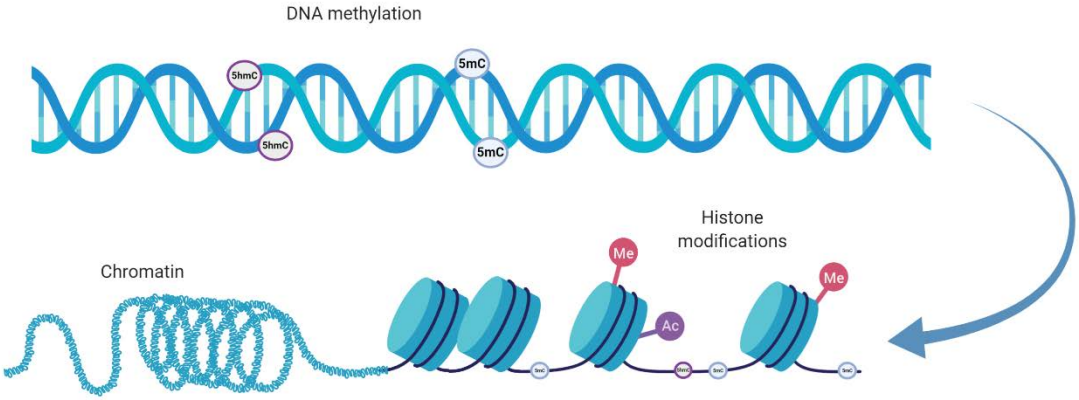

Figure 1: Epigenetic modifications of chromatin. Illustration of DNA methylation marks 5mC and 5hmC at CpG sites along the DNA strand, and histone modifications (methylation (Me) acetylation (Ac).

DNA demethylation can occur by two conceptually different pathways, passive and active demethylation (12). Passive DNA demethylation occurs in the context of DNA replication, when maintenance methyltransferases are inactive. This passive process results in unmethylated cytosine on the newly synthesized strand, and a dilution of the methylation signal. Active DNA demethylation is replication-independent and involves sequential oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) (13) and further, to 5-formyl- (5fC), and eventually to 5-carboxycytosine (5caC) (14, 15), by the ten eleven translocation (TET) family of enzymes (Figure 2). This is followed by subsequent thymine DNA glycosylase (TDG) action and base excision repair with reinsertion of an un-methylated cytosine. This reaction has recently attracted much attention, and may be positively or negatively associated with a cellular environment of oxidative stress, depending on context (16, 17), and may affect methylation states, e.g. in the lung (18).

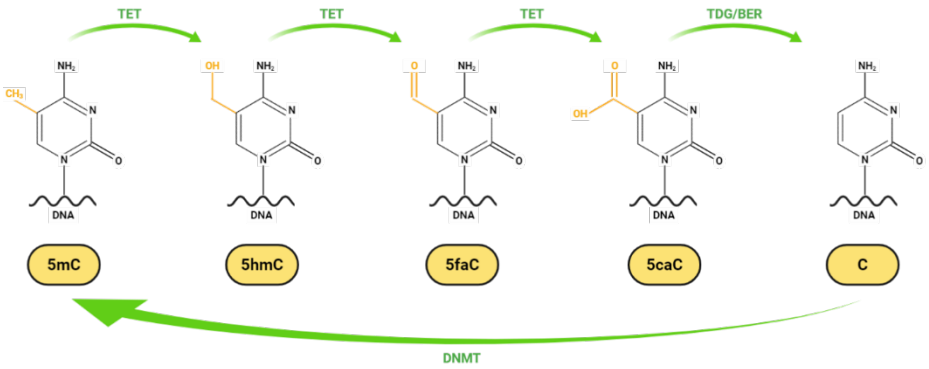

Figure 2: DNA demethylation pathway. Active DNA demethylation is replication-independent and involves sequential oxidation of 5mC by TET enzymes.

Hydroxymethylated cytosine (5hmC) is, however, not only an intermediary step of active demethylation, but has also been shown as a stable modification with environmental sensory and possible gene regulatory properties (19, 20), and indeed it functions as a recruitment mark in its own right for transcription factors and gene regulation (21, 22).

Previous studies have demonstrated a genome-wide presence of DNA methylation in a CpG context along gene bodies, other gene-related regions, as well as intergenic regions (1). While the majority of the dinucleotide CpGs are methylated, CpG islands (CGIs) are mainly unmethylated (1, 23). On average, CGIs are 1000 base pairs long and harbor an increased composition of C+G bases and show little CpG depletion, in contrast to other regions throughout the genome. Their location can be proximal to transcription start site (TSS), intragenically along gene bodies, or in intergenic regions between annotated genes (23). Of note, about 70% of annotated gene promoters have been linked to a CGI (24).

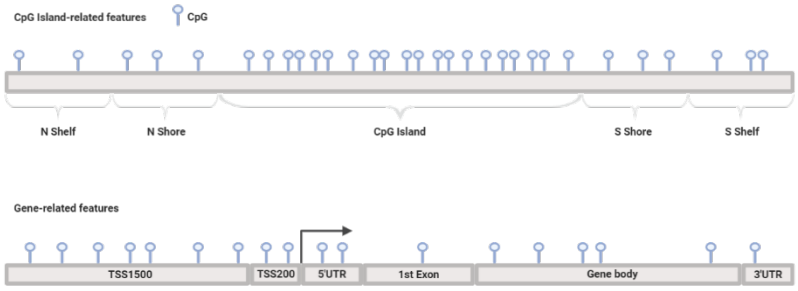

Figure 3: Features in relation to CpG islands and genes. CpG island-related features: Northern (N) and southern (S) shelves and shores surrounding a CpG island. Gene-related features including TSS1500, TSS200 (transcription start site), 5'UTR and 3'UTR (untranslated region), and 1st Exon and gene body.

Epigenetic modifications such as DNA methylation have received increased attention when studying pathogenesis of human disease, and analysis methods have advanced in both scale and resolution. In this way, the epigenome can be screened for biomarkers of e.g. disease risk or environmental exposure, and further investigated since epigenetic modifications may mediate environmental or genetic risk factors (25). Epigenetic modifications have been well described in cancer, but more recently, their role in common disease has also been investigated (26-30).

1.4 PULMONARY IMMUNOLOGY

The lung is a gas-exchange organ that provides all tissues of the human body with oxygen, while also removing carbon dioxide from the blood. It is organized into the upper and lower airways, with the upper airways located where the respiratory tract and digestive system cross. The lower airways (Figure 4) is a highly branched unit consisting of the trachea, and the bronchi dividing into bronchioles that eventually terminates as alveolar ducts. The lung epithelium and mucosal surface together with related immune cells serve as a physical barrier to the environment, and a first defense against airborne particles and pathogens. As a way to coordinate pulmonary innate immunity, epithelial cells direct a complex cross-talk between alveolar macrophages, T lymphocytes, and dendritic cells, which is required for a well-functioning lung (31).

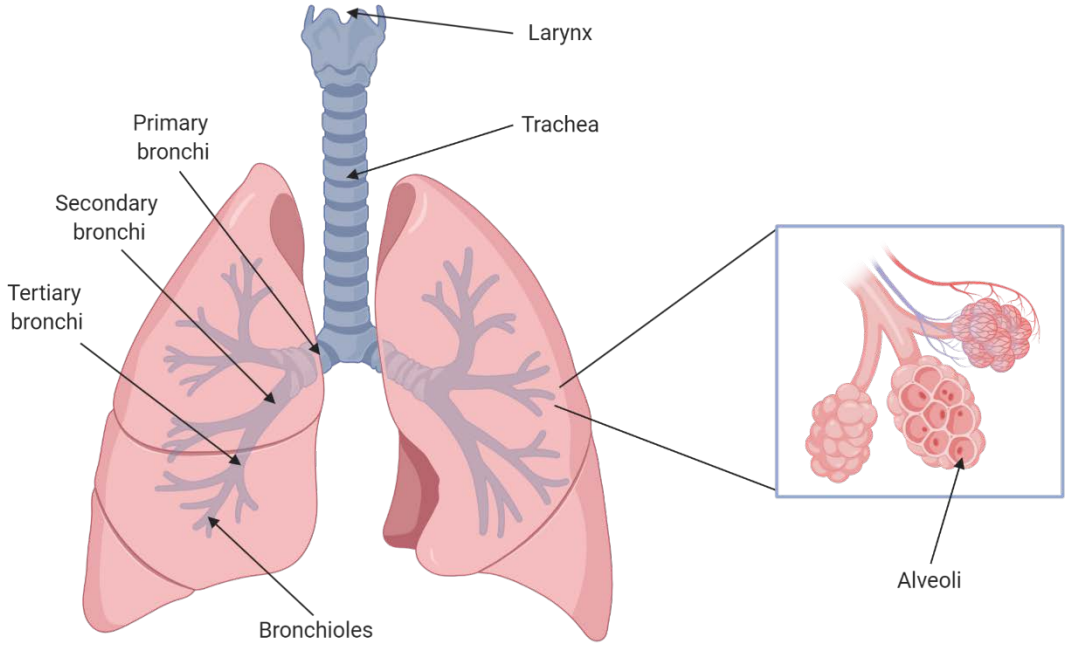

Figure 4: Lower respiratory tract. Illustration of the highly branched unit including the trachea, bronchi, and bronchioles that terminates as alveolar ducts

1.4.1 Alveolar Macrophages

Macrophages are a type of immune cells that exhibit plasticity by adapting to their surrounding microenvironment and in response to environmental cues. Their primary role is to ingest and kill invading pathogens, and to maintain healthy tissue by clearing dead host cells and debris (32, 33). Depending on their location and function, macrophages are divided into distinct subpopulations (32, 33). At least two populations of macrophages are recognized in the lung at homeostasis (34). The interstitial macrophages populate the interalveolar space where they e.g. present antigens, regulate dendritic cell maturation and migration, and aid in maintenance and tissue remodeling (34, 35). The other population is alveolar macrophages, which reside in the lumen of alveoli and are directly exposed to the environment. Alveolar macrophages are tissue-resident macrophages with phagocytic capacity, and responsible for clearance of debris, microbes, and inhaled substances such as particles in the lung (34). They are also important for structural integrity since they catabolize the pulmonary surfactant lining the alveolar surface, and thereby prevent alveolar collapse (36-38).

Most tissue-resident macrophages derive from embryonic precursors with different origins, namely the yolk-sac, bone marrow, and fetal liver (39, 40). A recent study of lung macrophages in mice describes three lineages of macrophages, all developmentally distinct and expressing different markers (38). The first lineage populates the peripheral and perivascular regions as primitive interstitial macrophages, and are derived from yolk sac hematopoietic cells. The

second population occupies the interstitium and become alveolar macrophages. The third lineage are derived from bone marrow, spread diffusely around in the lung parenchym, and become interstitial macrophages. The origin of alveolar macrophages remained elusive for a long time, but was recently described as cells that derive from fetal monocytes during lung development, and self-maintain as mature cells (41). However, it is still unclear how alveolar macrophages replenish during conditions disturbing homeostasis. When alveolar macrophages are depleted by e.g. lung transplantation, irradiation, or toxins, they may be replenished through progenitor cells from the circulation, possibly from one of the lineages (38, 42-45).

1.5 CIGARETTE SMOKING

Tobacco smoking with subsequent inhalation of harmful chemicals has long been known to impose serious health issues, yet it is still common and remains a global health challenge. The composition of cigarette smoke is complex and contains over 6000 chemicals, such as carbon monoxide, cadmium, nicotine, and oxidative agents (46). Many of them will reach the lower airways where they can influence the respiratory system and disturb lung homeostasis. Smoke exposure can thereby have a devastating impact on the body, both through local pro-inflammatory and oxidative responses in the lung, and by exerting systemic effects (7).

1.5.1 Smoking and Autoimmune Disease

Tobacco smoking or smoke exposure increases the risk of developing several diseases, such as chronic obstructive pulmonary disease (COPD), cardiovascular disease, asthma, and multiple types of cancer (47, 48). It is also an established risk factor for several autoimmune disorders, particularly RA and MS, with a recognized interaction between HLA genes and environmental factors (49-52). The leading hypothesis is that RA and MS have part of their etiological basis in the lungs and that risk and severity is exacerbated by smoking and exposure to other noxious airway exposures. The exact events and triggers involved are likely dependent on disease, but early inflammation due to smoke may be important during pathogenesis of many autoimmune diseases. Furthermore, the triggers likely vary from person to person, further adding to the complexity of autoimmune etiology. Smoking and the accompanying inflammation of the lung epithelium and subsequent production of autoantibodies represents the first step in a process that first may cause lung problems and subsequently severe disease in other organs such as joints (RA) and brain (MS).

1.5.2 Smoking and DNA methylation

It is well known that cigarette smoke has a devastating impact on health, and that mucosal tissues can contribute to autoimmune diseases. Smoking is associated with an increased risk of disease in both MS and RA, particularly in genetically susceptible individuals (51, 52). However, it is unclear how the epigenome of lung compartments is affected and relates to pathogenesis. Studying the epigenome is highly relevant, since it integrates signals from genes and the environment, and thereby shapes the transcriptome.

Epigenetic modifications may partly mediate effects of cigarette smoke through alterations in DNA methylation, which may induce transcriptional changes and contribute to smoking-associated disease. With the introduction of genome-wide DNA methylation arrays, smoking-associated signatures in blood cells from adults has been extensively studied (53-55). Cord blood and blood obtained from newborns whose mothers smoked during pregnancy also show pronounced smoking-associated effects (56, 57).

1.6 AUTOIMMUNE DISEASE

The immune system is a complex network of cells, tissues, and organs that function together to prevent infection. In autoimmune diseases, the body's own immune system mistakenly identify own antigens as foreign invaders and starts to attack self-tissue. In predisposed individuals, a reduced self-tolerance could more easily lead to this immune dysregulation and thereby initiation of autoimmune disease (58).

Pathogenesis of autoimmune diseases is a complex matter, and often involves an interaction between susceptibility genes and the environment (7, 8, 49-52). This gene-environment interaction can further be mediated by epigenetic modifications of the genome, and environmental factors can provoke altered epigenetic states. Alterations in DNA methylation patterns has been reported for the development of a range of prevalent autoimmune diseases, such as Sjögren's syndrome, systemic lupus erythematous (SLE), multiple sclerosis (MS) and rheumatoid arthritis (RA) (59-62).

1.7 MULTIPLE SCLEROSIS

1.7.1 Prevalence and Incidence

MS manifests as a chronic inflammatory disease affecting the central nervous system (CNS) leading to neurological defects and disability. A hallmark of MS is the destruction of myelin producing cells (oligodendrocytes) and demyelination of neurons in the central nervous system, leading to destabilization of the myelin sheath and subsequent axonal damage (63).

MS affects millions of people worldwide, but the distribution of affected individuals is asymmetrical, with high prevalence at northern latitudes such as northern Europe and America (64). In Sweden, the prevalence is 0.2% and females are affected more than twice as often as men (65).

1.7.2 Etiology

The cause of MS is still unknown, but several factors are known to contribute to MS susceptibility. It is likely that the disease develops through an interplay of several factors such as genes, environment and epigenome.

1.7.3 Genetic Risk Factors of MS

Genetic factors has been established in studies looking at familial aggregation, and later when detecting an increased risk of MS when comparing monozygotic twin concordant rates to

dizygotic twin concordant rates (66). Many immunologically relevant genes are associated with MS, the strongest being the *HLA-DRB1*1501* which increases the risk of developing disease (67), but also *HLA-A*0201*, a protective Class I allele that decreases the risk (68). In a large international study, over 100 non-HLA risk susceptibility variants were identified, further confirming the polygenic nature of MS pathobiology (69).

1.7.4 Environmental Risk Factors of MS

As mentioned earlier, environmental factors contribute to MS etiology. The geographical location is important but the reason behind this effect is unclear. Some location based theories include the 'hygiene hypothesis' and sun exposure, with a possible connection to vitamin D (70, 71). Other non-geographical environmental factors include early-life obesity, infection, and cigarette smoking (72-74).

1.7.5 Smoking and Citrullination in MS Pathogenesis

Cigarette smoking increases the risk of developing MS, and high cigarette usage cumulatively increases the risk further (74). Just as in RA, citrullination has been shown to play a role the pathogenesis of MS, and this is of interest in MS since smoking promotes citrullination and disequilibrium in the lung mucosa. Even though the mechanisms behind the smoking-associated impact in MS pathogenesis is unknown, inflammation and irritation from smoke exposure has been described as contributors (75). Whether citrullination induces auto-reactivity by activation of T cells is unclear, but could also be a consequence of the inflammatory response (76). The role of lung mucosa and associated cells in MS needs further exploration.

PAD4 is predominantly expressed in neutrophils and eosinophils, while PAD2 is expressed in a wide range of tissues (77). Both of these PAD enzymes are expressed in the brain, and PAD2 targets a constituent of the myelin sheath, the myelin basic protein (MBP). High levels of citrullination has previously been reported in white matter of MS patients compared to controls, and especially in areas of ongoing demyelination (78, 79). The MBP protein is known for its importance when inducing experimental autoimmune encephalomyelitis (EAE), an animal model for MS (80). Importantly, there is a growing body of evidence in the EAE model, that the lung is a priming site for immune cells prime before the infiltrate the CNS (81-84).

1.7.6 DNA Methylation and MS

Epigenetic modifications, such as DNA methylation, have been proposed to mediate interactions between genetic and environmental risk factors of autoimmune disease. In MS, it was recently demonstrated that DNA methylation is a potential mediator of genetic risk by *HLA-DRB1*15:01* variant in MS through changes in *HLA-DRB1* gene expression (85). In a recent study on the association between smoking and the risk of developing MS, smoking was shown to affect DNA methylation in peripheral blood cells, and that an exposure-response relationship exists (86).

1.8 RHEUMATOID ARTHRITIS

1.8.1 Prevalence and Incidence

Rheumatoid arthritis (RA) is a chronic autoimmune disease characterized by inflammation in synovial joints, leading to functional disabilities, and comorbidities such as cardiovascular disease and osteoporosis (87, 88). The prevalence is about 0.5-1% in developed countries and is 2-3 times as frequent among women (89).

1.8.2 Anti-Citrullinated Protein Antibodies

The disease can be divided into two subgroups, defined by the presence or absence of anti-citrullinated protein antibodies (ACPAs). These autoantibodies target peptides containing the amino-acid citrulline. ACPA is an important early classification biomarker and can be detected in approximately 70% of RA patients and has a very high specificity (94-98%) for the disease. Presence of ACPA is also associated with a worse prognosis, including a more aggressive disease course and more severe joint destruction (90). It was recently shown that ACPA induced activation of synovial fibroblasts and osteoclasts, suggesting an important role of ACPAs in RA pathogenesis (91, 92).

1.8.3 Etiology

Although the etiology of RA is unclear, studies demonstrate that ACPA can be present in sera many years before disease onset, before clinical joint manifestations (93). Triggering of immune reactions that lead to production of ACPAs is therefore suggested to originate distant from the joints, at mucosal surfaces such as gum, gastrointestinal tract and lungs (93).

1.8.4 Genetic Risk Factors of RA

Twin studies have shown that genetic factors play an important role in RA pathogenesis, contributing to about 40-60% of the risk (94, 95). The strongest genetic risk factor, HLA-DRB1, is a locus that codes for a region of the HLA class II molecules. There are several alleles of HLA-DRB1 (*DRB1*0401, *0404, *01*) and they all share a highly similar amino acid sequence, termed the “shared epitope”. This sequence codes for the peptide binding part of HLA II and is likely important for antigen presentation capabilities and relevant in RA pathogenesis (96).

1.8.5 ACPA-positive RA and Citrullination

Posttranslational modifications (PTMs) such as glycosylation, phosphorylation, citrullination, and carbamylation are common biological processes that have critical impact on the structure and functionality of a protein. As previously mentioned, ACPA target citrullinated epitopes. These epitopes are created through an enzymatic process called citrullination, where peptidylarginine deiminases (PAD) enzymes catalyze the conversion of an arginine residue into citrulline. This is a physiological process and occurs in various tissues in both health and disease (97). In RA, both smokers and ACPA-positive RA non-smokers have an increased presence of citrullinated proteins in the lungs (98). Since any inflammation can induce

citrullination, the citrullinated peptides themselves may not be enough to trigger an autoimmune response. However, the recognition and tolerance towards citrullinated peptides might be lost, either due to an increase in citrullination or through susceptibility genes, or possibly a combination of them both.

1.8.6 Environmental Exposure

One of the prevailing hypothesis when it comes to the pathogenesis of ACPA-positive RA, is that initial citrullination and ACPA-formation may occur in mucosal surfaces such as the lungs promoting systemic autoantibody production and subsequent arthritis development (99, 100). Environmental factors such as silica dust, cigarette smoke, and other irritants activate the innate immune system by local inflammation, and thereby induces the conversion of arginine to citrulline through PAD enzyme activity. Antigen-presenting cells can then present these citrullinated peptides by loading them onto MHC II molecules (101). Subsequent activation of auto-reactive T and B cells leads to the production of ACPAs by plasma cells, which further can enter the blood stream without resulting in any clinical symptoms (99, 102). Increased concentrations of ACPAs have been found in the lung compartment of individuals at risk of developing RA as well as untreated newly diagnosed RA patients (98). Further, inflammation and inducible bronchus-associated lymphoid tissue (iBALT) structures have been found in lung biopsies of untreated ACPA-positive individuals with early RA but no associated lung disease (99). The hypothesis of ACPAs importance in the pathogenesis of ACPA-positive RA is strengthened by the fact that the lungs and joints share common citrullinated targets (103, 104).

1.8.7 Smoking and PAD Enzyme Activity

As mentioned, tobacco smoking is strongly linked to risk of disease and to disease phenotype of ACPA-positive RA, further suggesting that the initial autoimmune responses could be initiated at a mucosal site, such as the lungs (105-107). In fact, more than 20% of all RA, and 33% of the more severe joint destructive disease, ACPA-positive form of RA is attributed to smoking (49). In addition, the gene-environment interaction is strong, with smoking and having the “shared epitope”, increasing the risk (21-f(108)old) of developing ACPA-positive RA (109). The pathogenesis appears to partly involve citrullination, either by smoking or other air pollutants.

PAD enzymes are expressed in many different tissues and includes five different isoforms, where PAD2 and PAD4 specifically has been associated to RA and other autoimmune diseases such as MS (77). In RA, PAD2 and PAD4 are the only isotypes expressed in the synovium of RA patients (110). In the lungs, healthy smokers also has an increased expression of PAD2. Smoking increases PAD2 enzyme expression and citrullination in BAL cells, but only PAD2 expression in bronchial biopsies (111). However, identical citrullinated peptides are found in synovial and bronchial tissues, further supporting a lung-joint connection in RA pathogenesis (112).

1.8.8 DNA Methylation in RA

Epigenome-wide analysis of ACPA-positive RA patients has shown that DNA methylation may mediate genetic risk for RA, and is located mainly in the MHC region (29). One of the identified CpGs was further reported to mediate an interaction between genotype and smoking status in the development of ACPA-positive RA (113). Also, DNA methylation clusters of variable methylated regions (VMRs) found inside MHC and other regions are associated with SNPs, and smoking-associated DNA methylation is 30-fold overrepresented in such clusters (114).

2 AIMS OF THE THESIS

The overall aim of my thesis was to characterize genome-wide DNA methylation patterns in health and autoimmune disease, more specifically in tissues associated with immunopathogenesis of Rheumatoid Arthritis (RA) and Multiple Sclerosis (MS).

The specific aims of the studies included were:

Study I: To investigate DNA methylation associated with ACPA and the pathogenesis of ACPA-positive RA while nullifying genetic factors.

Study II: To investigate the impact of cigarette smoke on the methylome and transcriptome on lung cells collected from bronchoalveolar lavage (BAL).

Study III: To acquire knowledge for determining reactions associated with inflammation caused by smoke on lung cells and find specific effects in the lung tissue from patients with autoimmune disease.

The expected outcomes are to acquire knowledge for determining reactions associated with inflammation caused by smoke on lung cells and find specific effects in the lung tissue from patients with autoimmune disease.

3 MATERIAL AND METHODS

For a more detailed description of methodologies, see the Material and Methods section for each individual study.

3.1 STUDY COHORTS AND SAMPLE COLLECTION

Each individual gave their written consent for study participation, and all studies were approved by the Regional Ethical Review Board in Stockholm, Sweden.

Two different cohorts were included in this thesis. In **Study I**, we included 24 twin pairs belonging to a population-based twin cohort (Twingene), which is part of the Swedish Twin Registry (115, 116). We obtained DNA from 5 healthy MZ twin pairs discordant for the presence of ACPA, and 7 MZ twin pairs discordant for ACPA-positive RA. Replication by pyrosequencing was performed on an additional 6 MZ twin pairs discordant for ACPA presence, and 6 MZ twin pairs discordant for ACPA-positive RA. ACPA presence was determined using CCP2 ELISA assays (Immunoscan CCPlus) according to manufacturer's instructions. Twins participating in **Study I** donated blood at the clinic, and sera and whole blood were transported to Karolinska University Laboratory and then to KI Biobank.

In **Study II** and **Study III**, we obtained BAL samples collected during bronchoscopy of healthy individuals and MS patients, as previously described (117). We only included individuals with no clinically relevant airway infection, and only those without asthma, lung diseases, COPD, or other inflammatory conditions. Individuals smoking >5 pack years (pack years = (cigarettes smoked per day / 20) x years of smoking), or at least 5 cigarettes/day, were defined as smokers. The non-smoking groups included individuals who had never smoked, or quit smoking >15 years ago (one subject in **Study III**). All subjects have been included in a previously published study (118).

3.2 DNA AND RNA EXTRACTION

In **Study I**, extraction of DNA was performed using Puregene extraction kit (Gentra Systems), and extracted DNA was stored at -20°C (KI Biobank). Aliquots of sera was stored at -180 °C (KI Biobank).

In **Study II** and **Study III**, BAL samples were centrifuged (400×g at 4°C, 10 min) in order to separate BAL cells from fluid. BAL fluid was immediately aliquoted and stored at -80°C, and cells were collected and resuspended in RPMI 1640 (no supplements). Extraction of DNA and RNA was performed with AllPrep DNA/RNA/miRNA Universal Kit (Qiagen) according to the manufacturer's protocol.

3.3 EPIGENOME-WIDE ARRAYS

A range of techniques have been developed for the purpose of investigating DNA methylation. In this thesis, two different array-based technologies were used for determining genome-wide DNA methylation as described in the following sections.

3.3.1 DNA Preparation and CHARM

For **Study I**, DNA methylation was profiled using the "comprehensive high-throughput arrays for relative methylation" (CHARM) technology, covering 2.1 million CpG sites (119). First, DNA was prepared by shearing, McrBC-digestion, and gel separation before being labelled and hybridized to arrays according to protocol (120). ACPA-positive healthy discordant twins and ACPA-positive RA discordant twins were processed separately in two batches. All samples within each batch were processed together. Due to apparent batch effects, the two twin sets were not statistically compared.

3.3.2 Bisulfite Treatment and EPIC

In **Study II** and **Study III** we aimed to investigate 5mC and 5hmC as separate DNA modifications. Since conventional bisulfite treatment converts both 5mC and 5hmC, the readout will be a combination of both DNA modifications. DNA extracted from BAL cells was processed using the TrueMethyl conversion kit (Cambridge Epigenetix). This workflow allows investigation of conventional bisulfite-treated samples (BS, 5mC+5hmC), oxidative bisulfite-treated samples (oxBS, 5mC), and the difference between them (5hmC).

DNA methylome in **Study II** and **Study III** was profiled using the Infinium HumanMethylationEPIC BeadChip Kit (Illumina), which covers over 850,000 CpG sites of the human genome. The samples were sent to the National Genomics Infrastructure (NGI), Science for Life Laboratory at Uppsala University, where they were further processed. In order to reduce bias towards any group, samples were randomized according to age, gender, smoking status, and macrophage content. IDAT format files containing the raw intensities were used for subsequent array analysis.

3.4 BIOINFORMATIC APPROACHES

We performed bioinformatics analyses in all studies included in this thesis. The following sections briefly describe some of the methodologies used.

3.4.1 Pre-processing and Analyses of CHARM Methylation Signals

In **Study I**, we used the CHARM platform to study the DNA methylation pattern of two phases of disease progression in ACPA-positive RA.

3.4.1.1 Pre-processing

In order to access the performance of the CHARM arrays (**Study I**), we considered numerous quality controls according to CHARM Bioconductor package (121) including: the signal of background probes; the standard deviation of untreated channel signals; the difference between the medians of control and non-control probes; the probe quality for inclusion/exclusion of samples. Preprocessing and subsequent analyses of each twin set was performed separately due to the occurrence of batch effects. Sub-quantile normalization was used to normalize between samples.

3.4.1.2 Differential Methylation at Single Probes (DMPs) and Across Regions (DMR)

A linear model, dmrFind from the CHARM Bioconductor package (121), was used for analysis of both differentially methylated probes (DMPs) and differentially methylated regions (DMRs). No individual probe was significantly differentially methylated at false discovery rate (FDR, Benjamini-Hochberg approach (122)) <0.2. Family-wise error rate (FWER) was computed for each region of differentially methylated consecutive probes (DMRs).

For the cell type-specific profiles of DNA methylation we isolated CD4+ and CD8+ T lymphocytes, CD56+ NK cells, and neutrophils from peripheral blood from five healthy male individuals as previously described (123), followed by CHARM analysis as mentioned above. To estimate cell proportion in the MZ twin samples, we utilized a methodology for cell deconvolution originally developed for Infinium 450K array (124), and adapted it to CHARM arrays using cell type-specific methylation profiles.

As a way to check if DMRs from ACPA-positive healthy discordant MZ twins were also candidate DMRs in ACPA-RA discordant MZ twins, we performed projection analysis by computing permuted p-values (125) using test statistics from the DMR finding.

3.4.2 Pre-processing and Analyses of EPIC Methylation Signals

In **Study II** and **Study III**, we generated DNA methylation datasets using the Infinium HumanMethylationEPIC BeadChip arrays (Illumina), which covers approximately 4% of all CpG dinucleotides of the human genome. In **Study II**, we investigated smoking-associated patterns of true 5mC and 5hmC in BAL cells from healthy individuals. In **Study III** we deciphered MS-associated changes to BAL cells in smokers and non-smokers, also by studying both 5mC and 5hmC.

3.4.2.1 Pre-processing

Raw data were imported as IDAT files, then processed using minfi (126, 127) and ChAMP (128) packages in the R environment. BS and oxBS-treated samples from the same individual were processed together and run on the same array. Normalization was done using stratified quantile normalization (SQN) from the minfi package, a normalization strategy that stratifies probes by region (e.g. CpG island, shores).

3.4.2.2 Differential Methylation at Single Probes (DMPs) and across Regions (DMRs)

DMP analyses on combined 5mC+5hmC and 5mC were performed on M-values transformed from β-values ($M = \log2(\beta/(1-\beta))$ as recommended by Du et al. (129). Linear modelling (limma) with empirical Bayes was used to compute test statistics, with non-smokers as reference group and adjustment for covariates. A DMP was considered significantly differentially methylated with a p-value < 0.05 after multiple testing correction (FDR, Benjamini-Hochberg). In **Study II**, an additional methylation difference cutoff (15%) was used to investigate differences with high effect size. Before performing 5hmC DMP analysis (**Study II** and **Study III)**, normalized oxBS β-values (5mC) were first subtracted from normalized

5mC+5hmC (BS methyl) β-values to calculate the hydroxymethyl level (Δβ) at each CpG. As above, limma with empirical Bayes was used when calculating test statistics.

In **Study II**, we performed DMR analysis in addition to the DMP analysis by applying the DMRcate package (130) to profile differences between smokers and non-smokers. A linear model of methylation values was fitted at each probe and adjusted for covariates (sex, age). DMRs were considered significant with Stouffer-transformed p-values < 0.05 and mean absolute $\Delta\beta > 15\%$.

3.5 VALIDATION AND REPLICATION BY BISULFITE PYROSEQUENCING

Primers for **Study I** and **Study II** were designed using the PyroMark Assay Design 2.0 software and optimized for best annealing temperature. PCR amplification was performed using the PyroMark PCR kit (Qiagen). The resulting PCR product was used with streptavidin Sepharose high performance beads (GE Healthcare), sequencing primers, and PyroMark Gold Q96 reagent kit (Qiagen), and sequenced on a PSQ 96 system (Qiagen). Signal peaks were analyzed using the PyroMark Q96 software.

Bisulfite pyrosequencing was performed as both technical validation and replication in **Study I**. Technical validation was performed on the same samples used in the CHARM analyses, and the percentage of agreement between bisulfite pyrosequencing and CHARM results were computed. Next, a replication analysis was implemented by profiling a new set of individuals. Additionally, a meta-analysis was performed including bisulfite pyrosequencing data from both (unpaired) technical verification and the replication cohort. Two different methods were used for the meta-analysis and yielded similar results. One was based on p-values ("summation of p-value"-method (131)), and the other on effect size (132).

In **Study II**, oxidative and conventional non-oxidative bisulfite pyrosequencing was performed to validate 5mC and 5hmC results from Infinium HumanMethylationEPIC BeadChip arrays (Illumina).

3.6 TRANSCRIPTION FACTOR ANALYSIS

In **Study II** we performed transcription factor analysis using eFORGE (experimentally derived Functional element Overlap analysis of ReGions from EWAS) (133) as a way to identify cell-type specific signals and to investigate transcription factor motif associations. As input, we used the top 1000 significant BS-DMPs and 5mC-DMPs, and all 67 significant 5hmC-DMPs. Each dataset was examined with enrichment analysis of DNase I hypersensitive sites (DHSs) and histone marks (H3K27me3, H3K36me3, H3K4me3, H3K9me3, and H3K4me1). Transcription factor motifs were linked back to a transcription factor gene list and further analyzed using PANTHER pathway analysis (134).

3.7 RNA SEQUENCING

Differential expression was investigated in both **Study II** and **Study III**. Briefly, RNA sequencing reads were quality filtered and trimmed for adapters using TrimGalore. The

Kallisto algorithm was used in order to pseudoalign the reads to a reference transcriptome (GENCODE v24 comprehensive transcript). Only samples with RIN value > 7 and genes with normalized read count > 10 were included for downstream analysis. Transcriptional differences were calculated using the DESeq2 package in the R environment, and adjusted for covariates sex and age. In **Study II**, genes with BH-adjusted (FDR) p-value < 0.05 and absolute log2 fold change > 1 were considered significant. In **Study III**, genes with a nominal (unadjusted) p-value < 0.05 were considered suitable for gene ontology (GO) analysis.

3.8 NETWORK ANALYSIS

To gain insights into the biological relevance of differential methylation and expression in **Study II** and **Study III**, we performed gene ontology (GO) analyses using Ingenuity Pathway Analysis (IPA, Qiagen) and overrepresentation analysis (ORA, www.webgestalt.org) (135). IPA and ORA analyses were applied using unbiased parameters for all criteria. Right-tailed Fisher's exact test was used to calculate p-values, and a p-value < 0.05 was considered statistically significant. The REVIGO tool (136) was used when visualizing GO data. STRING database was used to generate networks with a minimum level of confidence > 0.4.

3.9 STATISTICAL METHODS

This section covers methods used for statistical analyses in **Study I-III**. All analyses were performed in the R environment unless stated otherwise.

3.9.1 P-values

A p-value is the observed significance level of data under the condition that the null hypothesis (H_0) is true, representing the probability of getting the result or more extreme than your observed result. For example, a p-value of 0.05 would mean that there is a 5% chance of getting your observed result. If the p-value is below a set threshold, H_0 is rejected in favor of the alternative hypothesis (H_1). The significance level was set to <0.05 (*p < 0.05, **p < 0.01, ***p < 0.001) if not stated otherwise.

3.9.2 Wilcoxon Rank Sum Test

Mann-Whitney U/Wilcoxon rank sum test is a nonparametric test to compare two unmatched groups. This statistical method was used in **Study II** when comparing distribution profiles related to genomic features and gene locations between smokers and non-smokers.

3.9.3 Multiple Testing

When performing multiple statistical tests, some will have p-values < 0.05 due to chance, even though all of the null hypotheses are true. This means that a fraction of the results yielded after multiple testing will be false positives. The goal when correcting for multiple comparisons is to reduce the number of false positives, but it also comes at the cost of an increased number of false negatives.

In **Study I-III**, we used the Benjamini-Hochberg (BH) procedure to control the false discovery rate (FDR) of array and sequencing data. When the BH-adjusted p-value was smaller than the FDR, the test was considered significant. Multiple testing correction of DMRs in **Study I** was performed through bumphunter (137), which controls the family-wise error rate (FWER). The stricter Bonferroni correction, which also controls the FWER, was used when correcting for multiple testing in distribution analysis of β-values and in enrichment/depletion analysis.

3.9.4 Linear Models

In our analyses, we wanted to discover which features (CpGs or transcripts) were different between two groups. For this purpose we used the R package limma, deploying linear regression with the aim to model the relationship between a dependent variable and one or more independent variables.

3.9.5 Wilcoxon Signed-Rank Test

When analyzing feature-specific distribution analysis of β-values in **Study II**, we used Wilcoxon signed-rank test with multiple testing correction by the Bonferroni approach.

3.9.6 Pearson and Spearman Correlation Coefficients

Correlation analysis in **Study II** was calculated using Pearson correlation coefficient for normally distributed data, and Spearman correlation coefficient when not normally distributed.

3.9.7 Chi-square

In **Study II**, a Pearson's Chi-squared test was performed on contingency tables of count data, and adjusted for multiple comparison using Bonferroni.

4 RESULTS AND DISCUSSION

The studies included in this thesis cover genome-wide methylome and transcriptome analyses in tissues involved in Rheumatoid Arthritis and Multiple Sclerosis immunopathogenesis, as well as smoking-associated signatures. This section includes a short summary of the results and discussion from each study included in the thesis. The results of the individual studies are discussed more extensively in the respective paper or manuscript.

4.1 STUDY I

In this study, we explored the genome-wide DNA methylation using the CHARM platform, and describe differentially methylated regions (DMRs) of two groups of monozygotic twin pairs (Study I; Table 1). One set of twin pairs (5 pairs) were discordant for a generally presumed pre-stage of RA, i.e. healthy with high titer of anti-citrullinated peptide antibodies (ACPAs) vs. low ACPA titer, and the second group (7 pairs) was discordant for ACPA-positive RA vs. ACPA-negative healthy individuals. Raised ACPA titer, which indicates an abnormal immunity reaction taking place, can exist long before any clinical symptoms of RA are evident. The two discordant groups may thus serve as an epigenetic disease trajectory model representing different phases of disease development, without involving genetic confounding effects.

We implemented a statistical method aimed at investigating genome-wide DNA methylation in a small number of samples, and by employing paired analyses of the genetically identical discordant MZ twins, we were able to neutralize any genetic influence while profiling epigenetic patterns. Since DNA was extracted from whole blood, we also applied a mathematical cell deconvolution algorithm adapted to CHARM that adjust for epigenetic alterations related to differences in cell type proportions. These DNA methylation patterns were only available after additional profiling DNA methylation patterns of physically sorted cell types on the CHARM platform.

Analysis of the CHARM array data was conducted in two steps. First, the analysis was performed without including cell proportion profiles in the statistical model. The following results reflect changes from both the phenotype of interest, and differences in cell proportion. Since these results are not affected by possible errors introduced from statistical corrections, they were also used for technical validation.

We identified 17 genome-wide significant DMRs (FWER <0.2) in the ACPA discordant twin pairs (pre-RA stage) after adjusting for cell type proportions. In the ACPA-positive RA discordant group, we found 36 DMRs where only one remained after cell type correction, suggesting that most differential methylation identified in whole blood from ACPA-positive RA patients are mainly due to altered cell distributions. Indeed, cell-specific analysis of ACPA-positive RA vs. their ACPA-negative healthy siblings, identified neutrophils and CD4+ cells as the main drivers of the DMRs. A DMR identified in ACPA-positive healthy individuals remained after projection analysis to the discordant ACPA-positive RA twin pair set, suggesting that the DMR may be involved with onset of ACPA-positive RA. The DMR is

related to the protocadherin beta gene cluster, including important cell surface recognition factors (138) and may be of interest for hypothesis generation and further research.

Notably, none of the DMRs overlapped with previously known genetically associated genes within the MHC region, suggesting a successful nullifying of genetic factors in our MZ twin study design.

Besides the clinical relevance of the study, we proposed a novel and robust methodological framework adapted to identify changes in DNA methylation with high-specificity, reducing the number of false positives and problems of low sensitivity. In conclusion, our biostatistical methodology optimized for a low-sample twin design revealed differential methylation in non-genetically linked genes associated with two distinct phases of RA development.

4.2 STUDY II

Tobacco smoking is a major health problem known to associate with development of multiple diseases including cancer and autoimmune disease. Smoke exposure can alter DNA integrity, leading to altered gene expression, partly through changes to the DNA methylome. In this study we analyzed smoking-associated changes in 5-methylcytosine (5mC) and its oxidized form 5-hydroxymethylcytosine (5hmC) in alveolar macrophage-dense BAL cells from 14 smokers and 21 non-smokers (Study II; Table 1).

We profiled both the methylome and hydroxymethylome of BAL cells by using both conventional bisulfite-treated (BS) and oxidized bisulfite-treated samples in combination with the latest Illumina EPIC BeadChip. A large majority of identified smoking-associated BS (5mC+5hmC) methyl (1,639/1,667), 5mC (1,738/1,756), and 5hmC (67/67) DMPs have not previously been reported in alveolar macrophages (or BAL cells), neither in 27K (139) nor 450K BeadChip (140). When compared to previously published results from 450K BeadChip analyses on alveolar macrophages, we could confirm 60% (18/30) of reported DMPs (140).

Smoking-associated BS methyl DMPs were predominantly hypomethylated, a signature that was even more striking among 5mC DMPs. In contrast, the vast majority of 5hmC DMPs were hypermethylated, and thereby opposing the hypomethylated 5mC DMP signature. These findings support the hypothesis of a DNA demethylation process initiated by smoking-induced oxidative stress, with 5hmC as the first step of a sequential oxidation. Interestingly, oxidative stress can lead to 8-oxoG lesions on DNA (141), and it is tempting to speculate that smoking-induced oxidative stress also introduce the discovered lesions. In this context, the OGG1 protein which has affinity for the oxidized modification of guanine, is in turn required for TET1 binding and initiation of the cytosine DNA demethylation pathway (142). This OGG1-TET1 complex could thereby lead to hypomethylation of CpG sites in smokers.

Notably, a strong enrichment of enhancer sites was identified among our DMPs (Study II: Fig. 2d), especially in hypomethylated BS-DMPs and 5mC (from 3.3% of EPIC background probes to 18.1% and 20.1% respectively). What this enrichment means functionally remains to be investigated, but the enhancer landscape of tissue macrophages are dependent on the

surrounding microenvironment (143), and we show that DNA methylation at enhancer sites in alveolar macrophages is seemingly affected by cigarette smoke. Further, we used RNA-seq to investigate smoking-associated transcriptional changes in the BAL cells. We revealed a negative correlation between CpG site-specific DNA methylation and gene expression at both promoter regions and enhancer sites (Study II; Fig. 4a, Supplementary Table 10). It is already established that interaction between enhancers and promoters plays an essential role in regulating gene expression (144), but here we not only report that enhancer-associated DMPs are enriched in smokers, but also that they negatively correlate with gene expression. This substantially increases our knowledge of gene regulation by inflammatory processes in the lung.

Pathway analysis revealed a strong overlap of biological processes (GO terms) in differentially expressed genes and DMP-annotated genes that converge to immune-related processes such as cell adhesion, migration, and leukocyte recruitment (Study II; Fig. 6a-b). The migratory profile may explain the increased amount of cells found in BAL of smokers.

In conclusion, we report novel epigenetic biomarkers, relevant disease-associated genes and biological processes related to smoking, and thereby increase and refine the knowledge of the molecular mechanisms underlying the effect of smoking. Our findings suggest that tobacco smoking modifies the epigenetic landscape of alveolar macrophages (BAL cells), possibly involving a continuous active demethylation and subsequent increased activity of immune-related processes in the lungs.

4.3 STUDY III

The work in Study III is an extension of Study II, providing an investigation of molecular changes occurring in pulmonary immune cells from MS patients. For this purpose, we profiled the methylome of BAL cells from 17 female MS patients (8 non-smokers, 9 smokers) and 22 healthy individuals (HC; 12 non-smokers, 10 smokers) using the Infinium EPIC BeadChip. In addition, we investigated the transcriptomic profile of MS patients compared to healthy individuals.

In order to statistically determine differential methylation and hydroxymethylation in BAL cells, we used a linear model with eBayes and adjusted for covarites (age, sex, macrophage content). The most noticeable changes were detected in relation to smoking, both in MS and HC. DMP analysis in MS patients revealed 1376 BS, 131 5mC, and four 5hmC-DMPs (BH-$P_{adj} < 0.05$) between smokers and non-smokers (Study III; Fig. 1, Table 2, Supplementary Table 2). In healthy individuals, we identified 1821 BS, 234 5mC, and one 5hmC-DMPs smoking-associated changes. No DMPs were significantly associated with MS compared to healthy individuals, neither in smokers nor non-smokers. Interestingly, only a fraction (~1/3) of the smoking-associated DMPs found in MS were also differentially methylated in healthy individuals. This difference may partly be due to our cohort characteristics (e.g. sample size), but may also represent distinct smoking-associated profiles between MS patients and healthy individuals.

To gain insight into the shared and distinct smoking-associated methylome profiles in MS patients and healthy individuals, we performed GO analyses on smoking-associated genes in MS (827 genes) and in healthy individuals (1036 genes). Enrichment analysis for *Biological functions and Diseases* revealed that the MS and HC profiles share most categories, including the top significant terms cytoskeleton rearrangement, immune cell trafficking, and cellular movement (Study III; Fig. 3a). Likewise, analysis of smoking-associated enrichment for *Canonical Pathways* in MS and HC revealed many shared processes, including integrin and actin cytoskeleton signaling (Study III; Fig. 3b). Further exploration of pathways that were specifically enriched in MS after smoking revealed a distinct neuronal signature, including synaptogenesis signaling as the most enriched process (Study III; Fig. 3c). This profile may appear strange for immune cells residing in the bronchoalveolar space, but might be highly relevant in other tissues. Systemic smoking-induced alterations in the methylome could potentially have functional consequences in specialized cells in other compartments. However, whether this smoking-associated neuronal signature is reflected in other tissue macrophages, blood monocytes, or other cells and tissues should be further explored. Notably, some of the BS-DMPs from our cohort overlapped with the smoking-associated profile in whole blood from MS patients (86), but with increased effect size in our study for the majority (9/13) of candidate BS-DMPs (unadjusted p-value < 0.05). Due to the limited sample size of our cohort we were unable to investigate whether MS disease modifies the effect of smoking in BAL cells by interacting with smoking load, as reported for whole blood (86).

Motivated by the distinct smoking-associated profile of MS in BAL cells, and with accumulating evidence that the lung is involved in MS immunopathogenesis, we profiled the methylome and transcriptome of MS patients compared to healthy individuals. Neither DMPs nor transcripts passed significant threshold after correction for multiple testing. This could reflect a profile comprising only minor biological differences between MS patients and healthy individuals, but may also be a result of the undeniable lack of power due to our relatively small and heterogeneous cohort. However, subtle differences between MS and healthy controls (unadjusted p-value < 0.001) were detected in both the smoking and non-smoking group, and converged to cytoskeleton dynamics and cellular mobility, transcription, and neuronal processes (Study III, Fig. 4a). These changes support studies in EAE animals that demonstrate motility of circulating immune cells homing into lung tissue (81-84). Thus, our study may provide novel insights into MS pathogenesis, and support the hypothesis of a relationship between the CNS and inflammatory processes in the lung.

5 CONCLUSIONS AND FUTURE PERSPECTIVES

The underlying theme in this thesis has been genome-wide investigation of DNA methylation patterns in patients with autoimmune disease and healthy individuals. We have utilized two different genome-wide methods to measure DNA methylation. The main conclusions are presented in the following sections.

In **Study I**, we adapted a statistical framework to empower low-sample twin design and present non-genetically linked genes of relevance for development of ACPA-positive RA. These results should be of interest for further studies of epigenetic mechanisms influencing disease progression of RA. Our optimized statistical framework may also prove useful in twin studies. Since both RA (and MS) pathogenesis have genetic components, twin studies are important to nullify the confounding effect of different gene setups. Our study included a limited number of monozygotic twins, but with the adapted statistical framework, we aimed to minimize the number of false positives.

In **Study II**, we report smoking-associated methylome and transcriptome signatures in alveolar macrophage-dense BAL cells from otherwise healthy patients. We thereby provide insights into the local impact of smoking-associated effects in the lung, which may be reflected systematically through peripheral immune functions. These findings may also be of relevance in the immune events leading up to disease. We reveal many new smoking-associated methylation sites that especially map to regions not covered be previous epigenome-wide methodologies, but could also confirm previously reported differential methylation in alveolar macrophages. In addition, we present pathways that are affected and altered by smoke, along with associated gene activities.

It would be interesting to further explore the DNA demethylation pathway in the context of cigarette smoking and oxidative stress, since we found an opposing effect of 5mC and 5hmC. The DNA demethylation process may be induced by oxidative stress caused by cigarette smoking, and the interactions and mechanisms involved may be relevant to investigate using *in vitro* models.

The etiology and immunopathogenesis of MS remains elusive but involves multiple factors that increase the risk of developing disease. Individuals with genetic predisposition, and that are exposed to triggering environmental factors such as cigarette smoke, may initiate a cascade of immune system events that lead to demyelination of nerve cells. In **Study III**, we demonstrate that BAL cells from the lungs of MS patients display molecular patterns that are distinct to those from healthy individuals, both in smokers and non-smokers. Our findings support the hypothesis that there is a bidirectional autoimmune link between the lungs and CNS. These findings may contribute to the complex picture of MS disease pathogenesis, and highlight the importance of lifestyle factors in prevention and treatment of MS. It would be of interest to stratify MS patients into *HLA-DRB1*15* carriers and non-carriers, since the combination of smoking and having the *HLA-DRB1*15* allele is one of the major risk factors for MS. A similar study of BAL cells from smoking and non-smoking RA patients would be

interesting, especially when stratified for ACPA-positivity, since ACPA-positivity increases the risk of developing RA.

Overall, it would be valuable to expand the studies in this thesis. Collecting the samples from healthy individuals for Study II, and especially MS patients in Study III, was a tremendous task since BAL is usually only performed to diagnose lung disease. This would have allowed us to detect smaller differences in effect size, such as 5hmC patterns or methylation differences between non-smoking MS and healthy individuals.

For the future, it would be interesting to characterize BAL cell populations and investigate if there are in fact differences in surface expression molecules, ultimately representing different cell populations of macrophages. Whether this is a smoking signature of resident alveolar macrophages, or reflects the influx of another macrophage population with a different methylation profile remains to be investigated. Another way to further strengthen our findings would be to use sorted cell populations for genome-wide analysis, in addition to total BAL cells, but this would ultimately mean multiplying the number of processed samples many times and a steep increase in cost.

Studies of epigenetics make it possible to better understand how environment and life style influence disease, and indeed are interacting with the genome. While we have identified genes and pathways associated with inflammation induced by cigarette smoking, it would be out of interest to study functional aspects of specific genes identified in our studies. Specifically, mechanisms involving environmental-induced changes of the functional genome, thus act through epigenetic modifications determining the functional state of a gene, or genetic element. We have already adopted an organotypic 3D model of lung tissue and implemented alveolar macrophages from healthy patients to further investigate the response to cigarette smoke extract *in vitro*.

DNA methylation is partially recognized for its association to many diseases, but its functional role in pathogenesis still needs to be clarified in most cases. It would be interesting to expand our analyses by simultaneous studies of other epigenetic marks, i.e. histone modifications that can affect DNA methylation (145). Single cell techniques would also add additional layers of information to the studied cell populations both in health and disease.

To summarize, our findings add interesting and valuable information about RA immunopathogenesis and MS disease, in addition to novel insights of smoking-associated impact on the lung. The leading hypothesis is that the autoimmune disease MS (and also RA) has part of its etiological basis in the lungs and that risk and severity is exacerbated by smoking. In the future we would like to acquire in-depth knowledge about reactions associated with inflammation caused by smoke on lung cells and find specific effects in the lung tissue from patients with autoimmune disease. We would like to investigate and find ways to interfere with epigenetic regulation of key players using therapeutic approaches, that could be used locally (e.g. by inhalation), targeting the epigenome and specific disease implicated

genes. New knowledge may also be used to accurately communicate the public of risks and desirable lifestyles for preventing the appearance of the disease in the first place.

6 ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all people involved in the process of making this thesis. Thank you for all your support, assistance, and hard work. It has been an exciting journey filled with high points and low points, but in the end it was a great experience!

In particular I would like to acknowledge all patients and healthy volunteers that contributed to the studies, your participation made this research possible.

To my main supervisor **Tomas Ekström**, for giving me the opportunity to start in your lab, for letting me develop along with the projects, and for believing in me after countless of hours learning about bioinformatics. Thank you for sharing your passion, it is truly inspiring.

To my co-supervisor **Malin Almgren**, for all your support and encouragement, and for introducing me to the methodologies when I was new to the lab. Your enthusiasm has helped me through setbacks at the lab, and to move on and focus on the next step.

To my co-supervisor **Heidi Vähämaa**, for your endless positivity and for being encouraging and helping out whenever needed. I have always looked forward to our meetings and lunches.

To my co-supervisor **Anca Catrina**, for being welcoming, for your valuable guidance and good support whenever needed, and for helping me to refocus and with getting back on the right track.

To my co-supervisor **Yun Liu**, for introducing me to the field of bioinformatics, it certainly was an eye-opener. Also, thank you for a fantastic research visit to Shanghai, and say hi to everyone at the lab!

I would also like to thank my mentor **Catharina Lavebratt** for all advice given during my thesis, they have been very helpful.

To all my fantastic colleagues at floor L8:00 for creating a nice atmosphere with lots of laughter and interesting discussions.

To former colleagues of the Ekström group, for interesting discussions on epigenetics and every other topic we could think of. **Lollo**, you are real joy-spreader and having you in the office made it easy to stay positive, and thanks for all the delicious desserts. **Joëlle**, thanks for all the nice chats, for being including and for all your fun parties. **Oskar**, it was great having you at the lab and you always had a topic to discuss. To **Atosa**, thanks and hope everything is great over in the US.

To **Vincent**, it has been great having you around. We have been following each other from the first few months of our PhD journey, helped each other out whenever needed, whether it was moving or talking, and thanks for all your statistical input. I will come and visit in Vienna!

To **Sahl**, you are such as kind and inclusive person. It has been great to see you finishing your thesis, and you have been really helpful when I have come to you with questions. Thanks for being good company during our countless visits to various lunch restaurants.

To **Parvin**, for being a constant source of information, especially all the random facts that came out of nowhere. And I think it is time for the moving-in party.

To KIgene, **Annika**, **Katarina**, **Malin**, for being helpful and a nice company during lunch breaks or at fika times. Also, I would like to thank **Anna-Lee**, and good luck with your new adventures (soon it will be my turn to reach the top of Kili).

To current and past members of the Neurogenetics group, **Liu**, **Miranda**, **Jakob**, **Iris**, **Binbin**, **Jiajia**, **Paschalis**, **Paulina**, **Laura**, and **Elisabeta**. Thanks for all nice chats in the lunch room or outside in the grass, it has been great having you around.

To current and previous members of the hematology group, **Yujiao**, **Jingya**, **Chuanyou**, **Xiaotian**, **Bignan**, and **Jenny**.

To members of the Rheumatology Unit, **Bence**, **Vijay**, **Akilan**, **Meng**, and **Marianne**, for your help and discussions about anything that has to do with ACPA.

To all co-authors that I have not yet mentioned.

To **Ida**, for all the years that we have been sharing office. It has been fun and nice to follow your research from front row.

I would also like to thank **Tina**, **Anna**, **Elin**, and **Karin**, for all the nice breaks and inspiring talks and dissertation defenses. I did not expect it to be my turn this soon…

Thank you **Anna M** and **Ingegerd**, you have always been so friendly and welcoming. Also many thanks to **Anna F-H**, our discussions on immunology has helped me to improve my understanding of immunological events. Many thanks also to N**icky** and **Nastya**.

To **Michael**, I could not have done it without your effort with the BAL cells. Also, thank you **Johan** for the smoke incubator and discussion around BAL.

Many thanks to **Lara**, you are a great researcher that has helped me on the last stretch, supporting me and helping me improve my writing. I would also like to thank **Maria** for your bioinformatic and statistical expertise, it has been invaluable. Thank you **Maja** for all your help, you are truly inspiring researcher.

Thank you to everyone in the Epigenetic working group: **David**, **Ewoud**, **Leif**, and **Tojo**. It was a great introduction to bioinformatics, epigenome-wide studies, and statistics. Also, thank you **Sunjay** for all discussions and laughs about data analysis and everything else.

Thank you to past and present colleagues working at CMM, I could not have wished for a better research environment.

Many thanks to my family for the endless love and support, I am very happy I get to share this with you. To my **mum** and **dad**, you are always there for me, and to **Hans** and **Gun**, that are always there for them. To my amazing sisters **Jenny** and **Cecilia**, and to **Owen** and **Sara**, your support has been invaluable. To **Aurore**, you have been my motivation and made the last stretch feel so much easier than expected.

7 REFERENCES

1. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13(7):484-92.

2. Li E, Zhang Y. DNA methylation in mammals. Cold Spring Harb Perspect Biol. 2014;6(5):a019133.

3. Holliday R. Epigenetics: an overview. Dev Genet. 1994;15(6):453-7.

4. Cavalli G, Heard E. Advances in epigenetics link genetics to the environment and disease. Nature. 2019;571(7766):489-99.

5. Nicoglou A, Merlin F. Epigenetics: A way to bridge the gap between biological fields. Stud Hist Philos Biol Biomed Sci. 2017;66:73-82.

6. Bird A. Perceptions of epigenetics. Nature. 2007;447(7143):396-8.

7. Lee J, Taneja V, Vassallo R. Cigarette smoking and inflammation: cellular and molecular mechanisms. J Dent Res. 2012;91(2):142-9.

8. Lacasse Y, Martin S, Simard S, Desmeules M. Meta-analysis of silicosis and lung cancer. Scand J Work Environ Health. 2005;31(6):450-8.

9. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. Science. 1974;184(4139):868-71.

10. Kornberg RD, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell. 1999;98(3):285-94.

11. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011;21(3):381-95.

12. Chen ZX, Riggs AD. DNA methylation and demethylation in mammals. J Biol Chem. 2011;286(21):18347-53.

13. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930-5.

14. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333(6047):1303-7.

15. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300-3.

16. Delatte B, Jeschke J, Defrance M, Bachman M, Creppe C, Calonne E, et al. Genome-wide hydroxymethylcytosine pattern changes in response to oxidative stress. Sci Rep. 2015;5:12714.

17. Wei H, Feng Y, Liang F, Cheng W, Wu X, Zhou R, et al. Role of oxidative stress and DNA hydroxymethylation in the neurotoxicity of fine particulate matter. Toxicology. 2017;380:94-103.

18. Rithidech KN, Jangiam W, Tungjai M, Gordon C, Honikel L, Whorton EB. Induction of Chronic Inflammation and Altered Levels of DNA Hydroxymethylation in Somatic and Germinal Tissues of CBA/CaJ Mice Exposed to (48)Ti Ions. Front Oncol. 2016;6:155.

19. Dao T, Cheng RY, Revelo MP, Mitzner W, Tang W. Hydroxymethylation as a Novel Environmental Biosensor. Curr Environ Health Rep. 2014;1(1):1-10.

20. Jin SG, Wu X, Li AX, Pfeifer GP. Genomic mapping of 5-hydroxymethylcytosine in the human brain. Nucleic Acids Res. 2011;39(12):5015-24.

21. Iurlaro M, Ficz G, Oxley D, Raiber EA, Bachman M, Booth MJ, et al. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. Genome Biol. 2013;14(10):R119.

22. Mendonca A, Chang EH, Liu W, Yuan C. Hydroxymethylation of DNA influences nucleosomal conformation and stability in vitro. Biochim Biophys Acta. 2014;1839(11):1323-9.

23. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev. 2011;25(10):1010-22.

24. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A. 2006;103(5):1412-7.

25. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12(8):529-41.

26. Feinberg AP, Tycko B. The history of cancer epigenetics. Nat Rev Cancer. 2004;4(2):143-53.

27. Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. Genome Res. 2010;20(2):170-9.

28. Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D, et al. Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. PLoS Genet. 2011;7(9):e1002300.

29. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol. 2013;31(2):142-7.

30. Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. Nucleic Acids Res. 2019;47(D1):D983-D8.

31. Whitsett JA, Alenghat T. Respiratory epithelial cells orchestrate pulmonary innate immunity. Nat Immunol. 2015;16(1):27-35.

32. Murray PJ, Wynn TA. Protective and pathogenic functions of macrophage subsets. Nat Rev Immunol. 2011;11(11):723-37.

33. Gordon S, Taylor PR. Monocyte and macrophage heterogeneity. Nat Rev Immunol. 2005;5(12):953-64.

34. Franke-Ullmann G, Pfortner C, Walter P, Steinmuller C, Lohmann-Matthes ML, Kobzik L. Characterization of murine lung interstitial macrophages in comparison with alveolar macrophages in vitro. J Immunol. 1996;157(7):3097-104.

35. Bedoret D, Wallemacq H, Marichal T, Desmet C, Quesada Calvo F, Henry E, et al. Lung interstitial macrophages alter dendritic cell functions to prevent airway allergy in mice. J Clin Invest. 2009;119(12):3723-38.

36. Shibata Y, Berclaz PY, Chroneos ZC, Yoshida M, Whitsett JA, Trapnell BC. GM-CSF regulates alveolar macrophage differentiation and innate immunity in the lung through PU.1. Immunity. 2001;15(4):557-67.

37. Yoshida M, Ikegami M, Reed JA, Chroneos ZC, Whitsett JA. GM-CSF regulates protein and lipid catabolism by alveolar macrophages. Am J Physiol Lung Cell Mol Physiol. 2001;280(3):L379-86.

38. Tan SY, Krasnow MA. Developmental origin of lung macrophage diversity. Development. 2016;143(8):1318-27.

39. Epelman S, Lavine KJ, Randolph GJ. Origin and functions of tissue macrophages. Immunity. 2014;41(1):21-35.

40. Ginhoux F, Guilliams M. Tissue-Resident Macrophage Ontogeny and Homeostasis. Immunity. 2016;44(3):439-49.

41. Guilliams M, De Kleer I, Henri S, Post S, Vanhoutte L, De Prijck S, et al. Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. J Exp Med. 2013;210(10):1977-92.

42. Blusse van Oud Alblas A, van der Linden-Schrever B, van Furth R. Origin and kinetics of pulmonary macrophages during an inflammatory reaction induced by intravenous administration of heat-killed bacillus Calmette-Guerin. J Exp Med. 1981;154(2):235-52.

43. Landsman L, Jung S. Lung macrophages serve as obligatory intermediate between blood monocytes and alveolar macrophages. J Immunol. 2007;179(6):3488-94.

44. Godleski JJ, Brain JD. The origin of alveolar macrophages in mouse radiation chimeras. J Exp Med. 1972;136(3):630-43.

45. Thomas ED, Ramberg RE, Sale GE, Sparkes RS, Golde DW. Direct evidence for a bone marrow origin of the alveolar macrophage in man. Science. 1976;192(4243):1016-8.

46. Rodgman A, Perfetti TA. The Chemical Components of Tobacco and Tobacco Smoke. 2nd Edition ed. Boca Raton: CRC Press; 2013.

47. Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. Lancet. 2003;362(9387):847-52.

48. Arnson Y, Shoenfeld Y, Amital H. Effects of tobacco smoke on immunity, inflammation and autoimmunity. J Autoimmun. 2010;34(3):J258-65.

49. Kallberg H, Ding B, Padyukov L, Bengtsson C, Ronnelid J, Klareskog L, et al. Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. Ann Rheum Dis. 2011;70(3):508-11.

50. Hedstrom AK, Sundqvist E, Baarnhielm M, Nordin N, Hillert J, Kockum I, et al. Smoking and two human leukocyte antigen genes interact to increase the risk for multiple sclerosis. Brain. 2011;134(Pt 3):653-64.

51. Hedstrom AK, Baarnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. Neurology. 2009;73(9):696-701.

52. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum. 2006;54(1):38-46.

53. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet. 2011;88(4):450-7.

54. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet. 2016;9(5):436-47.

55. Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics. 2016;8(5):599-618.

56. Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ Health Perspect. 2012;120(10):1425-31.

57. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. Am J Hum Genet. 2016;98(4):680-96.

58. Sakaguchi S, Sakaguchi N, Asano M, Itoh M, Toda M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. J Immunol. 1995;155(3):1151-64.

59. Altorok N, Coit P, Hughes T, Koelsch K, Stone D, Rasmussen A, et al. Genome-wide DNA methylation patterns in naive CD4+ T cells from patients with primary Sjögren's syndrome. Arthritis Rheumatol. 2014:66(3):731-9.

60. Absher D, Li X, Waite L, Gibson A, Roberts K, Edberg J, et al. Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations. PLoS Genet. 2013:9(8):e1003678.

61. Bos S, Page C, Andreassen B, Elboudwarej E, Gustavsen M, Briggs F, et al. Genome-wide DNA methylation profiles indicate CD8+ T cell hypermethylation in multiple sclerosis. PLoS One. 2015:10(3):e0117403.

62. Whitaker J, Shoemaker R, Boyle D, Hillman J, Anderson D, Wang W, et al. An imprinted rheumatoid arthritis methylome signature reflects pathogenic phenotype. Genome Med 2013:30;5(4):40.

63. Compston A, Coles A. Multiple sclerosis. Lancet. 2002:359(9313):1221-31.

64. Milo R, Kahana E. Multiple sclerosis: geoepidemiology, genetics and the environment. . Autoimmun Rev. 2010:9(5):A387-94.

65. Ahlgren C, Oden A, Lycke J. High nationwide prevalence of multiple sclerosis in Sweden. Mult Scler. 2011:17(8):901-8.

66. Willer C, Dyment D, Risch N, Sadovnick A, Ebers GTCCSGTCfHG. Twin Concordance and Sibling Recurrence Rates in Multiple Sclerosis. Proceedings of the National Academy of Sciences of the United States of America. 2003:100(22), 12877–82.

67. Sawcer Sea. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011:476, 214-9.

68. Brynedal B, al. e. HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. PLoS One. 2007:2:e664.

69. Beecham AH, Patsopoulos NA, Xifara DK. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet. 2013:45(11):1353-60.

70. Knippenberg S, Damoiseaux J, Bol Y, Hupperts R, Taylor B, Ponsonby A, et al. Higher levels of reported sun exposure, and not vitamin D status, are associated with less depressive symptoms and fatigue in multiple sclerosis. Acta Neurol Scand. 2014:129(2):3-31.

71. Ascherio A, Munger K, White R, Köchert K, Simon K, Polman C, et al. Vitamin D as an early predictor of multiple sclerosis activity and progression. JAMA Neurology. 2014:71(3):306–14.

72. Hedström A, Olsson T, Alfredsson L. High body mass index before age 20 is associated with increased risk for multiple sclerosis in both men and women. . Multiple Sclerosis Journal. 2012:18(9):1334–6.

73. Ascherio A, Munger K. Environmental risk factors for multiple sclerosis. Part I: the role of infection. Ann Neurol. 2007:61(4):288–99.

74. Hedström AK, Baarnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. Neurology. 2009:73(9):696-701.

75. Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. Nat Rev Neurol. 2017;13(1):25-36.

76. Faigle W, Cruciani C, Wolski W, Roschitzki B, Puthenparampil M, Tomas-Ojer P, et al. Brain Citrullination Patterns and T Cell Reactivity of Cerebrospinal Fluid-Derived CD4(+) T Cells in Multiple Sclerosis. Front Immunol. 2019;10:540.

77. Baka Z, György B, Géher P, Buzás E, Falus A, Nagy G. Citrullination under physiological and pathological conditions. Joint Bone Spine. 2012:79(5):431-6.

78. Bradford CM, Ramos I, Cross AK, Haddock G, McQuaid S, Nicholas AP, et al. Localisation of citrullinated proteins in normal appearing white matter and lesions in the central nervous system in multiple sclerosis. J Neuroimmunol. 2014;273(1-2):85-95.

79. Moscarello MA, Mastronardi FG, Wood DD. The role of citrullinated proteins suggests a novel mechanism in the pathogenesis of multiple sclerosis. Neurochem Res. 2007;32(2):251-6.

80. Sospedra M, Martin R. Immunology of multiple sclerosis. Annu Rev Immunol. 2005;23:683-747.

81. Odoardi F, Sie C, Streyl K, Ulaganathan VK, Schlager C, Lodygin D, et al. T cells become licensed in the lung to enter the central nervous system. Nature. 2012;488(7413):675-9.

82. Glenn JD, Smith MD, Xue P, Chan-Li Y, Collins S, Calabresi PA, et al. CNS-targeted autoimmunity leads to increased influenza mortality in mice. J Exp Med. 2017;214(2):297-307.

83. Glenn JD, Liu C, Whartenby KA. Frontline Science: Induction of experimental autoimmune encephalomyelitis mobilizes Th17-promoting myeloid derived suppressor cells to the lung. J Leukoc Biol. 2019;105(5):829-41.

84. Chen Q, Liu Y, Lu A, Ni K, Xiang Z, Wen K, et al. Influenza virus infection exacerbates experimental autoimmune encephalomyelitis disease by promoting type I T cells infiltration into central nervous system. J Autoimmun. 2017;77:1-10.

85. Kular L, Liu Y, Ruhrmann S, Zheleznyakova G, Marabita F, Gomez-Cabrero D, et al. DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. Nat Commun. 2018;9(1):2397.

86. Marabita F, Almgren M, Sjoholm LK, Kular L, Liu Y, James T, et al. Smoking induces DNA methylation changes in Multiple Sclerosis patients with exposure-response relationship. Sci Rep. 2017;7(1):14589.

87. Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. Lancet. 2009;373(9664):659-72.

88. Klodzinski L, Wislowska M. Comorbidities in rheumatic arthritis. Reumatologia. 2018;56(4):228-33.

89. Alamanos Y, Drosos A. Epidemiology of adult rheumatoid arthritis. Autoimmun Rev. 2005:4:130–6.

90. Klareskog L, Rönnelid J, Lundberg K, Padyukov L, Alfredsson L. Immunity to citrullinated proteins in rheumatoid arthritis. Annu Rev Immunol 2008:26:651-75.

91. Krishnamurthy A, Joshua V, Haj Hensvold A, al e. Identification of a novel chemokine-dependent molecular mechanism underlying rheumatoid arthritis-associated autoantibody-mediated bone loss Annals of the Rheumatic Diseases 2016:75:721-9.

92. Sun M, Rethi B, Krishnamurthy A, Joshua V, Circiumaru A, Hensvold AH, et al. Anticitrullinated protein antibodies facilitate migration of synovial tissue-derived fibroblasts. Ann Rheum Dis. 2019.

93. Catrina A, Ytterberg A, Reynisdottir G, Malmström V, Klareskog L. Lungs, joints and immunity against citrullinated proteins in rheumatoid arthritis. Nat Rev Rheumatol 2014 10(1):645-53.

94. MacGregor A, Snieder H, Rigby A, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 2000:43(1):30-7.

95. Hensvold A, Magnusson P, Joshua V, Hansson M, Israelsson L, Ferreira R, et al. Environmental and genetic factors in the development of anticitrullinated protein antibodies (ACPAs) and ACPA-positive rheumatoid arthritis: an epidemiological investigation in twins. Ann Rheum Dis. 2015:74(2):375-80.

96. Gregersen P, Silver J, Winchester R. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum 1987:30(11):1205-13.

97. Mowen K, David M. Unconventional post-translational modifications in immunological signaling. Nat Immunol. 2014:15(6):512-20.

98. Reynisdottir G, Karimi R, Joshua V, Olsen H, Hensvold A, Harju A, et al. Structural changes and antibody enrichment in the lungs are early features of anti-citrullinated protein antibody-positive rheumatoid arthritis. Arthritis Rheumatol. 2014:66(1):31-9.

99. Catrina A, Joshua V, Klareskog L, Malmström V. Mechanisms involved in triggering rheumatoid arthritis. Immunol Rev. 2016:269(1):162-74.

100. Roos Ljungberg K, Joshua V, Skogh T, Eklund A, Skold CM, Karimi R, et al. Secretory anti-citrullinated protein antibodies in serum associate with lung involvement in early rheumatoid arthritis. Rheumatology (Oxford). 2019.

101. Scally S, Petersen J, Law S, Dudek N, Nel H, Loh K, et al. A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. J Exp Med. 2013:210: 2569-82.

102. van de Stadt LA, de Koning MH, van de Stadt RJ, Wolbink G, Dijkmans BA, Hamann D, et al. Development of the anti-citrullinated protein antibody repertoire prior to the onset of rheumatoid arthritis. Arthritis Rheum. 2011;63(11):3226-33.

103. Catrina AI, Ytterberg AJ, Reynisdottir G, Malmstrom V, Klareskog L. Lungs, joints and immunity against citrullinated proteins in rheumatoid arthritis. Nat Rev Rheumatol. 2014;10(11):645-53.

104. Ytterberg A, Joshua V, Reynisdottir G, Tarasova N, Rutishauser D, Ossipova E, et al. Shared immunological targets in the lungs and joints of patients with rheumatoid arthritis: identification and validation. Ann Rheum Dis. 2015:74(9):1772-7.

105. Källberg H, Ding B, Padyukov L, Bengtsson C, Rönnelid J, Klareskog L, et al. Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. Ann Rheum Dis. 2011:70(3): 508-11.

106. Klareskog L, Gregersen P, Huizinga T. Prevention of autoimmune rheumatic disease: state of the art and future perspectives. Ann Rheum Dis. 2010:69(12):2062-6.

107. Joshua V, Chatzidionisyou K, Catrina AI. Role of the lung in individuals at risk of rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2017;31(1):31-41.

108. Catrina AI, Svensson CI, Malmstrom V, Schett G, Klareskog L. Mechanisms leading from systemic autoimmunity to joint-specific disease in rheumatoid arthritis. Nat Rev Rheumatol. 2017;13(2):79-86.

109. Klareskog L, Stolt P, Lundberg K, Källberg H, Bengtsson C, Grunewald J, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum 2006:54(1):38-46.

110. Foulquier Cea. Peptidyl arginine deiminase type 2 (PAD-2) and PAD-4 but not PAD-1, PAD-3, and PAD-6 are expressed in rheumatoid arthritis synovium in close association with tissue inflammation. . Arthritis Rheum. 2007:56, 3541–53

111. Makrygiannakis D, Hermansson M, Ulfgren A, Nicholas A, Zendman A, Eklund A, et al. Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. Ann Rheum Dis. 2008:67(10):1488-92.

112. Ytterberg AJ, Joshua V, Reynisdottir G, Tarasova NK, Rutishauser D, Ossipova E, et al. Shared immunological targets in the lungs and joints of patients with rheumatoid arthritis: identification and validation. Ann Rheum Dis. 2015;74(9):1772-7.

113. Meng W, Zhu Z, Jiang X, Too CL, Uebe S, Jagodic M, et al. DNA methylation mediates genotype and smoking interaction in the development of anti-citrullinated peptide antibody-positive rheumatoid arthritis. Arthritis Res Ther. 2017;19(1):71.

114. Liu Y, Li X, Aryee MJ, Ekstrom TJ, Padyukov L, Klareskog L, et al. GeMes, clusters of DNA methylation under genetic control, can inform genetic and epigenetic analysis of disease. Am J Hum Genet. 2014;94(4):485-95.

115. Hensvold AH, Magnusson PK, Joshua V, Hansson M, Israelsson L, Ferreira R, et al. Environmental and genetic factors in the development of anticitrullinated protein

antibodies (ACPAs) and ACPA-positive rheumatoid arthritis: an epidemiological investigation in twins. Ann Rheum Dis. 2015;74(2):375-80.

116. Magnusson PK, Almqvist C, Rahman I, Ganna A, Viktorin A, Walum H, et al. The Swedish Twin Registry: establishment of a biobank and other recent developments. Twin Res Hum Genet. 2013;16(1):317-29.

117. Olsen HH, Grunewald J, Tornling G, Skold CM, Eklund A. Bronchoalveolar lavage results are independent of season, age, gender and collection site. PLoS One. 2012;7(8):e43644.

118. Ockinger J, Hagemann-Jensen M, Kullberg S, Engvall B, Eklund A, Grunewald J, et al. T-cell activation and HLA-regulated response to smoking in the deep airways of patients with multiple sclerosis. Clin Immunol. 2016;169:114-20.

119. Irizarry RA, Ladd-Acosta C, Carvalho B, Wu H, Brandenburg SA, Jeddeloh JA, et al. Comprehensive high-throughput arrays for relative methylation (CHARM). Genome Res. 2008;18(5):780-90.

120. Ladd-Acosta C, Aryee MJ, Ordway JM, Feinberg AP. Comprehensive high-throughput arrays for relative methylation (CHARM). Curr Protoc Hum Genet. 2010;Chapter 20:Unit 20 1 1-19.

121. Aryee MJ, Wu Z, Ladd-Acosta C, Herb B, Feinberg AP, Yegnasubramanian S, et al. Accurate genome-scale percentage DNA methylation estimates from microarray data. Biostatistics. 2011;12(2):197-210.

122. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J R Stat Soc B. 1995;57(1):289-300.

123. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One. 2012;7(7):e41361.

124. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.

125. Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman & Hall; 1993. xvi, 436 p. p.

126. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-9.

127. Fortin JP, Triche TJ, Jr., Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics. 2017;33(4):558-60.

128. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics. 2017;33(24):3982-4.

129. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010;11:587.

130. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, R VL, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin. 2015;8:6.

131. Edgington ES. Normal Curve Method for Combining Probability Values from Independent Experiments. J Psychol. 1972;82(1):85-+.

132. Hedges LV, Olkin I. Statistical methods for meta-analysis. Orlando: Academic Press; 1985. xxii, 369 p. p.

133. Breeze CE, Paul DS, van Dongen J, Butcher LM, Ambrose JC, Barrett JE, et al. eFORGE: A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data. Cell Rep. 2016;17(8):2137-50.

134. Mi H, Thomas P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. Methods Mol Biol. 2009;563:123-40.

135. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. 2005;33(Web Server issue):W741-8.

136. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6(7):e21800.

137. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol. 2012;41(1):200-9.

138. Yagi T. Clustered protocadherin family. Dev Growth Differ. 2008;50 Suppl 1:S131-40.

139. Philibert RA, Sears RA, Powers LS, Nash E, Bair T, Gerke AK, et al. Coordinated DNA methylation and gene expression changes in smoker alveolar macrophages: specific effects on VEGF receptor 1 expression. J Leukoc Biol. 2012;92(3):621-31.

140. Monick MM, Beach SR, Plume J, Sears R, Gerrard M, Brody GH, et al. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. Am J Med Genet B Neuropsychiatr Genet. 2012;159B(2):141-51.

141. Schieber M, Chandel NS. ROS function in redox signaling and oxidative stress. Curr Biol. 2014;24(10):R453-62.

142. Zhou X, Zhuang Z, Wang W, He L, Wu H, Cao Y, et al. OGG1 is essential in oxidative stress induced DNA demethylation. Cell Signal. 2016;28(9):1163-71.

143. Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell. 2014;159(6):1312-26.

144. Mora A, Sandve GK, Gabrielsen OS, Eskeland R. In the loop: promoter-enhancer interactions and bioinformatics. Brief Bioinform. 2016;17(6):980-95.

145. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol. 2015;16(9):519-32.