Thesis for doctoral degree (Ph.D.)
2020

Exploring the genomic and transcriptomic landscape of immune cells in Multiple Sclerosis: Towards better biomarkers, diagnosis and treatment

Sunjay Jude Fernandes

Karolinska Institutet

From Department of Medicine, Solna
Karolinska Institutet, Stockholm, Sweden

EXPLORING THE GENOMIC AND TRANSCRIPTOMIC LANDSCAPE OF IMMUNE CELLS IN MULTIPLE SCLEROSIS: TOWARDS BETTER BIOMARKERS, DIAGNOSIS AND TREATMENT

Sunjay Jude Fernandes

Karolinska Institutet

Stockholm 2020

Published by Karolinska Institutet.
Printed by E-Print AB 2020

ISBN 978-91-7831-707-3

Exploring the genomic and transcriptomic landscape of immune cells in Multiple Sclerosis: Towards better biomarkers, diagnosis and treatment.

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Sunjay Jude Fernandes

Public defence
Center for Molecular Medicine, Lecture hall L8:00,
Karolinska Universitetssjukhuset, Solna
Friday 28th February 2020 at 9.00am

Principal Supervisor:
Professor Jesper Tegnér
Karolinska Institutet
Department of Medicine, Solna
Unit of Computational Medicine

Co-supervisor(s):
Professor Ingrid Kockum
Karolinska Institutet
Department of Clinical Neurosciences
Genetic Epidemiology of Multiple Sclerosis

Assistant Professor David Gomez-Cabrero
Karolinska Institutet
Department of Medicine, Solna
Unit of Computational Medicine

Opponent:
Associate Professor Calliope Dendrou
University of Oxford
Wellcome Trust Centre for Human Genetics
Nuffield Department of Medicine
Cross-Autoimmune Disease Functional Genomics

Examination Board:
Associate Professor Carsten Daub
Karolinska Institutet
Department of Biosciences and Nutrition
Clinical Transcriptomics

Professor Ann-Christine Syvänen
Uppsala University
Department of Medical Sciences
Genetics and Epigenetics in Leukemia and Autoimmune Diseases

Associate Professor Magnus Andersson
Karolinska Institutet
Department of Clinical Neurosciences
Head of Department of Neurology, Karolinska University Hospital

To Dad.

To my family.

Thank you for your love, support and encouragement.

ABSTRACT

The overall aim of this thesis was to determine the changes in gene regulation taking place in immune cells during the course of Multiple Sclerosis. Over 200 MS-associated SNPs have been identified from GWAS studies. These regions were found to be primarily in the non-coding regions of the genome and point to the vast immune system as the leading cause of MS. Inferring their function therefore has been a challenge. In addition, a complex interaction of genetics and environment has been proposed. This leads to the unresolved question associated with specific changes in the immune system that can lead to disease.

In order to resolve the role of the immune system in MS, we applied an array of high-throughput genomic and transcriptomic profiling techniques to identify specific changes in specific immune cells. MS being a complex immune mediated neurological disease, makes inference of regulation dependent changes in gene expression a challenge. By integrating different layers of evidence we are able to propose multiple interactions taking place within and across immune cells. We also find evidence that confirms previous findings in MS related to the increased activity of T and B cells. In addition, we identify multiple new genes, chromatin regions and DNA-methylated regions with differential activity primarily in T and B cells.

Collectively the results from these studies highlight the multiple factors leading to the dysregulation of the immune system in MS and the specific cells associated with pathogenesis. These studies also provide a resource for hypothesis building, validation of other studies and application of newer integration methodologies in a complex immune disease such as MS.

LIST OF SCIENTIFIC PAPERS

I. **Non-parametric combination analysis of multiple data types enables detection of novel regulatory mechanisms in T cells of Multiple Sclerosis patients**

Fernandes SJ, Morikawa H, Ewing E, Ruhrmann S, Joshi RN, Lagani V, Karathanasis N, Khademi M, Planell N, Schmidt A, Tsamardinos I, Olsson T, Piehl F, Kockum I, Jagodic M, Tegnér J, Gomez-Cabrero D.
Sci Rep. 2019 Aug 19;9(1):11996. doi: 10.1038/s41598-019-48493-7. PMID: 31427643.

II. **Paired analysis of chromatin and expressed genes in four immune cell-types in the blood of Multiple Sclerosis patients**

Fernandes SJ, Ericsson M, Khademi M, Olsson T, Gomez-Cabrero D, Kockum I, and Tegnér J.
Manuscript

III. **Single cell transcriptomics of paired blood and cerebrospinal fluid of multiple sclerosis patients with special focus on the immune repertoire**

Fernandes SJ, Radpour S, Thimappa M, Al Nimer F, Gyllenberg A, Piehl F, Gomez-Cabrero D, Kockum I and Tegnér J.
Manuscript

IV. **Combining evidence from four immune cell types identifies DNA methylation patterns that implicate functionally distinct pathways during Multiple Sclerosis progression.**

Ewing E, Kular L, **Fernandes SJ**, Karathanasis N, Lagani V, Ruhrmann S, Tsamardinos I, Tegner J, Piehl F, Gomez-Cabrero D, Jagodic M.
EBioMedicine. 2019 May;43:411-423. doi: 10.1016/j.ebiom.2019.04.042. Epub 2019 Apr 30. PMID: 31053557.

ADDITIONAL PUBLICATIONS:

I. **Feedforward regulation of Myc coordinates lineage-specific with housekeeping gene expression during B cell progenitor cell differentiation.**
Ferreirós-Vidal I, Carroll T, Zhang T, Lagani V, Ramirez RN, Ing-Simmons E, Gómez-Valadés AG, Cooper L, Liang Z, Papoutsoglou G, Dharmalingam G, Guo Y, Tarazona S, **Fernandes SJ**, Noori P, Silberberg G, Fisher AG, Tsamardinos I, Mortazavi A, Lenhard B, Conesa A, Tegner J, Merkenschlager M, Gomez-Cabrero D. PLoS Biol. 2019 Apr 12;17(4):e2006506. doi: 10.1371/journal.pbio.2006506. eCollection 2019 Apr. PMID: 30978178.

II. **Phosphatase inhibitor PPP1R11 modulates resistance of human T cells toward Treg-mediated suppression of cytokine expression.**
Joshi RN, **Fernandes SJ**, Shang MM, Kiani NA, Gomez-Cabrero D, Tegnér J, Schmidt A. J Leukoc Biol. 2019 Aug;106(2):413-430. doi: 10.1002/JLB.2A0618-228R. Epub 2019 Mar 18. PubMed PMID: 30882958.

III. **Impact of genetic risk loci for multiple sclerosis on expression of proximal genes in patients.**
James T, Lindén M, Morikawa H, **Fernandes SJ**, Ruhrmann S, Huss M, Brandi M, Piehl F, Jagodic M, Tegnér J, Khademi M, Olsson T, Gomez-Cabrero D, Kockum I. Hum Mol Genet. 2018 Mar 1;27(5):912-928. doi: 10.1093/hmg/ddy001. PMID: 29325110.

IV. **Time-resolved transcriptome and proteome landscape of human regulatory T cell (Treg) differentiation reveals novel regulators of FOXP3.**
Schmidt A, Marabita F, Kiani NA, Gross CC, Johansson HJ, Éliás S, Rautio S, Eriksson M, **Fernandes SJ**, Silberberg G, Ullah U, Bhatia U, Lähdesmäki H, Lehtiö J, Gomez-Cabrero D, Wiendl H, Lahesmaa R, Tegnér J.
BMC Biol. 2018 May 7;16(1):47. doi:10.1186/s12915-018-0518-3. PMID: 29730990.

V. **Hypermethylation of MIR21 in CD4+ T cells from patients with relapsing-remitting multiple sclerosis associates with lower miRNA-21 levels and concomitant up-regulation of its target genes.**
Ruhrmann S, Ewing E, Piket E, Kular L, Cetrulo Lorenzi JC, **Fernandes SJ**, Morikawa H, Aeinehband S, Sayols-Baixeras S, Aslibekyan S, Absher DM, Arnett DK, Tegner J, Gomez-Cabrero D, Piehl F, Jagodic M.
MultScler. 2017 Aug. PMID: 28766461.

VI. **Multiplexed next-generation sequencing and de novo assembly to obtain near full-length HIV-1 genome from plasma virus.**
Aralaguppe SG, Siddik AB, Manickam A, Ambikan AT, Kumar MM, **Fernandes SJ**, Amogne W, Bangaruswamy DK, Hanna LE, Sonnerborg A, Neogi U.
J Virol Methods. 2016 Oct. PMID: 27448822.

CONTENTS

LIST OF ABBREVIATIONS

APC	Antigen presentation cell
ATAC	Assay for transposase-accessible chromatin
BCR	B-cell receptor
CIS	Clinically isolated syndrome
CMV	Cytomegalovirus
CNS	Central nervous system
CSF	Cerebrospinal fluid
DBR	Differentially bound regions
DC	Dendritic cells
DEG	Differentially expressed genes
DMP	Differentially methylated probes
DNA	Deoxyribonucleic acid
EAE	Experimental autoimmune disease
EBV	Epstein-Barr virus
eQTL	Expression quantitative trait loci
HLA	Human leukocyte antigen
MAIT	Mucosa-associated invariant T cell
MBP	Myelin basic protein
meQTL	Methylation quantitative trait loci
MOG	Myelin oligodendrocyte glycoprotein
MRI	Magnetic resonance imaging
NPC	Non-parametric combination
PBMC	Peripheral blood mononuclear cells
RNA	Ribonucleic acid
RRMS	Relapse-remitting multiple sclerosis
rSNP	Risk single nucleotide polymorphism
SCT	Single cell transcriptomics
SNP	Single nucleotide polymorphism
SPMS	Secondary-progressive multiple sclerosis
TCR	T-cell receptor

1 INTRODUCTION

1.1 MULTIPLE SCLEROSIS

Multiple Sclerosis (MS) is a complex autoimmune-mediated neurodegenerative disease. Disseminated demyelination of nerve fibers of the brain and spinal cord are characteristic of the disease. Acute inflammatory injury of axons and glia leads to disability associated with movement and cognition (1). 2.3 million individuals are affected worldwide (2). Immune cells cross the blood-brain barrier and infiltrate brain tissue promoting inflammation, demyelination and gliosis which cause the formation of lesions (3). The early appearance of T cells in MS lesions along with the presence of myelin reactive T cells in the blood has led to them being considered strong drivers of the disease. Sex and age of onset of the disease have been shown to be a determining factor in the risk of disease and progression (4). The gender prevalence of MS is in a ratio of 3: 1 (female to male) with an average age of onset being 34. In 85% of patients the onset of the disease is in the form of Relapse Remitting MS (RRMS). In RRMS there are episodes of active disease during which demyelinating lesions form in the central nervous system (CNS) followed by episodes of remission where remyelination and healing take place. The remaining 15% of patients directly present with gradually worsening disability without clear relapses known as primary progressive MS (PPMS). A disease course resembling PPMS affects approximately half of all untreated patient with RRMS after 10 years. Environmental factors such as Epstein-Barr Virus (EBV), low levels of vitamin D and smoking have also shown to increase susceptibility to MS (5,6). Genetic factors have also been implicated in susceptibility to this condition. The human leukocyte antigen *(HLA)-DRB1*15:01* and other HLA alleles) affect the risk of MS, along with over 200 single nucleotide polymorphisms (SNPs) (7). Of these identified SNPs a large number have been found in the vicinity of immune-related genes. Current MS therapies provide only partial protection against relapses and primarily target the immune system but are ineffective against progressive symptoms.

1.1.2 Multiple Sclerosis: Pathology and Progression

Clinically Isolated Syndrome (CIS): CIS is recognized as the first clinical presentation of MS. CIS fulfils characteristics of inflammatory demyelination but not criteria for the dissemination of inflammatory lesions in time (8). The revised McDonald MS diagnosis criterion of 2010 (9) allowed for better diagnosis of MS from a single scan criterion for patients presenting with as little as 1 single clinical episode. However, in the latest update of McDonald MS diagnosis criteria 2019, CIS has been included as RRMS (10).

Relapse Remitting Multiple Sclerosis (RRMS): Characterized by initial episodes of neurological dysfunction followed by periods of remission and recovery. Using magnetic resonance imaging (MRI) can visualize the characteristic lesions caused by inflammation and demyelination in the white matter. As the disease progresses the recovery from neurological damage decreases and disability increases.

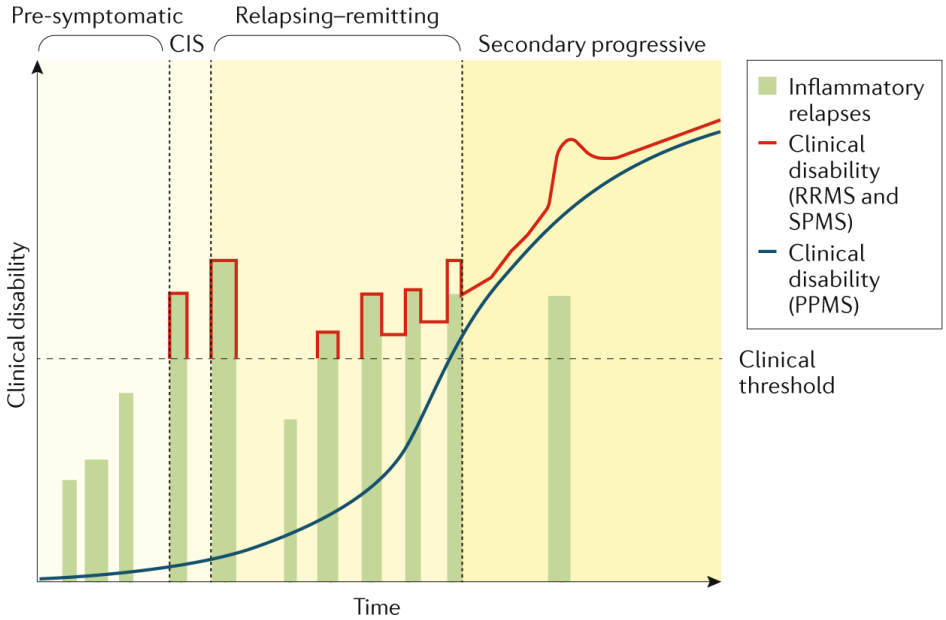

Figure 1: Clinical course of MS

Reprinted by permission: Filippi, M., Bar-Or, A., Piehl, F. *et al.* Multiple sclerosis. *Nat Rev Dis Primers* **4,** 43 (2018) doi:10.1038/s41572-018-0041-4

Secondary Progressive Multiple Sclerosis (SPMS): Characterized by gradual worsening of symptoms from an initial relapsing disease course (RRMS), with or without acute relapses. Inflammatory lesions are less frequent and progressive neurological decline is accompanied by the decrease in brain volume or CNS atrophy. Addressing specific clinical, pathologic and immunological criteria is tough due to the slow nature of conversion from RRMS to SPMS, as a result, little progress has been made in the field of biomarkers and imaging.

Primary Progressive Multiple Sclerosis (PPMS): Progressive decline in neurological function is characteristic of PPMS. Following a disease course similar to SPMS the pathogenesis is associated less with inflammation and more with increasing neurological decline from CNS atrophy. Though PPMS lacks a strong immunological component, clinical, imaging and genetic data suggest it is a part of the progressive spectrum of MS (11). Lesions or plaques show evidence of gradual expansion at the margin of the lesions.

1.1.3 Susceptibility to Multiple Sclerosis

MS is a multifactorial disease. Many factors such as genetic, immunological and environmental have been shown to contribute to its susceptibility. Genetic susceptibility to MS has been shown to account for about 30% of the overall risk to MS and the familial risk was estimated at about 60% (12,13). This leaves a substantial contribution to be explained by other factors such as immunological heterogeneity, and environmental interactions (14). Viral triggers such as EBV and CMV have been associated with molecular mimicry (15–18). Environmental

factors such as vitamin D deficiency have also been shown to increase the risk of MS (19). Finally, with increasing age, MS affected individuals show a gradual decline in muscular and cognitive abilities which is a direct consequence of the progression of the disease.

1.1.4 Immune cells in Multiple Sclerosis

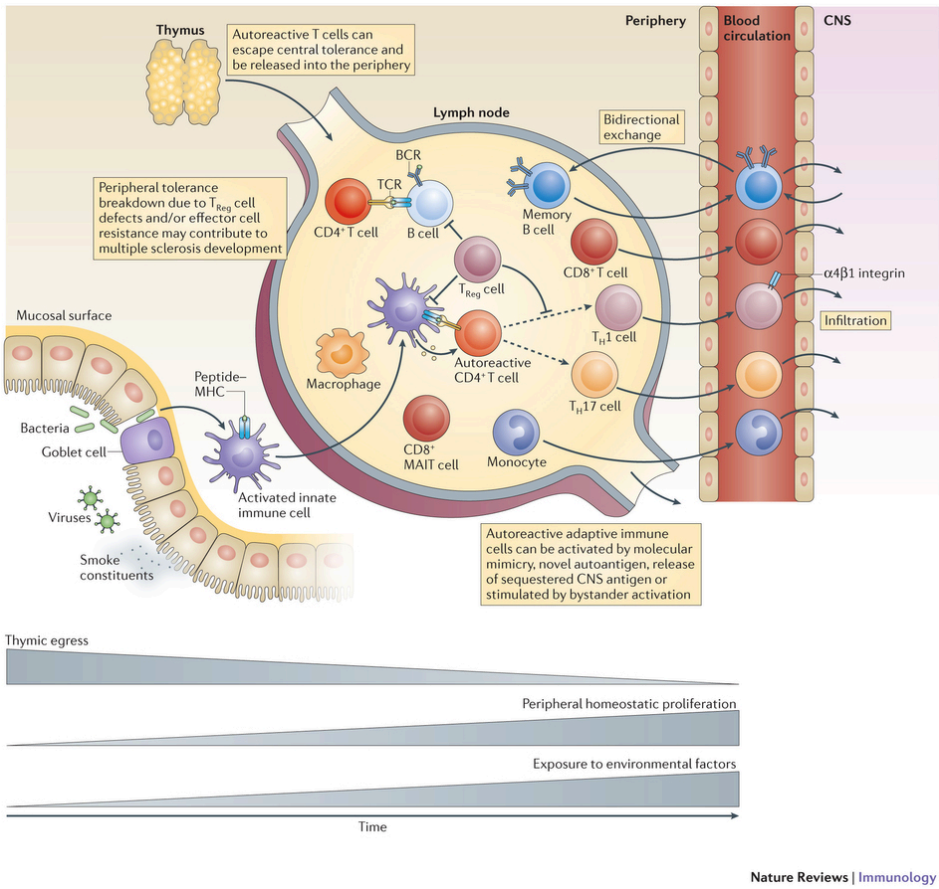

Figure 2: Immune cells involved in peripheral immune dysregulation

Reprinted by permission: Dendrou, C., Fugger, L. & Friese, M. Immunopathology of multiple sclerosis. *Nat Rev Immunol* **15,** 545–558 (2015) doi:10.1038/nri3871

Animal models such as Experimental Autoimmune encephalomyelitis (EAE) have significantly shaped our understanding of the inflammatory response in MS (20). Myelin proteins including myelin basic protein (MBP) and myelin oligodendrocyte glycoprotein (MOG) are targets of CD4+ T cell dependent inflammatory demyelination in EAE (21–23). CD4+ T cells recognizing MOG and MBP have been detected in MS patients. T cells express chemokine receptors, adhesion molecules and integrins that allow them to cross the choroid plexus and menegial venules (24). T cells specific for myelin are activated in the periphery and

after gaining access to the CNS are reactivated by APCs including CD11c dendritic cells (DCs) that present self-antigens. Reactivation of T cells leads to the production of soluble mediators which recruit other immune cells like monocytes and naïve CD4+ T cell that can be activated through epitope spreading (25). However, activation of T cells in the periphery is still poorly understood. While MBP is found in lymphoid tissue, other myelin-specific protein are synthesized by CNS residing oligodendrocytes (26). These antigen recognizing CD4 T cells are also present in healthy individuals however regarding their frequency and avidity there is inconsistent evidence (27,28). Molecular mimicry has been proposed as a potential mechanism. In EAE, a subset of activated T cells express CCR6 (CC-chemokine receptor 6) known play a role in facilitating the entry of T cells into the CNS (29). CCR6 is a ligand of CCL20 which is constitutively expressed by epithelial cells of the choroid plexus in humans and mice adding to the idea T cells cross the blood-brain barrier.

Pro-inflammatory Th1 and Th17 are the main CD4+ T cell subtypes implicated in MS. CD4+ T cells expressing CCR6 have increased expression of *IFNγ* and *IL17A*, both of which are Th1 and Th17 signature cytokines, while some lesions show an intermediate phenotype of these cells simultaneously expressing *IFNγ* and *IL17A* (30). *GM-CSF* (granulocyte-macrophage colony-stimulating factor) producing Th17 cells contribute to chronic inflammation in EAE (31), while Th1 cells have been shown to be the primary producers of this cytokine in humans (32). As a result, the relative importance of these in MS is highly contended due to the predominance of one cell type over another either between phases of disease or in human vs the disease model EAE (33,34). Overall, since the implication of pro-inflammatory Th1 and Th17 cells in MS many therapeutic concepts have been to skew it towards the more anti-inflammatory Th2 phenotype. This is the mode of treatment used in first-line therapies such as *IFNβ*, glatiramer acetate and dimethyl fumarate (35–37)

CD4+ Regulatory T cells or Tregs which are associated with suppression of inflammation, are thought to be defective in MS (38). Some reasons for this dysfunction of peripheral Tregs may be due to dysregulation of APCs (39) and variation in the *BACH2* transcription factor as identified from non-HLA genetic associations which is essential to the development of Tregs and T cell identity (13,40,41). Along with reduced suppressive capacity, reduced expansion of the memory Treg cell populations has also been reported (42,43). Skewing of the Treg population towards Th1 cell-like phenotype in patients has been observed but reversible on *IFNγ* therapy (44). Finally, a resistance to suppression by effector T cells has been proposed which is mediated by IL6 induced signal transducer and activator of transcription (*STAT3*)-mediated signalling contributing to resistance (45,46).

Higher frequency of CD8+T cells has been identified in MS lesions compared to CD4+T cells and show correlation with axonal damage (47). In EAE, CD8 T cells were also shown to be activated by epitope spreading aided by cross-presentation by monocyte-derived DCs in the CNS even though EAE is primarily a CD4+ T cell-driven model (48). In MS lesions, up to one-fourth of CD8+ T cells produce IL17 which leads to the belief that they may be Mucosa-associated invariant T (MAIT) cells (49). CD8+ T cells require more detailed study when

considered as targets of therapy since broad-spectrum drugs such as natalizumab and alemtuzumab have an unclear role in targeting these cells. CD8+ Treg cells display less regulatory phenotype in MS while cytotoxic CD8+ Tregs show enhanced function following glatiramer acetate therapy (50).

Our understanding of the role of B cells or CD19 cells is relatively new in the field of MS. Treatment depleting B cells such as anti-CD20 mediated depletion (rituximab and ocrelizumab) have proven to be effective in preventing relapses in MS. These treatments do not however deplete plasma B cells (51). Memory B cells were seen in higher proportion in CSF of MS patients. Proposed mechanisms for their activity has been as antigen-presenting to T cells since they possess MHC class II molecules like APCs. Reciprocal activation by activated T cells of B cells has also been studied (52). A hallmark of diagnostic findings in MS has been the presence of oligoclonal bands in CSF which also point to the presence of plasma B cells in MS (53). However, evidence suggests they are less likely to play a role in driving inflammation since antibody levels or OCBs do not change rapidly with disease relapses.

1.2 GENOMICS AND BIOINFORMATICS IN DISEASE

Individual cell types have unique gene expression patterns that lead to characteristic properties. This unique gene expression is primarily determined by the regulation of the genome through modification either directly or indirectly of DNA. Direct modifications of DNA such as methylation of DNA at specific sites in the genome can lead to the silencing of these regions. Similarly proteins that bind DNA and help in the efficient folding to form chromatin can undergo various types of modification such as acetylation, methylation and phosphorylation leading to the unwinding of tightly packed heterochromatin allowing these regions to be transcribed or expressed. This unwound or "open" DNA can now undergo additional levels of regulation to expression through binding of either single transcription factors or protein complexes consisting multiple expression regulating proteins (54,55).

Normal regulation of expression can be altered in disease through changes in DNA nucleotides in the form of single nucleotide polymorphisms (SNPs) or deletion/insertions or translocation of large fragments of DNA across the genome. High-throughput methods to study these states primarily in cancer have varied from DNA sequencing, genotyping, methylation, DNase hypersensitivity, ChIP seq and Hi-C. These are powerful methods in order to detect a majority of changes in a particular cell type in a single assay.

In immune-mediated diseases such as multiple sclerosis, due to the heterogeneity of cells and signals involved, interpretation of data can be complex. CD4 and CD8 T cells consist of ~10 suspected cell types which have been shown individually to contribute to the pathology of MS (56). Therefore, combining data types can be very informative since it increases the chance of detection of changes in regulation while adding more comprehensive evidence for that change.

Analysis of this high-throughput data has been a considerable challenge. Detecting robust changes that are both significantly detected in the data and have biological relevance requires

a combination of expertise in experimental design, high-throughput data methodologies for data generation and data analysis.

1.2.2 Genetic associations in Multiple Sclerosis

HLA class I and class II molecules are essential for antigen recognition by CD8+ and CD4+ T lymphocytes, respectively. Using high-throughput genotyping arrays and traditional PCR based genotyping, class II alleles such as *DRB1*15:01*, *DRB1*03:01*, and *DRB1*13:03* have been associated with increased risk of MS while HLA class I allele *A2* is associated with a decreased risk to MS (57). Additionally, genome-wide association studies (GWAS) have identified more than 200 common genetic variants (single nucleotide polymorphism, SNPs) associated with multiple sclerosis. These were found primarily in gene loci related to the adaptive immune system (7,13).

1.2.3 Gene expression studies in Multiple Sclerosis

Whole transcriptome sequencing or RNA-Seq uses the poly A tail present on post-translationally modified mature messenger RNA to fish out a majority of the coding RNA and couples it with high-throughput sequencing (58). This allows for the quantitative and qualitative analysis of all the expressed genes in a given sample. Very few studies have been carried out in MS using RNA-Seq however a few studies have been done on expression microarrays and mostly address PBMCs (59). Using cDNA microarrays, PBMCs from MS patients and healthy controls, subtle changes were detected in gene expression between groups with some of the detected genes that have previously associated with the disease (60).

Similarly, one study was done in CD4+ and CD8+ T cells showed pronounced changes in multiple genes in CD8+T cells however CD4+ T cells showed fewer changes. These results confirmed differential expression in genes associated with MS pathogenesis (61). Other studies have been carried out with similar results primarily in RRMS (62–65). The heterogeneity of PBMCs and T cells can lead to the reduced signal from relevant genes and more specific cell types are required for expression studies. But, integration of complementary data such as methylation, transcription factor binding can help in increased significance and lead to the detection of more relevant targets.

1.2.4 Chromatin binding and its implications in Multiple Sclerosis

DNase 1 Hypersensitivity Assay which exploits the susceptibility of open DNA regions to DNase1 enzyme cleavage was combined with high throughput sequencing to develop DNase Seq. This assay allows the determination of open or accessible DNA across the whole genome in native bound chromatin (66,67). ATAC Seq, an assay that works on a similar principle to DNase seq was developed, which allowed the use of far fewer cells (50,000) as input compared to DNase Seq (5 Million) (68). Reducing the requirement for input material makes this protocol ideal for use in studying complex diseases such as Multiple Sclerosis where input material is limiting. Inferring transcription factor binding post DNase seq was made possible through the careful characterization of transcription factors and their binding sites which was done over

numerous chromatin immunoprecipitation assays. The curation of these results has enabled rich resources such as JASPAR and TRANSFAC for TF-binding in multiple cell types and species (69,70).

A large number of disease-associated SNPs have been identified from genome-wide association studies (GWAS). A majority of these SNPs have been found to lie outside of gene coding regions with as much as 45% in intron regions and 43% in intergenic regions (71). Additionally, regulatory elements were shown to be associated with SNPs potentially determining chromatin states (72–74). MS-associated SNPs have been implicated in altered gene expression constituting expression quantitative traits (eQTLs) (75,76). A majority of these genetic variants are present in the vicinity of immune-related genes (77). Additionally, it was shown that these MS-associated SNPs are present or associated with regulatory elements such as transcription factors specific to the cell subtypes of CD4+ and CD8+ T cells, namely Th1, Th17 and cytotoxic CD8+ T cells. However, these studies were done using data that was collected from healthy individuals. Application of chromatin binding assays to a complex immune disease such as MS would give us additional insights into the altered regulation that may lead to disease development and progression.

1.2.5 Single cell genomics and its potential in Multiple Sclerosis

Understanding the heterogeneity of immune cells in MS has been challenging in the development and progression. Using a pool of cells as done in bulk RNA Seq gives rise to an expression signal which is the average gene expression of many cell types in different states of regulation and expression. Single cell genomics allows for comprehensive yet specific determination of expression patterns per cell type. Frequency and strength of transcriptional bursts in gene expression (78–80), paternal and maternal allelic expression which may play a role in disease development (78,80,81) and gene regulatory interactions and networks (82,83) can be inferred. The role of CD4+ and CD8+ myelin-specific T cells has been studied in blood and CSF in MS patients (27,28). Understanding the clonal expansion of these specific T cells and linking them to their functional phenotypes can be extremely important in our understanding of MS pathogenesis. Methods in single cell transcriptomic analysis also allow for reconstruction of the expressed T and B cell repertoire (84). Two studies have been performed which give us a better understanding of the role of immune cells in MS (85,86). Adding to these would give us a larger sample size to assess more stable change and infer additional cell-cell interactions.

1.2.6 DNA methylation studies in Multiple Sclerosis

DNA Methylation involves the addition of a methyl group to the carbon in the 5'position of cytosine residues in CpG dinucleotides. DNA methyltransferases (DNMTs) such as DNMT1 (maintains methylation during replication), DNMT2, DNMT3A and DNMT3B (denovo-methylation) are responsible for DNA methylation (87). In the mammalian genome, approximately half of all genes are associated with CpG islands, which are regions of high content of CpG dinucleotides. In these genes, DNA hypomethylation is associated with gene

activation while hypermethylation is associated with gene inactivation. The second type of methylation modification, Hydroxymethylation of DNA, where a hydroxymethyl group is added to the 5' position of cytosine by TET proteins was first described in 2009 (88,89). Hydroxymethylation is thought to be a signal for chromatin factors. Methylation and Hydroxymethylation of DNA are thought to be an efficient mechanism of deregulation since it can result in the lower binding affinity of regulatory proteins to DNA, resulting in altered gene expression. Stress, environmental factors and individual habits can induce altered methylation of DNA which can, in turn, contribute to the establishment and maintenance of autoimmune diseases (90).

Methylation studies carried out in blood are very few and have ranged from addressing peripheral blood mononuclear cells to T cells. These studies have shown that many changes occur in MS that are associated with immune dysregulation. (91–94).

1.2.7 Data integration

Given the heterogeneous nature of MS, the development of stable markers of disease prognosis or therapeutic response are very clinically significant. With multiple applications of omics technologies now being available that measure DNA, gene expression, regulation and proteins, any given method can be very informative (95–97). However, the enormously complex machinery that determines the regulation and expression of genes and their subsequent dysregulation in disease can be complicated to infer (98,99). The integration of data allows for the combination of multiple layers of data to infer co-ordinated changes taking place across the same system (100,101). This holistic approach towards inferring coordinated changes across a system is at the heart of the field of systems biology.

These coordinated changes in turn give us higher confidence that they are stable and can be interrogated as potential mechanisms of dysregulation. Identifying stable mechanisms of dysregulation are important and the first step in considering them as biomarkers for better diagnosis and even as new therapeutic targets.

2 THESIS AIMS

The overall aim of this thesis was to determine the changes in gene regulation taking place in immune cells during the course of Multiple Sclerosis.

Study 1: To characterize the dysregulated T cells in RRMS and SPMS.

Study 2: To characterize chromatin accessibility associated with inflammation in newly diagnosed RRMS patients.

Study 3: To characterize immune cell populations at the single cell level in newly diagnosed patients sampled after first relapse.

Study 4: To characterize the epigenetic changes in immune cells in RRMS and SPMS.

3 METHODOLOGICAL CONSIDERATIONS

For detailed description of methodologies, the materials and methods sections from individual manuscripts can be referred.

3.1 COHORT INFORMATION AND SAMPLE COLLECTION

Study participants gave their written consent for inclusion in each study. All studies were approved by the Regional Ethical Review Board, Stockholm, Sweden.

Three cohorts were used in the 4 studies. Cohort 1, used in **Study 1** and **Study 4** included a total of 17 HC, 12 RRMS and 12 SPMS samples. The patients included in this study were primarily (85%) newly diagnosed with MS while the rest were on a medication wash-out period of up to 6 months. All patients were recruited at the Neurology clinic at Karolinska University Hospital in Stockholm. In **Study 2** (cohort 2) and **Study 3** (cohort 3), all patients were RRMS and newly diagnosed with MS, with no medication being administered prior to sampling. HCs for all 4 studies were age-matched within 5 years of patients and were under no medication prior to sampling. **In Study 3**, HCs were additionally genotyped and only HLA DRB1*15 positive individuals were included.

Study 1, **2** and **4** consisted of PBMCs isolated from blood that was freshly drawn from study participants. **Study 3** consisted of PBMCs and CSF cells freshly drawn from study participants.

3.2 IMMUNE CELL ISOLATION FROM BLOOD AND CEREBROSPINAL FLUID

In all 4 studies, PBMCs were isolated using the ficol method (102). In Study 1 and Study 4, CD14 was isolated using MACS microbeads (Miltenyi Biotec). CD4, CD8 and CD19 were isolated using flow cytometry by binding each cell with CD3 and CD4 or CD8 for T cells and CD19 for B cells followed by high speed sorting on the MoFlo cell sorter (Beckman Coulter Inc.). In **Study 2**, each cell-type was isolated using MACS microbeads (negative selection). In **Study 3**, CSF cells were isolated by centrifuging freshly isolated CSF at 300rcf for 10 min. The supernatant was removed, and cells resuspended in 2ml of PBS and centrifuged again at 300rcf for 10 min and used immediately.

3.3 RNA AND DNA EXTRACTION

In **Study 1**, total RNA was extracted using the miRNeasy Mini Kit (Qiagen) and **Study 2** it was extracted using the RNeasy mini kit (Qiagen) as per the manufacturer's protocol. The integrity of RNA was measured using the Bioanalyzer (Agilent Inc.) RNA with an RNA integrity number (RIN) above 9 was used. In **Study 4**, genomic DNA was extracted using the MinElute Mammalian Genomic DNA miniprep kit (Qiagen). Quantity and purity of DNA and RNA was measured using the NanoDrop ND-1000 (Nanodrop Technologies Inc.).

3.4 TRANSCRIPTOMIC LIBRARY PREPARATION AND ANALYSIS

In **Study 1** and **Study 2**, transcriptomic (RNA-Seq) libraries were prepared using the Illumina TruSeq mRNA stranded library preparation kit (Cat.No. RS-122-1203) according to the manufacturers protocol. To reduce the effect of protocol and batch related confounders, samples for library preparation and sequencing were arranged non-sequentially separating cell type, sample group and gender. Quantity and quality of the libraries were measured using the Qubit (Invitrogen Inc.) and Bioanalyzer (Agilent Inc.). Molarity for sequencing was calculated using the Kapa library quantification kit (Cat. No. KK4827, Roche) as per the manufacturer's protocol. Sequencing was performed on the Illumina HiSeq 2500 as per the manufacturer's protocol with paired-end reads of length 75bp.

Data quality was assessed using FastQC. Low quality reads and adapter trimming was carried out using Cutadapt v 1.9.1. Alignment of reads was done using TopHat2 v2.1.1 with GRCh37 as reference (103). Read count per gene was done using Ht-seq . Only genes with a count per million (cpm) over 1 were used for downstream analysis. Normalization was done using CQN correcting for biases associated with gene length, library size and GC content (104). COMBAT was used to correct library and sequencing related batch effect (105). LIMMA was used for differential expression with linear models including disease or experimental group, age and gender as explanative variables (106).

3.5 CHROMATIN ACCESSIBILITY (ATAC-SEQ) LIBRARY PREPARATION AND ANALYSIS

In **Study 2**, ATAC-Seq libraries were prepared from 50,000 cells as per the protocol developed by Buenostro et.al. (68). Libraries were sequenced using the Illumina HiSeq 2500 as per the manufacturer's protocol generating 42bp single-end reads.

Data quality was assessed using FastQC and Cutadapt v1.9.1 was used for adapter trimming. Using GRCh38 as the genome reference, reads were aligned using Bowtie2 v2.2.6 (107). Open chromatin regions were identified using HOMER (108). Narrow and broad regions/peaks were identified separately and merged if they were within a distance of 100bp. Only regions present in more than 3 samples per cell type were chosen in order to obtain a set of consensus regions for downstream analysis. Regions with uncharacteristic high enrichment known as 'blacklist' regions as defined by ENCODE were discarded along with mitochondrial genes. After read count normalization per sample, only regions with a CPM greater than 1 were retained. Subsequent analysis was carried out using CQN, and COMBAT followed by differential binding using LIMMA as described in the transcriptomic analysis.

3.6 CHROMATIN FOOT-PRINTING

In **Study 2**, to determine the chromatin footprint associated with DNA binding protein we used Wellington (settings: -fp 10, 26, 2 -sh 11,36,1, FDR 0.01) on ATAC-Seq reads shifted by +4bp and -5bp on the positive and negative strand respectively (109). Identified regions with an FDR of <0.2 were used for motif scanning (FIMO) and annotation of motifs using TRANSFAC (69,110) to determine the transcription factors (TFs) bound at that respective

region. The identified TFs were filtered using the paired RNA-Seq data retaining only the TFs expressed on the gene level.

3.7 SINGLE CELL TRANSCRIPTOMIC LIBRARY PREPARATION AND ANALYSIS

In **Study 3**, freshly isolated CSF cells and PBMCs were loaded onto the Chromium Single Cell Controller using the Chromium Single Cell 5' library and gel bead kit (Cat.No. PN-1000006 and PN-1000014). Single cell library and TCR and BCR amplification was carried out using the Chromium Single Cell 5' Library Construction Kit, (Cat. No. PN-1000020), Chromium Single Cell V(D)J Enrichment Kit for Human T Cell, (Cat. No. PN-1000005), Chromium Single Cell V(D)J Enrichment Kit for Human B Cell, (Cat. No. PN-1000016) and Chromium i7 Multiplex Kit (Cat. No. PN-120262) as per the manufacturers protocol. Sequencing was carried out on the Illumina Next Seq and Miseq platforms. Read length was as per the 10X genomics protocol for single cell, TCR and BCR libraries.

Demultiplexing of data, initial data quality and alignment were carried out using the Cell Ranger pipeline v 3.1.0 (10X Genomics) according to the manufacturer's instruction. The alignment of data was done to the GRCh38 human genome reference. Read filtering, normalization, feature selection, scaling, data integration, anchoring, dimensionality and clustering of data were performed using Seurat 3 (111).

3.8 T AND B CELL RECEPTOR ANALYSIS

In **Study 3**, Demultiplexing, alignment, and initial quality control were performed using the cell ranger pipeline (10X Genomics). Rearrange clonetypes were annotated using VDJ tools with the VDJdb as reference (112,113).

3.9 DNA METHYLATION ARRAYS AND ANALYSIS

In **Study 4**, DNA methylation data was generated using the Infinium Human Methylation 450K bead chip arrays as per the manufacturer's protocol (Illumina). The analysis was carried out using the MinFi and ChaMP packages (114,115). BMIQ was used for normalization between type 1 and type 2 array probes. Probes with known SNPs, X and Y chromosome and not passing a detection p-value of 0.01 were excluded for downstream analysis. Differential methylation was carried out using LIMMA as described in the transcriptomic analysis above.

3.10 NON-PARAMETRIC COMBINATION ANALYSIS

In **Study 1** and **Study 4**, in order to combine data across celltypes, we implemented the 'omicsNPC' R function which is part of the STATegra R package (116).

3.11 SIRNA BASED GENE SILENCING OF T CELLS

In **Study 1**, to elucidate the function of SH3YL1, we performed an siRNA mediated silencing of it. PBMCs were isolated as described in section 3.2. from buffy coats. The negative selection of CD4 cells was performed (Miltenyi Biotec). 12M CD4+CD25- T cells from

individual donors were resuspended in 100ul of Nucleofection buffer solution form human primary T cells 12 Mio CD4+ CD25− T cells from individual donors were resuspended in 100 µl of Nucleofection® buffer solution for human primary T cells (Nucleofector™ Kits for Human T Cells, Lonza) containing 2 µM of ON-TARGETplus SH3YL1 siRNA pool or ON-TARGETplus non-targeting control pool (Dharmacon, GE). The cells were transfected using program U-014 of the Nucleofector™ 2b device using manufacturer's recommendations. After this the cells were transferred to pre-warmed X-VIVO 15 medium (Lonza) and incubated for 4.5 days. Post incubation for 5 hours, medium was changed. The cells were equally distributed for 3 time points; resting (0 hours), 6 hours and 24 hours. The cells for the 6 and 24 hour time points were stimulated with antibodies against CD3 (0.2 µg/ml, clone OKT3; Biolegend, LEAF grade; Cat. No. 317315) and CD28 (2 µg/ml, clone 15E8, Miltenyi Biotec, functional grade, Cat. No. 130-093-375) with goat anti-mouse Ig antibody as a cross-linker (2 µg/ml, Southern Biotech, cat no. 1010-01) mimicking TCR and co-stimulation at incubated at 37 °C and 5% CO2.

4 RESULTS AND DISCUSSION

The studies presented in this thesis cover genome and transcriptome-wide profiling techniques and analysis applied to immune cells known to play a role in Multiple Sclerosis. This section provides a brief summary of the results and discussion from each of the studies included in this thesis. Detailed results and discussion section can be found in the individual manuscripts.

4.1 STUDY 1

In this study, we explored the transcriptome-wide changes in gene expression using the Illumina platform. Using CD4 and CD8 T cells in 3 groups of samples, namely HC, RRMS and SPMS, we determined the changes in gene expression between HC and RRMS and, RRMS and SPMS to determine MS progression related changes. Due to a small sample size we detected few changes across these groups. To mitigate this, we adapted a methodology for non-parametric combination (NPC) of data, ie: we integrated the differential analysis output from each cell type to determine shared changes in the progression of MS. A strong justification for this integration came from both biological and data-driven reasons. NPC identified 149 differentially expressed genes. A majority of these genes fell into four groups depending on their expression pattern in progression, either being upregulated or downregulated from RRMS to SPMS or increasing or decreasing in expression from RRMS to SPMS. Overlapping these 149 genes with differentially methylated sites from the same samples and the same type of analysis, we obtain 24 and 18 pairs in CD4 and CD8 within a distance of 1Mb. Of these pairs, 1 overlapped between CD4 and CD8 and was associated with the gene SH3YL1, with the methylated region residing in the promoter of the same gene. The expression-methylation pattern suggested a downregulation of this gene from RRMS to SPMS. To determine the role of SH3YL1 in CD4 T cells, the gene was silenced and the cells were activated and harvested at 0, 6 and 24 hours. Using qPCR, *IL2* and *IFNG* gene expression were found to be upregulated post silencing between 0 and 6 hours. Transcriptome sequencing was performed on all the three time points from 4 donors using the same Illumina platform. DE analysis and gene-set enrichment analysis suggested SH3YL1 promotes activation and is a novel regulator of TCR-induced cytokine expression.

This profile determined by the biological processes, genes and coordinated epigenetic changes shows an evident dysregulation of T cells in MS. The genes detected in this analysis are both novel and confirmatory in nature. Importantly, within this work we show that the integration of multiple data-types can be a powerful method to determine novel changes in disease contexts which are complex and where sample numbers are limiting.

4.2 STUDY 2

In this study we profiled the paired chromatin accessibility (ATAC-Seq) and transcriptome (RNA-Seq) of CD4, CD8, CD14 and CD19 in patients newly diagnosed with MS and HCs. Differential chromatin accessibility analysis revealed 106, 30, 13 and 203 differentially bound (DB) regions in CD4, CD8, CD14 and CD19 respectively. Due to the small sample number there was limited statistical power to identify differences in chromatin accessibility. To

confirm the relevance of these regions in comparison to previously reported regions, we overlapped them with i) MS-associated SNPs and the corresponding SNPs in linkage disequilibrium with them, and ii) MS-specific DNA methylation data from the same immune cell types. MS-associated SNPs were enriched in regions associated with CD4 and CD8 T cells, while in MS-specific DNA methylation data CD4 and CD19 associated regions were enriched. The enrichment in these regions from previously published data confirmed the relevance of the DB regions identified in MS. Finally, the genes 1Mb upstream and downstream of each DBRs were identified per cell type. DE of these genes revealed 42, 2, 0, 1 DEGs in CD4, CD8, CD14 and CD19 respectively. Of the identified DBRs and DEGs in CD4, 25 genes were found to correlate between the two data types within a region of 1Mb suggesting regulation-dependent changes between the two.

Overall, the open-chromatin and transcriptomic profiles identify regions and genes that are MS-specific, some of these have been previously reported in MS and others are novel. Interestingly, the most chromatin and expression activity is observed in CD4 cells. B cells, on the other hand, have higher chromatin activity but little detectable activity on the gene expression level in this data. Specifically, we identify SERTAD1 and CCDC114 being differentially regulated in MS in CD4 and both had transcription factor binding sites recognized by EGR1. This study advances our knowledge of the factors that lead to the dysregulation of the immune system in MS.

4.3 STUDY 3

In this study, we profiled the immune cells in patients newly diagnosed with MS after relapse from blood and CSF. HCs were used for comparison, with all HCs being *HLA-DRB*15:01*. The CSF cells and PBMCs were profiled using 5' single cell transcriptomics (SCT). In addition, this was paired with T and B cell receptor profiling of the same cells. Using SCT, we define distinct transcriptomic cell clusters in CSF and PBMCs. Some of these clusters were found to vary in proportion between HC and MS. Most strikingly, plasma B cells were seen to increase in both compartments while memory B cells were seen to increase in CSF in MS. A few T cell clusters were also seen to increase in both CSF and PBMCs in MS. These cells are known to play a significant role in the pathogenesis of MS. Next, we performed differential expression analysis on each cluster between MS and HC. We identified some clusters having DE genes while others did not, indicating specific cell types play an important role in the pathogenesis of MS. Analysis of the paired T and B cell receptor identified an increased diversity of T and B cell receptors in the CSF in MS. These TCRs were found to recognize primarily CMV, EBV and *H Sapiens* antigens. All three of these have been studied extensively in the context of MS and are known to influence the pathogenesis of MS. Finally, leveraging the paired transcriptome and receptor data, we identified a specific subset of TCRs with more than 2 receptor chains and are elevated in a specific subset of immune cells in MS. Antigens binding these TCRs were found to be enriched for the IE gene of CMV and EBNA4 gene of EBV.

The results of this study demonstrate the power of unbiased techniques such as SCT in identifying cell types expanded in a complex disease such as MS where multiple immune cells play a role in pathogenesis. Profiling immune cells from CSF and blood demonstrates clear differences in activity and populations of certain immune cells in MS. In addition, using paired TCR and BCR data we can identify specific cells populations that are enriched for a specific receptor and the epitope they recognize. The results of this study are both of a confirmatory nature and novel.

4.4 STUDY 4

In this study we profiled the DNA-methylation status of immune cells (CD4, CD8, CD14, CD19) in HC, RRMS and SPMS. RRMS patients were primarily newly diagnosed with a few being on a drug wash-out period of up to 6 months. Differential methylation analysis detected 1511, 666 and 30 regions in CD19, CD14 and CD8 between RRMS, SPMS and HC. To increase the statistical power and use the multiple cell types as evidence, we integrated the output of the differential methylation test with a permutation-based non-parametric combination methodology. This analysis revealed 1976 DM probes (DMP) in all four cell types, 1273 DMPs in lymphocytes (CD4, CD8, CD19), 423 DMPs from T cells and 2782 DMPs in cells with antigen presentation capabilities. Some of these regions have been reported in MS previously while others were novel. To determine the relevance of these regions in MS, we overlapped them with previously published MS-associated SNP loci and found a significant enrichment in lymphocytes, T cells and CD14+ cells. Also, of 1976 regions differentially methylated regions in all four cell types, 13.2% of regions showed meQTL effect suggesting genetic roles for their presence.

Overall, using the adapted methodology we were able to identify disease-relevant methylated regions shared by 4 cell immune cell types in MS. This gives us evidence for the presence of shared regulation dependent changes in MS. In addition, this methodology allows us to leverage small sample cohorts to infer additional information.

5 CONCLUSIONS AND FUTURE PERSPECTIVES

The underlying theme of this thesis was to identify changes in the immune system that are consistent and reproducible in MS. The primary purpose was to better understand the mechanisms that are active in the development and progression of MS. This could, in turn, be used to provide better biomarkers for diagnosis and treatment of MS. The world health organization (WHO) defines biomarkers in its broadest sense as "any measurement reflecting an interaction between a biological system and a potential hazard, which may be chemical, physical, or biological. The measured response may be functional and physiological, biochemical at the cellular level, or a molecular interaction" (117). We used measurements at the DNA and RNA level since they are easy to extract and are quite stable for relatively long periods of time under easily accessible storage conditions. In addition, we primarily looked for these changes in blood since it is easily accessible from patients. Using blood we could identify disease-relevant changes since MS is thought to be triggered in the periphery. In addition, we could use it as a proxy for what is happening in the central nervous system where we know there is demyelination of neurons caused by immune cells which in turn leads to neurodegeneration. With each study, we then refined the sample selection process with the intention of increasing reproducibility or excluding some factors of variation in sampling such as time of sampling, age of patients and healthy controls, disease stage and medication prior to sampling.

Using high-throughput methodologies for profiling nucleic acids, we could then obtain genome and transcriptome-wide snap-shots of the state of immune cells in MS. This allowed us to identify changes that take place at multiple loci from a single assay. These changes could again be identified on multiple levels since any change in the activity of immune cells would require a complex interaction of multiple factors. For example, DNA methylation changes could alter transcription factor binding which would, in turn, alter gene expression or the presence of a certain SNP could affect the binding affinity of DNA binding proteins such as enhancers which again could lead to changes in gene expression. These resulting changes in gene expression could then carry forward to the protein level and affect protein activity. In order to adequately answer these question, we performed profiling methods that covered some aspects of gene regulation such as chromatin accessibility assays, DNA methylation status and gene expression. By pairing these data such as chromatin accessibility and transcriptomics and DNA-methylation and transcriptomics we were able to identify certain changes that appear to function in a coordinated fashion in the same sample. Such as in **Study 1** and **Study 2** we identify among many others, *SH3YL1*, *SERTAD1* and *CCDC114*.

Adequately combining multiple levels of data is a challenge since we see different changes from different cohorts. This can be a result of the variation seen in the disease itself and sampling. Overlapping the finding from these studies with previous studies, we were able to get a better understanding of why results do not obviously overlap. Gene regulation is a complex process with multiple factors affecting multiple targets. For example, a given enhancer can affect multiple genes but for the enhancer to function it needs to be within a

specific distance (1Mb) from the target genes or DNA methylation in open chromatin regions can prevent or enhance chromatin binding thus leaving certain genes in its vicinity more capable of being transcribed or silenced. To identify these regions of overlap in multiple factors we performed a co-localization analysis and find that there is an enrichment of loci obtained from different data types in close proximity to each other as seen in **Study 2** and **Study 4**.

MS involves multiple immune cells interacting, it was therefore only logical to study multiple immune cells in the same patients. MS is thought to be triggered in CD4 cells which subsequently recruit other immune cells which amplify the immune reaction and lead to demyelination. This would lead to changes in multiple immune cell types. Integrating the same data type across different cell types, we found that there were common loci across different immune cell types changing in MS. These shared loci give us insights into the susceptibility of the immune system to changes in MS as seen in **Study 1** and **Study 4**. A common finding across all the studies and all the cell types is the increased level of transcription related loci which confirms the increased activity of the immune system in MS.

The complexity of the interactions between immune cells is hard to adequately understand. As a first step towards this, we profiled immune cells in MS from PBMCs and CSF in **Study 3** at the single cell level. This profile primarily allowed us to answer the question of what cells were present in MS in an unbiased manner. In addition, we were able to connect the TCR and BCR to the immune cell subtype allowing us to identify potential triggering mechanisms and the subtypes of cells triggered. The specific mechanisms of the interaction of these cells, however, remain to be determined. Interestingly, from **Study 1** and **Study 3**, we see the presence of B cell related genes in T cells. Suggesting a closer yet unexplained interaction.

Having gathered this rich array of data primarily in newly diagnosed patients, we firstly understand sample size is a big limitation in identifying strong reproducible changes. Overcoming this with multiple layers of data requires further work. Secondly, changes in MS as a disease have to have an underlying commonality in mechanisms and a chain of events associated with them, which we presently miss between different cohorts associated with different studies. Thirdly, in spite of sampling newly diagnosed patients, we cannot take into consideration genetic and environmental factors since they are far too complex to decipher from a small sample size. Fourth, being newly diagnosed patients doesn't necessarily mean it is their first bout of disease but as a result we should have a small continuum of samples (Figure 1). To identify the changes associated with this continuum requires further work. Finally, since all regulation dependent changes would be connected, profiling at any level would allow inference of another level. Thereby leveraging these different levels of data towards a common set of mechanisms. All of this requires the application of existing and development of new methods of data analysis that are unsupervised and allow multiple combination possibilities. Together, this would give us an opportunity to understand the system better and work towards the development of robust biomarkers for diagnosis and therapy. Hopefully, now we are a little closer.

6 ACKNOWLEDGMENTS

What an incredible journey! It's been a long but very fulfilling one. In this time I have had the privilege of meeting many people that have taught me so much.

To all the MS patients and healthy controls who have been a part of the various studies.

My main supervisor, **Jesper Tegnér**, from the very first, very informal chat we had about science to our continued interactions, I have managed to imbibe some of your knowledge, experience and reluctantly also some of your behavioral traits. It's hard not to, you are quite unique in that sense. I am happy that I had the chance to work with you because you have given me the opportunity to grow into a more mature and independent researcher with a continued thirst for knowledge. As part of the unit of computational medicine we were afforded many luxuries which made it a special place to work and you facilitated that.

To my co-supervisor **Ingrid Kockum**, thank you for 'adopting' me. You became a larger part of my Ph.D. journey at a critical phase. Thereafter we worked more closely. You always made yourself available in spite of being terribly busy. It's been a pleasure learning and working with you and I am grateful that I had the opportunity to do so.

To my co-supervisor, **David Gomez-Cabrero**, my connection to the world of all things data and paella. With you I have managed to learn and do things I would never have imagined putting my foot into voluntarily. It's been a great learning experience and I look forward to the future it unearths.

Tomas, Maja and Fredrik, you'll are truly inspiring researchers. Your willingness to share your knowledge and create opportunities has made collaborating with you'll a truly enriching experience.

To the members of the unit of computational medicine, so many over the years but some of you had a bigger hand in my well-being and progress. **Pernilla**, as far as you were concerned I was your PhD student, always keeping a watchful eye. Navigating the new Swedish system and the PhD administrative landscape was made so much easier by the information and advice you shared. It was and will continue to be greatly appreciated. **Peri**, you made life in the lab so much easier. Always making sure we had everything we required so that we could focus on our experiments. We were all lucky to have you as our lab manager. My fellow PhD students (former and future) **Soudabeh**, **Rubin**, **Szabolcs** and **Giorgios**. We shared many triumphs and tribulations and are hopefully the better for it today. To a successful future. Thanks to **Matilda**, **Hiromasa**, **Angelika**, **Gordon**, **Venki**, **Francesco** and **Gilad** who collaborated, advised, shared experiences, knowledge or just sat down for coffee and a chat.

To the members of the neuroimmunology unit. **Ewoud**, what can I say that I never said before. Last donut? Thank you for the great scientific collaboration and even greater banter. It's been a pleasure. All the best in your coming endeavors. **Mohsen**, thanks for making it so much easier to work with patient samples and understand all the complexities associated with adequate

sampling. **Alexandra**, for help with collecting healthy controls. Thanks to **Faiez**, **Tojo** and **Gunn** for your input and efforts towards the different projects. Thanks to **Pernilla**, **Maria**, **Lara**, **Jesse** and **Thomas** for the scientific discussions and good times.

To **CMM IT** and administration, **Timmy**, **Daniel**, **Christer**, **Henrik**, **Ingela** and **Annika**. You guys are superstars. Thanks for all the help over the years.

To **Ujjwal** and **Anantha,** thank you for drawing me a map to Karolinska and the Unit of Computational Medicine.

The journey toward the Ph.D. has been possible because of to the opportunities and encouragement given to me by **Dr. Betty Daniel**, **Prof. Mitradas Panicker**, **Dr. Raja Mugasimangalam**, **Dr. Sudha Rao** and **Dr. Ashok Gopinath**. Thank you.

Over the years, I have made some great friends, be it in CMM, KI or Stockholm, who have been with me through thick and thin. You have enriched my life in more ways than I can adequately acknowledge. Know that I am truly grateful that I had a chance to meet and create lasting friendships with you.

To my family. You'll have given me so much. Dad, you were always an inspiration to reach for the stars. Quietly encouraging and always concerned. I miss you. Mum, you are such a strong person. You have taught me despite what life throws at you, to focus on what is best. Sunil and Sundeep my dear brothers, I have learned so much from you'll over the years. Looking forward to the day we are reunited on the same continent.

7 REFERENCES

1. Mallucci G, Peruzzotti-Jametti L, Bernstock JD, Pluchino S. The role of immune cells, glia and neurons in white and gray matter pathology in multiple sclerosis. Prog Neurobiol [Internet]. Europe PMC Funders; 2015 Apr [cited 2020 Jan 19];127–128:1–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25802011

2. GBD 2016 Multiple Sclerosis Collaborators MT, Culpepper WJ, Nichols E, Bhutta ZA, Gebrehiwot TT, Hay SI, et al. Global, regional, and national burden of multiple sclerosis 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol [Internet]. Elsevier; 2019 Mar 1 [cited 2020 Jan 19];18(3):269–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30679040

3. Filippi M, Bar-Or A, Piehl F, Preziosa P, Solari A, Vukusic S, et al. Multiple sclerosis. Nat Rev Dis Prim [Internet]. Nature Publishing Group; 2018 Dec 8 [cited 2019 Feb 4];4(1):43. Available from: http://www.nature.com/articles/s41572-018-0041-4

4. Harbo HF, Gold R, Tintoré M. Sex and gender issues in multiple sclerosis. Ther Adv Neurol Disord [Internet]. SAGE Publications; 2013 Jul [cited 2016 Aug 8];6(4):237–48. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23858327

5. Ascherio A, Munger KL, Simon KC. Vitamin D and multiple sclerosis. Lancet Neurol. 2010;9(6):599–612.

6. Tselis A. Evidence for Viral Etiology of Multiple Sclerosis. Semin Neurol [Internet]. © Thieme Medical Publishers; 2011 Jul 30 [cited 2016 Aug 8];31(03):307–16. Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0031-1287656

7. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science (80-) [Internet]. 2019 Sep 27 [cited 2019 Nov 21];365(6460):eaav7188. Available from: http://www.sciencemag.org/lookup/doi/10.1126/science.aav7188

8. Miller D, Barkhof F, Montalban X, Thompson A, Filippi M. Clinically isolated syndromes suggestive of multiple sclerosis, part I: natural history, pathogenesis, diagnosis, and prognosis. Lancet Neurol. 2005;4(5):281–8.

9. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. Ann Neurol. 2011;69(2):292–302.

10. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol [Internet]. Elsevier; 2018 Feb 1 [cited 2019 Nov 23];17(2):162–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29275977

11. Lassmann H, Brück W, Lucchinetti CF. The Immunopathology of Multiple Sclerosis: An Overview. Brain Pathol [Internet]. Blackwell Publishing Ltd; 2007 Apr [cited 2016 Aug 11];17(2):210–8. Available from: http://doi.wiley.com/10.1111/j.1750-3639.2007.00064.x

12. Westerlind H, Ramanujam R, Uvehag D, Kuja-Halkola R, Boman M, Bottai M, et al. Modest familial risks for multiple sclerosis: a registry-based study of the population of Sweden. Brain [Internet]. 2014 Mar [cited 2019 Nov 21];137(3):770–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24441172

13. Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kemppinen A, Cotsapas C, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet [Internet]. 2013;45(11):1353–60. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3832895&tool=pmcentrez&rendertype=abstract

14. Brodin P, Jojic V, Gao T, Bhattacharya S, Lopez Angel CJ, Furman D, et al. Variation in the Human Immune System Is Largely Driven by Non-Heritable Influences. Cell [Internet]. 2015 [cited 2016 Aug 11];160:37–47. Available from: http://dx.doi.org/10.1016/j.cell.2014.12.020

15. Harkiolaki M, Holmes SL, Svendsen P, Gregersen JW, Jensen LT, McMahon R, et al. T Cell-Mediated Autoimmune Disease Due to Low-Affinity Crossreactivity to Common Microbial Peptides. Immunity. 2009;30(3):348–57.

16. Münz C, Lünemann JD, Getts MT, Miller SD. Antiviral immune responses: triggers of or triggered by autoimmunity? Nat Rev Immunol [Internet]. Nature Publishing Group; 2009 Apr [cited 2016 Aug 11];9(4):246–58. Available from: http://www.nature.com/doifinder/10.1038/nri2527

17. Olson JK, Croxford JL, Calenoff MA, Dal Canto MC, Miller SD, Ota K, et al. A virus-induced molecular mimicry model of multiple sclerosis. J Clin Invest [Internet]. American Society for Clinical Investigation; 2001 Jul 15 [cited 2016 Aug 11];108(2):311–8. Available from: http://www.jci.org/articles/view/13032

18. Tengvall K, Huang J, Hellström C, Kammer P, Biström M, Ayoglu B, et al. Molecular mimicry between Anoctamin 2 and Epstein-Barr virus nuclear antigen 1 associates with multiple sclerosis risk. Proc Natl Acad Sci U S A [Internet]. National Academy of Sciences; 2019 Aug 20 [cited 2019 Nov 24];116(34):16955–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31375628

19. Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. Nat Rev Neurol [Internet]. 2017 Jan 9 [cited 2018 Dec 15];13(1):25–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27934854

20. Korn T, Mitsdoerffer M, Kuchroo VK. Immunological Basis for the Development of Tissue Inflammation and Organ-Specific Autoimmunity in Animal Models of Multiple Sclerosis. In Springer Berlin Heidelberg; 2009 [cited 2016 Aug 8]. p. 43–74. Available from: http://link.springer.com/10.1007/400_2008_17

21. Zamvil SS, Mitchell DJ, Moore AC, Kitamura K, Steinman L, Rothbard JB. T-cell epitope of the autoantigen myelin basic protein that induces encephalomyelitis. Nature [Internet]. Nature Publishing Group; 1986 Nov 20 [cited 2016 Aug 8];324(6094):258–60. Available from: http://www.nature.com/doifinder/10.1038/324258a0

22. Tuohy VK, Sobel R a, Lees MB. Susceptibility to PLP-induced EAE is regulated by non-H-2 genes. Ann N Y Acad Sci [Internet]. Blackwell Publishing Ltd; 1988 Nov [cited 2016 Aug 8];540(1 Advances in N):709–11. Available from: http://doi.wiley.com/10.1111/j.1749-6632.1988.tb27221.x

23. Mendel I, Kerlero de Rosbo N, Ben-Nun A. A myelin oligodendrocyte glycoprotein peptide induces typical chronic experimental autoimmune encephalomyelitis in H-2b mice: fine specificity and T cell receptor V beta expression of encephalitogenic T cells. Eur J Immunol [Internet]. WILEY-VCH Verlag GmbH; 1995 Jul [cited 2016 Aug

8];25(7):1951–9. Available from: http://doi.wiley.com/10.1002/eji.1830250723

24. Ransohoff RM, Kivisäkk P, Kidd G. Three or more routes for leukocyte migration into the central nervous system. Nat Rev Immunol [Internet]. Nature Publishing Group; 2003 Jul [cited 2016 Sep 18];3(7):569–81. Available from: http://www.nature.com/doifinder/10.1038/nri1130

25. McMahon EJ, Bailey SL, Castenada CV, Waldner H, Miller SD. Epitope spreading initiates in the CNS in two mouse models of multiple sclerosis. Nat Med [Internet]. Nature Publishing Group; 2005 Mar 27 [cited 2016 Sep 18];11(3):335–9. Available from: http://www.nature.com/doifinder/10.1038/nm1202

26. Pender MP, Tabi Z, Nguyen KB, McCombe PA. The proximal peripheral nervous system is a major site of demyelination in experimental autoimmune encephalomyelitis induced in the Lewis rat by a myelin basic protein-specific T cell clone. Acta Neuropathol [Internet]. Springer-Verlag; 1995 [cited 2016 Sep 18];89(6):527–31. Available from: http://link.springer.com/10.1007/BF00571507

27. Bielekova B, Sung MH, Kadom N, Simon R, McFarland H, Martin R. Expansion and functional relevance of high-avidity myelin-specific CD4+ T cells in multiple sclerosis. J Immunol (Baltimore, Md 1950) [Internet]. 2004 [cited 2016 Aug 8];172(6):3893–904. Available from: http://www.jimmunol.org/content/172/6/3893

28. Hellings N, Bare M, Verhoeven C, D'Hooghe MB, Medaer R, Bernard CCA, et al. T-cell reactivity to multiple myelin antigens in multiple sclerosis patients and healthy controls. J Neurosci Res [Internet]. John Wiley & Sons, Inc.; 2001 Feb 1 [cited 2016 Aug 8];63(3):290–302. Available from: http://doi.wiley.com/10.1002/1097-4547%2820010201%2963%3A3%3C290%3A%3AAID-JNR1023%3E3.0.CO%3B2-4

29. Reboldi A, Coisne C, Baumjohann D, Benvenuto F, Bottinelli D, Lira S, et al. C-C chemokine receptor 6–regulated entry of TH-17 cells into the CNS through the choroid plexus is required for the initiation of EAE. Nat Immunol [Internet]. Nature Publishing Group; 2009 May 22 [cited 2016 Sep 18];10(5):514–23. Available from: http://www.nature.com/doifinder/10.1038/ni.1716

30. Kebir H, Ifergan I, Alvarez JI, Bernard M, Poirier J, Arbour N, et al. Preferential recruitment of interferon-gamma-expressing TH17 cells in multiple sclerosis. Ann Neurol [Internet]. 2009 Sep [cited 2016 Sep 24];66(3):390–402. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19810097

31. Codarri L, Gyülvészi G, Tosevski V, Hesske L, Fontana A, Magnenat L, et al. RORγt drives production of the cytokine GM-CSF in helper T cells, which is essential for the effector phase of autoimmune neuroinflammation. Nat Immunol [Internet]. 2011 Jun [cited 2016 Sep 24];12(6):560–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21516112

32. Noster R, Riedel R, Mashreghi M-F, Radbruch H, Harms L, Haftmann C, et al. IL-17 and GM-CSF expression are antagonistically regulated by human T helper cells. Sci Transl Med [Internet]. 2014;6(241):241ra80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24944195

33. Frisullo G, Nociti V, Iorio R, Patanella AK, Marti A, Caggiula M, et al. IL17 and IFNgamma production by peripheral blood mononuclear cells from clinically isolated syndrome to secondary progressive multiple sclerosis. Cytokine [Internet]. 2008 Oct

[cited 2016 Sep 24];44(1):22–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18793860

34. Tzartos JS, Friese MA, Craner MJ, Palace J, Newcombe J, Esiri MM, et al. Interleukin-17 production in central nervous system-infiltrating T cells and glial cells is associated with active disease in multiple sclerosis. Am J Pathol [Internet]. 2008 Jan [cited 2016 Sep 24];172(1):146–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18156204

35. Kozovska ME, Hong J, Zang YC, Li S, Rivera VM, Killian JM, et al. Interferon beta induces T-helper 2 immune deviation in MS. Neurology [Internet]. 1999 Nov 10 [cited 2016 Sep 24];53(8):1692–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10563614

36. Miller A, Shapiro S, Gershtein R, Kinarty A, Rawashdeh H, Honigman S, et al. Treatment of multiple sclerosis with copolymer-1 (Copaxone): implicating mechanisms of Th1 to Th2/Th3 immune-deviation. J Neuroimmunol [Internet]. 1998 Dec 1 [cited 2016 Sep 24];92(1–2):113–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9916886

37. Zoghi S, Amirghofran Z, Nikseresht A, Ashjazadeh N, Kamali-Sarvestani E, Rezaei N. Cytokine secretion pattern in treatment of lymphocytes of multiple sclerosis patients with fumaric acid esters. Immunol Invest [Internet]. 2011 [cited 2016 Sep 24];40(6):581–96. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21510778

38. Venken K, Hellings N, Broekmans T, Hensen K, Rummens J-L, Stinissen P. Natural naive CD4+CD25+CD127low regulatory T cell (Treg) development and function are disturbed in multiple sclerosis patients: recovery of memory Treg homeostasis during disease progression. J Immunol [Internet]. 2008 May 1 [cited 2016 Sep 24];180(9):6411–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18424765

39. Yogev N, Frommer F, Lukas D, Kautz-Neu K, Karram K, Ielo D, et al. Dendritic cells ameliorate autoimmunity in the CNS by controlling the homeostasis of PD-1 receptor(+) regulatory T cells. Immunity [Internet]. 2012 Aug 24 [cited 2016 Sep 24];37(2):264–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22902234

40. Roychoudhuri R, Hirahara K, Mousavi K, Clever D, Klebanoff CA, Bonelli M, et al. BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis. Nature [Internet]. 2013 Jun 27 [cited 2016 Sep 24];498(7455):506–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23728300

41. Vahedi G, Kanno Y, Furumoto Y, Jiang K, Parker SCJ, Erdos MR, et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. Nature [Internet]. 2015;520(7548):558–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25686607

42. Venken K, Hellings N, Thewissen M, Somers V, Hensen K, Rummens J-L, et al. Compromised CD4+ CD25(high) regulatory T-cell function in patients with relapsing-remitting multiple sclerosis is correlated with a reduced frequency of FOXP3-positive cells and reduced FOXP3 expression at the single-cell level. Immunology [Internet]. Wiley-Blackwell; 2008 Jan [cited 2016 Sep 24];123(1):79–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17897326

43. Feger U, Luther C, Poeschel S, Melms A, Tolosa E, Wiendl H. Increased frequency of CD4+ CD25+ regulatory T cells in the cerebrospinal fluid but not in the blood of

multiple sclerosis patients. Clin Exp Immunol [Internet]. Wiley-Blackwell; 2007 Mar [cited 2016 Sep 24];147(3):412–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17302889

44. Dominguez-Villar M, Baecher-Allan CM, Hafler DA. Identification of T helper type 1-like, Foxp3+ regulatory T cells in human autoimmune disease. Nat Med [Internet]. 2011 Jun [cited 2016 Sep 24];17(6):673–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21540856

45. Schneider A, Long SA, Cerosaletti K, Ni CT, Samuels P, Kita M, et al. In Active Relapsing-Remitting Multiple Sclerosis, Effector T Cell Resistance to Adaptive Tregs Involves IL-6–Mediated Signaling. Sci Transl Med. 2013;5(170).

46. Bhela S, Kempsell C, Manohar M, Dominguez-Villar M, Griffin R, Bhatt P, et al. Nonapoptotic and extracellular activity of granzyme B mediates resistance to regulatory T cell (Treg) suppression by HLA-DR-CD25hiCD127lo Tregs in multiple sclerosis and in response to IL-6. J Immunol [Internet]. NIH Public Access; 2015 Mar 1 [cited 2016 Sep 24];194(5):2180–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25637022

47. Frischer JM, Bramow S, Dal-Bianco A, Lucchinetti CF, Rauschka H, Schmidbauer M, et al. The relation between inflammation and neurodegeneration in multiple sclerosis brains. Brain [Internet]. 2009 May [cited 2016 Aug 8];132(Pt 5):1175–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19339255

48. Sasaki K, Bean A, Shah S, Schutten E, Huseby PG, Peters B, et al. Relapsing-remitting central nervous system autoimmunity mediated by GFAP-specific CD8 T cells. J Immunol [Internet]. 2014 Apr 1 [cited 2016 Sep 24];192(7):3029–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24591371

49. Willing A, Leach OA, Ufer F, Attfield KE, Steinbach K, Kursawe N, et al. $CD8^+$ MAIT cells infiltrate into the CNS and alterations in their blood frequencies correlate with IL-18 serum levels in multiple sclerosis. Eur J Immunol [Internet]. 2014 Oct [cited 2016 Sep 24];44(10):3119–28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25043505

50. Tennakoon DK, Mehta RS, Ortega SB, Bhoj V, Racke MK, Karandikar NJ. Therapeutic induction of regulatory, cytotoxic CD8+ T cells in multiple sclerosis. J Immunol [Internet]. 2006 Jun 1 [cited 2016 Sep 24];176(11):7119–29. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16709875

51. Sospedra M. B cells in multiple sclerosis. Curr Opin Neurol [Internet]. 2018 [cited 2019 Nov 22];31(3):256–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29629941

52. Harp CT, Lovett-Racke AE, Racke MK, Frohman EM, Monson NL. Impact of myelin-specific antigen presenting B cells on T cell activation in multiple sclerosis. Clin Immunol [Internet]. 2008 Sep [cited 2020 Jan 19];128(3):382–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18599355

53. Obermeier B, Mentele R, Malotka J, Kellermann J, Kümpfel T, Wekerle H, et al. Matching of oligoclonal immunoglobulin transcriptomes and proteomes of cerebrospinal fluid in multiple sclerosis. Nat Med [Internet]. 2008 Jun 18 [cited 2020 Jan 19];14(6):688–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18488038

54. Mendenhall EM, Bernstein BE. Chromatin state maps: new technologies, new insights. Curr Opin Genet Dev [Internet]. 2008 Apr [cited 2016 Sep 22];18(2):109–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18339538

55. Defossez P-A, Stancheva I. Biological functions of methyl-CpG-binding proteins. Prog Mol Biol Transl Sci [Internet]. 2011 [cited 2016 Sep 22];101:377–98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21507359

56. Fletcher JM, Lalor SJ, Sweeney CM, Tubridy N, Mills KHG. T cells in multiple sclerosis and experimental autoimmune encephalomyelitis. Clinical and experimental immunology [Internet]. 2010 Oct [cited 2014 May 25];162(1):1–11. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2990924&tool=pmcentrez&rendertype=abstract

57. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature [Internet]. 2011 Aug 11 [cited 2016 Sep 24];476(7359):214–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21833088

58. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Kitagawa H, Ishikawa Y, Li Q, Watanabe C, editors. Nat Methods [Internet]. Nature Publishing Group; 2008;5(7):621–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18516045

59. Gandhi KS, McKay FC, Cox M, Riveros C, Armstrong N, Heard RN, et al. The multiple sclerosis whole blood mRNA transcriptome and genetic associations indicate dysregulation of specific T cell pathways in pathogenesis. Hum Mol Genet [Internet]. 2010 Jun 1 [cited 2018 Sep 16];19(11):2134–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20190274

60. Bomprezzi R, Ringnér M, Kim S, Bittner ML, Khan J, Chen Y, et al. Gene expression profile in multiple sclerosis patients and healthy controls: identifying pathways relevant to disease. Hum Mol Genet [Internet]. Oxford University Press; 2003 Sep 1 [cited 2016 Sep 24];12(17):2191–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12915464

61. Ratzer R, Søndergaard HB, Christensen JR, Börnsen L, Borup R, Sørensen PS, et al. Gene expression analysis of relapsing-remitting, primary progressive and secondary progressive multiple sclerosis. Mult Scler [Internet]. 2013;19(14):1841–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24085340

62. Gurevich M, Tuller T, Rubinstein U, Or-Bach R, Achiron A. Prediction of acute multiple sclerosis relapses by transcription levels of peripheral blood cells. BMC Med Genomics [Internet]. 2009 [cited 2016 Sep 24];2:46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19624813

63. Ottoboni L, Keenan BT, Tamayo P, Kuchroo M, Mesirov JP, Buckle GJ, et al. An RNA profile identifies two subsets of multiple sclerosis patients differing in disease activity. Sci Transl Med [Internet]. 2012 Sep 26 [cited 2016 Sep 24];4(153):153ra131. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23019656

64. Satoh J, Misawa T, Tabunoki H, Yamamura T. Molecular network analysis of T-cell transcriptome suggests aberrant regulation of gene expression by NF-kappaB as a biomarker for relapse of multiple sclerosis. Dis Markers [Internet]. 2008 [cited 2016 Sep 24];25(1):27–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18776589

65. Satoh J, Nakanishi M, Koike F, Onoue H, Aranami T, Yamamoto T, et al. T cell gene expression profiling identifies distinct subgroups of Japanese multiple sclerosis patients. J Neuroimmunol [Internet]. 2006 May [cited 2016 Sep 24];174(1–2):108–18. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16564577

66. Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc [Internet]. 2010 Feb [cited 2014 May 28];2010(2):pdb.prot5384. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3627383&tool=pmcentrez&rendertype=abstract

67. John S, Sabo PJ, Canfield TK, Lee K, Vong S, Weaver M, et al. Genome-Scale Mapping of DNase I Hypersensitivity. In: Current Protocols in Molecular Biology [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2013 [cited 2016 Sep 23]. p. 21.27.1-21.27.20. Available from: http://doi.wiley.com/10.1002/0471142727.mb2127s103

68. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Meth [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013 Dec;10(12):1213–8. Available from: http://dx.doi.org/10.1038/nmeth.2688

69. Matys V, Kel-Margoulis O V, Fricke E, Liebich I, Land S, Barre-Dirrie A, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res [Internet]. 2006 Jan 1 [cited 2016 Sep 23];34(Database issue):D108-10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16381825

70. Mathelier A, Zhao X, Zhang AW, Parcy FO, Worsley-Hunt R, Arenillas DJ, et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles.

71. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A [Internet]. 2009 Jun 9 [cited 2016 Sep 23];106(23):9362–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19474294

72. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. 2011;

73. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science (80-). 2012;337(6099).

74. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, et al. A map of open chromatin in human pancreatic islets. Nat Genet [Internet]. 2010 Mar [cited 2016 Sep 23];42(3):255–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20118932

75. Cheung VG, Spielman RS. Genetics of human gene expression: mapping DNA variants that influence gene expression. Nat Rev Genet [Internet]. Nature Publishing Group; 2009 Sep 28 [cited 2016 Sep 23];10(9):595–604. Available from: http://www.nature.com/doifinder/10.1038/nrg2630

76. James T, Lindén M, Morikawa H, Fernandes SJ, Ruhrmann S, Huss M, et al. Impact of genetic risk loci for multiple sclerosis on expression of proximal genes in patients. Hum Mol Genet [Internet]. Oxford University Press; 2018 Mar 1 [cited 2018 Sep 13];27(5):912–28. Available from: https://academic.oup.com/hmg/article/27/5/912/4792999

77. Hawkins RD, Larjo A, Tripathi SK, Wagner U, Luu Y, Lönnberg T, et al. Global chromatin state analysis reveals lineage-specific enhancers during the initiation of human T helper 1 and T helper 2 cell polarization. Immunity [Internet]. 2013 [cited 2016 Aug 9];38(6):1271–84. Available from: http://dx.doi.org/10.1016/j.immuni.2013.05.011

78. Borel C, Ferreira PG, Santoni F, Delaneau O, Fort A, Popadin KY, et al. Biased allelic expression in human primary fibroblast single cells. Am J Hum Genet [Internet]. Elsevier; 2015 Jan 8 [cited 2016 Sep 23];96(1):70–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25557783

79. Kim J, Marioni JC, Mortazavi A, Williams B, McCue K, Schaeffer L, et al. Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. Genome Biol [Internet]. BioMed Central; 2013 [cited 2016 Sep 23];14(1):R7. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-1-r7

80. Ginart P, Kalish JM, Jiang CL, Yu AC, Bartolomei MS, Raj A. Visualizing allele-specific expression in single cells reveals epigenetic mosaicism in an H19 loss-of-imprinting mutant. Genes Dev [Internet]. Cold Spring Harbor Laboratory Press; 2016 Mar 1 [cited 2016 Sep 23];30(5):567–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26944681

81. Kim JK, Kolodziejczyk AA, Ilicic T, Teichmann SA, Marioni JC. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. Nat Commun [Internet]. The Author(s); 2015 Oct 22;6:8687. Available from: http://dx.doi.org/10.1038/ncomms9687

82. Padovan-Merhar O, Raj A. Using variability in gene expression as a tool for studying gene regulation. Wiley Interdiscip Rev Syst Biol Med [Internet]. [cited 2016 Sep 23];5(6):751–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23996796

83. Xue Z, Huang K, Cai C, Cai L, Jiang C, Feng Y, et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA[thinsp]sequencing. Nature [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013 Aug 29;500(7464):593–7. Available from: http://dx.doi.org/10.1038/nature12364

84. Stubbington MJT, Lönnberg T, Proserpio V, Clare S, Speak AO, Dougan G, et al. T cell fate and clonality inference from single-cell transcriptomes. bioRxiv [Internet]. 2015;13(4):025676. Available from: http://biorxiv.org/content/early/2015/08/28/025676.abstract

85. Schafflick D, Xu CA, Hartlehnert M, Cole M, Lautwein T, Schulte-Mecklenbeck A, et al. Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. bioRxiv [Internet]. Cold Spring Harbor Laboratory; 2019 Apr 23 [cited 2020 Jan 19];403527. Available from: https://www.biorxiv.org/content/10.1101/403527v2.full

86. Schafflick D, Cole M, Hartlehnert M, Lautwein T, Buscher K, Wolbert J, et al. Single-cell transcriptomics identifies drivers of local inflammation in multiple sclerosis. bioRxiv [Internet]. Cold Spring Harbor Laboratory; 2018 Aug 29 [cited 2019 Nov 22];403527. Available from: https://www.biorxiv.org/content/10.1101/403527v1.full

87. Denis H, Ndlovu MN, Fuks F. Regulation of mammalian DNA methyltransferases: a route to new mechanisms. EMBO Rep [Internet]. 2011 Jul [cited 2016 Sep 22];12(7):647–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21660058

88. 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 [Internet]. 2009 May 15 [cited 2016 Sep 22];324(5929):930–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19372391

89. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science [Internet]. 2009 May 15 [cited 2016 Sep 22];324(5929):929–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19372393

90. Hewagama A, Richardson B. The genetics and epigenetics of autoimmune diseases. J Autoimmun [Internet]. 2009 Aug [cited 2016 Sep 22];33(1):3–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19349147

91. Kumagai C, Kalman B, Middleton FA, Vyshkina T, Massa PT. Increased promoter methylation of the immune regulatory gene SHP-1 in leukocytes of multiple sclerosis subjects. J Neuroimmunol. 2012;246:51–7.

92. Graves MC, Benton M, Lea RA, Boyle M, Tajouri L, Macartney-Coxson D, et al. Methylation differences at the HLA-DRB1 locus in CD4+ T-Cells are associated with multiple sclerosis. Mult Scler [Internet]. SAGE Publications; 2014 Jul [cited 2016 Sep 22];20(8):1033–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24336351

93. Mastronardi FG, Noor A, Wood DD, Paton T, Moscarello MA. Peptidyl argininedeiminase 2 CpG island in multiple sclerosis white matter is hypomethylated. J Neurosci Res [Internet]. Wiley Subscription Services, Inc., A Wiley Company; 2007 Jul [cited 2016 Sep 22];85(9):2006–16. Available from: http://doi.wiley.com/10.1002/jnr.21329

94. Calabrese R, Valentini E, Ciccarone F, Guastafierro T, Bacalini MG, Ricigliano VAG, et al. TET2 gene expression and 5-hydroxymethylcytosine level in multiple sclerosis peripheral blood cells. 2014;

95. Tumani H, Hartung H-P, Hemmer B, Teunissen C, Deisenhammer F, Giovannoni G, et al. Cerebrospinal fluid biomarkers in multiple sclerosis. Neurobiol Dis [Internet]. 2009 Aug [cited 2016 Sep 23];35(2):117–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19426803

96. Fitzner B, Hecker M, Zettl UK. Molecular biomarkers in cerebrospinal fluid of multiple sclerosis patients. Autoimmun Rev [Internet]. Elsevier B.V.; 2015;14(10):903–13. Available from: http://dx.doi.org/10.1016/j.autrev.2015.06.001

97. Housley WJ, Pitt D, Hafler DA. Biomarkers in multiple sclerosis. Clin Immunol. 2015;161:51–8.

98. Sieberts SK, Schadt EE. Moving toward a system genetics view of disease. Mamm

Genome [Internet]. Springer; 2007 Jul [cited 2016 Sep 23];18(6–7):389–401. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17653589

99. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L. The human disease network. Proc Natl Acad Sci U S A [Internet]. National Academy of Sciences; 2007 May 22 [cited 2016 Sep 23];104(21):8685–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17502601

100. Gomez-Cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merkenschlager M, Gisel A, et al. Data integration in the era of omics: current and future challenges.

101. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2015 Feb;16(2):85–97. Available from: http://dx.doi.org/10.1038/nrg3868

102. Böyum A. Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g. Scand J Clin Lab Invest Suppl [Internet]. 1968 [cited 2020 Jan 17];97:77–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/4179068

103. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol [Internet]. BioMed Central; 2013 [cited 2017 Feb 16];14(4):R36. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-4-r36

104. Hansen KD, Irizarry RA, WU Z. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics [Internet]. Oxford University Press; 2012 Apr 1 [cited 2017 Feb 16];13(2):204–16. Available from: https://academic.oup.com/biostatistics/article-lookup/doi/10.1093/biostatistics/kxr054

105. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics [Internet]. Oxford University Press; 2007 Jan 1 [cited 2017 Feb 16];8(1):118–27. Available from: https://academic.oup.com/biostatistics/article-lookup/doi/10.1093/biostatistics/kxj037

106. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res [Internet]. Oxford University Press; 2015 Apr 20 [cited 2017 Feb 16];43(7):e47–e47. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv007

107. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods [Internet]. Nature Publishing Group; 2012 Apr 4 [cited 2019 Jan 7];9(4):357–9. Available from: http://www.nature.com/articles/nmeth.1923

108. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell [Internet]. 2010 May 28 [cited 2019 Jan 2];38(4):576–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20513432

109. Piper J, Elze MC, Cauchy P, Cockerill PN, Bonifer C, Ott S. Wellington: a novel

method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Res [Internet]. Wiley, Hoboken, New Jersey, USA; 2013 Nov 1 [cited 2017 Aug 23];41(21):e201–e201. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt850

110. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics [Internet]. Oxford University Press; 2011 Apr 1 [cited 2019 Jan 9];27(7):1017–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21330290

111. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol [Internet]. Nature Publishing Group; 2018 May 2 [cited 2019 Nov 16];36(5):411–20. Available from: http://www.nature.com/articles/nbt.4096

112. Shugay M, Bagaev D V., Turchaninova MA, Bolotin DA, Britanova O V., Putintseva E V., et al. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. Gardner PP, editor. PLOS Comput Biol [Internet]. Public Library of Science; 2015 Nov 25 [cited 2019 Nov 16];11(11):e1004503. Available from: https://dx.plos.org/10.1371/journal.pcbi.1004503

113. Shugay M, Bagaev D V, Zvyagin I V, Vroomans RM, Crawford JC, Dolton G, et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. Nucleic Acids Res [Internet]. Oxford University Press; 2018 [cited 2019 Nov 16];46(D1):D419–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28977646

114. 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 [Internet]. 2014 May 15 [cited 2018 Dec 11];30(10):1363–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24478339

115. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics [Internet]. 2014 Feb 1 [cited 2018 Dec 11];30(3):428–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24336642

116. Karathanasis N, Tsamardinos I, Lagani V, Wang J, To ST, Cuartas I, et al. omicsNPC: Applying the Non-Parametric Combination Methodology to the Integrative Analysis of Heterogeneous Omics Data. Fröhlich H, editor. PLoS One [Internet]. Springer; 2016 Nov 3 [cited 2017 Mar 16];11(11):e0165545. Available from: http://dx.plos.org/10.1371/journal.pone.0165545

117. WHO Task Group on Biomarkers and Risk Assessment: Concepts and Principles., Robinson A, United Nations Environment Programme., International Labour Organisation., World Health Organization., International Program on Chemical Safety. Biomarkers and risk assessment : concepts and principles. World Health Organization; 1993. 82 p.