UNIVERSITA' VITA-SALUTE SAN RAFFAELE

CORSO DI DOTTORATO DI RICERCA INTERNAZIONALE
IN MEDICINA MOLECOLARE

Curriculum in Neuroscienze e Neurologia Sperimentale

CHARACTERIZATION OF THE GENOMIC PROFILE AND IMMUNE REPERTOIRE IN MULTIPLE SCLEROSIS PATIENTS TO PREDICT DISEASE ACTIVITY

Federica Esposito

DOS: Dr. Federica Esposito
Second Supervisor: Prof. An Goris

Tesi di DOTTORATO di RICERCA di Laura Ferrè
matr. 013916
Ciclo di dottorato XXXIV
SSD: MED/26

Anno Accademico 2020/2021

CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

La sottoscritta Laura Ferrè, n° matr. 013916

nata a Monza il 13/02/1987

autore della tesi dal titolo:
"CHARACTERIZATION OF THE GENOMIC PROFILE AND IMMUNE REPERTOIRE IN MULTIPLE SCLEROSIS PATIENTS TO PREDICT DISEASE ACTIVITY"

☐ **NON AUTORIZZA**

la consultazione della tesi stessa per 12 mesi a partire dalla data di conseguimento del titolo e precisamente dal 11/05/2022 al 10/05/2023 poiché ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di pubblicazione.

Data 10/04/2022 Firma .....

Declaration

This thesis has been composed by myself and has not been used in any previous application for a degree.

Throughout the text I use both ‘I’ and ‘We’ interchangeably.

All the analyses presented here were obtained by myself, except for:

1) The differential expression analysis was performed in collaboration with Dr. Ferdinando Clarelli, from the Laboratory of Human Genetics of Neurological Disorders, IRCCS San Raffaele Scientific Institute, Milan, Italy

2) The construction of the predictive model was performed by Prof Giorgio Valentini and Dr. Tommaso Fontana from the University of Milan,, Milan, Italy

All sources of information are acknowledged by means of reference.

Acknowledgements

I am forever grateful to my director of study, Dr. Federica Esposito, for the constant supervision and encouragement and the stimulating discussions; thank you for guiding me during all these years.

I sincerely thank my second supervisor, Prof. An Goris, for the time she devoted to counselling me and for her inspiring suggestions.

All my gratitude to Prof. Massimo Filippi, head of the Neurology, Neurorehabilitation and Neurophysiology Unit, for his full support and guidance in the clinic as well as in the research field.

Many special thanks to all the members of the Laboratory of Human Genetics of Neurological Disorders at OSR, and in particular to Dr. Melissa Sorosina, Dr. Silvia Santoro, Dr. Elisabetta Mascia and Dr. Martina Tosi that were in charge of the generation of genotyping data, RNA and TCR sequencing. Sincere thanks to Dr. Ferdinando Clarelli for the assistance in statistical analyses and mainly for the gene expression substudy.

A thank you to Prof. Giorgio Valentini and to Dr. Tommaso Fontana from University of Milan, who collaborated to the present project conducting the machine learning analyses.

Warm thanks to all the neurologists, nurses and administrative staff of the Neurology department and MS center for their help and support.

I sincerely thank all the teachers, secretariats and colleagues of the PhD course; it has been a stimulating and rewarding experience.

Finally, I want to acknowledge Fondazione Centro San Raffaele (FCSR)-Fronzaroli for supporting my PhD fellowship.

Moreover, the project on which the present PhD thesis relies upon has been financed by the Italian Ministry of Health [grant project: GR-2019-12368672].

Abstract

Multiple Sclerosis has a highly heterogeneous clinical course and, given the large number of available therapies, there is a strong need to identify prognostic markers of disease activity that can guide patients' management towards a more personalized approach. In the present thesis we investigated clinical, genomic and immunological parameters associated with inflammatory activity. To this aim, we enrolled two cohort of multiple sclerosis patients: an "Extended" cohort of ~1,000 subjects that started a first-line drug, with available clinical and genetic data, and a "Core" dataset of ~200 patients with genetic, transcriptomic and immune repertoire information obtained before treatment. The patients were observed for 4 years and classified according to the no evidence of disease activity status and the time to first relapse criteria. We then tested the relationship between the different –omics data and disease outcomes and we integrated the different layers of information into a prognostic model of disease activity. Our results confirmed that a younger age at onset, a shorter disease duration and female gender strongly correlate with higher inflammatory activity during follow-up. Besides, the genetic study highlighted some suggestive associations near to genes implicated in the regulation of coagulation system and vascular permeability as well as in antioxidant processes, regulation of oligodendrocyte differentiation and remyelination. These findings were strengthened by the results of the transcriptomic and pathway analyses that prioritized biological paths involved not only in immune functions but largely in cell homeostasis and death as well as neurodegeneration. Moreover, we also found that a higher immunological diversity seems to correlate with MS reactivation during follow-up. Finally, we applied machine learning algorithms to integrate clinical, genetic and immunological data and found that the best predictive performance was obtained using clinical information, while the addition of molecular data only slightly improved the prediction of disease activity. In conclusions, our findings demonstrated that genetic, transcriptomic and immune repertoire data can help in deciphering biological processes underlying Multiple Sclerosis pathophysiology and clinical expression but the predictive power of models integrating the different layers of information is currently not enough for application in clinical practice, probably because of the limited sample size of the studied cohort compared also to the number of thousands of features that are tested in the predictive model.

Table of contents

Acronyms and Abbreviations

MS: Multiple Sclerosis

CNS: central nervous system

MRI: magnetic resonance imaging

VitD: Vitamin D

IFNβ: interferon beta

CUA: combined unique active lesions

Th1: T helper 1 cells

Th11: T helper 11 cells

Treg: regulatory T cells

EBV: Epstein Barr Virus

sNfL: serum neurofilament light chain

EBNA: EBV nuclear antigens

RR: relapsing remitting

SP: secondary progressive

IL10: interleukin 10

MHC: Major histocompatibility complex

HLA: human leukocyte antigen

LD: linkage disequilibrium

GWAS: genome wide association study

MAF: minor allele frequency

IMSGC: International MS Genetic Consortium

SNP: single nucleotide polymorphism

NK: natural killer cells

DC: dendritic cells

eQTL: expression quantitative trait loci

BBB: blood brain barrier

WM: white matter

GM: gray matter

NAWM: normal appearing white matter

NAGM: normal appearing grey matter

CIS: clinically isolated syndrome

PP: primary progressive

Pr: progressive relapsing

DIT: dissemination in time

DIS: dissemination in space

CSF: cerebrospinal fluid

EP: evoked potential

OCBs: oligoclonal bands

DMTs: disease modifying treatments

RCTs: randomized controlled trials

ARR: annualized relapse rate

EDSS: expanded disability status scale

GA: glatiramer acetate

Teri: teriflunomide

DMF: Dimethylfumarate

PML: progressive multifocal leukoencephalopathy

NTZ: natalizumab

JCV: John Cunningham virus

FTY: fingolimod

S1P: sphingosine 1 phosphate

AAO: age at onset

Gd+: Gadolinium-enhancing

RNFL: retinal nerve fiber layer

OCT: optical coherence tomography

MSSS: multiple sclerosis severity score

TFR: time to first relapse

NEDA: no evidence of disease activity

CC: core cohort

TCR: T cell receptor

EC: extended cohort

OSR: Ospedale San Raffaele

EDA: evidence of disease activity

OR: odds ratio

DD: disease duration

HR: hazard ratio

MDS: multidimensional scaling

SNPs: single nucleotide polymorphisms

QQ: quantile-quantile

Chr: chromosome

QC: quality control

MAF: minor allele frequency

eQTL: expression quantitative trait loci

EAE: experimental autoimmune encephalomyelitis

KEGG: Kyoto Encyclopedia of Genes and Genomes

DEGs: differentially expressed genes

LE: number of leading edges

ES: enrichment score

NES: normalized enrichment score

FDR: false discovery rate

CDR3: complementarity-determining region 3

SC: Simpson clonality

ASCT: autologous stem cell transplantation

ML: machine learning

LSVM: linear support vector machine

DTs: decision trees

RFs: random forests

MLP: multi-layer perceptron

AUROC: area under the receiver operating characteristics

AUPRC: area under the precision-recall curve

Ensemble WA: weighted average ensemble

Ensemble PR: perceptron-based ensemble

AI: artificial intelligence

HWE: Hardy-Weinberg Equilibrium

PCs: principal components

List of Figures and Tables

Figures

Tables

1 Introduction

1.1 Multiple Sclerosis

Multiple sclerosis (MS) is an inflammatory demyelinating disease of the central nervous system (CNS), with more than 2 million cases worldwide (Feigin *et al*, 2017). It is considered an immune-mediated disease and is characterized by widespread CNS inflammation, demyelination and axonal loss (Trapp *et al*, 1998). It typically affects adults between 20 and 40 years of age, mainly females (female-to-male ratio 3:1 (Dilokthornsakul *et al*, 2016)), and represents the leading cause of non-traumatic neurological disability in young population in Western countries.

Although its etiology has not yet been defined, it is nowadays acknowledged that both genetic and environmental factors play an important role in its pathogenesis. Moreover, MS is characterized by a high heterogeneity in term of clinical presentation, disease course, magnetic resonance imaging (MRI) and pathological features and response to treatment, that stress the need for a highly personalized approach in its management.

1.2 Epidemiology and environmental risk factors

As already mentioned, both genetic and environmental risk factors contribute to MS susceptibility and clinical presentation. The importance of the environment is demonstrated by the particular trend in disease prevalence (Kurtzke *et al*, 1979), that increases with latitude (Vukusic *et al*, 2007; Simpson *et al*, 2011) (Figure 1.1). Indeed, northern Europe and North America show the highest prevalence with >30 cases per 100.000 inhabitants, while southern Europe and southern United States of America have a medium prevalence (5-30 per 100.000) and Asia and South America have fewer cases (<5 per 100.000). This geographical distribution can partly be due to genetic factors however several studies showed that MS incidence is associated with the country of residence during childhood (Elian *et al*, 1990; Detels *et al*, 1977; Dean & Kurtzke, 1971; Alter *et al*, 1962) and that subjects moving from high-risk to low-risk areas early in life show a reduced risk (Kurtzke *et al*, 1985; Gale & Martyn), stressing the role of acquired, environmental factors.

Figure 1.1: Geographical distribution of MS. *Map showing areas of high versus low prevalenceMS (from https://multiplesclerosis.net/, last accessed 2018)*

Furthermore, another aspect supporting the key role of the environment is the dramatic increase in the female-to-male ratio (Koch-Henriksen & Sørensen, 2010; Orton *et al*, 2006), that raised from about 1.5:1 in 1950s to up to 3.1:1 (Dilokthornsakul *et al*, 2016), most likely due to female-specific changes in smoking habit.

The following paragraphs summarize the available evidences supporting a role for environmental factors in MS.

1.2.1 Sun exposure and Vitamin D levels

Exposure to sunlight, and specifically to ultraviolet B radiation, is the major determinant of Vitamin D (VitD) levels that consequently decrease with increasing latitudes. Hence, VitD was suggested to be a possible mediator of the "latitude effect" in MS, able to explain its peculiar geographical distribution.

The association between VitD levels and MS was proved by two nested case-control study that showed a higher incidence of MS among subjects with low VitD levels, measured before disease onset (Salzer *et al*, 2012; Munger *et al*, 2006). Moreover, different studies showed a correlation of VitD levels with disease activity (Fitzgerald *et al*, 2015; Runia *et al*, 2012; Ascherio *et al*, 2014), suggesting a beneficial effect throughout disease course. For this reason, a recent randomized controlled study (Hupperts *et al*, 2019) evaluated the effect of VitD supplementation when added to interferon beta (IFNβ) treatment versus IFNβ alone: it failed to demonstrate any effect on relapse rate and disability worsening, but showed benefit in terms of neuro-radiological disease activity; specifically, there was a 30% reduction in the number of

combined unique active (CUA) lesions – defined as the sum of new gadolinium enhancing (Gd+) and/or new/enlarging T2-weighted lesions - at the brain MRI scans and a 50% decrease in the occurrence of new, hypointense T1 lesions at 48 weeks in the supplementation group. Moreover, additional support came from a mendelian randomization analysis that described an increased MS risk in presence of genetic variant linked to low VitD levels (Mokry *et al*, 2015).

The mechanisms through which VitD lower MS risk are not yet clear, but the active form of VitD, 1,25-dihydroxycholecalciferol, has been shown to modulate immune functions, mainly on CD4+ T cells, by reducing T helper 1 (Th1) and T helper 17 (Th17) activity and inducing a shift towards T regulatory (Treg) cells (Hayes *et al*, 2015).

1.2.2 Epstein Barr Virus infection

Due to the immune-mediated pathogenesis of MS, several kinds of pathogens have been investigated as possible triggers for disease onset. Among the many infectious agents hypothesized to play a role in MS, Epstein Barr Virus (EBV) is the most robustly and consistently associated. Late EBV infection in adolescence and early adulthood, and a history of infectious mononucleosis have been associated with an increased MS risk (Haahr *et al*, 2004) and, noteworthy, almost 100% of MS patients are seropositive for EBV. A very recent longitudinal study that prospectively evaluated around 10 million individuals in the US military, of whom 955 developed MS, showed that MS risk increases 32-times after EBV infection and that serum neurofilament light chain (sNfL) levels also increase after seroconversion, suggesting an active role for the virus in MS(Bjornevik *et al*, 2022). Indeed, high titers of Immunoglobulin G (IgG) antibodies directed against the EBV nuclear antigens (EBNA) seem to be most predictive of MS development (Munger *et al*, 2011). The mechanism by which EBV infection increase MS risk is not clear but molecular mimicry and presence of cross-reactive T cells and antibodies have been proposed. A direct EBV infection of the CNS has also been hypothesized, but evidence to support it are still missing (Lassmann *et al*, 2011).

Overall, this data are consistent with the so called "*hygiene hypothesis*" (Bach, 2002), which postulates that exposure to infectious stimuli early in life participates to the

correct development of the immune system, while a more hygienic setting and later infections lead to a dysregulation in Th1/Th2 balance.

1.2.3 Gut microbiota

Recent researches also suggested that the gut microbiota, that is, the combination of bacteria that physiologically colonize the human intestine, is altered in MS patients and could regulate T cell activity, possibly contributing to disease pathogenesis (Hindson, 2017). In the experimental autoimmune encephalomyelitis (EAE), the animal model of MS, transgenic mice raised in germ-free conditions appeared to be protected against disease development until the introduction of commensal microbiota, suggesting a key role played by the commensal gut flora(Berer *et al*, 2011). Indeed, several studies have shown that the microbiota of MS patients is altered compared to controls(Chen *et al*, 2016; Tremlett & Waubant, 2018; Jangi *et al*, 2016), and these differences have also been correlated with changes in the immune transcriptome(Jangi *et al*, 2016). However, further studies are needed to prove the causal role of these changes and to investigate possible therapeutic interventions aimed at modulating the gut microbiome.

1.2.4 Smoking status

The risk of MS is higher in cigarette smokers compared to non-smokers, with an estimated relative risk of 1.48 [C.I. 1.35 - 1.63](Handel *et al*, 2011). Moreover, smoking habit has also been linked to a higher risk of conversion from a relapsing remitting (RR) to a secondary progressive (SP) form of the disease (Healy *et al*, 2009; Hernán *et al*, 2005). A direct toxic effect of some smoke components, as well as an indirect systemic effect mediated by the peribronchial lymphatic tissue, have been hypothesized.
Interestingly, the change in MS female-to-male ratio in the last decades is thought to reflect gender-specific changes in the smoking behavior, further supporting a causal relationship. Hence, an active campaign supporting smoke cessation could potentially help reducing MS burden.

1.2.5 Other environmental risk factors

Several studies demonstrated that obesity during childhood and early adolescence is associated with a 2-fold increase in MS risk (Hedström *et al*, 2012), with

an effect that is more evident for women than men likely due to smaller cohorts and reduced statistical power in male studies (Gianfrancesco *et al*, 2014; Munger *et al*, 2013). Among the possible explanations, a reduction in VitD levels in obese individuals (Pereira-Santos *et al*, 2015) has been proposed, but a role for leptin cannot be excluded (Matarese *et al*, 2010).

A role for melatonin has also been suggested in modulating MS disease course (Farez *et al*, 2015b); melatonin reduces activation of proinflammatory Th17 cells also promoting secretion of anti-inflammatory interleukin 10 (IL10) and has been reported to ameliorate the murine model of MS (Álvarez-Sánchez *et al*, 2015).

Finally, a diet with a high salt intake was proposed to predispose to MS exacerbations (Farez *et al*, 2015a), but the finding was not confirmed by further studies.

1.3 Genetic background

1.3.1 Familial aggregation

The first concerns about the presence of a genetic predisposition in MS arose with studies showing familial aggregation (Sadovnick & Baird, 1988; Robertson *et al*, 1996; Prokopenko *et al*, 2003; Carton *et al*, 1997): first, second, and third-degree relatives of people with MS have a higher risk to develop the disease compared to the general population, which increases with the degree of relatedness.

Nevertheless, MS is not a classic mendelian disease, where a single gene is responsible for disease segregation in families; in fact, the increase in disease risk does not grow linearly with the amount of shared DNA, pointing to a multigenic model of inheritance (Risch, 1990). The extent of familial aggregation can be best expressed in term of λs, obtained dividing the risk of affected siblings by the population risk, which in MS is estimated between 20 and 40 (Oksenberg *et al*, 1996).

However, a limitation of family studies is that it is often not easy to separate the effect of genetics from the influence of a shared environment. For this reason, MS incidence in adoptees has been investigated (Ebers *et al*, 1995): despite identical environmental exposure, MS risk in adopted relatives is the same as in general population. Obviously, these data do not deny the influence of environmental factors on disease susceptibility but demonstrate that familial aggregation is due to a genetic component.

Indeed, studies in identical twins led to the same conclusion: concordance rate in monozygotic twins is 25%, while in dizygotic twins is 2-5%, confirming the significant contribution of genetic factors (Willer *et al*, 2003; Hansen *et al*, 2005), that can also influence disease phenotype (Sadovnick *et al*, 2009). Conversely, the remaining, relevant proportion of monozygotic twins are discordant for the disease, stressing the role of acquired factors.

1.3.2 Genetic discoveries in MS

1.3.2.1 Major histocompatibility complex (MHC) region

The first genetic factor associated with MS was discovered through a linkage study in 1972, in the human leukocyte antigen (HLA) class I region. The HLA cluster is located on chromosome 6p21.3 and encompass more than 200 genes within 4.5 megabases, with important roles in maturation, maintenance and regulation of the T cell repertoire, as well as in other immunological processes.

The first studies pointed to HLA class I antigens A3 and B7 (Naito *et al*, 1972; Jersild *et al*, 1972) and were followed by observation that also HLA class II polymorphisms are associated with the disease (Jersild *et al*, 1973). Specifically, the strongest association involves the "HLA-DR15 haplotype"(Hauser *et al*, 1989), including HLA DRB1*1501 and DQB1*0602 that are almost invariantly found together in individual of European ancestry. For this reason only with a study in a mixed African and American population (Oksenberg *et al*, 2004) it was possible to demonstrate that the association signal was driven by the HLA DRB1*1501, that confers a 3-fold increased risk of MS, and that other associations were mostly due to linkage disequilibrium (LD) with it.

Nevertheless, when accounting for the association signal in this class II antigen, other HLA variants resulted to be independently associated with the disease, such as the protective HLA A*0201 (Fogdell-Hahn *et al*, 2000) as well as some non-HLA variants in the locus (Sadovnick & Baird, 1988).

The most recent genome-wide association study (GWAS) in MS (Patsopoulos *et al*, 2019), which evaluated 47,351 MS patients and 68,284 healthy controls, identified up to 32 independent variants associated with MS in the MHC region and also highlighted the existence of interaction effects of HLA DRB1*1501 mainly with HLA*DQB1 and HLA*DQA1.

Overall, the MHC locus alone was estimated to explain 20% of the narrow-sense heritability in MS (Patsopoulos *et al*, 2019), that is, the proportion of MS susceptibility variance due to genetic factors.

1.3.2.2 Non MHC regions

Thanks to technical improvements that allowed to easily genotype the whole genome of thousands of individuals, the knowledge of MS genetic architecture has advanced extraordinarily. Several case-control GWAS have been performed that discovered hundreds of additional non-MHC genetic risk variants (Patsopoulos *et al*, 2019). GWAS are particularly successful in identifying new disease variants, because they do not require "a priori" assumptions on disease pathogenesis, as candidate gene studies do. In these kind of investigations, single nucleotide polymorphisms (SNPs) spread all over the genome are typed and compared in cases and controls to identify allelic differences between the two groups: due to the LD structure that links alleles in different genetic markers, this allow to identify interesting regions associated with the disease; subsequently, fine mapping of the selected regions is used to investigate the truly causal variants. Furthermore, GWAS are particularly powerful in detecting common alleles (minor allele frequency (MAF)>5%) with modest effect, that are likely to play a role in complex diseases.

The first MS GWAS (Hafler *et al*, 2007) was performed in 2007 on 931 trio families – consisting of an MS case and both parents - using 334,923 SNPs: it identified a SNP in the interleukin-7 receptor (*IL7R*) locus (p-value 2.94E−07, OR 1.18) and an additional signal in the interleukin-2 receptor (*IL2RA*) locus (p-value 2.96E-08 OR 1.25) as associated with the disease (Lundmark *et al*, 2007; Gregory *et al*, 2007). Since then, several other GWAS have been performed in MS; in particular, the collaborative efforts of the International MS Genetic Consortium (IMSGC) to put together MS cohorts from several countries in order to increase the statistical power for detection of significantly associated loci, led to a dramatic increase in the number of MS-associated variants. In 2011, a study that analyzed almost 30,000 individuals, including a discovery and a replication cohort (Sawcer *et al*, 2011), allowed to confirm most of the already known risk variants and reported 29 new associated single nucleotide polymorphisms (SNPs). Noteworthy, the prioritized genes were enriched in genes involved in lymphocyte

functions, such as T cell activation and proliferation, as well as in VitD metabolism. Genes representing known targets of MS therapies (e.g. *VCAM1* for Natalizumab (NTZ) and *IL2RA* for Daclizumab) were also selected, stressing the benefit of a collaborative strategy.

Interestingly, a significant proportion of MS risk variants are shared with other autoimmune diseases (Richard-Miceli & Criswell, 2012), such as type 1 diabetes, rheumatoid arthritis and intestinal bowel diseases, implying the existence of common genetic basis predisposing to immune dysregulation. Nevertheless, in some instances, genetic variant show opposite effect in different autoimmune disorders, increasing the risk of a specific disease while being protective for another one.

This insight prompted the development of a genetic array platform specifically created to interrogate ~200,000 immune-related loci, known as ImmunoChip; it was used to genotype more than 80,000 individuals of European ancestry and enabled the detection of 48 additional MS loci (Beecham *et al*, 2013).

Finally, the greatest advancement in the knowledge of MS genetic basis was made possible by the recent meta-analysis from the IMSGC (Patsopoulos *et al*, 2019), that involved more than 100,000 individuals and almost doubled the amount of known MS risk loci.

The circos plot in Figure 1.2 gives an overview of the association signals with MS: 200 autosomal SNPs outside the MHC region and one variant on the X chromosome reached the genome-wide threshold and were validated as MS risk variants in the meta-analysis.

Noteworthy, as previously observed, the prioritized variants are enriched in SNPs which, according to gene expression atlas, play a key role in immune cells of both the adaptive (T and B cells) and the innate (natural killer (NK) and dendritic cells (DC)) compartments; on the other hand no enrichment was found in genes expressed in CNS tissues and resident cells apart from microglial cells, suggesting a role for the local immune compartment. Similarly, the study of gene expression regulation (eQTL effect) and the results of a pathway analysis supported the central role of immune cells.

In conclusion, the last decade has witnessed an impressive progress in the knowledge of MS genetic background but, despite the increasing number of detected risk variants, we are still not able to explain the whole amount of estimated heritability in MS.

This so called “missing heritability” issue is well known in MS (Lill, 2014) and is shared by other complex disorders (Manolio *et al*, 2009), thus stressing the need for complementary approaches to be applied next to GWAS in the coming years.

Indeed, studies investigating the role of rare/low frequency variants that are not adequately captured by GWAS are already ongoing as well as studies investigating the presence of gene-gene or gene-environment interactions, that will help broaden our understanding of MS genetic complexity.

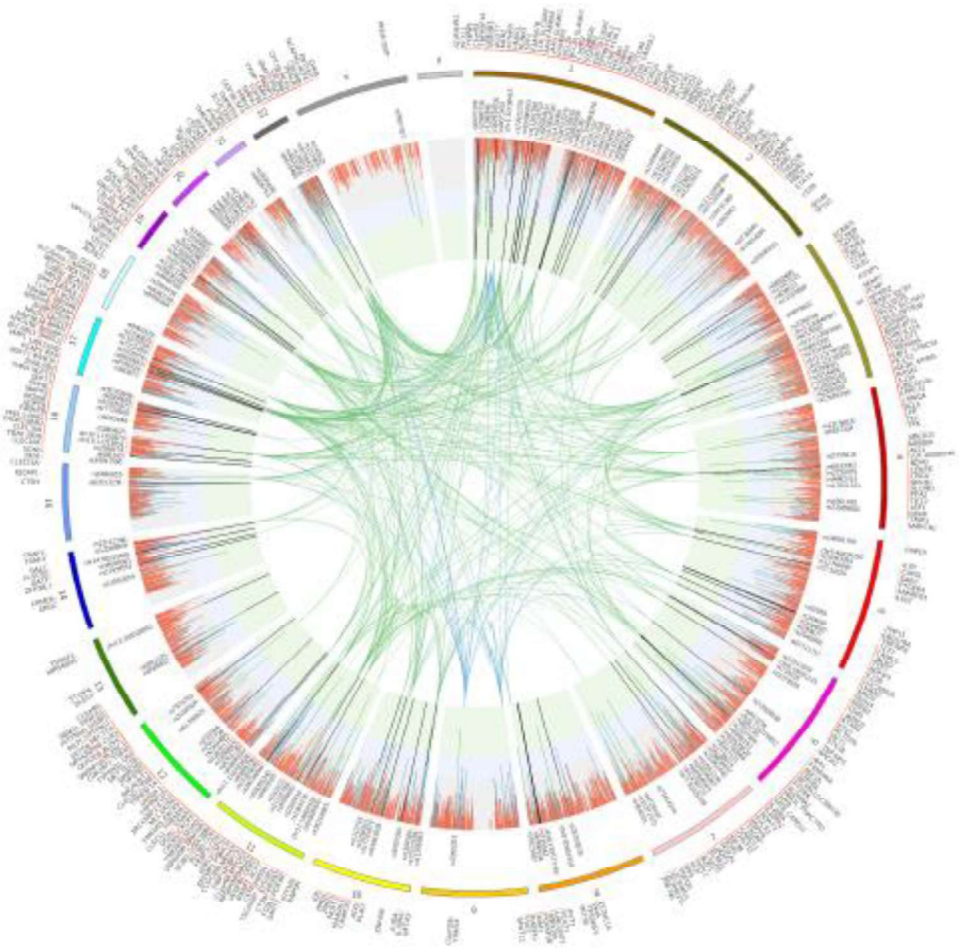


Figure 1.2: MS associated signals throughout the genome from Patsopoulos et al. (Patsopoulos *et al*, 2019)**.** *The circus plot gives an idea of the distribution of MS associated regions throughout the whole genome. Chromosomes are arranged circularly and are annotated with associated genes (outer layer) while the inner circle shows the corresponding manhattan plot (genome wide threshold defined by the green background).*

1.4 MS pathology and pathogenesis

1.4.1 Histopatological features

MS is a chronic demyelinating disease of the CNS for whom an autoimmune pathogenesis has been proposed and characterized, at the pathological level, by inflammation and neurodegeneration (Frohman *et al*, 2006; Lassmann, 2018).

Since the first description in the nineteen century (Charcot, 1880), it is known that acute MS lesions in the CNS consist of areas of demyelination, blood brain barrier (BBB) breakdown, T, B lymphocytes and plasma cell infiltration as well as oligodendrocyte loss; activated astrocytes are also present in acute lesions while, in later phases, variable degrees of reactive gliosis and axonal degeneration are found. Besides, after the acute stage a certain amount of remyelination is also possible (Nait-Oumesmar *et al*, 2007), but highly variable.

Demyelinating plaques are preferably located in myelin-rich regions such as the white matter (WM) of optic nerves, periventricular regions, brainstem and spinal cord; lesions usually have an ovoid shape and typically show a perivenous distribution perpendicular to the ventricles that give rise to the so called "*Dawson fingers*". Furthermore, gray matter (GM) demyelination has been shown in the spinal cord (Gilmore *et al*, 2006), cerebral (Kidd *et al*, 1999) and cerebellar cortex(Kutzelnigg *et al*, 2007) as well as at subpial level, where it is prevalent in progressive stages of the disease.

Alongside with focal lesions, widespread damage in normal appearing white matter (NAWM) and gray matter (NAGM)(Klaver *et al*, 2015) is seen, consisting of perivascular inflammatory infiltrates, diffuse microglial activation, diffuse axonal injury and astrocytic gliosis.

Noteworthy, four main histopathological patterns of WM lesions have been identified in MS patients (Lucchinetti *et al*, 2000), that remain constant throughout disease stages. These findings suggest that at least part of MS phenotypic heterogeneity can be explained by distinct pathological backgrounds.

1.4.2 Inflammation

Inflammation is the key feature of the disease and is present throughout its course. During disease reactivation, perivascular inflammation is characterized by increased BBB permeability with consequent infiltration of lymphocytes (mostly CD8+

T cell and to a lesser degree CD4+ T cells, B cells and plasma cells) as well as macrophages and activated microglia. The acute phase is also associated with the expression of inflammatory cytokines, such as interleukin 12 (IL12) and tumor-necrosis factor (TNF)(Hofman *et al*, 1989; Windhagen *et al*, 1995), and of chemokines and adhesion molecules that promote immune cells migration (Holman *et al*, 2011; Cannella & Raine, 1995).

Owing to the pathological features of acute lesions and their cytokine profile, MS has classically been considered a Th1 cell-mediated disease. Nonetheless, recently other cell types from both the adaptive immune compartment (e.g. Th17 and B cells) and the innate immunity (e.g. NK cells, DCs, and microglia) have been implicated in the disease. Indeed, meningeal follicle-like structures associated with prominent B cell infiltration and severe demyelination of the underlying cortex have been detected in MS brains (Magliozzi *et al*, 2006). Moreover, chronic active and slowly expanding lesions characterized by central demyelination surrounded by a rim of activated microglia and macrophages were described in progressive forms (Prineas *et al*, 2001).

1.4.3 Neurodegeneration

The other hallmark of MS is neurodegeneration, characterized by axonal loss and primary demyelination without acute inflammatory infiltrates (Henderson *et al*, 2009). Whether neurodegeneration is independent or secondary to CNS inflammation is still to be elucidated; some works showed evidence of mitochondrial dysfunction, oxidative damage and iron brain accumulation in MS suggesting that they could represent the common final steps in neurodegeneration (Fischer *et al*, 2012; Campbell *et al*, 2011; Lassmann *et al*, 2012). Therefore, additional studies are needed to disentangle the relationship between neuronal injury and CNS inflammation.

A comprehensive description of MS histopathological features can be found elsewhere (Lassmann *et al*, 2007; Kutzelnigg & Lassmann, 2014; Lassmann, 2018) and we will not further comment on it because it is beyond the scope of the present thesis.

1.5 Clinical presentation

1.5.1 Clinical manifestations

Typically, MS presents with acute neurological symptoms, lasting at least 24 hours and followed by a variable degree of spontaneous recovery. This first clinical event, termed clinically isolated syndrome (CIS), can affect one or more sites in the CNS and can represent a single, monophasic episode; however, when accompanied by WM lesions evident on brain MRI scans it confers a 80% risk of developing a second MS attack within 20 years (Fisniku *et al*, 2008).

Hence, if the disease is not recognized and effectively treated, several acute relapses can follow in the so called RR phase leading to accumulation of neurological impairment; in addition eventually, in most people (80%) a more progressive and insidious accumulation of disability takes over, independently of relapses, termed SP phase (Lublin & Reingold, 1996). A small proportion of MS cases (10-15%) shows a Primary Progressive (PP) course with slow and constant disability accumulation since the onset, with (progressive relapsing (PR)) or without superimposed relapses(Thompson *et al*, 1997). Whether RRMS/SPMS and PPMS represent separate entities or the extreme phenotypes of the same pathological process is still disputed. Though, the huge overlap in MRI, histopathological (Lassmann *et al*, 2012) and genetic characteristic (Vyshkina *et al*, 2005; Guaschino *et al*, 2014) suggests that the latter hypothesis might be true.

As for the clinical presentation, MS is highly heterogeneous and almost every kind of neurological symptoms can occur during an MS relapse, including pyramidal motor dysfunction, spasticity, ataxia, optic neuritis, sensory impairment, bladder and bowel disturbances, fatigue and cognitive impairment. Conversely the progressive phase is characterized by a predominant motor and sensory impairment, usually starting in the legs and eventually affecting the upper limbs and bulbar region; predominant cerebellar features are also possible.

We will not enter into further details about the clinical symptoms of MS, because it is not in the purpose of the present thesis.

1.5.2 Clinical phenotypes

Recently, the 1996 classification (Lublin & Reingold, 1996) of in RR, SP, PR and PP-MS has been revised to include information on disease activity and progression during the last year prior to observation (Lublin *et al*, 2014).

In the 2013 revision, CIS were included in the spectrum of MS phenotypes while PRMS definition was dropped; besides, for each MS subtype, a clinical and neuro-radiological assessment of disease activity is now requested at least annually to classify the disease into "active" or "not active" forms (Figure 1.3).

For SPMS and PPMS an evaluation of whether a progression of disability has occurred over the previous year should also be performed, to classify the disease as "with progression" or "without progression" (Figure 1.4)

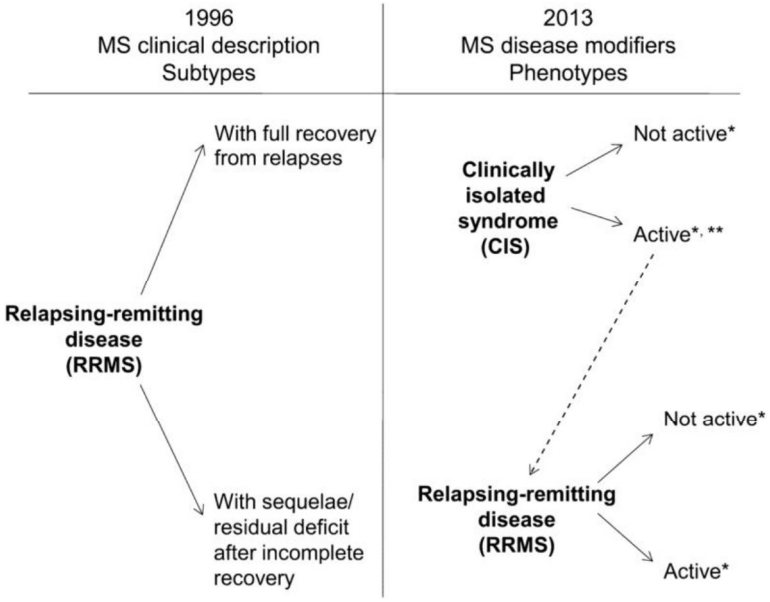


Figure 1.3: The 1996 vs 2013 MS phenotype descriptions for relapsing disease from Lublin et al.(Lublin et al, 2014). **Activity is determined by clinical relapses and/or MRI activity (contrast-enhancing lesions; new or unequivocally enlarging T2 lesions assessed at least annually); if assessments are not available, activity is "indeterminate." CIS, if subsequently clinically active and fulfilling current MS diagnostic criteria, becomes relapsing-remitting MS (RRMS).*

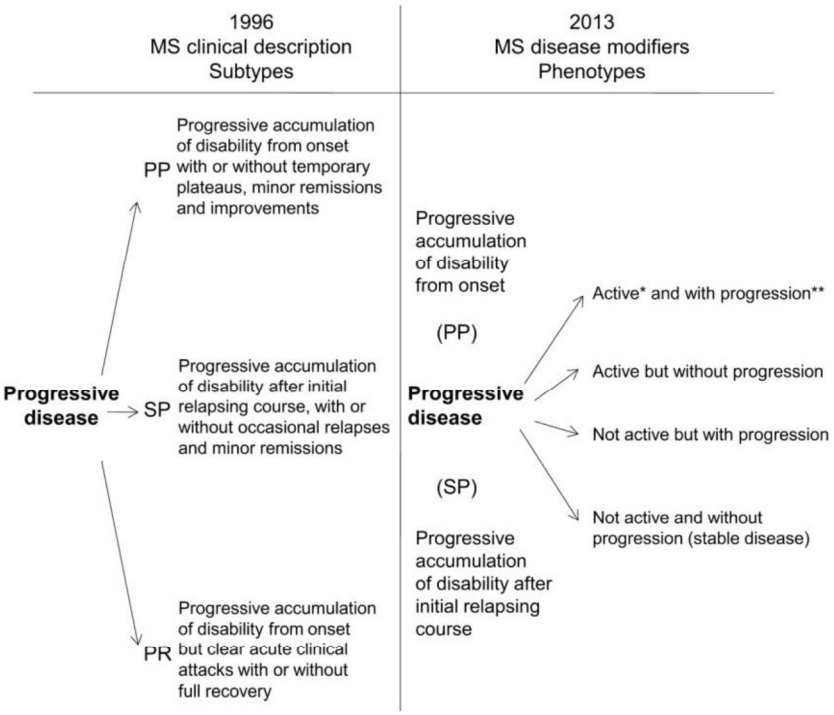


Figure 1.4: The 1996 vs 2013 MS phenotype descriptions for progressive disease from Lublin et al(Lublin et al, 2014). **Activity determined by clinical relapses assessed at least annually and/or MRI activity (contrast-enhancing lesions; new and unequivocally enlarging T2 lesions). **Progression measured by clinical evaluation, assessed at least annually. If assessments are not available, activity and progression are "indeterminate." MS = multiple sclerosis; PP = primary progressive; PR = progressive relapsing; SP = secondary progressive.*

1.5.3 Diagnostic criteria

Since the first version of the diagnostic criteria published by Schumacher in 1956 (Schumacher *et al*, 1965), MS diagnosis relied on the evidence of dissemination in time (DIT) and space (DIS): two or more distinct episodes of neurological dysfunction involving different parts of the CNS should be established.

Originally this could be done only demonstrating the occurrence of two or more clinical relapses affecting distinct functional systems but, over the subsequent revisions, paraclinical tools such as cerebrospinal fluid (CSF) analysis, MRI and evoked potentials (EP) were incorporated into the diagnostic criteria.

CSF is particularly useful to rule out differential diagnoses and can show the presence of liquor specific oligoclonal bands (OCBs) in up to 95% of patients with MS (Link & Huang, 2006), representing restricted classes of antibodies that are identified as discrete bands on agarose gel.

Nevertheless, due to its high specificity and sensitivity, brain MRI is fundamental to obtain and support MS diagnosis: MRI can detect new MS lesions 5 to 10 times more frequently than clinical evaluation alone (Barkhof *et al*, 1992). Typical MS lesions are located in the periventricular regions, corpus callosum, iuxtacortical and infratentorial regions, and less frequently in deep WM and basal ganglia; on MRI their aspect is hyperintense on T2-weighted images and active lesions show contrast enhancement on T1-weighted sequences after Gd injection, a marker of increased BBB leakage. Brain MRI scans can also display persistently hypointense lesions in T1-weighted images (the so called "*black holes*") that are associated with severe tissue damage (Filippi & Rocca, 2011).

Besides, neurophysiological studies may detect increased conduction latencies thus revealing subclinical damage in the motor, sensory, auditory and visual pathways and confirming its demyelinating nature (Leocani & Comi, 2014).

In order to allow an earlier diagnosis, MS diagnostic criteria have been simplified over time. The 2010 revised McDonald criteria (Polman *et al*, 2011), that were applied in the present study, are detailed in Table 1.1. However, while patients enrolment was ongoing, the latest edition of the MS criteria (2017)(Thompson *et al*, 2018) has been published, which mostly retrace the previous version with further simplifications: demonstration of CSF-specific OCBs can now be used to demonstrate DIT; both symptomatic and asymptomatic MRI lesions can be considered to fulfill DIS and DIT, with the exception of optic nerve lesions in patients presenting with ON; cortical lesions are considered in lesion count, together with juxtacortical lesions.

Finally, the criterion of "*no of better explanation*" is fundamental for MS diagnosis, since several neurological diseases might mimic MS. Clinical and paraclinical features not suggestive of MS, whenever present, should be considered "red flags" and should alert clinicians to reconsider the differential diagnosis (Geraldes *et al*, 2018).

Table 1.1: The 2010 revised McDonald Criteria for diagnosis of MS adapted from Polman et al.(Polman et al, 2011)

Clinical presentation	**Additional data needed for MS diagnosis**
≥ 2 attacks; **objective clinical evidence of ≥ 2 lesions or objective clinical evidence of 1 lesion with reasonable historical evidence of a prior attack**	None
≥ 2 attacks; **objective clinical evidence of 1 lesion**	DIS, demonstrated by one of: ✓ ≥ 1 T2-hyperintense lesion in at least 2 of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial or spinal cord)*; ✓ a further clinical attack implicating a different CNS site
1 attack; **objective clinical evidence of ≥ 2 lesions**	DIT, demonstrated by one of: ✓ simultaneous presence of asymptomatic Gd-enhancing and nonenhancing lesions at any time; ✓ a new T2-hyperintese and/or Gd-enhancing lesion(s) on FU MRI, irrespective of its timing with reference to a baseline scan; a second clinical attack
1 attack; **objective clinical evidence of 1 lesion (CIS)**	DIS and DIT, demonstrated by: For DIS one of: ✓ ≥ 1 T2-hyperintense lesion in at least 2 of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial or spinal cord)*; ✓ a second clinical attack implicating a different CNS site; For DIT one of: ✓ simultaneous presence of asymptomatic Gd-enhancing and nonenhancing lesions at any time; ✓ a new T2-hyperintense and/or Gd-enhancing lesion(s) on FU MRI, irrespective of its timing with reference to a baseline scan; ✓ a second clinical attack
Insidious neurological progression suggestive of MS (PPMS)	1 year of disease progression (retrospectively or prospectively determined) plus 2 of 3 of the following criteria: ✓ Evidence for DIS in the brain based on ≥ 1 T2-hyperintense lesion in the MS-characteristic (periventricular, juxtacortical, or infratentorial)* regions ✓ Evidence for DIS in the spinal cord based on ≥ 2 T2-hyperintense lesion in the spinal cord ✓ Positive CSF (isoelectric focusing evidence of OCBs and/or elevated IgG index)

**Gd enhancing lesions are not required. Symptomatic lesions are excluded from consideration in subject with spinal cord or brainstem syndromes*

1.6 MS therapeutic strategies

Currently, no cure for MS exists and available treatments typically intend to prevent new relapses and to slow disability progression.

Nonetheless, therapeutic opportunities have recently dramatically increased and more than 10 drugs are now licensed for MS treatment, which target distinct biological pathways and have different safety profiles. Alongside with these advancements in treatment strategies, new challenges arose, such as to identify the best option for individual patients; indeed, no validated markers exist to guide treatment selection (Teunissen *et al*, 2015).

At present, therapeutic choice is mostly empirical, based on the assessment of MS activity (attack frequency and severity, MRI parameters), drug availability and cost, comorbidities, concomitant medications and tolerability as well as patient's preferences. Moreover, treatment failure is defined by the evidence of breakthrough disease during drug intake, thus leading to the risk of disability accumulation.

In this perspective, in the last years the need for a more personalized approach to MS management became urgent and paved the way to several studies, which investigated the role of biomarkers in predicting treatment response.

In the next paragraphs we will give an overview of the available treatments for RRMS.

1.6.1 Management of MS relapses

Acute treatment for MS relapses is based on use of corticosteroids, usually administered as high doses such as 1 g methylprednisolone given intravenously once a day for 3 to 5 days, with or without subsequent tapering. Intramuscular and oral formulations of steroids are also available (Martinelli *et al*, 2009), but intravenous therapy is preferred for severe symptoms (Beck *et al*, 1992). In case of poor recovery, intravenous immunoglobulins (2 g/kg in five days) or plasma exchange (3 to 5 courses) can also be used (Weinshenker *et al*, 1999).

1.6.2 Disease-modifying treatments

As already mentioned, more than 10 disease modifying treatments (DMTs) are currently approved for RRMS; their use is supported by randomized controlled trial (RCTs) and observational studies that demonstrated their efficacy in reducing relapse

frequency and MRI activity or in delaying the transition to progressive phases. Several studies also demonstrated the utility of an early treatment, since the phase of CIS (Comi *et al*, 2012), in order to prevent further disease evolution.

Classically, two main strategies are applied in clinical practice: an escalating approach, that consists in starting a first-line treatment, which is generally safe, switching to a more aggressive therapy in case of breakthrough activity; or a more aggressive induction strategy, when more powerful treatments are preferred since the beginning of the disease due to the presence of negative prognostic factors, such as a high annualized relapse rate (ARR) or a high MRI lesion load (Degenhardt *et al*, 2009; Filippi *et al*, 1994).

1.6.2.1 First-line treatments

Four main compounds are included in the first-line treatment category, due to their good tolerability and safety profile:

1.6.2.1.1 Interferons

Six IFNßs preparations are approved for RRMS: IFNß 1a by intramuscular [Avonex®] or by subcutaneous injection [Rebif-22® and Rebif-44®, Plegridy®] and IFNß 1b by subcutaneous injection [Betaferon®, Extavia®].

The mechanism of action of IFNß is not fully understood, but seems to imply a reduction in BBB permeability and the regulation of adaptive immunity (T and B cells) and cytokine secretion (Dhib-Jalbut, 2002).

IFNß reduces clinical relapse rate by about 30%, decreases the number of Gd enhancing and new/enlarging T2 lesions, and slows worsening at the expanded disability status scale (EDSS) in RRMS (Jacobs *et al*, 2000; Comi *et al*, 2012). IFNß 1b was also shown to be effective in SPMS, though the benefit was minimal (Kappos & European Study Group on interferon beta-1b in secondary progressive MS, 1998).

IFNß is extremely safe and requires minimal monitoring, but is burdened with a high incidence of associated flu-like symptoms and the need for frequent injections that decrease patients' adherence (Giovannoni *et al*, 2012). Moreover the development of IFNß neutralizing antibodies sometimes occurs, that reduces drug efficacy.

1.6.2.1.2 Glatiramer acetate

Glatiramer acetate (GA) preparation for subcutaneous injection [Copaxone®] are available for daily (20mg) or 3-times-a-week (40mg) administration.

It is hypothesized that the immunomodulatory action of GA is dependent on the stimulation of regulatory T cells and the shift from Th1 to Th2 phenotype (Dhib-Jalbut, 2002).

Its efficacy is comparable to IFNß (Cadavid *et al*, 2009) and it is equally safe, but usually better tolerated. Similarly to IFNß preparation, treatment compliance to GA is affected by the need for frequent injections and injection site reactions.

1.6.2.1.3 Teriflunomide

Teriflunomide (Teri)[Aubagio®] is a once-daily oral drug that works by inhibiting the dihydroorotate dehydrogenase, an enzyme required for de novo pyrimidine synthesis in proliferating cells (Claussen & Korn, 2012).

It was evaluated in three phase III, placebo-RCTs (TEMSO (O'Connor *et al*, 2011a), TOWER (Confavreux *et al*, 2014) and TOPIC (Miller *et al*, 2014)) and in one active comparator study (TENERE (Vermersch *et al*, 2014)) that demonstrated it efficacy in reducing the ARR and disability accumulation by about one third as well as the occurrence of new active lesions at MRI.

Teri has a good tolerability profile, asymptomatic liver enzyme elevation being the most reported adverse events. Its prolonged half-life represents a potential advantage in case of poor compliance but requires an accelerated elimination procedure with cholestyramine in case of adverse events and/or switch to other drugs.

1.6.2.1.4 Dimethylfumarate

Dimethylfumarate (DMF) [Tecfidera®] is a twice-daily oral DMT whose mechanism of action has not yet been completely elucidated, but it is thought to reduce oxidative stress by activation of the nuclear-related factor 2 transcriptional pathway, and to have anti-inflammatory properties (Linker *et al*, 2011).

DMF was evaluated in two pivotal RCTs and proved effective in reducing ARR and MRI activity but not disability progression (Fox *et al*, 2012; Gold *et al*, 2012).

The main side effects of DMF are flushing and gastrointestinal complaints but, overall it is well tolerated. Noteworthy, a proportion of patients treated with DMF develop moderate-to-severe lymphopenia (Miclea *et al*, 2016), mainly due to reduction in CD8+ cells, that is potentially linked to higher risk of infections and, particularly, of progressive multifocal leukoencephalopathy (PML)(Lehmann-Horn *et al*, 2016).

1.6.2.2 Second-line treatments

1.6.2.2.1 Natalizumab

Natalizumab (NTZ) [Tybsari®] is a humanized monoclonal antibody that selectively targets the α4 subunit of the cell adhesion molecule "very late antigen 4" expressed on the surface of lymphocytes and monocytes, thus preventing their transmigration across the BBB.

Two RCTs (AFFIRM (Polman *et al*, 2006) and SENTINEL(Rudick *et al*, 2006)) showed a remarkable efficacy of NTZ on all outcomes, with a 70% reduction in relapse rate, a 80% decrease in the occurrence of MRI active lesions and a 40% lower probability of disability worsening measured using EDSS, compared to placebo.

NTZ is usually well tolerated but long-term treatment is hampered by the risk of PML, a CNS infection caused by the John Cunningham virus (JCV), for which no cure is available. Risk factors for PML include evidence of prior JCV exposure (assessed testing the presence of anti-JCV antibodies in serum), duration of NTZ therapy and prior use of immunosuppressants (McGuigan *et al*, 2015). For this reason, a careful assessment of the benefit-to-risk ratio is required before starting or continuing NTZ treatment, that also take into consideration the risk of disease reactivation after NTZ discontinuation (Sangalli *et al*, 2014; Papeix *et al*, 2016; Jokubaitis *et al*, 2014; O'Connor *et al*, 2011b).

1.6.2.2.2 Fingolimod

Fingolimod (FTY) [Gylenia®] was the first oral drug to be approved for the treatment of RRMS patients with high disease activity and/or that failed first-line drugs.

It is a sphingosine 1 phosphate (S1P) analogue administered once a day that acts as a functional antagonist inducing lymphocytes sequestration into lymph nodes and thus reducing their infiltration into the CNS (Cohen & Chun, 2011).

Three phase III RCTs (FREEDOMS (Kappos *et al*, 2010), FREEDOMS II (Calabresi *et al*, 2014) and TRANSFORMS)(Thompson *et al*, 2018) demonstrated its effectiveness in reducing relapse rate by 54% and 48% compared to placebo and by 52% compared to IFNβ 1a. FTY also showed a significant benefit on MRI parameters of disease activity, reducing the number of new/enlarging T2 and Gd-enhancing lesions and brain atrophy (Kappos *et al*, 2015).

FTY is overall well tolerated but requires some special cautions and a 6 hours monitoring period after the first administration due to the risk of bradycardia. Moreover liver enzyme elevation, macular edema and a higher risk of infections have been associated with FTY.

1.6.2.2.3 Ozanimod

Ozanimod is another S1P modulator more selective compared to FTY that specifically targets S1P receptors 1 and 5. It has been evaluated in the RADIANCE (Cohen *et al*, 2019) and SUNBEAM(Comi *et al*, 2019) trials that demonstrated its safety and effectiveness on clinical and MRI parameters, with results similar to FTY. Besides, due to its receptor selectivity, ozanimod does not require clinical observation after the first dose, unless pre-existing cardiac issues are present.

1.6.2.2.4 Mitoxantrone

Mitoxantrone [Novantrone®] is an immunosuppressive drug approved for rapidly worsening RRMS and SPMS. At standard doses, its use is limited to 2 years because of cumulative dose-related cardiomyopathy. Moreover, it is associated with a high risk of treatment-related acute leukemia that drastically reduced its use in MS (Martinelli *et al*, 2011).

1.6.2.2.5 Alemtuzumab

Alemtuzumab [Lemtrada®] is a humanized monoclonal antibody directed against the CD52 antigen, that has recently been approved for patients with active RRMS. It is administered intravenously once daily for 5 days and then repeated for 3 days after one year and acts by inducing a long-term depletion of T and B lymphocytes.

Three studies evaluated its efficacy (CAMMS223 (Coles *et al*, 2008), CARE-MS I (Cohen *et al*, 2012) and CARE-MS II (Coles *et al*, 2012)) showing a significant reduction in relapse rate and occurrence of MRI activity in the alemtuzumab arm compared to IFNß.
Its main side-effects include infections and autoimmune adverse events occurring in approximately one third of patients, mainly thyroid disorders and less frequently immune thrombocytopenia or nephropathy (Havrdova *et al*, 2015).

1.6.2.2.6 Cladribine

Cladribine [Mavenclad®] is an oral treatment approved for highly active RRMS. It is administered over two short courses one year apart and acts by depleting circulating lymphocytes.
It was evaluated in three RCTs (CLARITY(Giovannoni *et al*, 2010), CLARITY Extension (Giovannoni *et al*, 2017) and ORACLE MS (Leist *et al*, 2014)) that demonstrated a high proportion of relapse free (80%) patients at 2 years as well as a significant reduction in MRI activity.
The most clinically relevant adverse events reported in clinical trials were lymphopenia and herpes zoster infections (Cook *et al*, 2011).

1.6.2.2.7 Ocrelizumab

Ocrelizumab [Ocrevus®] is a humanized monoclonal antibody directed against CD20-positive B cells. According to two RCTs (Hauser *et al*, 2017), OPERA I and OPERA II, ocrelizumab is superior to IFNß 1a in decreasing relapse rate (45% reduction), disability accumulation (40% reduction) and neuro-radiological disease activity (95% reduction in the occurrence of new lesions on brain MRI) in RRMS.

1.6.2.3 Treatments for progressive MS

Up to date no specific treatments with neuroprotective and remyelinating properties are available for the treatment of the progressive forms of MS (Villoslada & Steinman, 2020; Allanach *et al*, 2022).
However three drugs are currently approved for the treatment of PPMS and SPMS patients that also display signs of concomitant inflammatory disease activity.

Specifically, ocrelizumab was evaluated for the treatment of PPMS in the ORATORIO trial (Montalban *et al*, 2017) and proved effective in reducing the risk of disability progression by 24% compared with placebo. Hence it is now the only compound approved for the treatment of PPMS patients.
On the other end, siponimod [Mayzent®], a S1P modulator with enhanced penetration in the CNS, was evaluated in the EXPAND trial (Kappos *et al*, 2018) on SPMS where was associated to a 21% reduction in disability progression compared to placebo.
Besides, a preparation of IFNß 1b [Betaferon®] is also approved for the treatment of SPMS patients but not often used due to the possible worsening of spasticity symptoms correlated with flu-like side effects.

In addition to the approved treatments listed above, other agents are used off-label to treat RR and progressive MS, such as azathioprine, cyclophosphamide [Endoxan®] and rituximab [Mabthera®].

1.7 Prognostic marker in MS

As already mentioned, the amazing increase in the number of therapeutic options in MS that occurred in the last decade brought with it a new challenge for MS clinicians, that is to identify the most effective treatment for individual patients. In fact, MS is highly heterogeneous in terms of clinical presentation, disease course and response to treatments. Hence, now more than in the past, there is a huge need for markers able to prognosticate disease evolution that can help to select patients to be early addressed to highly aggressive therapies.
Numerous factors can potentially influence disease course such as demographic and clinical parameters (age, gender, disease severity and duration, previous treatments) as well as underlying pathology and genetic predisposition; several efforts have been made to identify such prognostic parameters that are briefly summarized in the following paragraphs.

1.7.1 Clinical and demographic markers

Disease characteristics associated with long-term outcomes have recently been reviewed (Rotstein & Montalban, 2019) with the aim of promoting a more personalized management.

Among the demographic variables associated with disease outcomes is gender: as already mentioned MS is much more common in females compared to males and women also have a higher relapses rate (Kalincik *et al*, 2013), however male patients tend to reach the progressive phase earlier in life with an increased risk of disability accumulation (Confavreux *et al*, 2003).

An earlier age at onset (AAO) has been associated to a higher risk of conversion from CIS (Tintore *et al*, 2015) to MS and to a more active disease (Fromont *et al*, 2008); moreover, patients that are younger at the time of first relapse tend to reach functional limitations earlier in life. On the other hand an older AAO is associated to a more rapid disability accumulation (Confavreux *et al*, 2003) and a higher incidence of PPMS.

Among the other clinical factors associated with a worse prognosis are an incomplete remission after the first episode, a short interval between the first and the second relapse, multisystemic, cerebellar and spinal cord involvement and the PPMS course (Runmarker & Andersen, 1993; Confavreux *et al*, 2003).

Moreover, also some environmental risk factors already known to predispose to MS development have been correlated to a more severe course of the disease; in particular low vitamin D levels were correlated to a higher relapse rate (Munger *et al*, 2006), a higher incidence of new MRI lesions as well as a higher risk of disability progression (Ascherio *et al*, 2014). Similarly, smoking was associated to higher disability and conversion to SPMS (Rodgers *et al*, 2021; Ramanujam *et al*, 2015) as well as to brain atrophy (Graetz *et al*, 2019).

1.7.2 MRI parameters

Several MRI measures, both conventional and advanced, have been studied as prognostic parameters; first of all the number of T2 lesions on brain MRI is one of the main indicators of a high risk of conversion to MS in CIS patients (Brodsky *et al*, 2008; Tintore *et al*, 2015), moreover lesion load also correlate with disability up to 20 years after onset (Fisniku *et al*, 2008). Likewise, the presence of active Gadolinium-

enhancing (Gd+) lesions at basal MRI was also predictive of higher risk of long-term disability accumulation and conversion to SPMS (Brownlee *et al*, 2019).

In addition, the distribution of MRI lesions, whether symptomatic or not, has also been linked to disease outcome, infratentorial and spinal cord lesions predicting a poorer prognosis (Minneboo *et al*, 2004; Sombekke *et al*, 2013).

Among the more advanced neuro-radiological variables associated with disease severity, brain atrophy and the presence of cortical lesions are the most consistently associated (Lavorgna *et al*, 2014; Calabrese *et al*, 2012; Pérez-Miralles *et al*, 2013). However, their use in clinical practice is limited by the technical requirements and the possible confounding factors that can affect their correct assessment.

1.7.3 Paraclinical variables

1.7.3.1 Oligoclonal bands

The presence of OCBs detected trough CSF analysis, and even their number, has been associated to a higher risk of conversion from CIS to definite MS (Arrambide *et al*, 2018) and also to an increased risk of disability accumulation (Tintore *et al*, 2015; Avasarala *et al*, 2001). For this reason OCBs have been included in the diagnostic criteria of MS as a surrogate for DIT (Thompson *et al*, 2018).

1.7.3.2 Neurofilament light chain

Among the serum and CSF biomarkers that have been evaluated as prognostic factors in MS (Magliozzi & Cross, 2020), NfLs are the most supported.

NfL levels in the CSF correlates with neuronal damage and recent technical advancements have allowed their detection in the serum with a good correlation to the CSF values, thus facilitating their possible implementation in clinical practice.

Serum NfLs have been associated with relapse occurrence, MRI lesion load and brain volume loss (Kuhle *et al*, 2016; Barro *et al*, 2018); moreover a study also showed that higher serum levels of NfLs predicted disability progression in the following year (Barro *et al*, 2018).

1.7.3.3 Retinal nerve fiber layer thickness

Neurophysiological measures can also provide useful prognostic information; in particular, thickness of the retinal nerve fiber layer (RNFL) can be easily measured using optical coherence tomography (OCT) and has been proposed as a measure of neurodegeneration. A low RNFL thickness has been associated with a higher risk of motor disability increase and of cognitive impairment (Toledo *et al*, 2008; Martinez-Lapiscina *et al*, 2016).

1.7.4 Genetic factors

Fewer studies investigated genetic variants associated with disease severity compared to MS susceptibility, but they represent a completing approach to better disentangle MS pathological basis. Initially, candidate gene studies investigated the effect of known variants and genes implicated in MS susceptibility on disease progression but failed to identify any significant correlation (Jensen *et al*, 2010) apart for the HLA-DRB1*1501 allele (Hauser *et al*, 2000; Barcellos *et al*, 2003).

Similarly, a following GWAS (Baranzini *et al*, 2009) on disease progression showed no enrichment of immune-related genes. Indeed, when considering genetic markers associated with MRI parameters such as T2 lesion load and brain volume there was an enrichment of genes related to neuronal functions (e.g. *GRIN2A, NLGN1*).

Likewise, another GWAS that evaluated brain glutamate concentrations as endophenotype (Baranzini *et al*, 2010) followed by a network analysis, prioritized a gene module including genes implicated in glutamate biology that was associated with brain volume loss and decrease in N-Acetylaspartate over time, suggesting a possible role for excitotoxicity in modulating MS progression.

Finally, an additional GWAS (Briggs *et al*, 2011) evaluated polymorphisms associated to disability level assessed using the MS severity score (MSSS) and performed a pathway analysis that showed an over-representation of genes implicated in neuronal processes, axon guidance and signaling as well as with binding of interferon α/β receptor and antigen processing and presentation.

In conclusions, findings from genetic studies investigating markers of MS severity suggested a role for neurons-related factors in addition to immune mechanisms in

determining disease course; however no single marker was found with enough prognostic power to be used in clinical practice.

1.7.5 Defining MS outcome

One of the biggest problems when trying to identify prognostic markers is to have an agreement on the definition of the desired outcome.

As for today, no consensus exists in this respect and different studies used distinct criteria to classify patients. Overall, most studies investigated the prognostic values of parameters associated with variable measures of disability progression over time. However, disability accumulation frequently occurs in later stages of the disease thus requiring studies with large populations and long follow-ups. For this reason, other clinical and paraclinical measures of disease inflammatory activity that are associated with an increased risk of long-term disability accumulation have also been used as surrogate outcomes.

1.7.5.1 Clinical measures

Clinical parameters indicative of inflammatory activity are mainly based on presence of relapses, changes in ARR and time to first relapse (TFR). Conversely, disability measures like the EDSS score reflect the occurrence of functional and structural damage that can be secondary to inflammatory flares as well as a consequence of more insidious disease progression independent from disease activity (Río *et al*, 2006; Waubant *et al*, 2003).

Unfortunately the ability of these parameters to predict long term outcomes is not clear. Various studies demonstrated that a high rate of early relapses positively correlates with a high risk of long-term disability (Weinshenker *et al*, 1989; Confavreux *et al*, 2003), while other works found no clear association (Runmarker & Andersen, 1993; Eriksson *et al*, 2003).

An additional level of complexity, when considering changes in EDSS, is that patients' classification strongly depends on the thresholds and timing used to define disease progression (Liu & Blumhardt, 2000).

1.7.5.2 MRI measures

MRI is much more sensitive than clinical evaluations in detecting breakthrough disease activity (Barkhof *et al*, 1992) and it is a reliable proxy for underlying disease pathology. It has been shown that an active MRI is a risk factor for additional subsequent MRI and clinical activity (Jacobs *et al*, 1996; Simon *et al*, 1998; Li & Paty, 1999; Paty & Li, 1993) and some studies also suggested that the frequency of Gd-enhancing lesions in the early stages of MS correlates with brain atrophy progression and future disability status (Fisher *et al*, 2000; Filippi *et al*, 1995). However, other studies found only little correlation between classical MRI parameters and clinical outcomes (Martinelli Boneschi *et al*, 2004).

Nowadays, in addition to the number of new T2 and Gd enhancing lesions, other MRI parameters can be measured, that potentially better describe the extent of CNS damage occurring in MS. Among them, measurement of brain atrophy is a promising tool that has already been included as a secondary outcome in randomized control trials (RCTs); though, assessment of brain atrophy is not yet been implemented in everyday clinical practice.

1.7.5.3 Combined outcomes

Usually, measures of clinical and MRI activity are integrated to obtain a more comprehensive clinical outcome. In the case of MS, different classification systems have been proposed (Río *et al*, 2006, 2009; Rudick *et al*, 2004; Kappos *et al*, 2001), mainly applied to patients treated with first line therapy, but none of them has been validated in a long-term follow-up.

In recent years, the notion of no evidence of disease activity (NEDA) as a target for MS management have gained attention.

NEDA status, also called NEDA-3, is defined by the absence of clinical relapses, active lesions at MRI and disease progression. NEDA status at 2 years was shown to be predictive of no disability worsening up to 7 years (Rotstein *et al*, 2015). However, its reliability to predict clinical outcome at longer follow-up is still lacking (Goodin *et al*, 2018).

Nonetheless, given its widespread use as main outcome in MS clinical trials and the simplicity of its calculation, in the present study we used the NEDA-3 criterion as the main endpoint in all the analysis and also considered TFR as a complementary outcome.

2 Aims of the study

Due to the recent increase in the therapeutic opportunities for RRMS, there is a strong need to identify clinical and biological parameters that can predict disease outcome and help in the treatment choice, towards a more personalized approach taking into account patient's specific features.

To this end, the present study aimed to explore genetic, gene expression and immunological markers associated with inflammatory activity in RRMS patients and to integrate them into a predictive model that could support clinicians in classifying patients according to their clinical and molecular characteristics and in unravelling the biological bases of the disease.

Specifically, we investigated:

- The presence of demographic, clinical and neuro-radiological parameters associated with MS activity at 4 years after the start of a first-line treatment in 2 cohorts of RRMS patients;
- The existence of genetic markers associated with the explored clinical outcomes, by means of a GWAS meta-analysis of the 2 above-mentioned cohorts;
- Genes differentially expressed by RRMS patients with or without inflammatory disease activity during follow-up;
- Biological pathways enriched of signals associated with disease severity;
- The correlation between immune repertoire characteristics and the clinical endpoints of interest;
- The integration of all these different layers of information to build a predictive model of disease activity.

The first point is elaborated in the clinical analysis, described in section 3.2 and 5.1, the second, third and fourth tasks are detailed in the genomic analysis in section 3.3 and 5.2, while the fifth aim is illustrated in the immune repertoire analysis (section 3.4 and 5.3) and the last one in the following predictive model section (section 3.5 and 5.4).

The present project was made possible thanks to the funding and support of the Italian Ministry of Health [grant project: GR-2019-12368672]

3 Results

3.1 Overall study design

Two cohorts of RRMS patients were included in the present study:

- a Core cohort (CC) of around 200 individuals who were sampled before the start of a first-line drug in order to generate genetic, gene expression and T cell receptor (TCR) sequencing data and who were followed for up to 4 years after treatment start;
- a larger Extended cohort (EC) of around 1000 RRMS patients with available genetic data that were generated in the context of previous projects; this cohort was included in order to increase the statistical power of the genetic analysis. For this cohort we retrospectively collected demographic and clinical information at the time of a first-line treatment start and for the following 4 years.

Patients of the two cohorts were classified according to the occurrence of disease activity during the 4-year observation period. Figure 3.1 synthetically illustrates the study design.

The study protocol was approved by the local Ethical Committee of Ospedale San Raffaele (OSR) and all patients signed a written informed consent before undergoing blood sampling.

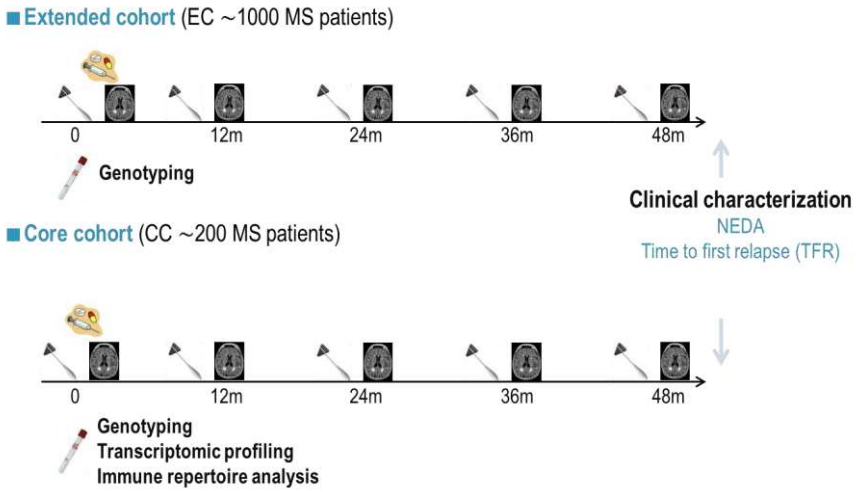


Figure 3.1: Study design. *The study included a larger Extended cohort of about 1000 RRMS patients, with available clinical and genetic data, and a Core cohort of about 200 RRMS patients, for whom also transcriptomic and immune repertoire data were generated. All patients were characterized according to occurrence of disease activity at the end of the 4-year observation period.*

3.2 Clinical analysis

In the *clinical analysis* we explored the association between baseline demographic, clinical and MRI parameters and the occurrence of MS inflammatory activity at 4 years in RRMS patients treated with a first-line drug.

We enrolled in the CC 208 RRMS patients who were sampled before the start of a first-line therapy between September 2011 and March 2016.

Moreover, genetic information of 1086 MS patients treated with first-line drugs were already available at the Laboratory of Genetics of Human Neurological Disorder, OSR, generated in the context of previous projects, and were included in the EC. Due to its retrospective nature, patients included in the EC started a first line treatment between January 1994 and February 2016, in an earlier period compared to the CC.

Baseline clinical and demographic characteristics of the two cohorts are detailed in Table 3.1.

Table 3.1: Baseline clinical and demographic characteristics of RRMS patients included in the study, divided into CC and EC

	CC **n = 208**	**EC** **n = 1086**	**P-value**
Gender (Females:Males)	133:75	762:324	ns
Age at onset	30.62 ± 9.28	28.37 ± 8.85	0.001
Age at baseline	36.25 ± 9.71	33.77 ± 9.69	0.0006
Disease duration	5.64 ± 6.55	5.4 ± 5.90	ns
EDSS at baseline	1.6 ± 0.7	1.8 ± 0.9	0.007
First line therapy, n (%)	IFNβ: 95 (45.7%) GA: 68 (32.7%) DMF: 43 (20.7%) Teri: 2 (0.9%)	IFNβ: 807 (74.3%) GA: 277 (25.5%) DMF: 2 (0.2%)	<0.001

Baseline demographic and clinical characteristics of patients included in the study, divided into CC and EC. For continuous variables, the mean value ± standard deviation (SD) is reported. P-value refers to the comparison between EC and CC. EDSS: Expanded Disability Status Scale; IFN-b: interferon beta; GA: glatiramer acetate; DMF: dimethyl fumarate; Teri: teriflunomide; No tp: no therapy.

As shown in the table, patients in the CC had a slightly later AAO and higher age at the baseline visit compared to the EC, as well as a lower mean EDSS; moreover, the

distribution of first-line drugs started in the 2 cohorts is very different, due to the later period of enrolment for the CC. Indeed, patients belonging to the EC started a therapy before the marketing authorization of DMF and Teri and were mostly treated with IFNβ or GA. Conversely, around 20% of patients in the more recent CC were treated with DMF, that was approved in 2014.

In order to identify clinical parameters associated with the occurrence of inflammatory disease activity during follow-up, that will be used as covariates in the genetic analyses, we tested the correlation between clinico-demographic variables and NEDA-3 at 4 years and TFR, both in the CC and EC; a meta-analysis was then performed to integrate the results obtained in the two datasets.

3.2.1 NEDA-3

Overall, 205 patients in the CC and 954 in the EC were classifiable according to the NEDA-3 criterion. As expected for patients treated with a first-line drug and due to the medium term follow-up, most of them showed some degree of disease reactivation during the analyzed period; exactly 69% of patients in the CC and 84% of patients in the EC had evidence of inflammatory activity. The slightly higher proportion of active patients in the EC compared to the CC is explained by the different period of treatment start. Indeed, the majority of patients belonging to the EC started a first-line treatment when few second-line drugs were available on the market; for this reason also subjects with a medium-to-high level of baseline disease activity started a first-line treatment leading to a higher rate of breakthrough disease activity during the observation period. On the contrary, due to the availability of highly effective and well-tolerated second-line drugs, such as fingolimod (Cohen *et al*, 2010), starting from 2012 patients showing a moderate level of disease activity were rather addressed to more aggressive treatments and were not included in the present analysis.

Table 3.2 shows the clinical variables associated with evidence of disease activity (EDA-3) during the 4 year observation period in the CC and EC, according to a univariable and multivariable setting, as well as the results of the meta-analysis.

AAO was the parameter more strongly and consistently associated to EDA during follow-up, both in CC and EC (odds ratio (OR): 0.94, p-value: 0.0007 and OR: 0.96, p-value <0.0001, respectively); precisely, a younger AAO was associated with a higher

risk of breakthrough inflammatory activity with an OR of 0.95 (p-value<0.0001) in the meta-analysis). Similarly, a shorter disease duration (DD) was also significantly correlated with a higher risk of disease activity in the CC and the meta-analysis (OR: 0.93, p-value: 0.0031 and OR: 0.95, p-value: 0.0007, respectively) and there was also a nominally significant association in the EC (OR: 0.97, p-value: 0.033). Finally, a trend for association was also detected for gender, females having an almost 50% increased risk of reactivation (OR: 1.45, p-value: 0.028), and baseline EDSS, with increasing likelihood of EDA for higher scores (OR: 1.29, p-value: 0.017).

Table 3.2: Baseline characteristics associated to EDA-3 in the CC, EC and the meta-analysis.

	CC			EC		
	Univariable analysis					
	OR	**95% CI**	**P Value**	**OR**	**95% CI**	**P Value**
Gender (F)	1.37	0.74-2.51	0.313	1.4	0.97-2.00	0.071
AAO	**0.95**	**0.92-0.98**	**0.004**	**0.97**	**0.95-0.98**	**0.0003**
Disease duration	0.95	0.91-1	0.030	0.99	0.97-1.02	0.652
Baseline EDSS	0.71	0.47-1.06	0.099	1.28	1.04-1.61	0.025
	Multivariable analysis					
	OR	**95% CI**	**P Value**	**OR**	**95% CI**	**P Value**
Gender (F)	1.28	0.66-2.43	0.460	1.52	1.03-2.23	0.034
AAO	**0.94**	**0.9-0.97**	**0.0007**	**0.96**	**0.94-0.98**	**<0.0001**
Disease duration	**0.93**	**0.88-0.97**	**0.003**	0.97	0.93-1.00	0.033
Baseline EDSS	0.89	0.57-1.42	0.623	**1.42**	**1.13-1.81**	**0.003**

	Meta-analysis		
	OR	**95% CI**	**P Value**
Gender (F)	1.45	1.04-2.02	0.028
AAO	**0.95**	**0.94-0.97**	**<0.0001**
Disease duration	**0.95**	**0.93-0.98**	**0.0007**
Baseline EDSS	1.29	1.05-1.58	0.017

Table 3.2 shows the association of different baseline parameters with evidence of disease activity (EDA-3) at 4 years. Results are shown for the CC and EC (upper section) and the meta-analysis (bottom panel). Odds ratio (OR) with 95% confidence intervals (CI) and relative p-value are reported for associated baseline variables. AAO: age at onset; EDSS: Expanded Disability Status Scale.

3.2.2 Time to first relapse (TFR)

In the EC, 640 patients faced at least a clinical relapse during the 4 years observation period while 80 patients experienced new or worsening symptoms during follow up in the CC.

The analysis of clinical parameters associated with the TFR (Table 3.3) yielded similar results as for the NEDA-3 outcome.

Table 3.3: Baseline characteristics associated to TFR in the CC, EC and the meta-analysis.

	CC			EC		
	Univariable analysis					
	HR	**95% CI**	**P Value**	**HR**	**95% CI**	**P Value**
Gender (F)	1.68	1.03-2.74	0.039	**1.29**	**1.08-1.53**	**0.005**
AAO	0.98	0.96-1	0.11	**0.99**	**0.98-1**	**0.003**
Disease duration	0.95	0.92-0.99	0.021	0.99	0.97-1	0.047
Baseline EDSS	1.22	0.93-1.6	0.159	1.04	0.95-1.14	0.402
	Multivariable analysis					
	HR	**95% CI**	**P Value**	**HR**	**95% CI**	**P Value**
Gender (F)	1.51	0.92-2.48	0.101	**1.35**	**1.1-1.67**	**0.005**
AAO	**0.97**	**0.94-0.99**	**0.008**	**0.98**	**0.97-0.99**	**0.0001**
Disease duration	**0.93**	**0.9-0.97**	**0.002**	**0.97**	**0.96-0.99**	**0.002**
Baseline EDSS	1.41	1.07-1.84	0.013	1.09	0.99-1.2	0.072

	Meta-analysis		
	HR	**95% CI**	**P Value**
Gender (F)	**1.38**	**1.15-1.67**	**0.001**
AAO	**0.98**	**0.97-0.99**	**<0.0001**
Disease duration	**0.97**	**0.96-0.98**	**<0.0001**
Baseline EDSS	1.12	1.03-1.23	0.012

Table 3.3 shows the association of different baseline parameters with the time to first relapse (TFR). Results are shown for the CC and EC (upper section) and the meta-analysis (bottom panel). Hazard ratio (HR) with 95% confidence intervals (CI) and relative p-value are reported for associated baseline variables. AAO: age at onset; EDSS: Expanded Disability Status Scale.

Specifically, a younger AAO was again linked to a higher risk of early relapse during the observation period in the meta-analysis (hazard ratio (HR): 0.98, p-value <0.0001)

and when considering the CC and EC separately (OR: 0.97, p-value: 0.008 and OR: 0.98, p-value: 0.0001, respectively). Besides, a longer DD was associated with later relapses in the CC (HR: 0.93, p-value: 0.002), EC (HR: 0.97, p-value: 0.002) and meta-analysis (HR: 0.97, p-value<0.0001). Female gender was correlated with a worse clinical outcome in the EC (HR: 1.35, p-value: 0.005) and not in the CC, even if the direction and magnitude of the effect was consistent in the two datasets (HR: 1.51, p-value: 0.101); when meta-analyzing the two cohorts the association was still present and statistically significant (HR: 1.38, p-value: 0.001).

Noteworthy, a trend of association was present also for the baseline EDSS, higher scores being correlated to a higher risk of early relapse (HR of 1.12, p-value: 0.012), similarly to what already described for the NEDA-3 outcome.

3.2.3 Considerations

In this first part of the study, we performed the characterization of the two enrolled cohorts of RRMS and completed the clinical classification according to the selected endpoints NEDA-3 and TFR.

The two datasets had slightly different baseline characteristics in terms of AAO, age at baseline visit and EDSS and, most importantly, differed according to the type of first-line treatment started at baseline. This imbalance is explained by the fact that the CC has been included in a later period, when more first-line drugs were available. The same reason also explains the variability in the proportion of patients showing disease reactivation during follow-up in the two cohorts; the accessibility of new drugs in more recent years has allowed for a more accurate selection of patients to be addressed to highly-active treatments, leading to an improvement in the rate of treatment response which is reflected by the better outcomes obtained in the CC compared to the EC.

Taking these differences into consideration, we decided to perform the analyses separately in the two datasets and then meta-analyze the results. By doing so, we were able to identify demographic and clinical parameters strongly associated with breakthrough inflammatory activity during follow-up. Specifically, a younger AAO and a shorter disease duration were associated with a higher risk of disease reactivation measured using both the NEDA-3 criterion and the TFR. Moreover, female gender was strongly associated to a higher likelihood of experiencing a clinical relapse during the observation period and there was also a trend of association when considering the

NEDA-3 outcome. These data are in line with previous reports and what expected by clinical practice; indeed, previous studies reported that an older AAO correlated with less inflammatory flares in patients treated with IFN beta (Fromont *et al*, 2008), while an early onset was associated to an increased risk of evolution from CIS (Tintore *et al*, 2015) or RIS (Lebrun-Frénay *et al*, 2021) to definite MS. Moreover, another study (Kalincik *et al*, 2013) showed that female gender was linked to a higher level of inflammatory activity and that disease activity tended to decrease with age.

In conclusion, these results suggest that our cohorts are representative of the overall MS population and that our meta-analytic approach is valid and allows to identify meaningful association with the outcomes of interests. The clinical parameters that were found to be strongly associated with EDA and TFR in this first substudy were included as covariates in the genetic analyses, in order to adjust for potential confounding factors.

3.3 Genomic analysis

In this study section we went on to investigate the presence of genetic variants and genes associated with the clinical activity endpoints and to explore the biological pathways they are involved in, with the purpose of highlighting functional paths that play an important role in determining disease activity.

Indeed, the genetic predisposition to MS susceptibility is well established, with more than 200 autosomal and 30 MHC variants associated with the disease. Nonetheless, fewer studies have also shown a possible contribution of genetic variation, mainly focusing on the MHC, with disease course and phenotype (Isobe *et al*, 2016; Hauser *et al*, 2000; Barcellos *et al*, 2003); hence it is reasonable to hypothesize that the genetic background, together with epigenetic changes, could also influence disease activity.

We believe that the identification of genetic polymorphisms and/or transcriptional changes correlated with breakthrough inflammation in MS holds a great potential to guide in patients stratification, toward a more personalized management. Moreover, it could also possibly lead to the detection of new, interesting therapeutic targets.

3.3.1 Genetic study

3.3.1.1 Genome-wide association study on NEDA-3 outcome

3.3.1.1.1 CC

Among the 208 patients of the CC included in the clinical study, 8 patients were excluded due to low call rate leading to poor imputation quality, 1 subject was discarded for high heterozygosity, 1 for sex-mismatch and 3 were excluded after the multidimensional scaling (MDS) analysis. As a results, after stringent per sample and per single nucleotide polymorphisms (SNPs) filtering, a total of 195 individuals and 6,952,445 SNPs were evaluated for association with NEDA-3

Figure 3.2 illustrates the results of the association analysis with NEDA-3 in the CC.

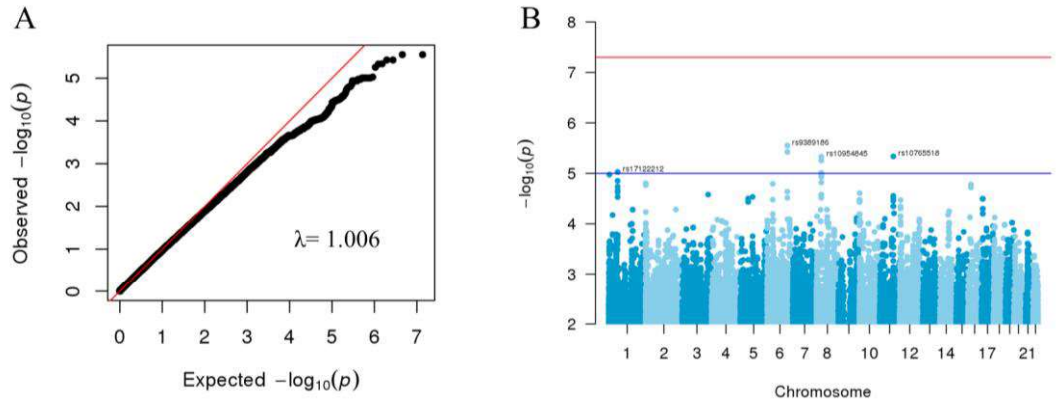


Figure 3.2: QQ plot and Manhattan plot showing results of the association with NEDA-3 in the CC. *A: Quantile-quantile plot plotting the –log10(p-value) distribution of our analysis versus a theoretical normal distribution. B: Manhattan plot showing association p-value for each SNPs across different chromosomes. λ: genomic inflation factor.*

The quantile-quantile (QQ) plot on the left panel, where the distribution of actual p-values from the analysis are plotted against expected values, reports a genomic inflation factor (λ) (Devlin & Roeder, 1999) of 1.006 suggesting a negligible amount of population stratification; moreover it also shows a downward deflection suggesting that the analysis is likely underpowered, as expected due to the low number of patients. Indeed, the Manhattan plot in Figure 3.2 B displays no genome-wide significant associations, however 4 signals exhibited some evidence of association with EDA, one on chromosome (Chr) 6, Chr 8, Chr 11 and Chr 1 each. We will not explore these associations further because we intended to concentrate on signals prioritized by the meta-analysis as supported in both cohorts; however, it is interestingly to note that

among the top variants with suggestive association are SNPs intronic to neuregulin 1, *NRG1*, coding for a myelin-related growth factor which promotes proliferation of oligodendrocytes and has been found to be downregulated in MS lesions, suggesting a possible role of *NRG1* in MS pathogenesis (Viehover *et al*, 2001; Kataria *et al*, 2021).
The top-10 associated SNPs for this analysis are shown in Table 3.4.

Table 3.4: Top 10 results of the analysis of association with NEDA-3 in the CC

SNP	Chr	Gene	BP	A1	MAF	OR	95%-CI	P Value
rs2144219	6	*RP11-557H15.3*	134817135	T	0.20	0.21	0.11-0.4	2.82E-06
rs9389186	6	*RP11-557H15.3*	134813324	T	0.20	0.21	0.11-0.4	2.82E-06
rs57424335	6	*RP11-557H15.3*	134812488	A	0.20	0.21	0.11-0.41	3.76E-06
rs72980343	6	*RP11-557H15.3*	134810991	A	0.20	0.21	0.11-0.41	3.76E-06
rs10765518	11	*FAT3* (5')	91800709	G	0.44	0.25	0.14-0.45	4.62E-06
rs10954845	8	*NRG1*	32319842	G	0.22	0.22	0.12-0.43	4.67E-06
rs4733124	8	*NRG1*	32371798	C	0.18	0.22	0.12-0.43	5.56E-06
rs17122212	1	*NFIA*	61903554	T	0.19	6.94	2.94-16.34	9.42E-06
rs12549090	8	*NRG1*	32373385	C	0.20	0.24	0.13-0.46	9.91E-06
rs13277678	8	*NRG1*	32373112	C	0.20	0.24	0.13-0.46	9.91E-06

SNP: variant name; Chr: chromosome; Gene: gene the variant is mapped to, when variant is intergenic the name of the closest coding gene and direction is reported; BP: physical position of the SNP in build 37; A1: minor allele; MAF: minor allele frequency in the European population of 1000 Genome project; OR: odds-ratio, 95%-CI: confidence interval for the odds ratio; P: association p-value.

3.3.1.1.2 EC

For the EC, we started from 1086 patients included in the clinical analysis and excluded 162 samples from the imputation due to poor coverage. We then applied standard quality checks as for the CC and rejected additional 14 patients due to high heterozygosity and 8 subjects who were outliers at the MDS analysis; eventually, 902 individuals and 6,952,978 passed filters and quality controls (QCs) and entered the analysis.
Figure 3.3 shows the results for the NEDA-3 analysis in this cohort.

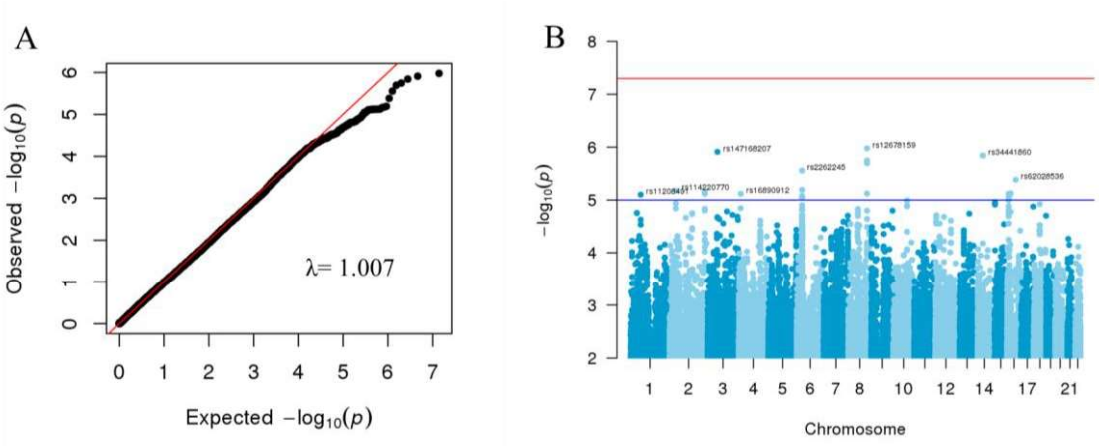


Figure 3.3: QQ plot and Manhattan plot showing results of the association with NEDA-3 in the EC. *A: Quantile-quantile plot plotting the –log10(p-value) distribution of our analysis versus a theoretical normal distribution. B: Manhattan plot showing association p-value for each SNPs across different chromosomes. λ: genomic inflation factor.*

Again, the QQ plot in Figure 3.3 A rules out the presence of unusual enrichment due to genetic stratification, with a lambda value of 1.007, while the Manhattan plot in Figure 3.3 B shows the results of the association analysis: no genome wide significant hits were identified.

Table 3.5: Top 10 results of the analysis of association with NEDA-3 in the EC

SNP	Chr	Gene	BP	A1	MAF	OR	95%-CI	P Value
rs12678159	8	*SAMD12*	119436865	G	0.39	2.10	1.56-2.83	1.05E-06
rs147168207	3	*CADPS*	62717754	G	0.01	0.07	0.02-0.2	1.23E-06
rs34441860	14	*AL163953.3*	53839846	C	0.03	0.19	0.1-0.37	1.44E-06
rs12679873	8	*SAMD12*	119436711	T	0.38	2.08	1.54-2.81	1.79E-06
rs13269591	8	*SAMD12*	119436240	T	0.38	2.07	1.53-2.79	2.01E-06
rs2262245	6	*C6orf106*	34567694	C	0.33	0.50	0.37-0.67	2.78E-06
rs62028536	16	*MMP2*	55448622	C	0.02	0.24	0.13-0.44	4.14E-06
rs6912327	6	*UHRF1BP1*	34764922	C	0.24	0.49	0.36-0.67	6.42E-06
rs114220770	2	*SLC8A1*	40729223	A	0.02	0.14	0.06-0.33	6.71E-06
rs11689046	2	*SERPINE2* (5')	225022528	G	0.13	0.42	0.29-0.61	6.91E-06

SNP: variant name; Chr: chromosome; Gene: gene the variant is mapped to, when variant is intergenic the name of the closest coding gene and direction is reported; BP: physical position of the SNP in build 37; A1: minor allele; MAF: minor allele frequency in the European population of 1000 Genome project; OR: odds-ratio, 95%-CI: confidence interval for the odds ratio; P: association p-value.

3.3.1.1.3 Meta-analysis

We then integrated the results obtained in the two datasets by means of a fixed effect model which was deemed appropriate given the homogeneity of the cohorts including patients of European ancestry. Overall 6,521,446 common variants were present in both datasets and were explored in the meta-analysis.

Figure 3.4 displays the corresponding QQ and Manhattan plots.

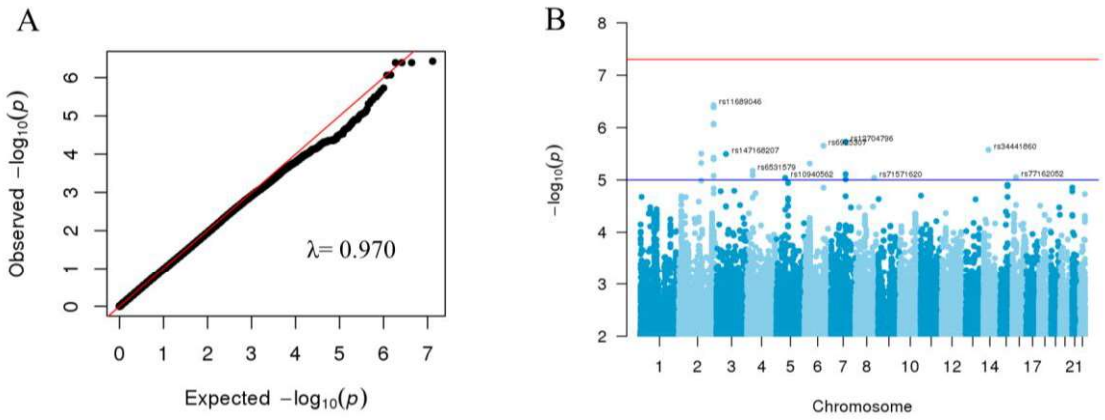


Figure 3.4: QQ and Manhattan plot showing results of the association with NEDA-3 in the meta-analysis. *A: Quantile-quantile plot plotting the –log10(p-value) distribution of our association analysis versus a theoretical normal distribution. B: Manhattan plot showing p-value of association for each SNPs across different chromosomes. λ: genomic inflation factor.*

As illustrated in the Manhattan plot (Fig 3.4 B and Table 3.6), no variants reached the Bonferroni genome-wide significant threshold; nonetheless, some suggestive signals of association were present, in particular on Chr 2 and 7.

The top-associated variant, rs11689046 (OR 0.42, p-value 3.6e-7), is an intergenic SNP mapped to Chr 2:225022528 around 118Kb at the 5' of the *SERPINE2* gene. Besides, other 5 variants among the top ones, are in very high LD (r^2~1) with it and represent the same association signal, as displayed in the corresponding regional plot (Figure 3.5).

Table 3.6: Top 10 results of the meta-analysis of association with NEDA-3

SNP	Chr	Gene	A1	MAF	OR_CC	95CI_CC	P_CC	OR_EC	95CI_EC	P_EC	OR_meta	95CI_meta	P_meta	I2
rs11689046	2	*SERPINE2* (5')	G	0.13	0.42	0.2-0.86	0.018	0.42	0.28-0.61	6.69E-06	0.42	0.3-0.58	3.60E-07	0
rs10933044	2	*SERPINE2* (5')	C	0.13	0.42	0.2-0.86	0.018	0.42	0.29-0.61	7.32E-06	0.42	0.3-0.58	3.93E-07	0
rs78224482	2	*SERPINE2* (5')	C	0.13	0.42	0.2-0.86	0.018	0.42	0.29-0.61	7.32E-06	0.42	0.3-0.58	3.93E-07	0
rs79017028	2	*SERPINE2* (5')	T	0.13	0.42	0.2-0.86	0.018	0.42	0.29-0.61	7.32E-06	0.42	0.3-0.58	3.93E-07	0
rs7573097	2	*SERPINE2* (5')	G	0.13	0.42	0.2-0.86	0.018	0.43	0.29-0.63	1.54E-05	0.43	0.31-0.6	8.19E-07	0
rs6436469	2	*SERPINE2* (5')	T	0.16	0.44	0.22-0.88	0.020	0.43	0.3-0.63	1.42E-05	0.43	0.31-0.6	8.41E-07	0
rs12704796	7	*PON2*	A	0.21	0.31	0.16-0.6	4.27E-04	0.57	0.41-0.77	3.43E-04	0.51	0.38-0.67	1.98E-06	61.9
rs6925307	6	*CLVS2* (5')	A	0.39	0.48	0.28-0.82	7.51E-03	0.58	0.44-0.76	7.50E-05	0.56	0.44-0.71	2.12E-06	0
rs34441860	14	*AL163953.3*	C	0.03	0.56	0.13-2.4	0.438	0.19	0.1-0.37	1.41E-06	0.23	0.13-0.42	2.61E-06	43.8
rs147168207	3	*CADPS*	G	0.01	0.46	0.05-4.16	0.491	0.07	0.02-0.2	1.21E-06	0.1	0.04-0.26	3.16E-06	56.8

SNP: variant name; Chr: chromosome; Gene: gene the variant is mapped to, when variant is intergenic the name of the closest coding gene and direction is reported; MAF: minor allele frequency in the European population of 1000 Genome project; A1: minor allele; OR_CC: odds-ratio in the core cohort, 95%-CI_CC: confidence interval for the odds ratio in the core cohort; P_CC: association p-value in the core cohort; OR_EC: odds-ratio in the extended cohort, 95%-CI_EC: confidence interval for the odds ratio in the extended cohort; P_EC: association p-value in the extended cohort; OR_meta: odds-ratio in the meta-analysis, 95%-CI_meta: confidence interval for the odds ratio in the meta-analysis; P_meta: association p-value in the meta-analysis.

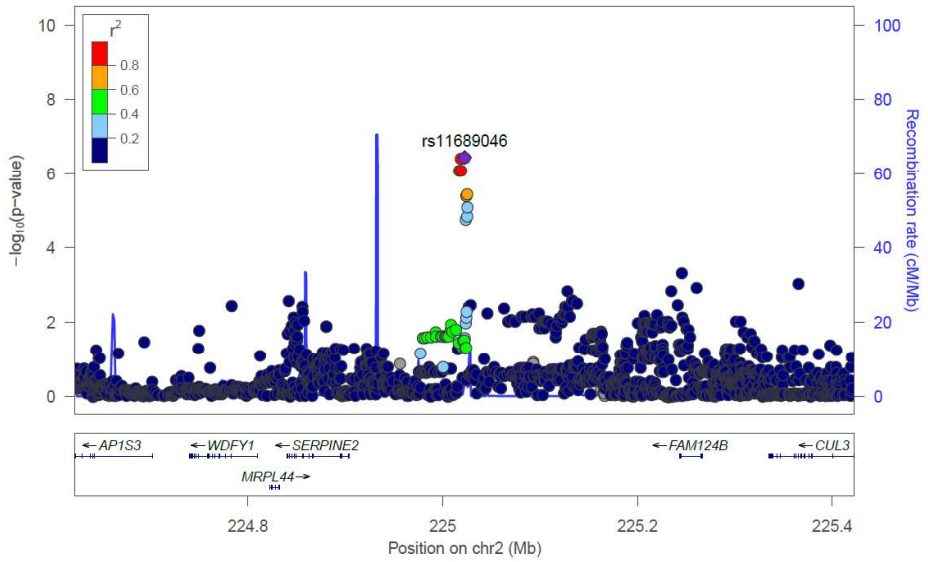


Figure 3.5: Regional association plot for the rs11689046 locus in the Meta-analysis. *Illustration of the chr2q36 region centered on rs11689046, with the local recombination rate plotted in light blue. Each point represents a SNP; the most associated SNP in the meta-analysis, rs11689046, is marked in purple. The color of each dot represents the extent of LD with rs11689046 according to the legend on the right. Physical positions are based on build 37 of the human genome.*

The rs11689046^{G} is a common allele (minor allele frequency (MAF) of 0.13 in the European population from 1000 Genome project) that seems to be associated with a protective effect towards the occurrence of disease activity (OR 0.42, p-value 0.018 and OR 0.42, p-value 6.69e-6 respectively in the CC and EC); indeed in both the CC and EC, a higher proportion of G carriers showed disease stability during the 4 years observation period compared to the TT subjects (Figure 3.6).

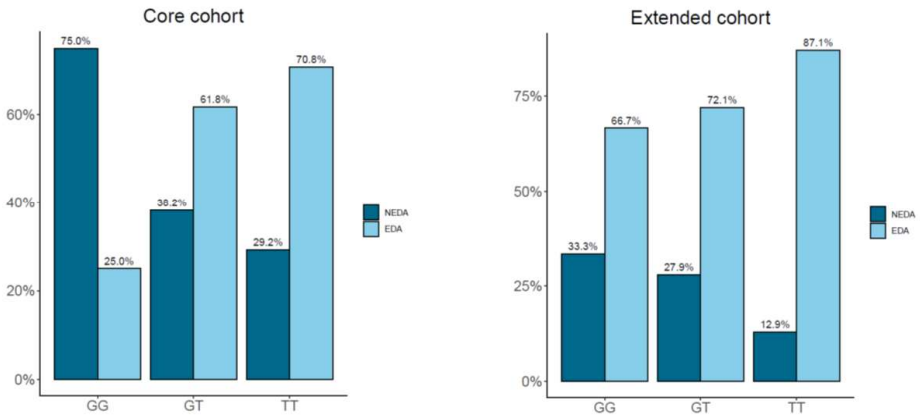


Figure 3.6: Proportion of patients with ED or stable disease NEDA at 4 years according to genotype at rs11689046 in the CC and EC. *On the y-axis is the proportion of patients with no evidence of disease activity (NEDA, blue) or EDA (light blue). On the x-axis are the genotypes at the rs11689046 variant.*

Leveraging the information available in public databases, we found that rs11689046 has an expression quantitative trait loci (eQTL) effect on *SCG2* gene in CD4 T cells (Schmiedel *et al*, 2018) (p-value 2.1e-5), *WDFY1* in monocytes (Ota *et al*, 2021) (p-value 0.002) and *SERPINE2* in whole blood (p-value 0.007), that is, genetic polymorphisms in this position are correlated with changes in the expression of the above mentioned genes. Among these genes *SERPINE2* also known as Nexin-1, is a particularly interesting candidate for association with MS disease activity. Indeed, it encodes for a glycoprotein which represents the most abundant thrombin inhibitor in the brain, being secreted by glial cells and neurons (Reinhard *et al*, 1994), and it has been demonstrated that the coagulation system is activated in the animal model of MS, the experimental autoimmune encephalomyelitis (EAE). Previous studies demonstrated that in the preclinical phase of EAE, at day 8 post-immunization, SERPINE2 is more expressed in brain homogenates from EAE mice compared to controls (Beilin *et al*, 2005) and other works reported a clinical improvement following treatment by coagulation inhibitors, such as heparin or dermatan sulfate (Chelmicka Szorc & Arnason, 1972; Inaba *et al*, 1999), suggesting a possible role of SERPINE2 in disease pathogenesis, even if it is not clear whether its action is neuroprotective or damaging (Meins *et al*, 2001; Houenou *et al*, 1995).

The second top-hit for the NEDA-3 analysis, is rs12704796 on Chr 7: 95065850 that maps to *PON2* gene (Figure 3.7).
The A allele at this variant has a MAF of 0.21 in the European population, and is associated with a lower risk of disease reactivation in both the CC, EC and meta-analysis (OR 0.31 and p-value 4.27e-4, OR 0.57 and p-value 3.43e-4, OR 0.51 and p-value 1.98e-6 respectively): Figure 3.8 illustrates a growing percentage of patients with NEDA during the observation period in subjects heterozygous or homozygous for the protective A allele, in the CC as well as the EC.

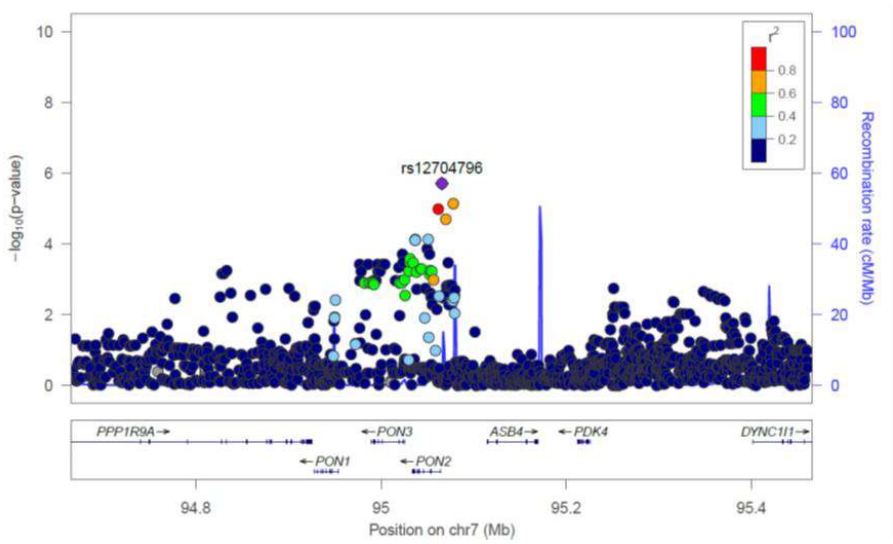


Figure 3.7: Regional association plot for the rs12704796 locus in the Meta-analysis*. Illustration of the chr2q36 region centered on rs12704796, with the local recombination rate plotted in light blue. Each point represents a SNP; the most associated SNP in the meta-analysis, rs12704796, is marked in purple. The color of each dot represents the extent of LD with rs12704796 according to the legend on the right. Physical positions are based on build 37 of the human genome.*

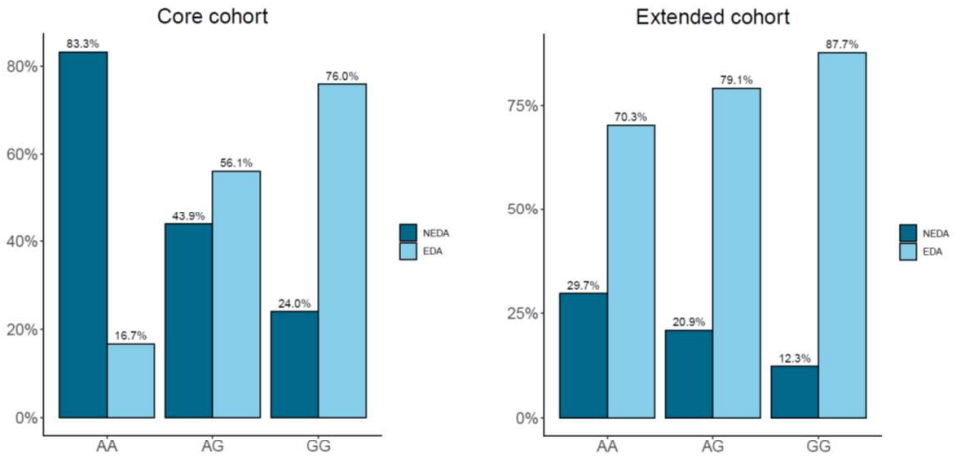


Figure 3.8: Proportion of patients with evidence of disease activity (EDA) or stable disease (NEDA) at 4 years according to genotype at rs12704796 in the CC and EC. *On the y-axis is the proportion of patients with no evidence of disease activity (NEDA, blue) or EDA (light blue). On the x-axis are the genotypes at the rs12704796 variant.*

Rs12704796 is located at the 5' of the coding region for *PON2*, in a regulatory region that acts as a weak enhancer in CD34+ primary cell, and regulates its expression in several tissue, among which nerve, muscle, pancreas and intestines (https://gtexportal.org (Lonsdale *et al*, 2013)); the A allele is associated with a higher expression of *PON2*. This gene codes for a member of the paraoxonase family, that has

antioxidant properties and has been associated with a higher risk of atherosclerosis (Devarajan *et al*, 2011) and vascular diseases (Slowik *et al*, 2007) as well as neurodegenerative disease (Erlich *et al*, 2006). Besides, it has been suggested that oxidative damage plays an important role in MS, active demyelinating lesions being enriched in oxidized lipids and DNA (Haider *et al*, 2011). For these reasons, the PON2 locus represent an interesting candidate for association with disease activity in MS, even if previous studies evaluating the association between its homologous PON1 and MS have shown negative results (Martínez *et al*, 2010).

3.3.1.2 Genome-wide association study on TFR

3.3.1.2.1 CC

We then moved on to analyze the genetic associations with TFR, starting with the 195 patients of the CC. As for the NEDA-3 analysis, we included in the model the clinical covariates Figure 3.9 illustrates the overall results of the GWA study in this smaller cohort and Table 3.7 lists the 10 most associated variants.

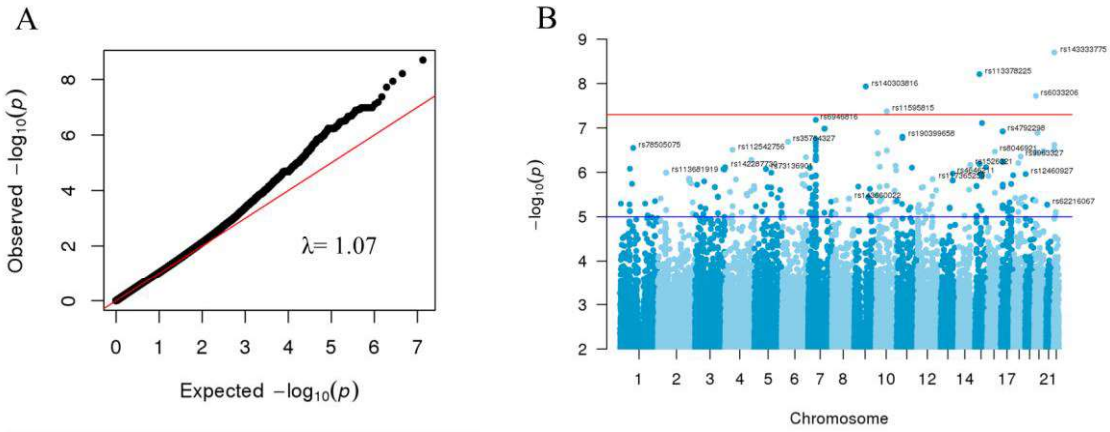


Figure 3.9: QQ plot and Manhattan plot showing results of the association with TFR in the CC. *A: Quantile-quantile plot plotting the –log10(p-value) distribution of our analysis versus a theoretical normal distribution. B: Manhattan plot showing association p-value for each SNPs across different chromosomes. λ: genomic inflation factor.*

Table 3.7: Top 10 results of the analysis of association with TFR in the CC

SNP	Chr	Gene	BP	A1	MAF	HR	95%-CI	P Value
rs143333775	22	*PI4KA, SERPIND1*	21142349	T	0.01	14.54	6.06-34.86	1.99E-09
rs113378225	15	*MYO5C*	52554644	T	0.02	11.14	4.94-25.13	6.14E-09
rs140303816	9	*GDA*	74770266	A	0.01	9.02	4.24-19.21	1.16E-08
rs6033206	20	*BTBD3 (5')*	11699643	C	0.02	27.34	8.63-86.63	1.89E-08
rs11595815	10	*COL13A1*	71568335	A	0.12	3.55	2.26-5.59	4.21E-08
rs6946816	7	*COBL (5')*	51523734	A	0.45	2.85	1.95-4.17	6.58E-08
rs150600742	15	*SPESP1*	69199776	A	0.01	23.23	7.37-73.19	7.75E-08
rs117167631	7	*C7orf66 (5')*	108803509	A	0.02	15.45	5.64-42.37	1.04E-07
rs118077222	7	*C7orf66 (5')*	108860728	A	0.02	15.45	5.64-42.37	1.04E-07
rs148930293	7	*C7orf66 (5')*	108886272	T	0.01	15.45	5.64-42.37	1.04E-07

SNP: variant name; Chr: chromosome; Gene: gene the variant is mapped to, when variant is intergenic the name of the closest coding gene and direction is reported; BP: physical position of the SNP in build 37; A1: minor allele; MAF: minor allele frequency in the European population of 1000 Genome project; HR: hazard-ratio, 95%-CI: confidence interval for the hazard ratio; P: association p-value.

Several variants passed the Bonferroni-corrected p-value threshold for association, however in most cases the signals were quite isolated (Figure 3.9 B) and involved low frequency variants, suggesting the possibility of spurious association. Also the QQ plot in Figure 3.9 A displayed an accentuated upward deflection suggesting an enrichment of positive association. Indeed, it is well known that the time-to-event analysis holds a higher power to identify truly significant associations compared to case-control studies (Hughey *et al*, 2019), however this type of approach has also been linked to a higher rate of type I error, mainly when dealing with rare variants and few events (Bi *et al*, 2020). Therefore, our meta-analytic approach can help mitigate this issue by selecting variants whose association with TFR is supported by two independent cohorts of RRMS patients.

3.3.1.2.2 EC

The QQplot and corresponding Manhattan plot for the TFR analysis in the EC are shown in Figure 3.10: also in this dataset, applying a cox regression model we were able to identify some genome-wide significant variants (Table 3.8). However, as expected due to the larger sample size and the corresponding number of events, the

association signals showed in Figure 3.10 B are more defined and not isolated, suggesting a true association. Seemingly, the distribution of the observed p-value is consistent with the presence of few associated causal polymorphism.

Noteworthy, among the top is a signal mapping to chromosome 3 to the region coding for SCL9A9, previously associated to response to IFNβ by a work from our group.

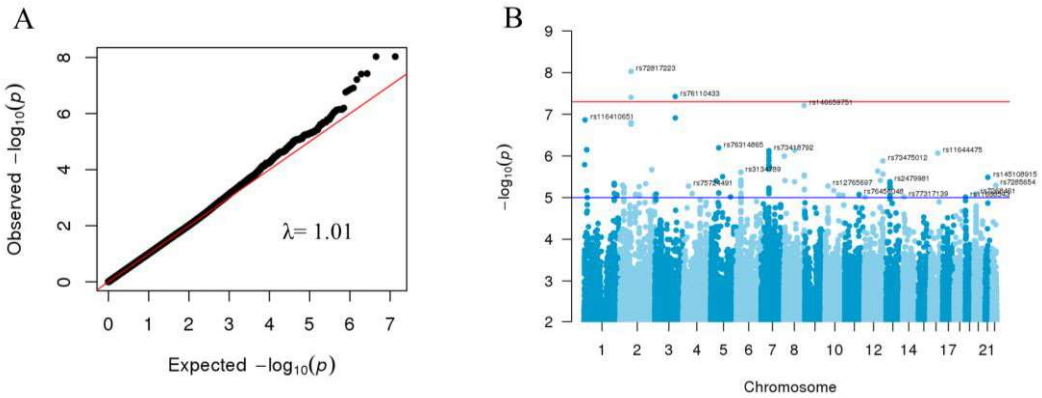


Figure 3.10: QQ plot and Manhattan plot showing results of the association with TFR in the EC. *A: Quantile-quantile plot plotting the –log10(p-value) distribution of our analysis versus a theoretical normal distribution. B: Manhattan plot showing association p-value for each SNPs across different chromosomes. λ: genomic inflation factor.*

Table 3.8: Top 10 results of the analysis of association with TFR in the EC

SNP	Chr	Gene	BP	A1	MAF	HR	95%-CI	P Value
rs62167313	2	*SNAR-H (5')*	78230161	A	0.01	2.98	2.05-4.33	9.33E-09
rs72817223	2	*SNAR-H (5')*	78216423	T	0.01	2.98	2.05-4.33	9.33E-09
rs76110433	3	*SLC9A9*	143292904	C	0.04	2.61	1.85-3.67	3.76E-08
rs62164743	2	*SNAR-H (5')*	78182091	T	0.01	2.91	1.99-4.25	3.92E-08
rs146659751	8	*CAMTA1*	140881285	C	0.02	3.17	2.09-4.81	6.16E-08
rs6784766	3	*SLC9A9*	143262448	T	0.04	2.49	1.78-3.49	1.22E-07
rs116410651	1	*CAMTA1*	7340079	T	0.02	4.22	2.47-7.2	1.37E-07
rs72815190	2	*AC073628.1*	78148685	C	0.01	2.83	1.92-4.17	1.56E-07
rs62163506	2	*SNAR-H (5')*	78131966	T	0.01	2.81	1.91-4.15	1.73E-07
rs76314865	5	*ACTBL2*	56863619	A	0.03	2.34	1.68-3.27	6.38E-07

SNP: variant name; Chr: chromosome; Gene: gene the variant is mapped to, when variant is intergenic the name of the closest coding gene and direction is reported; BP: physical position of the SNP in build 37; A1: minor allele; MAF: minor allele frequency in the European population of 1000 Genome project; HR: hazard-ratio, 95%-CI: confidence interval for the hazard ratio; P: association p-value.

3.3.1.2.3 Meta-analysis

We finally combined the results from the two datasets by means of a fixed-effect meta-analysis, whose results are shown in Figure 3.11 and Table 3.9.

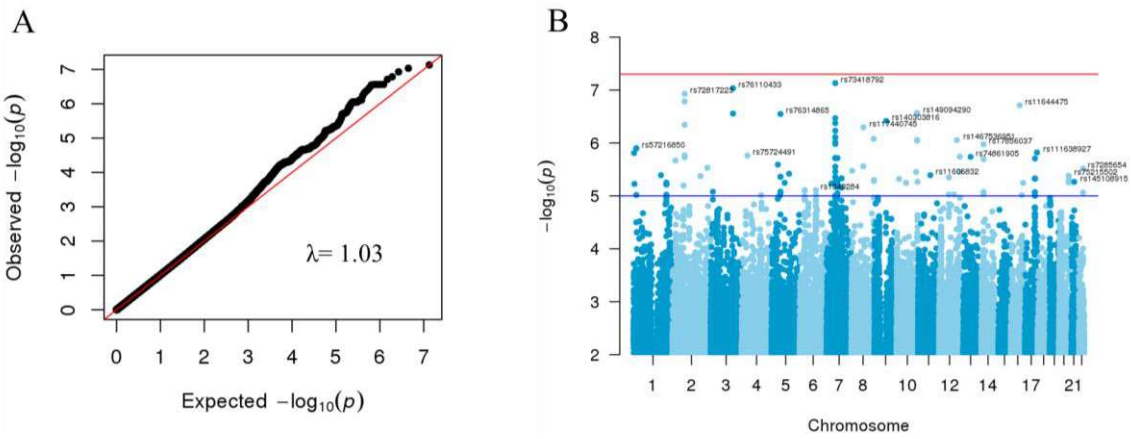


Figure 3.11: QQ plot and Manhattan plot showing results of the meta-analysis for TFR. *A: Quantile-quantile plot plotting the –log10(p-value) distribution of our analysis versus a theoretical normal distribution. B: Manhattan plot showing association p-value for each SNPs across different chromosomes. λ: genomic inflation factor.*

There are no variants beyond the genome-wide significant Bonferroni threshold, however some signals with a suggestive level of association are present.

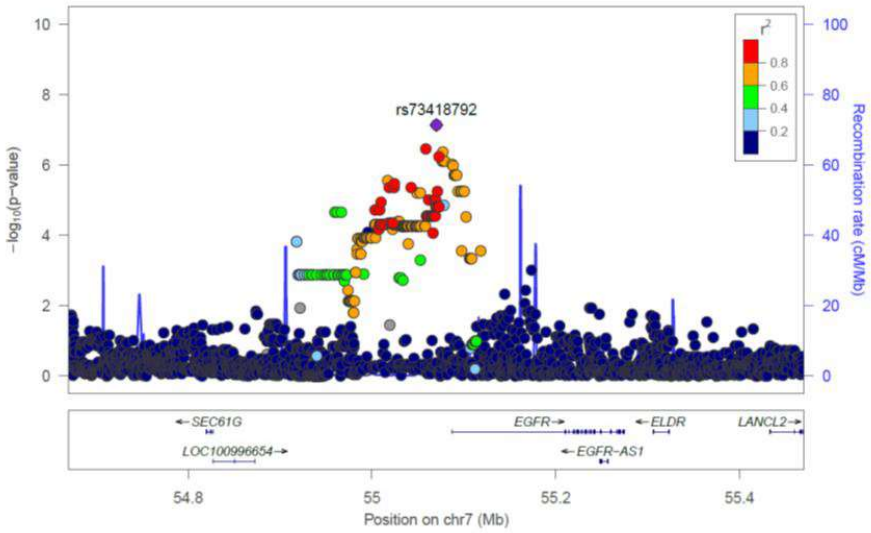


Figure 3.12: Regional association plot for the rs73418792 locus in the Meta-analysis. *Illustration of the chr7p11 region centered on rs73418792, with the local recombination rate plotted in light blue. Each point represents a SNP; the most associated SNP in the meta-analysis, rs73418792, is marked in purple. The color of each dot represents the extent of LD with rs73418792 according to the legend on the right. Physical positions are based on build 37 of the human genome.*

Specifically, the top-associated variant is rs73418792 on chromosome 7: 55070106 that shows a p-value very close to the genome-wide cut-off (p-value 7.39e-8, Figure 3.12); at this SNP the C allele is correlated with a higher risk of early relapse (HR 2.73 in the meta-analysis; HR 3.12 and p-value 0.033 in the CC; HR 2.68 and p-value 7.49e-7 in the EC, Figure 3.13).

Rs73418792 is a very interesting variant because it is located very close to the *EGFR* gene, about 17Kbp away from the coding region at the 5' (Figure 3.13), in a possible regulatory region.

In several reports, the *EGFR* has been implicate in remyelinating processes (Aguirre *et al*, 2007; Palazuelos *et al*, 2015) and a reduction in the expression of EGF receptors in MS brain compared to controls has been reported (Nicoletti *et al*, 2019). Besides, treatment with EGF was able to improve the clinical features of an EAE model, thus supporting its role in disease mechanisms. Indeed, EGFR seems to interact with and mediate the activity of LINGO1 (Lee *et al*, 2014), a molecule known to be involved in oligodendrocyte differentiation and myelination that has been proposed as a treatment to promote tissue repair in MS (Tran *et al*, 2014; Klistorner *et al*, 2018).

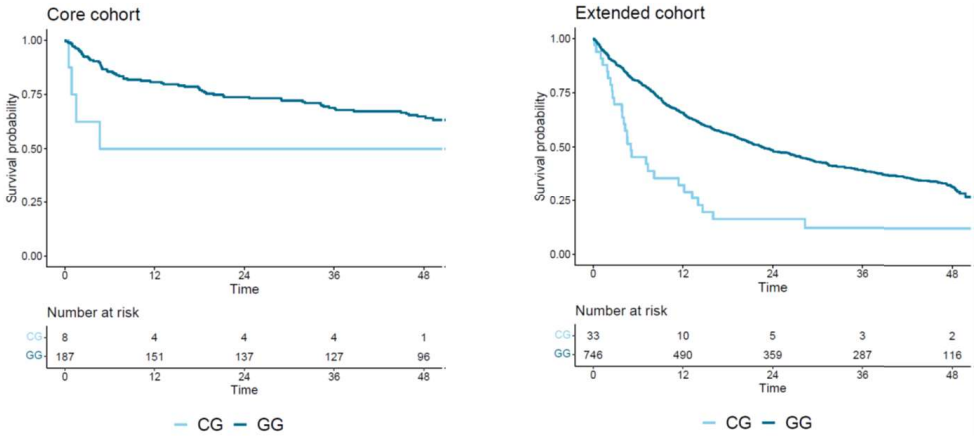


Figure 3.13: Kaplan Meier survival curves illustrating the proportion of patients free from clinical activity in the CC and EC, stratified according to rs73418792 genotype. *The x-axis indicates months from baseline visit; Y-axis indicate the proportion of relapse-free patients.*

Table 3.9: Top 10 results of the meta-analysis of association with TFR

SNP	Chr	Gene	A1	MAF	HR_CC	95CI_CC	P_CC	HR_EC	95CI_EC	P_EC	HR_meta	95CI_meta	P_meta	I2
rs73418792	7	EGFR (5')	C	0.01	3.12	1.1-8.89	0.033	2.68	1.82-3.97	7.49E-07	2.73	1.9-3.94	7.39E-08	0.00
rs76110433	3	SLC9A9	C	0.04	0.96	0.23-4.06	0.957	2.61	1.85-3.67	3.76E-08	2.47	1.77-3.45	9.24E-08	42.7
rs72817223	2	SNAR-H (5')	T	0.01	0.71	0.22-2.29	0.570	2.98	2.05-4.33	9.33E-09	2.61	1.83-3.73	1.18E-07	80.9
rs62167313	2	SNAR-H (5')	A	0.01	0.79	0.29-2.19	0.651	2.98	2.05-4.33	9.33E-09	2.55	1.8-3.62	1.64E-07	82.7
rs11644475	16	CETP	G	0.02	2.95	0.86-10.2	0.087	2.44	1.71-3.49	8.52E-07	2.48	1.76-3.49	1.94E-07	0.00
rs117923868	10	PTPRE	C	0.04	3.83	1.66-8.84	1.63E-03	1.95	1.44-2.64	1.52E-05	2.11	1.59-2.81	2.74E-07	54.9
rs140772206	10	PTPRE	C	0.04	3.83	1.66-8.84	1.63E-03	1.95	1.44-2.64	1.52E-05	2.11	1.59-2.81	2.74E-07	54.9
rs149094290	10	PTPRE	A	0.04	3.83	1.66-8.84	1.63E-03	1.95	1.44-2.64	1.52E-05	2.11	1.59-2.81	2.74E-07	54.9
rs76859393	10	PTPRE	G	0.04	3.83	1.66-8.84	1.63E-03	1.95	1.44-2.64	1.52E-05	2.11	1.59-2.81	2.74E-07	54.9
rs6784766	3	SLC9A9	T	0.04	0.96	0.23-4.06	0.957	2.49	1.78-3.49	1.22E-07	2.37	1.7-3.29	2.78E-07	37.1

SNP: variant name; Chr: chromosome; Gene: gene the variant is mapped to, when variant is intergenic the name of the closest coding gene and direction is reported; MAF: minor allele frequency in the European population of 1000 Genome project; A1: minor allele; HR_CC: hazard-ratio in the core cohort, 95%-CI_CC: confidence interval for the hazard-ratio in the core cohort; P_CC: association p-value in the core cohort; HR_EC: hazard-ratio in the extended cohort, 95%-CI_EC: confidence interval for the hazard-ratio in the extended cohort; P_EC: association p-value in the extended cohort; HR_meta: hazard-ratio in the meta-analysis, 95%-CI_meta: confidence interval for the hazard-ratio in the meta-analysis; P_meta: association p-value in the meta-analysis.

Another noteworthy result is represented by the signal on chromosome 3 in the region coding for *SLC9A9*, where the most associated variant is rs76110433 (HR 2.47 p-value 9.24E-08). This gene, that encodes a Na(+) -H(+) exchanger, has previously been implicated with the response to IFNbeta in RRMS by a previous work of our group (Esposito *et al*, 2015) in a cohort of patients partially overlapping with the EC. Interestingly, the identified signal does not correspond to the previously reported one but seems to support the involvement of this region in modulating disease activity rather than the response to a specific treatment. However the association signal was completely driven by the results in the EC and it was not present in the CC (HR 2.61 and p-value 3.76e-8, HR 0.96 and p-value 0.957 respectively), likely due to the small sample size and very low MAF (0.04), so that it is not fully supported by the results of the meta-analysis.

On the contrary, the signal on chromosome 16:57007652 leaded by rs11644475 is highly significant in the EC (HR 2.44 p-value 8.52e-7) but showed the same direction of effect and a trend towards significance in the CC (2.95 p-value 0.087). Therefore its association was reinforced by the meta-analysis (Figure 3.14) that confirmed a significantly increased risk of early relapse in patients carrying the G allele (2.48 p-value 1.94e-7, Figure 3.15).
This variant is intronic to the gene coding for cholesteryl ester transfer protein, *CETP,* that is involved in the transfer of lipids among lipoprotein particles, and rs11644475 also has an eQTL effect on its expression in adipose tissue. Interestingly, genetic variants in this gene have been associated with the level of circulating lipoproteins and also with Vitamin D levels (Sinnott-Armstrong *et al*, 2021). Specifically rs9939224^{G}, that is in high linkage disequilibrium (D'=1) with the identified risk variant rs11644475^{G}, has been associated with a decrease in serum Vitamin D levels that is a well-known risk factor for MS development and disease activity (Ascherio *et al*, 2014). Moreover, a small recent study also correlated genetic polymorphisms in CETP with the risk of optic neuritis development (Gedvilaite *et al*, 2019).

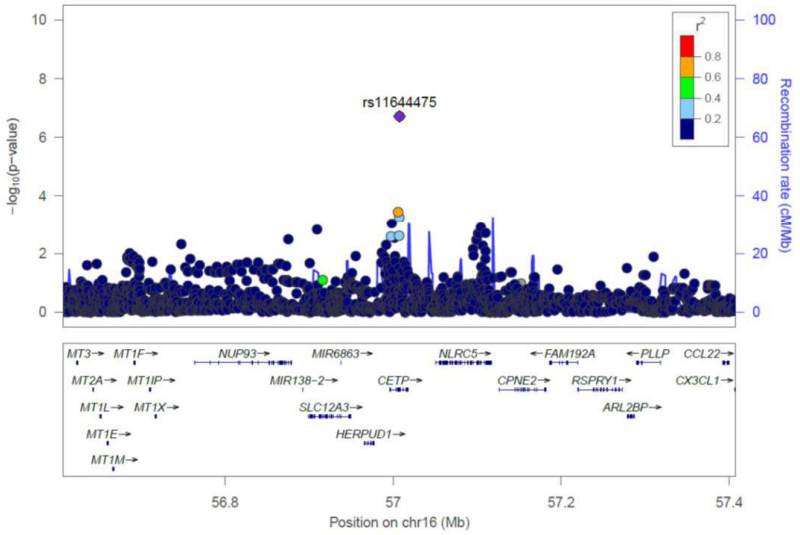


Figure 3.14: Regional association plot for the rs11644475 locus in the Meta-analysis. *Illustration of the chr16q13 region centered on rs11644475, with the local recombination rate plotted in light blue. Each point represents a SNP; the most associated SNP in the meta-analysis, rs11644475, is marked in purple. The color of each dot represents the extent of LD with rs11644475 according to the legend on the right. Physical positions are based on build 37 of the human genome.*

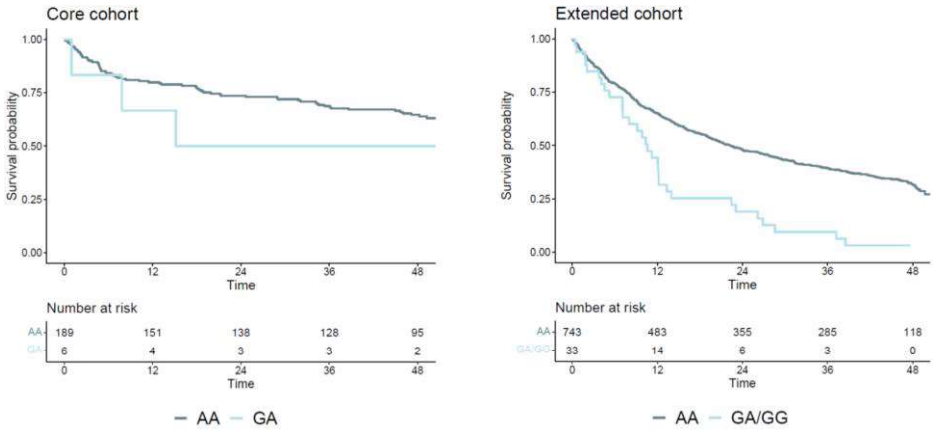


Figure 3.15: Kaplan Meier survival curves illustrating the proportion of patients free from clinical activity in the CC and EC, stratified according to rs11644475 genotype. *The x-axis indicates months from baseline visit; Y-axis indicate the proportion of relapse-free patients.*

3.3.1.3 Gene-based and pathway analysis

3.3.1.3.1 NEDA-3

In order to ease the interpretation of the genetic results and allow for the identification of possibly converging biological effects spread all over the genome, we proceeded with the gene-based and pathway analysis. First, we generated a gene-wise p-

value using the online tool VEGAS2 (Mishra & Macgregor, 2015) for 23,750 genes and we set the Bonferroni adjusted threshold to 2.1e-6. Table 3.10 lists the ten most associated genes and their relative p-value. There are no genes passing the corrected p-value threshold but, interestingly, the results seems to point to a region on chromosome 11, around the top-SNP rs4939517, that was not highlighted by the meta-analysis (p-value of genetic association 8.66e-5). This association with this region is remarkable given that it is a region coding for several genes among which also CD5 and CD6, previously associated with MS susceptibility (Patsopoulos *et al*, 2019). Indeed, rs4939517 is in LD (D' 0.5) with another SNPs, rs11230624 that shows an eQTL effect on CD5 in brain (Lonsdale *et al*, 2013).

Table 3.10: Top 10 genes associated with NEDA-3 in the meta-analysis

CHR	GENE	Gene_P	nSNPs	Leading-SNP
11	*TMEM138*	5	5.40E-05	rs4939517
11	*CYB561A3*	5	7.10E-05	rs4939517
5	*ACTBL2*	117	7.50E-05	rs10940562
6	*C6orf106*	268	7.50E-05	rs2262245
11	*DAK*	6	7.60E-05	rs4939517
11	*DDB1*	6	0.000176	rs35723406
6	*SNRPC*	170	0.000193	rs1998702
6	*UHRF1BP1*	301	0.000223	rs1998702
11	*VWCE*	8	0.000232	rs4274208
11	*TMEM216*	16	0.000251	rs7118316

CHR: chromosome; Gene_P: gene-wise p-value; nSNPs: number of SNPs included in the computation of gene-wise p-value; Leading_SNP: most associated SNP in the gene.

We then went further on to test the biological pathways that are more likely to be implicated with the occurrence of disease activity and, starting from the gene-based p-values we performed a gene-set enrichment analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) repositories as reference. Table 3.11 lists the top enriched pathways. Once again there are no results surviving multiple testing correction but among the topmost nominally associated pathways several are immune-related and include "NOD-like receptor signaling pathway", "Autoimmune thyroid disease", "Hepatitis B" and "Toll-like receptor signaling pathway" suggesting that, as expected,

immunological processes play a key role in determining breakthrough inflammation during disease course.

Table 3.11: KEGG biological pathways most enriched in genes associated with NEDA-3

KEGG pathway	Size	LE	ES	NES	PValue	FDR
NOD-like receptor signaling pathway	160	66	0.49	1.24	0.001	0.707
Autoimmune thyroid disease	48	18	0.59	1.40	0.006	0.598
Hepatitis B	140	46	0.50	1.24	0.006	0.742
Renin-angiotensin system	20	8	0.66	1.47	0.006	0.307
Riboflavin metabolism	8	2	0.77	1.51	0.012	0.322
Nicotinate and nicotinamide metabolism	29	10	0.61	1.40	0.016	0.451
Toll-like receptor signaling pathway	96	34	0.49	1.22	0.021	0.694
Bile secretion	70	35	0.51	1.25	0.024	0.765
Taste transduction	80	44	0.49	1.20	0.039	0.763
Cytosolic DNA-sensing pathway	59	19	0.51	1.24	0.043	0.690

Biological pathways enriched in genes associated with NEDA-3, based on the KEGG database. LE: number of leading edges; ES: enrichment score; NES: normalized enrichment score; FDR: false discovery rate. Results are listed according to the nominal p-value.

3.3.1.3.2 TFR

We then repeated the same process for the TFR analysis. Table 3.12 shows the 10 genes with the highest association to TFR and Table 3.13 lists the top pathways enriched for genes associated with TFR.

Again, among the top-associated pathways, most were related to immune functions such as “RIG-I-like receptor signaling pathway”, “Hepatitis B”, “Epstein-Barr virus infection”, “Toll-like receptor signaling pathway”, “Herpes simplex infection”, further supporting the main role played by inflammatory pathways in mediating inflammatory disease activity. Besides, there were also some results pointing to pathways involved in the metabolism of macromolecules and vitamins (“Inositol phosphate metabolism”, “Pantothenate and CoA biosynthesis” for TFR and “Riboflavin metabolism” and “Nicotinate and nicotinamide metabolism” for NEDA-3).

Table 3.12: Top 10 genes associated with TFR in the meta-analysis

CHR	GENE	Gene_P	nSNPs	Leading-SNP
19	*MUM1*	1.40E-05	79	rs11668543
3	*TOPAZ1*	1.70E-05	254	rs10510741
6	*LOC100506804*	8.60E-05	52	rs171263
2	*FBXO48*	1.02E-04	72	rs112407290
12	*LOH12CR1*	1.03E-04	376	rs35908046
19	*NDUFS7*	1.14E-04	72	rs11668809
11	*MIR4492*	1.17E-04	51	rs12282752
11	*BCL9L*	1.31E-04	70	rs12282752
3	*TCAIM*	1.98E-04	187	rs36051440
6	*LOC101929239*	2.33E-04	121	rs1830735

CHR: chromosome; Gene_P: gene-wise p-value; nSNPs: number of SNPs included in the computation of gene-wise p-value; Leading_SNP: most associated SNP in the gene.

Table 3.13: KEGG biological pathways most enriched in genes associated with TFR

KEGG pathway	Size	LE	ES	NES	PValue	FDR
RIG-I-like receptor signaling pathway	66	30	0.54	1.35	0.005	0.639
PI3K-Akt signaling pathway	339	116	0.43	1.16	0.008	0.679
Hepatitis B	140	55	0.47	1.23	0.008	0.687
Epstein-Barr virus infection	192	73	0.46	1.20	0.009	0.670
Gastric cancer	147	49	0.47	1.22	0.01	0.687
Toll-like receptor signaling pathway	96	35	0.47	1.22	0.028	0.683
Herpes simplex infection	171	68	0.44	1.17	0.029	0.683
Inositol phosphate metabolism	69	26	0.49	1.22	0.033	0.686
Protein digestion and absorption	84	41	0.47	1.20	0.044	0.669
Pantothenate and CoA biosynthesis	19	10	0.59	1.35	0.056	0.600

Biological pathways enriched in genes associated with TFR, based on the KEGG database. LE: number of leading edges; ES: enrichment score; NES: normalized enrichment score; FDR: false discovery rate. Results are listed according to the nominal p-value.

3.3.2 Transcriptomic study

Four out of 187 samples with available gene expression data were excluded because they failed quality control checks on extracted RNA, whereas another sample was discarded for the suboptimal quality of library preparation. In the end, gene expression information was available for 182 patients with an overall high quality of the sequencing data, as shown in Figure 3.16 with a median depth of 54.7 million reads (range 39.1 - 602.2). On average 28.4 million reads per patients were mapped to GENCODE features, for a total of 55,765 features identified. After filtering, 40,207 features remained for further processing.

Figure 3.16: Quality controls results for RNA sequencing data. *The picture illustrates the distribution of Phred scores for the sequenced bases for each sample. Phred score is a quality score that is logarithmically related to the base-calling error probabilities (Pe), so that when Phred score = 20, Pe = 10^{-2}, when Phred score = 30, Pe = 10^{-3}. The x-axis reports the Phred score and the y-axis the count of bases with that corresponding score.*

A hierarchical clustering and PC analysis were performed to rule out the presence of significant impact of library preparation batches, previous treatments and age on gene expression signature. On the contrary, we found a strong segregation between females and males patients (Figure 3.17), thus we included gender as a covariate in the subsequent analyses. Moreover, the analysis of the Cook distance allowed the identification of 5 additional samples classified as outliers who were excluded from further processing.

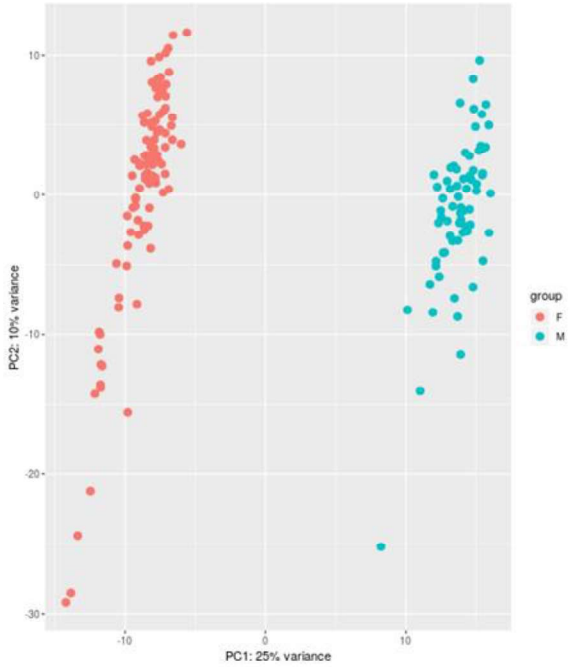


Figure 3.17: Principal component analysis to evaluate the impact of gender on gene expression data. *The diagram shows a plot of the first 2 principal components (PCs) colored according to gender (F= females, M= males) . x-axis and y-axis reports the proportion of variance explained by the 1st and 2nd PC respectively.*

3.3.2.1 NEDA-3

Among the 177 patients considered for the differential expression analysis, 60 (34%) had a stable disease during follow-up while 117 (66%) experienced disease reactivation (EDA). Table 3.14 shows the 6 differential expressed genes (DEGs) that were detected by the analysis.

Table 3.14: Differentially expressed genes according to NEDA-3 outcome

DEG	Log_2FC	p-value	FDR
XIST	-3,42	1,56E-19	4,34E-15
HBA2	-1,50	4,28E-10	5,94E-06
HBA1	-1,35	8,64E-07	7,08E-03
UTS2	-1,05	2,95E-06	1,64E-02
PCDHGB2	-0,57	5,11E-06	2,36E-02
MTRNR2L12	-1,28	6,33E-06	2,51E-02

DEG: differentially expressed gene; Log2FC: base 2 logarithm of the fold-change; p-value: nominal association p-value; FDR: false discovery rate.

Among the significant DEGs it is to note PCDHGB2, Protocadherin Gamma Subfamily B2, member of a large family of cadherin-related genes that are highly expressed in brain; it has been hypothesized to play a role le in cell adhesion, cellular interactions and regulation of the brain vasculature (Gabbert *et al*, 2020).

3.3.2.2 Clinical activity

Similarly to what done for the NEDA-3 outcome, we compared gene expression profiles in 65 patients (37%) that experienced at least one relapse during the 4-year observation period and 112 subjects who did not (63%). DEGs that survived multiple testing correction are reported in Table 3.15.

Table 3.15: Differentially expressed genes according to clinical activity during follow-up

DEG	Log_2FC	p-value	FDR
RP11-632C17_A.1	-0,98	1.32E-06	0.03
CLTB	0,25	2.77E-06	0.03
HIST1H3J	0,62	3.24E-06	0.03
RP11-359M6.3	-1,01	5.73E-06	0.04
HIST1H3G	0,68	8.24E-06	0.04
HIST1H2AB	0,59	8.67E-06	0.04
HIST1H2BB	0,58	1.16E-05	0.04
HIST1H3A	0,31	1.25E-05	0.04
HIST1H1B	0,51	1.27E-05	0.04
HIST1H2AI	0,38	1.65E-05	0.05

DEG: differentially expressed gene; Log2FC: base 2 logarithm of the fold-change; p-value: nominal association p-value; FDR: false discovery rate.

It is easily noted that most of the DEGs codes for histones, highly conserved proteins that regulate chromatin state thus modulating genes accessibility and expression as well as DNA replication. Post-transcriptional histone modifications are key regulators of transcriptional activities involved in several biological processes and have also been suggested to influence MS pathophysiology (He *et al*, 2018).

3.3.3 Integrated pathway analysis

As a first step of integration of the different -omics information derived from our cohorts, we performed a pathway analysis starting from the results of the previous analyses. The underlying hypothesis is that genes that show a mild-to-moderate genetic association and/or differential expression but does not satisfy the stringent threshold for multiple testing correction, can possibly interact through common molecular and biological processes, thus contributing to disease inflammatory mechanisms.

Table 3.16 and 3.17 lists the top-10 associated pathways for the NEDA-3 and clinical relapse endpoint.

Interestingly, after multiple testing correction we found that the "Systemic lupus erythematosus" pathway was significantly enriched in genes derived from both outcomes, again stressing the key role of immune-relates processes in possibly influencing disease activity.

Table 3.16: KEGG biological pathways most enriched in genes associated with NEDA-3 in the genetic and/or transcriptomic study

KEGG pathway	Size	Ob	Ex	ER	p-value	FDR
Systemic lupus erythematosus	**133**	**31**	**14.09**	**2.20**	**1.59E-05**	**0.005**
Proteasome	45	13	4.77	2.73	0.001	0.080
Necroptosis	162	31	17.16	1.81	0.001	0.080
Natural killer cell mediated cytotoxicity	131	26	13.87	1.87	0.001	0.090
Renin-angiotensin system	23	8	2.44	3.28	0.002	0.092
Tuberculosis	179	32	18.96	1.69	0.002	0.092
Alcoholism	180	32	19.06	1.68	0.002	0.092
Phagosome	152	28	16.10	1.74	0.002	0.092
Pathways in cancer	526	76	55.71	1.36	0.003	0.092
Kaposi sarcoma-associated herpesvirus infection	186	32	19.70	1.62	0.004	0.120

Biological pathways enriched in genes associated with NEDA-3, based on the KEGG database. Size: number of genes in the pathway; Ob: number of input genes observed in common with the pathway; Ex: expected number of common genes; ER: enrichment ratio; FDR: false discovery rate.

Moreover, for the analysis of relapse activity we also found other 7 significantly enriched pathways: noteworthy, most of them are involved in processes of cell regulation, cellular stress response and cell death ("Proteasome", "Huntington disease",

"Alzheimer disease", "Necroptosis") as well as in signal transduction and response to exogenous stimuli ("Viral carcinogenesis", "ECM-receptor interaction", "Alcoholism").

Table 3.17: KEGG biological pathways most enriched in genes associated with relapse activity in the genetic and/or transcriptomic study

KEGG pathway	Size	Ob	Ex	ER	p-value	FDR
Systemic lupus erythematosus	133	59	16.24	3.63	0.00	**0.00**
Alcoholism	180	63	21.98	2.87	8.88E-16	**1.45E-13**
Viral carcinogenesis	201	48	24.54	1.96	2.52E-06	**2.74E-04**
Proteasome	45	15	5.49	2.73	1.76E-04	**0.014**
Huntington disease	193	41	23.57	1.74	2.29E-04	**0.015**
Alzheimer disease	171	37	20.88	1.77	3.13E-04	**0.017**
ECM-receptor interaction	82	21	10.01	2.10	0.001	**0.030**
Necroptosis	162	34	19.78	1.72	0.001	**0.039**
Human papillomavirus infection	339	59	41.39	1.43	0.003	0.098
Transcriptional misregulation in cancer	186	36	22.71	1.59	0.003	0.100

Biological pathways enriched in genes associated with relapse activity, based on the KEGG database. Size: number of genes in the pathway; Ob: number of input genes observed in common with the pathway; Ex: expected number of common genes; ER: enrichment ratio; FDR: false discovery rate.

3.3.4 Considerations

Through the genomic analysis exposed in this section we tried to highlight genes and biological pathways that are likely to influence the occurrence of inflammatory activity in RRMS patients. First of all, we searched for genetic markers associated with clinical relapses and overall disease activity (also considering MRI information and disability) and, even if we failed to identify genome-wide significant signals after multiple testing correction, we found some interesting loci with a suggestive level of association. In particular, when analyzing the NEDA-3 outcome we found an interesting association on chromosome 2 near the SERPINE2 gene, on which a putative eQTL effect has also been reported. Given the role of SERPINE2 in the brain, where it is highly expressed and secreted by neurons and glial cells, it represents an attractive candidate gene for implication in inflammatory disease activity. Recently, several reports have suggested that the coagulation cascade is activated in the brain of MS patients and can contribute to the pathogenesis of MS lesions (Parsons *et al*, 2017;

Göbel *et al*, 2016); indeed, thrombin, a serine-protease with pleiotropic functions that is generated as a results of the coagulation cascade activation, has been shown to regulate blood brain barrier (BBB) permeability (Sweeney *et al*, 2018) and can modulate immune activation (Jordan *et al*, 2021). A proteomic study that evaluated CSF markers in patients with MS and other neurological diseases found that among the differentially expressed proteins there was an enrichment of molecules involved in complement and coagulation cascades and platelet degranulation (Shi *et al*, 2021), further supporting its involvement in the disease. Finally, a recently published work found an association between genetic polymorphisms in prothromobotic genes and MS (Abbadessa *et al*, 2022). Thus, in our opinion, the effect observed on chromosome 2 support further investigations and a replication in independent cohorts of RRMS patients to confirm the association.

Another interesting variant that warrants further examination has been identified on chromosome 7, close to the *EGFR* gene, when exploring the TFR outcome. *EGFR* is a member of the ErbB receptors family that has been implicated in demyelination and neurodegeneration, also through the induction of necroptosis (Hu *et al*, 2021). EGFR has been shown to mediate neuregulin-1 activity in the regulation of oligodendrocytes differentiation (Ding *et al*, 2021) and also interact with LINGO1 (Lee *et al*, 2014), whose targeted blockage through monoclonal antibodies has been proposed as a possible treatment to enhance remyelination in optic neuritis and MS (Klistorner *et al*, 2018; Tran *et al*, 2014). If confirmed in larger independent datasets, this result suggests that local processes taking place in the CNS and not directly correlated with immune cells infiltration play a key role in modulating the occurrence of inflammatory disease activity and, in the long term, can support the evaluation of new therapeutic strategies for MS. Indeed, our result is in line with previous reports that investigated genetic association with MS phenotype and failed to identify the expected enrichment of immune-related genes outside the MHC locus (Jensen *et al*, 2010), while reporting association with genes related to neural processes such as glutamate signaling and axon guidance (Baranzini *et al*, 2009, 2010).

Similar results were obtained in the transcriptional study. After accounting for the effect of gender, that seemed to drive most of the variability of the transcriptomic information in our cohort, only few DEGs were identified for both outcomes: among the most

interesting elicited genes in the NEDA-3 analysis there was *PCDHGB2*, that is involved in cell-adhesion and regulation of brain vasculature (Gabbert *et al*, 2020), while DEGs from the relapse outcome were enriched in histone proteins, which are key genes for gene expression regulation and have been associated with MS susceptibility and neurodegeneration. Taken together these findings suggest an important role for non-immune mechanisms regulating cell survival, remyelination and BBB permeability in determining disease activity.

The same conclusions also derives from the output of the pathway analysis that integrated the results obtained from the genetic and transcriptomic studies; not only immune-related path, such as "systemic lupus erythematosus", but also several pathways involved in processes of cell regulation, cellular stress response and cell death ("Proteasome", "Huntington disease", "Alzheimer disease", "Necroptosis") were prioritized by the analysis, confirming the importance of these cellular mechanisms for MS pathophysiology.

3.4 Immune analysis

In the immune sub-study we further characterized our patients by evaluating their immunological features through TCR sequencing.

TCR sequencing is a quite novel high throughput technology based on multiplex PCR that allows for the detection of tens thousands of rearranged TCR sequences, allowing to retrieve highly detailed information on the characteristics of an individual's immune repertoire.

TCR consists of 2 chains, called alfa and beta in the vast majority of cases, each one having a constant (C) and a variable region; in turn, the variable region is made of gene segments called V, J and, for the beta chain only, D segments.

Specifically, 52 V, 2 D and 13 J gene segments exist for the beta chain, as well as almost 70 V segments and 61 J segments for the alfa chain. During their development in the thymus, T cells undergo a genetic rearrangement so that one segment of each V, D and J class is selected to be part of the variable region, thus allowing for a great level of variety due to the many combinatorial possibilities (Figure 3.18). Moreover, during the joining process additional nucleotides are randomly inserted or deleted at the rearrangement junctions, further increasing the level of diversity generated by the

process (Janeway *et al*, 2001). In particular, the greatest variability is concentrated in a region called hypervariable region 3, or complementarity-determining region 3 (CDR3), that is the main responsible for antigen recognition.

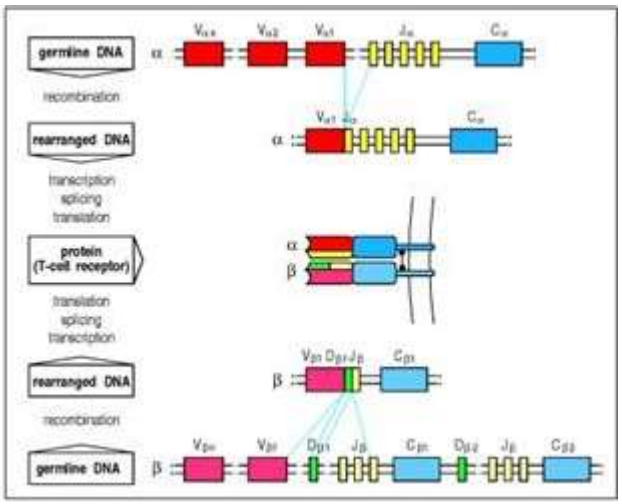


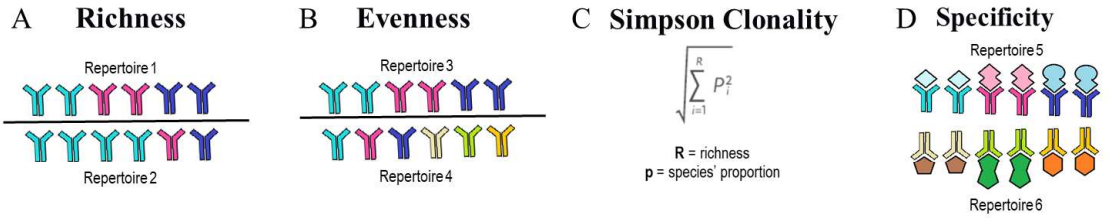
Figure 3.18: Schematic overview of the V(D)J recombination occurring in the alfa and beta chains from Janeway et al. (Janeway *et al*, 2001)***.*** *VJ segments are available for the TCR alfa chain while VDJ segments were present at the TCRbeta locus.*

It is estimated that this process can generate up to 10^{15}-10^{20} different TCR sequences, however the number of actual sequences produced in a single individual, that is, his immune repertoire, is not exactly known and is likely much lower, in the range of 10^{7}-10^{8}. Moreover, the number and characteristics of the TCR sequences are influenced by factors, such as previous infections and diseases (Goronzy *et al*, 2015), and in turn affect the spectrum of antigens that can be recognized by the specific subject (Arnaout *et al*, 2021).

By retrieving information on the variability of the immune repertoire, TCR sequencing represents a new and attractive tool to investigate the immunological status of people with MS and other autoimmune disorders (Wu *et al*, 2021). Several information can be derived through TCR sequencing, among which we identify two main types: 1) measures of overall repertoire diversity, that is defined using parameters such as richness (the number of different TCR sequences identified in a sample), evenness (a measure of the homogeneity in TCR sequencing distribution) or combination of these two, e.g. Simpson clonality (SC) (Figure 3.19 A, B and C); 2) information on TCR specificity, that is, on the exact sequence of the CDR3 region and its characteristics that are tightly linked to the spectrum of antigens recognized by the TCR (Figure 8.2 D).

In our analysis we focused on SC, as a summary measure of immune repertoire characteristics, which has already been associated with MS activity in patients treated with autologous stem cell transplantation (ASCT)(Muraro *et al*, 2014; Hayashi *et al*, 2021; Amoriello *et al*, 2020).

Figure 3.19: Schematic representation of measures of immune-repertoire characteristics



A: Richness is defined by the number of different TCRs identified in a sample, irrespective of their frequencies; in the example both repertoire 1 and 2 have a richness of 3. B: Evenness is a measure of how homogeneously the TCRs are distributed and ranges from 0, lowest evenness possible, to 1 when all the TCRs are equally numerically represented; in the example both repertoire 3 and 4 have the same evenness=1. C: simpson clonality is a diversity measure that takes into account both richness and evenness in the same index. D: Specificity is related to the nucleotidic and aminoacidic sequence and their properties to bind antigens; in the example repertoire 5 and 6 have the same richness and the same evenness but completely differ for antigen specificity.

Among 187 subject who were sampled before a first-line treatment start to generate TCR repertoire data, 4 failed the QCs; a total of 183 patients were included in further analyses. At the end, high quality sequencing data were obtained for the included samples (89.7% of clusters passing the QC filters).

Overall, we obtained 20,737,796 productive sequences, on average 113,321 per patient (range 33,842 – 289,477), that correspond to a mean of 76,423 unique clonotypes per sample (range 29,229 – 144,391)(Figure 3.20).

After the down-sampling procedure, 33,841 productive sequences for each patient were considered, corresponding to a mean of 26,762 unique clonotypes per patient (range 10,883 – 32,927).

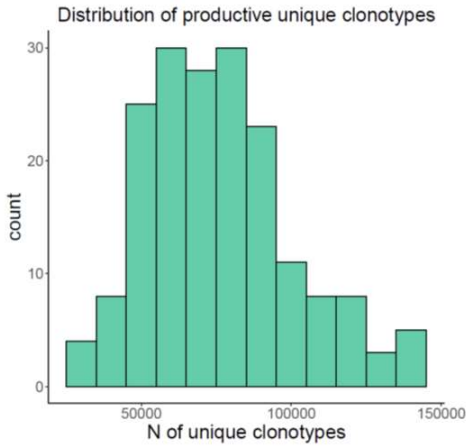


Figure 3.20: Distribution of the number of productive clonotypes per sample. *The x-axis shows the number of unique clonotypes; the y-axis shows the count of samples with a number of unique clonotypes included in the interval displayed in the x-axis.*

The mean value of SC calculated on the down-sampled dataset was 0.035 (range 0.006 – 0.22)(Figure 3.21): as expected from the general population, the value of SC was skewed towards 0. There was a high correlation between the SC calculated on the original dataset and the one calculated after down-sampling (r 0.99, p-value <0.0001), suggesting that the repertoire size reduction does not bias the diversity estimation.

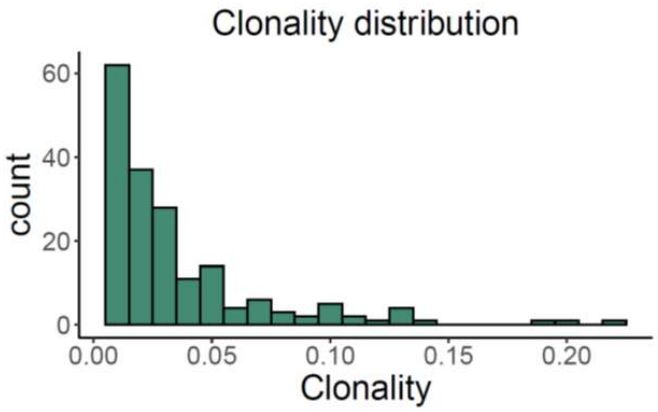


Figure 3.21: Simpson clonality distribution. *The x-axis shows the value of Simpson clonality; the y-axis shows the count of samples with clonality values included in the interval displayed in the x-axis.*

We then tested the correlation between SC and the main baseline variables, to identify the possible confounders to be included as covariates in the regression model.

As expected, we found a strong correlation between SC and the age at sampling (rho 0.37, p-value 2.4 e-7, Figure 3.22 A), with an increase in SC in older patients. We also found a significant difference in mean clonality in males (mean SC 0.043) compared to

females (mean SC 0.03, p-value 0.02)(Figure 3.22 B). Therefore, gender and age at sampling were included as covariates when analyzing the correlation between SC and the clinical outcomes.

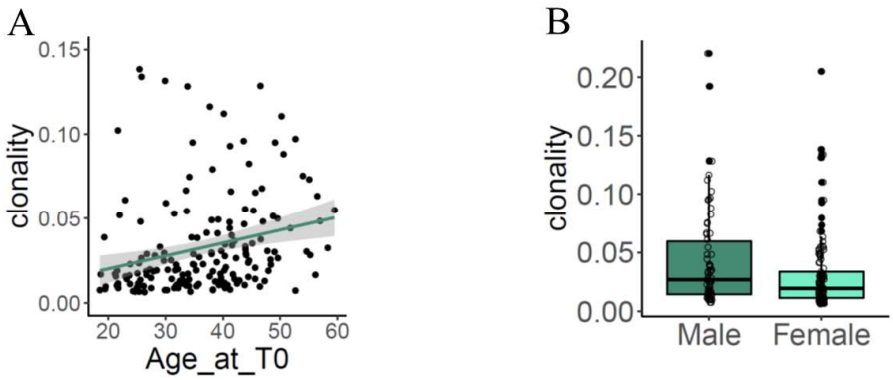


Figure 3.22: Correlation of Simpson clonality with age at sampling and gender. *A) SC tends to increase with increasing age at sampling; B) The mean value of SC was higher in males compared to females patients.*

3.4.1 NEDA-3

First, we evaluated if there was a difference in SC among patients who showed disease activity during the 4-year follow-up or had a stable disease and we found that NEDA patients had a significantly higher baseline SC compared to subjects with EDA (mean SC 0.041 vs 0.033 respectively, p-value 0.0036)(Figure 3.23 A).

In order to test the prognostic value of SC in predicting disease status at 4-years, we then divided the CC into two sets that we called “Low Clonality” and “High Clonality” groups by dividing patients according to the median value of SC. In a univariable regression model, SC was predictive of NEDA status at 4-years, the “High clonality” group having a lower risk of disease reactivation in the observation period (OR 0.33, p:0.001)(Figure 3.23 B).

In order to account for possible confounders, we the run a multivariable analysis including gender and age at sampling as covariates, together with the ARR in the year prior to sampling and the number of active lesions at baseline MRI that were associated with the clinical outcome. After correction for these parameters, the association of SC with the NEDA endpoint was reduced but still statistically significant (OR 0.43, p-value 0.028).

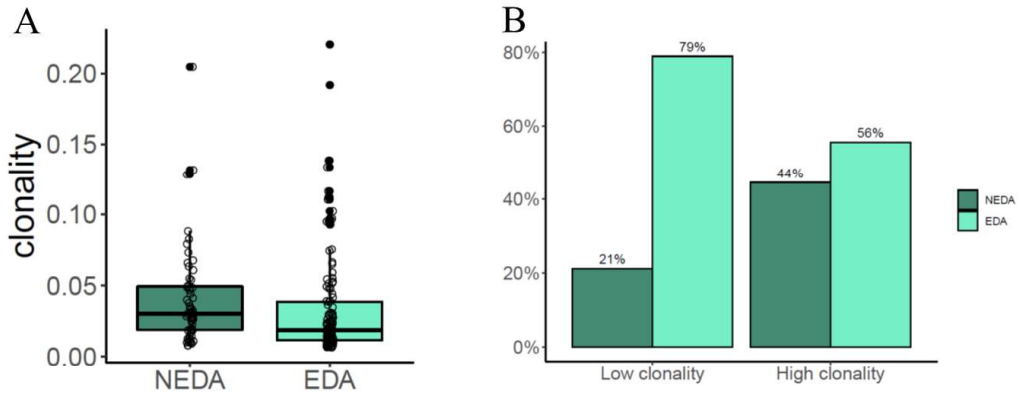


Figure 3.23: Correlation of Simpson clonality with NEDA-3 outcome. *A) Patients with NEDA have a higher SC compared to patients with disease reactivation during follow-up. B) When dividing patients according to the median value of SC, subjects with "Low clonality" have a higher risk of disease activity during follow-up compared to "High clonality" subjects.*

3.4.2 TFR

Similarly, when exploring the TFR endpoint we first tested whether patients with and without at least one clinical relapse during follow-up differed in terms of SC; no statistically significant differences were identified (SC value 0.031 vs 0.037 in patients with and without relapses, p-value 0.39).

Even if SC was not predictive of the occurrence and/or time to relapse, we did find a trend toward a shorter TFR in "Low clonality" patients (HR 0.76, p-value 0.27 in the univariable analysis)(Figure 3.24).

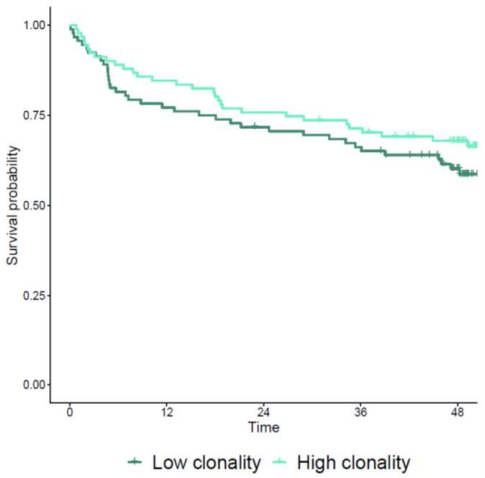


Figure 3.24: Trend towards an earlier TFR in low clonality. *Patients in the "Low clonality" group showed a non-significant trend towards an earlier time to first relapse compared to patients in the "High clonality" subset.*

3.4.3 HLADRB1*1501 association with Simpson clonality

Finally, given the known association of the HLADRB1*1501 with MS susceptibility (Patsopoulos *et al*, 2013) and disease activity (Barcellos *et al*, 2003), we tested its relationship with SC. We took advantage of the availability of genetic data for the CC and extracted genotype information on rs3135388, the tagging SNP for HLADRB1*1501. We then grouped the carriers of the A allele and compared their mean value of SC with that of GG subjects: A carriers showed a slightly lower repertoire diversity compared to non-carriers (SC value 0.04 vs 0.03 in A carriers vs CC individuals, p-value 0.05)(Figure 3.25).

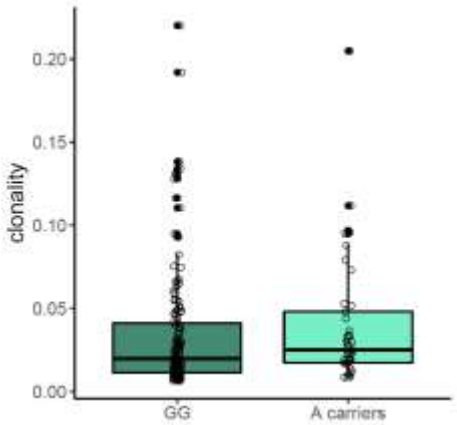


Figure 3.25: Association of Simpson clonality with HLADRB1*1501. *Carriers of the A allele in rs3135388 showed a higher mean value of SC compared to GG subjects*.

When including the HLADRB1*1501 in the prognostic model, the association between SC and NEDA-3 status was still significant (OR 0.42, p-value 0.027).

3.4.4 Considerations

In this study section we explored the relationship between immune diversity and MS inflammatory disease activity. TCR repertoire diversity indexes have been shown to correlate with MS status (Hayashi *et al*, 2021) as well as with the effect of MS treatments (Harris *et al*, 2020; Muraro *et al*, 2014; Amoriello *et al*, 2020). Hence, in the present work we focused on SC that is an established and concise measure of TCR repertoire heterogeneity.

As expected, we found that SC correlates with the age at the time of blood sampling; indeed, it is well known that clonality is higher in older individuals as a consequence of

reduced thymic output and immune senescence (Qi *et al*, 2014; Britanova *et al*, 2014). Moreover, we found a correlation with gender, females showing a more diverse repertoire, which was already reported in previous investigations (Britanova *et al*, 2016). Therefore, we took into account these possibly confounding factors and we found that repertoire diversity was associated with the occurrence of breakthrough disease activity at 4-years as measured using the NEDA-3 criterion. Indeed patients with clinical stability during observation retrospectively showed a higher mean value of SC compared to patients with EDA. Moreover, when evaluating the prognostic value of SC we confirmed that patients with lower clonality displayed a significantly higher risk of inflammation during follow-up (79% vs 56%), both in the univariable and multivariable analysis. On the other hand, no statistically significant associations were found between SC and TFR but, qualitatively, there was a trend towards a higher risk of early relapses in patients with more diverse repertoires, as shown in Figure 8.6. Besides, the association was still present when accounting for the presence or absence of the risk allele HLADRB1*1501, which in turn can influence immune repertoire diversity by restricting the spectrum of recognized antigens.

These results seem to suggest that immune repertoire characteristics are in some way related to the risk of disease reactivation and can potentially help in patients stratification. Noteworthy, our results are in line with previous reports showing a more diverse repertoire in MS patients compared to controls (Hayashi *et al*, 2021; Alves Sousa *et al*, 2019). Besides, other works that analyzed the immune repertoire of MS patients treated with different DMTs also showed a shrinkage in TCR diversity during treatment that is consistent with our findings (Jones *et al*, 2013; Ruck *et al*, 2018; Chiarini *et al*, 2015; Warnke *et al*, 2013). Indeed, treatment with highly active drugs that act with different mechanisms of action, such as alemtuzumab, natalizumab and fingolimod have all been demonstrated to induce a reduction in TCR repertoire expansion.

It has been suggested that the reduction in immune repertoire diversity can reflect a decrease in the recognition of CNS auto-antigens and antigen spreading induced by treatment leading to a reduced T-cell expansions both in CSF and peripheral blood.

Interestingly, for both NTZ (Warnke *et al*, 2013) and ALEM (Jones *et al*, 2013; Ruck *et al*, 2018), these changes in TCR repertoire dynamics have also been associated with the

risk of developing adverse events, such as PML during NTZ and secondary autoimmune disorders in ALEM; clonally restricted repertoires can be associated with a reduced immune surveillance increasing the risks of infections, or be linked to hyper-expansion of auto-reactive clones associated with autoimmune manifestations.

In conclusions, all these findings suggest that immunological profiling of RRMS could help in stratifying patients according to their risk of disease recurrence and can also aid in the identification of subjects that are at higher risk of developing side effects from specific drugs.

3.5 Predictive model

In this last part of the study, thanks to a collaboration with the University of Milan, we used machine learning (ML) techniques to integrate the different layers of information collected on the two cohorts, in order to test their contribution in predicting disease activity at 4 years.

The features selection procedure allowed to identify 63 and 66 variants that were selected in at least 3 of 5 hold-outs in the EC and CC respectively, which represented the input for the ML algorithm. Among these, 4 and 13 variants were selected in all the 5 hold-outs in the EC and CC respectively.

We then evaluated the performances obtained applying four classifiers: linear support vector machine (LSVM), decision trees (DTs), random forests (RFs) and Multi-layer perceptron (MLP) classifiers.

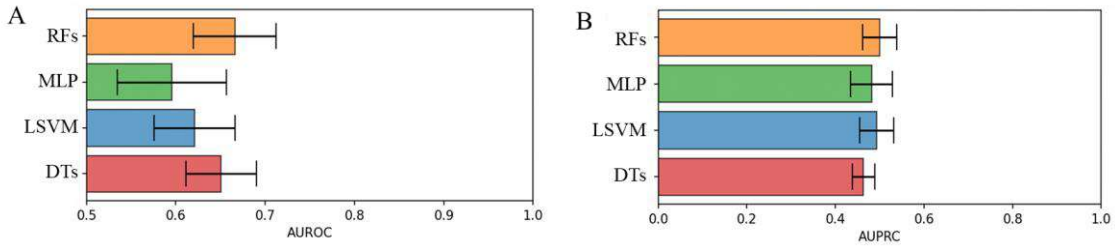


Figure 3.26: AUROC and AUPRC of the models tested on 5 hold-outs in the CC. *AUROC and AUPRC values calculated on 5 hold-outs of the CC using both clinical and genetic data.*

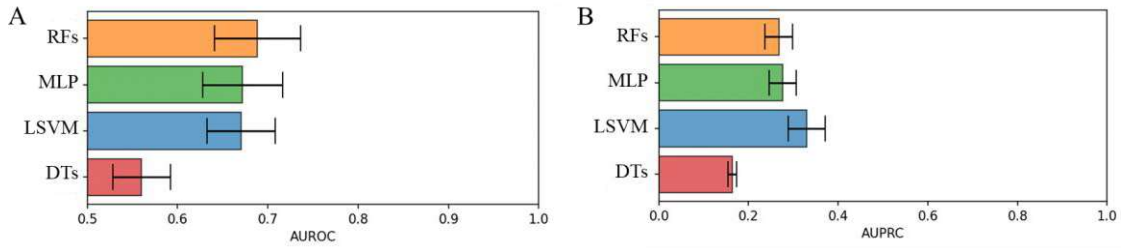


Figure 3.27: AUROC and AUPRC of the models tested on 5 hold-outs in the EC. *AUROC and AUPRC values calculated on 5 hold-outs of the EC using both clinical and molecular data.*

As expected, model predictivity in terms of area under the receiver operating characteristics (AUROC) and area under the precision-recall curve (AUPRC) was better in the larger EC compared to CC (Figure 3.26 and 3.27).

At first glance, RF models trained in clinical and molecular data seem to obtain the best predictivity according to the AUROC but the variance of estimated performance is quite large (Figure 3.26 A and 3.27 A); hence we selected as more predictive the LSVM classifier, based on its results in terms of AUPRC (Figure 3.26 B and 3.27 B).

Table 3.18: Performance of the LSVM classifier applied to clinical, genetic and integrated data

Cohort	Model	Accuracy	AUROC	AUPRC
CC	**Clinical**	0,62±0,019	0,66±0,019	0,55±0,022
	Genetic	0,54±0.034	0,56±0,032	0,49±0,026
	Ensemble WA	0,54±0,034	0,56±0,031	0,48±0,026
	Ensemble PR	0,62±0,093	0,57±0,028	0,46±0,023
	Combined	0,56±0,033	0,65±0,045	0,57±0,036
EC	**Clinical**	0,87±0,001	0,62±0,016	0,27±0,014
	Genetic	0,70±0,031	0,60±0,016	0,22±0,018
	Ensemble WA	0,70±0,031	0,60±0,015	0,22±0,018
	Ensemble PR	0,70±0,031	0,60±0,017	0,23±0,014
	Combined	0,70±0,033	0,61±0,016	0,24±0,019

The table displays predictivity metrics for the LSVM classifier applied on clinical, genetic and integrated data. The performance was evaluated on 5 hold-outs obtained splitting the cohorts with a 80-20% ratio. Ensemble WA: weighted average ensemble; Ensemble PR: perceptron-based ensemble; Combined: model obtained by simply aggregating clinical and molecular data.

Table 3.18 reports the predictive metrics of the LSVM classifier applied to clinical and genetic data as well as their combination, obtained through simple aggregation or ensembles approach.

The best predictive performance is obtained by applying LSVM to the clinical data alone while genetic data alone have a modest prognostic power. Also the integration of clinical with molecular data does not confer a significant increase in the predictive power, irrespective of the method used for data integration (aggregation vs ensembles approach).

However, it is worth noting that when aggregating clinical, immunological and genetic data (combined approach) in the CC dataset, for which we also have immune repertoire data, the overall predictivity is comparable to the clinical-only model but there is an increase in the AUPRC, suggesting a modest benefit from the integration of molecular data.

3.5.1 Considerations

In this last section of the study we tried to combine the different layers of clinical, genetic and immunological information into an integrated model to predict disease activity. To this end, we applied ML methodologies testing different statistical models and diverse ways to combine –omics data.

In our analysis simpler models such as LSVM, that perform linear classification, achieved the best predictive performance and the highest accuracy was obtained considering clinical parameters alone. Overall, the addition of molecular features seemed not to provide a benefit on model predictivity; however, in the CC, the aggregation of clinical and molecular data led to a slight increase in the performance evaluated using AUPRC compared to the clinical model, suggesting a possible advantage with the inclusion of genetic and immunological parameters.

Recently, due to the advancements in artificial intelligence (AI) methodologies, several studies have applied ML to the MS field, in particular to MRI data to perform lesion and tissue segmentation (Danelakis *et al*, 2018; Gabr *et al*, 2020) and to aid in the differential diagnosis with other CNS diseases (Eshaghi *et al*, 2015, 2016; Kim *et al*, 2020).

A few studies applied AI to clinical data, mainly derived from electronic health records, in order to predict disease activity and progression; these works evaluated different statistical models reporting predictive performances that are very similar to ours (AUROC ~0.65-0.70)(Ahuja *et al*, 2021; Walsh *et al*, 2022; De Brouwer *et al*, 2022), confirming the importance of taking into account disease history and characteristics

when assessing MS patients. A study that aimed to predict the conversion to SPMS and the accumulation of severe disability at 6 and 10 years from first evaluation (Pinto *et al*, 2020) reported a higher predictive performance, with an AUROC up to 0.86, considering longitudinal clinical data. Similarly, another paper investigating predictors of conversion to SPMS observed that considering longitudinal clinical information provided a significant benefit to the predictive power of the tested models (Seccia *et al*, 2020), with the drawback that these types of data are not always available and not applicable to newly diagnosed patients in a very early phase of the disease.

In addition to these clinical studies, a previous paper applied ML to investigate the correlation between 113 candidate gene variants and disease progression and selected 19 associated SNPs located near genes implicated in cytotoxicity of immune cells, complement activation and neuronal functions (Jackson *et al*, 2020), suggesting that AI can aid in the identification of meaningful biological variables associated with disease pathogenesis. Moreover, another work applied ML methods to genetic data in MS patients to predict treatment response in patients treated with GA (Ross *et al*, 2017): specifically, the authors implemented a multivariable bayesian modeling to forecast drug response and detected a 4 SNPs signature with a modest predictive performance (AUROC 0.66).

At the best of our knowledge, our study is the first one to apply ML techniques to integrate clinical, genetic and TCR repertoire information in MS. Even though our results in terms of predictive performance are comparable to previous studies that relied on clinical parameters alone, in the CC we identified a slight increase in the AUPRC metric when considering also clinical and immune-repertoire data.

Altogether, these findings suggest that molecular data, such as genetic polymorphisms and immune-repertoire metrics, hold the potential to be used together with clinical information in order to prognosticate disease course but currently the identified signatures does not confer significant advantages over clinical information alone and are not accurate enough to be applied in clinical practice. Further studies on larger cohorts, including independent datasets for training and testing of the predictive model, are required to investigate the presence of robust and reproducible signatures of molecular data to be used in the clinical setting. Moreover the inclusion of MRI data, which proved useful in discriminating MS subtypes with different evolution (Eshaghi *et al*,

2021; Wottschel *et al*, 2014), as well as the addition of neurophysiological (Bejarano *et al*, 2011), or other biological metrics such as sNfL (Thebault *et al*, 2022) and other immune repertoire metrics (Ostmeyer *et al*, 2017) could possibly further improve disease activity prediction.

4 Discussion

MS is a complex disorder, with substantial heterogeneity in terms of pathophysiological characteristics, clinical presentation and course and response to treatment,that would greatly benefit from a more patient-tailored management

Nowadays, 14 DMTs are approved for MS treatment and several drugs are being developed (Dargahi *et al*, 2017), which target distinct biological pathways and have different benefit-to-risk profiles. No definite recommendations are available to guide treatment selection and most patients are started on first-line drugs, shifting to more effective compounds when showing signs of disease reactivation. Obviously, this approach carries the risk of lesions accumulation and worsening disability.

Therefore, several attempts have been made to identify a prognostic algorithm to predict individual disease course that would allow to select the appropriate treatment based on the expected disease severity; consequently first-line treatment would be reserved for low-risk patients, thus avoiding possible adverse effects of highly active drugs, while more aggressive therapy would be started earlier in patients with poor prognostic indicators, maximizing treatment effectiveness. Most of the studies focused on clinical parameters which are known to be associated with disease outcomes (Rotstein & Montalban, 2019) and applied AI methodologies to construct a prognostic model of disease progression (Seccia *et al*, 2021). Some also considered other types of information, such as imaging features (Yoo *et al*, 2017; Wottschel *et al*, 2014) and neurophysiological parameters (Bejarano *et al*, 2011), suggesting that the integration of clinical data with other biologically relevant features can provide a significant benefit to disease prediction.

From this perspective, the recent improvements in -omics technologies offer an unique opportunity to evaluate the contribution of molecular biomarkers in modulating disease activity and to shed light into biologically important mechanisms implicated in inflammatory activity and/or neurodegeneration, without the need for “a priori” assumption.

In the present study, we applied a multi-omics approach to investigate clinical, genomic and immunological parameters associated with disease activity and to explore the biological processes underlying disease expression.

First, we explored clinical parameters associated with disease activity, defined using NEDA-3 criterion and TFR. As expected, we found that disease history and demographic variables are significantly correlated to subsequent disease course; indeed female gender, an earlier age at onset and a shorter disease duration were identified as significantly associated with breakthrough disease reactivation.

We then investigated the presence of genetic and transcriptomic markers of inflammatory activity: even though no genome-wide significant variants were identified, some signals with a suggestive level of association were detected by the GWAS meta-analysis. Specifically, when analyzing the NEDA-3 outcome, we found a very interesting association on chromosome 2, near the *SERPINE2* gene, which is implicated in the regulation of the coagulation system and the modulation of vascular permeability. Besides, when considering the TFR outcome, the top signal on chromosome 7 mapped in a possible regulatory region close to the *EGFR* gene which represents a very appealing candidate gene for implication with MS disease activity due to its role in regulating oligodendrocyte differentiation and remyelination. In addition, another interesting signal was found on chromosome 16, in the region coding for *CETP* that has been linked to blood VitD levels. Due to the limited sample size of the included cohorts, mainly the CC, these results need confirmation in additional independent populations, nonetheless they hint to a possible contribution of these processes in modulating disease reactivation.

Moreover, these findings were strengthened by the results observed in the transcriptomic and pathway analyses: among the genes that were found to be differentially expressed in patients with and without disease activity there was *PCDHGB2*, which, in fact, has been hypothesized to play a role in regulating brain vasculature.

Next, when integrating the results of the genetic and transcriptomic studies by means of a pathway analysis, we found that the immune-related pathway “Systemic lupus erythematosus” was the most associated with both clinical outcomes; nonetheless, several other pathway such as “Proteasome”, “Huntington disease”, “Alzheimer disease” and “Necroptosis” resulted significantly associated mainly with the relapse outcome, pointing to the involvement of biological path implicated not only in immune functions but largely in cell homeostasis and death and neurodegeneration. Though

preliminary, these data demonstrate that the integration of different –omics can increase the chance to identify and confirm meaningful biological information.

To further broaden our biological insight into disease activity, we also evaluated the impact of TCR repertoire characteristics and found that a higher immunological diversity measured using SC seems to correlate with MS reactivation during follow-up, mainy when analyzing the NEDA-3 outcomeThese findings are in line with previous reports and support the idea that immune repertoire features can represent a useful prognostic tool not only for predicting disease activity but also to stratify patients according to their risk of adverse events during second-line treatments (Ruck *et al*, 2018; Jones *et al*, 2013; Warnke *et al*, 2013).

Finally, encouraged by these results, we applied ML algorithms to clinical, genetic and immunological data hoping that a multi-layer model could better capture the complex mechanisms underlying MS phenotypic expression, thus performing better compared to single-modalities models. However, we found that the best predictive performance was obtained applying a LSVM model to clinical information alone (AUROC 0.66 and 0.62 in the CC and EC respectively) and that the addition of genetic and immunological data conferred slight improvement in prognosticating MS activity.

We are aware that the present study has some limitations that could have impacted our findings. First of all, the CC dataset that was used for the multi-omics characterization is quite small and could have reduced our statistical power to detect significant associations; indeed, for the genetic analysis we took advantage of a much larger available cohort to overcome this issue that is particularly important for genetic investigations, due to the huge number of considered features. On the other hand, the limited sample size allowed us to perform a highly detailed characterization of the cohort, both from a clinical and molecular point of view, possibly balancing the reduced power. Moreover, we did not include longitudinal data that could have improved the predictive accuracy of our model (De Brouwer *et al*, 2022; Seccia *et al*, 2020).

Finally, we did not find an overlap between the GWAS results obtained analyzing the NEDA-3 and the relapse activity endpoints but, in our opinion, this is not surprising and does not diminish the reliability of our findings. Indeed, even though both outcomes are used to assess disease activity, they evaluate quite different and complementary aspects: the NEDA-3 outcome, that takes into account MRI activity, is more sensitive but does

not discriminate the severity of disease reactivation, so that the occurrence of a unique, small, asymptomatic lesion at the brain MRI scan is classified in the same way as a serious relapse. On the other hand, when considering relapse activity we are indirectly taking into account those – still unknown – factors which cause a lesion to be symptomatic. Due to these substantial differences, we decided to consider both endpoints and we are not surprised by the lack of overlapping signals.

In conclusions, our findings have demonstrated that genetic, transcriptomic and immune repertoire data can help in deciphering biological processes underlying MS pathophysiology and clinical expression and that the integration of different –omics information can increase the chance to find significant results. On the other hand, the predictive power of models integrating the different layers of information is not enough for application in clinical practice.

4.1 Next steps

As mentioned before, the results obtained from the single –omics analyses are preliminary and further studies are needed to confirm them and to better clarify the biological meaning of our findings.

Specifically, we believe that the signals identified by the GWAS meta-analyses on chromosome 2 and 7 are particularly interesting due to the possible link with genes implicated in biological functions relevant for MS pathogenesis and are worth of a replication effort; first, we plan to increase our sample size with the inclusion of additional independent cohorts of MS patients, thanks to the collaboration with other groups, and to test the association of the identified regions with disease activity.

Then, in the event of a positive replication, we will perform some functional studies to better investigate the biological consequences of the identified variants. For the first signal on chromosome 2, if replicated in other datasets, we would test whether the eQTL effect previously reported on *SERPINE2* in blood is confirmed in our population; in fact, even if *SERPINE2* is more likely to act in the brain, we know that a great proportion of eQTL effects are shared among several tissues and, when the cell type of interest is not readily available, considering blood eQTL can ease data interpretation (Aguet *et al*, 2017; Liu *et al*, 2017). On the other hand, for the signal on chromosome 7, close to the *EGFR* gene, we would like to test the role of the identified SNP in

modulating its transcription level. Again, EGFR is supposed to play a role by regulating myelination in the brain and is not easily measurable at this level; thus, we would perform a luciferase reporter assay using cell lines, to evaluate if the variant of interest affects gene expression regulation.

Overall, these functional studies would allow us to better delineate the link between the genetic polymorphisms prioritized in the present study and the molecular processes implicated in MS pathogenesis.

Meanwhile, we will also expand the analysis of TCR sequencing data performing a repertoire architecture analysis that not only consider the number of TCR sequences present in a sample but also assesses their similarity, which reflects the breadth of antigens that can be recognized.

Finally, we already have ongoing collaborations with other European research groups that will allow to collect a larger number of MS patients with clinical and multi-omics information to test the application of ML algorithms in predicting disease outcomes.

5 Patients and Methods

5.1 Clinical analysis

5.1.1 Study population

Patients belonging to both the CC and the EC entered the clinical analysis.

The inclusion and exclusion criteria differed slightly for the two cohorts, due to the different omics to be considered, and are reported below:

5.1.1.1 Inclusion criteria for the CC

For enrollment in the CC, patients had to satisfy the subsequent criteria:

- Age $\geq$ 18 years;
- RRMS diagnosed according to the 2010 revised McDonald criteria (Polman *et al*, 2011) or subsequent revisions (Thompson *et al*, 2018) at the time of sampling;
- No ongoing DMT and planned start of a first-line therapy;
- Willingness to participate in the study and undergo blood sampling after signing a written informed consent.

5.1.1.2 Exclusion criteria for the CC

- Diagnosis of SPMS or PPMS at the time of enrolment;
- Steroid therapy in the month before blood sampling.

5.1.1.3 Inclusion criteria for the EC

To be included in the EC, patients had to satisfy the subsequent criteria:

- Availability of clinical data from the start of a first-line treatment up to 4 years;
- Availability of genetic information, generated in the context of previous studies;
- Diagnosis of RRMS at the time of first-line treatment start.

5.1.1.4 Exclusion criteria for the EC

- Diagnosis of SPMS or PPMS at the time of first-line treatment start.

5.1.2 Clinical data collection

Demographic and clinical data at baseline and during the 4 years of follow-up were collected through the revision of clinical reports and the iMED database in use at the MS center of OSR. Baseline was set at the time of blood sampling for the CC (that for the majority of patients corresponded to the day of drug initiation) and at the time of first-line treatment start for the EC; indeed, only genetic data were generated for the EC, allowing us to include also patients whose blood samples have been collected during treatment, given that treatment does not have an impact on individual genetic information on the contrary to the other omics data explored in the CC.

As for standard clinical practice at our MS center, during treatment patients underwent neurological examinations with EDSS assessment every 3 months and a brain MRI scan on average once a year. Additional evaluations were conducted in case of new clinical symptoms and suspected relapses.

During each visit, information regarding new neurological symptoms and side effects were collected, as well as the results of brain MRI scans in terms of number of new and/or enlarging T2 lesions and Gd+ lesions. A relapse was defined as the occurrence of new symptoms or worsening of pre-existing disturbances that lasted at least 24 hours without fever or signs of infection.

5.1.3 Disease activity outcomes

The following outcomes were evaluated at the end of the 4 years observation period, that take into account both the occurrence of clinical and neuro-radiological disease activity and the presence of disability worsening:

- NEDA-3 (Bevan & Cree, 2014) status, defined by the absence of relapses, new/enlarging T2 lesions and/or Gd+ lesions at brain MRI scans and disability progression. Patients with missing information and no evidence of disease activity or progression during the available follow up, were not included in this analysis since we were not able to correctly classify them;
- TFR (Sormani *et al*, 2013), defined as the time elapsed from baseline and the occurrence of the first clinical reactivation. For this outcome, patients with an observation period shorter than 4 years were censored at the date of the last available evaluation.

5.1.4 Statistical methods

In order to identify the clinical parameters associated with the outcomes of interest we applied a logistic regression model for the binary endpoint NEDA-3 and a Cox proportional hazards regression for the TFR; the analyses were performed in both univariable and multivariable settings. The regression analysis was performed separately in the two cohorts and the results were then meta-analyzed. A p-value threshold of 0.01 was considered strongly significant, in order to select variables that will be used as covariates in the genetic analyses.

Specifically, the following baseline parameters that were available in both the CC and EC were tested for association: gender, AAO, DD and baseline EDSS. Information regarding the number of relapses and of neuro-radiological activity before treatment start were not available for most patients in the EC and were therefore not considered in neither cohorts.

All analyses were carried out within R 3.6.1 statistical environment (www.R-project.org/); The *glm* function was used to fit logistic and linear regressions and the *survival* package (Therneau, 2020; Therneau & Grambsch, 2000) for the Cox model selection. The meta-analysis was performed using the *metafor* package (Viechtbauer, 2010).

5.2 Genomic analysis

5.2.1 Genetic study

5.2.1.1 Study population

For the genetic investigation we considered both the CC and EC, in order to take advantage of the largest available sample size and increase the power to detect statistically significant associations. As for the clinical analysis, a GWAS was performed separately in the two cohorts and the results were meta-analyzed.

5.2.1.2 Disease activity outcomes

Similarly to what done for the clinical analysis, the occurrence of inflammatory disease activity during follow-up was evaluated using the NEDA-3 criterion and TFR, as previously described.

5.2.1.3 Biological samples collection

Blood samples for DNA extraction and genotyping were obtained before the start of a first-line treatment in patients belonging to the CC and all patients signed a written informed consent before undergoing blood withdrawal. Biological samples were stored in the INSPE biobank at OSR and handled according to national legal regulations and applicable laws.

For the EC, genetic data had already been generated in the context of previous projects and no new biological samples were required from the patients.

5.2.1.4 Genotyping and quality controls

For the CC, DNA extraction has been performed using a standard phenol-chloroform protocol or the automated Maxwell® 16 Blood DNA Purification System, Promega; DNA quality has been assessed with the NanoDrop spectrophotometer and agarose gel electrophoresis to test purity and integrity. Genotyping was then performed using the HumanOmniExpress-24 BeadChip kit (Illumina®) and iSCan system.

On the contrary, for the EC genetic data had previously been generated on different Illumina® array platforms: Human-660 Quad array, HumanOmniExpress-12 BeadChip, HumanOmniExpress-24 BeadChip and HumanOmni- 2.5 BeadChip and deposited in the institutional storage server at OSR.

In order to harmonize the genetic data derived from different arrays and maximize the number of SNPs overlapping between the two cohorts and that will be object of the meta-analysis, an imputation step was performed on the Michigan Imputation Server (https://imputationserver.sph.umich.edu)(Das *et al*, 2016) using the Haplotype Reference Consortium panel as a reference; only SNPs imputed with high quality, defined by an $R^2 > 0.6$, were retained.

Standard per SNPs and per sample QC checks were performed on the two datasets using Plink v1.9beta (www.cog-genomics.org/plink/1.9)(Chang *et al*, 2015). Specifically, variants with MAF < 0.01, genotyping rate < 0.97 and p-value for Hardy-Weinberg Equilibrium (HWE) <1 e-4 were excluded, as well as subjects with a call rate < 0.95, showing excess cryptic relatedness, sex mismatch or who were outliers according to a MDS analysis.

5.2.1.5 Statistical analyses

The GWAS on the NEDA-3 endpoint was performed by applying a logistic regression model as implemented in Plink v1.9 beta (Chang *et al*, 2015) (--logistic function) separately in the CC and EC. For each datasets, the first 3 principal components (PCs) were included as covariates, together with the clinical parameters that were found to be strongly associated with the outcome in the clinical study, that is, AAO and DD. A fixed effect meta-analysis was finally performed in Plink v1.9 beta to combine the results obtained in the two cohorts. To note, the occurrence of inflammatory activity during the observation period was considered the event of interest so, when interpreting the results derived from this type of analysis, an OR > 1 identifies a factor associated with a higher risk of evidence of disease activity (EDA).

On the other hand, to test the association with TFR a survival analysis was executed within R3.6.1 statistical environment, using the *survival* package (Therneau, 2020; Therneau & Grambsch, 2000) to perform a cox proportional hazard regression (coxph function); an additive model was used, that tested the change in event probability over time conferred by each extra reference allele. Also in this case the first 3 principal components were included as covariates as well as AAO, DD and gender, that were found to be significantly associated in the clinical study. For the meta-analysis, we used the R package metaphor (Viechtbauer, 2010) (rma function), that integrates the results of different datasets using estimated coefficients and standard errors through an inverse-variance weighting.

For both analyses, we pre-defined the genome-wide significant threshold at P value < $5x10^{-8}$, which is the generally accepted cut-off for GWAS (Jannot *et al*, 2015), and also considered a suggestive threshold of association at P value < $1x10^{-5}$.

Finally, to ease data interpretation in a more systemic view, we performed a gene-based analysis followed by a pathway analysis to detect functional processes that are more likely to be involved in CNS inflammation. Specifically, we first integrated SNP-based results with the online tools VEGAS2 (Mishra & Macgregor, 2015), that combine p-values of single variants into a gene coding region (±20 Kb on each side to include putative regulatory sequences) to obtain a gene-wise p-value. Subsequently we performed a gene set enrichment analysis using the online tool WebGestalt (Wang *et al*,

2017) and testing biological processes included in KEGG repository. Genes were ranked based on the negative logarithm to the base 10 of the gene-base p-value. For each pathway, the software returns the following details: the number of genes included in the pathway; the ES, that indicate the degree to which a gene set is overrepresented at the top or bottom of a ranked list; the number of LE, that is the number of genes contributing most to the ES; the NES, that weights the size of the tested gene set and allows comparisons between different pathways; an enrichment p-value; FDR value, that corrects for the number of gene sets analyzed.

5.2.2 Transcriptomic study

5.2.2.1 Study population

In order to explore from different points of view the genes and biological processes implicated in disease activity, as an additional step we performed a transcriptomic study to evaluate DEGs between patients with and without evidence of inflammatory activity during the observation period.

Opposite to the genetic information that remain constant throughout the life of an individual, gene expression is a dynamic process that is markedly affected not only by genetic traits but also by physiological and environmental factors. In particular, it is well known that drugs and chemicals can possibly induce changes in the transcriptomic profile of tissues and cells. In order to account for these possible confounders, we included in the transcriptomic analysis only patients belonging to the CC that were sampled while untreated, before the start of a first-line drug, to rule out a possible interference of the ongoing treatment; moreover, we also excluded patients treated with steroids in the month before blood withdrawal and those that had taken other DMTs during the 3 months before sampling. Overall, 21 patients out of the 208 were excluded due to previous treatments and the remaining 187 were included in the transcriptional analysis.

5.2.2.2 Disease activity outcomes

In order to perform a differential expression analysis using the DESeq2 tool, as detailed below, a dichotomous outcome is required. Therefore the NEDA-3 outcome was applied as described for the clinical and genetic analysis, while the TFR endpoint

was not amenable to this type of study and was modified into a binary outcome by classifying patients into two groups: patients experiencing at least one clinical relapse during follow-up, irrespective of the time of onset, and clinically stable patients.

5.2.2.3 RNA extraction and alignment

Total RNA from whole blood was extracted using the PAXgene blood miRNA kit (Qiagen®) followed by RNA quantification and quality checks using Qubit assay (ThermoFisher®) and TapeStation (Agilent Technologies®). RNA libraries were prepared using the TruSeq Stranded Total RNA with Ribo-Zero Globin kit (Illumina®) and paired-end sequencing was performed on the Illumina NovaSeq6000 platform available at the "Center for Omics Sciences" at San Raffaele Hospital, Milan. Reads parameters were set as follows: length 100 bp, depth ≥ 45 million cluster per sample). Reads were then aligned to the hg19 genome assembly using the STAR software and alignment quality was tested using the MultiQC tool. Finally, reads assignment was performed with the *featureCounts* R function using the GENCODE hg19 build for annotation (www.gencodegenes.org/human/release_33lift37.html), that also includes pseudogenes and non-coding RNAs.

5.2.2.4 Statistical analysis

Transcripts with low reads counts (<100 reads in the whole dataset) were filtered out and a normalization procedure based on median ratios of transcript counts, to accommodate for different sequencing depth in samples, was performed; besides, samples which resulted to be outliers according to the Cook distance calculation were also discarded.

Before starting the differential gene expression analysis, we explored the presence of technical and clinical variables influencing the transcriptomic signature by means of PCA analysis and hierarchical clustering, in order to identify the parameters to be included in our statistical model. Finally we performed a differential gene expression analysis as implemented in the DESeq2 Bioconductor tool (www.bioconductor.org) (Love *et al*, 2014) that assumes a binomial negative distribution. DEGs with a FDR <5% were considered significant.

5.2.3 Integrated pathway analysis

5.2.3.1 Statistical analysis

In order to integrate the results from the genetic and trasncriptomic studies we performed an over-representation analysis by means of an hypergeometric test as implemented in WebGestalt online tool, using the KEGG pathway database as reference repository. For both the NEDA-3 and TFR outcome we selected those genes that were nominally associated with the outcome ($p<0.05$) in the genetic and/or the transcriptomic analysis and tested their functional enrichment using pathways derived from the KEGG repository.

5.3 Immune analysis

5.3.1 Study population

Similarly to the transcriptomic analysis, the investigation of immune repertoire characteristics can potentially be influenced by previous DMTs. Hence, for this analysis we only considered the 187 patients belonging to the CC that were sampled before the start of a first-line drug and at least a month apart from steroids administration.

5.3.2 Disease activity outcomes

Consistently to what done in previous analyses, patients were classified according to NEDA-3 criterion at 4 years of follow-up and to TFR, as already detailed elsewhere.

5.3.3 TCR sequencing

The ImmunoSeq® from Adaptive Biotechnologies was used for sequencing of the TCR as per manufacturer instructions; specifically, libraries were generated in triplicates starting from 3000 ng of genomic DNA per sample and quality checked using Bioanalyzer 2100 (DNA High Sensitivity or DNA 1000 kits, Agilent Biotechnologies) or TapeStation 4100 (DNA 1000 ScreenTape kit, Agilent Biotechnologies).
Sequencing data were transferred to the ImmunoSeq server and underwent QCs and demultiplexing for normalization. Data were then available for analysis on the Adaptive Analyzer 3.0.

5.3.4 Statistical analyses

TCR sequences data were downloaded from the Adaptive Analyzer 3.0 and statistical analysis were performed using the *immunarch* package in R.

In order to increase the comparability between samples with variable frequencies of TCR sequences, we performed a down-sampling to obtain for every sample an equal number of sequences (equal to the size of the smallest repertoire).

Only productive sequences were considered for further analyses, that is, sequences that are in-frame and does not contain stop-codon and lead to the generation of a functional TCR. Clonotypes were defined as the ensemble of all TCR sequences that produce the same aminoacidic structure. Simpson clonality was then calculated to measure immune diversity and its correlation with clinical outcomes was analyzed through regression models.

5.4 Predictive model

5.4.1 Study population

In this task we considered both the EC and CC separately; in particular, for the EC the combined contribution of clinical and genetic data was evaluated while for the CC also the repertoire diversity index was included among the input features.

5.4.2 Disease activity outcome

ML methods were applied to predict the occurrence of disease activity at 4-years defined by applying the NEDA-3 criterion. Non-binary outcomes are not suitable to this type of analysis so the TFR endpoint was not considered in this section.

5.4.3 Features selection

In consideration of the huge dimensionality of the genetic data and the resulting computational burden, before feeding the genetic features to ML algorithms we performed a features selection procedure. First, we performed a genetic pruning in order to exclude SNPs highly correlated due to LD; to do so PLINKv1.9 (Chang *et al*, 2015) tool was used to exclude variants in LD with a pairwise $r2 > 0.2$ considering a sliding windows of 50 SNPs at a time. By doing so we retained 107,803 SNPs for the EC and

117,397 for the CC respectively. Afterwards, we performed a further univariate selection by calculating 4 different statistics for each SNP to test the correlation with the NEDA-3 outcome: Pearson correlation, Spearman correlation, Chi-Square statistic and the Normalized Mutual Information score. We then selected the K most correlated genetic features, where the optimal number of features K was obtained through a grid search on the EC, evaluating the performance of both a MLP model with a hidden layer of 100 neurons and a LSVM model tested on K values ranging from 2 to 10,000. To obtain more robust results, the predictive performances were estimated on 5 hold-out samples splitting the dataset into a training and test set with a 80-20% ratio.

Finally, in addition to the selected genetic features we considered the clinical parameters (age, gender, DD, baseline EDSS and, for the CC, ARR in the previous year and basal brain MRI information) and, for the CC only, the SC value.

5.4.4 Predictive model construction

We tested different ML models, including simple linear models such as a LSVM as well as DT and RF classifiers and more complex models like a MLP neural network. Specifically, the following parameters were set for each model:

- LSVM: C=1.0, penalty= “L2”, loss= “squared hinge”, dual.
- DT: max depth = 5
- RF: number of tree = 100
- MLP: hidden layer size = 100 neurons, optimizer = “adam”, activation function = “rectified linear units”, regularization term = L2 penalty with alpha=0.0001, batch size = 200, learning rate=0.001, max iterations = 200.

The models were trained separately on the EC and CC and their performance was evaluated on 5 hold-out samples splitting the dataset into training e test sets with a 80-20% ratio. AUROC, AUPRC and accuracy were calculated to evaluate their predictive ability.

We then evaluated the predictive performance of these models applied on clinical data only, on genetic data only or by combining clinical and molecular information. When integrating the different layers of clinical and omics data, two approaches have been used, by simply aggregating the clinical and molecular features (*combined approach*), or by applying *ensemble* methods on models separately trained on clinical and genetic

data. Two different ensemble methods have been considered: a method that used a weighted average to combine the predictions from multiple models (ensemble WA), and a perceptron-based ensemble method (ensemble PR).

To avoid bias in the AUPRC interpretation that requires the minority class to be labelled as 1, the outcome was coded as NEDA = 1 and EDA = 0.

All the analyses were performed with Python, using the *scikit-learn, pandas* and *numpy* libraries.

References

Abbadessa G, Miele G, Di Pietro A, Sparaco M, Palladino R, Armetta I, D'Elia G, Trojsi F, Signoriello E, Lus G, *et al* (2022) Multiple sclerosis and genetic polymorphisms in fibrinogen-mediated hemostatic pathways: a case–control study. *Neurol Sci* 43: 2601

Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, Mohammadi P, Park YS, Parsana P, Segrè A V., *et al* (2017) Genetic effects on gene expression across human tissues. *Nat 2017 5507675* 550: 204–213

Aguirre A, Dupree JL, Mangin JM & Gallo V (2007) A functional role for EGFR signaling in myelination and remyelination. *Nat Neurosci* 10: 990–1002

Ahuja Y, Kim N, Liang L, Cai T, Dahal K, Seyok T, Lin C, Finan S, Liao K, Savovoa G, *et al* (2021) Leveraging electronic health records data to predict multiple sclerosis disease activity. *Ann Clin Transl Neurol* 8: 800

Allanach JR, Farrell JW, Mésidor M & Karimi-Abdolrezaee S (2022) Current status of neuroprotective and neuroregenerative strategies in multiple sclerosis: A systematic review. *Mult Scler* 28: 29–48

Alter M, Halpern L, Kurland LT, Bornstein B, Leibowitz U & Silberstein J (1962) Multiple sclerosis in Israel. Prevalence among immigrants and native inhabitants. *Arch Neurol* 7: 253–63

Álvarez-Sánchez N, Cruz-Chamorro I, López-González A, Utrilla JC, Fernández-Santos JM, Martínez-López A, Lardone PJ, Guerrero JM & Carrillo-Vico A (2015) Melatonin controls experimental autoimmune encephalomyelitis by altering the T effector/regulatory balance. *Brain Behav Immun* 50: 101–114

Alves Sousa A de P, Johnson KR, Ohayon J, Zhu J, Muraro PA & Jacobson S (2019) Comprehensive Analysis of TCR-β Repertoire in Patients with Neurological Immune-mediated Disorders. *Sci Rep* 9

Amoriello R, Greiff V, Aldinucci A, Bonechi E, Carnasciali A, Peruzzi B, Repice AM, Mariottini A, Saccardi R, Mazzanti B, *et al* (2020) The TCR Repertoire Reconstitution in Multiple Sclerosis: Comparing One-Shot and Continuous Immunosuppressive Therapies. *Front Immunol* 11: 559

Arnaout RA, Prak ETL, Schwab N & Rubelt F (2021) The Future of Blood Testing Is the Immunome. *Front Immunol* 12

Arrambide G, Tintore M, Espejo C, Auger C, Castillo M, Río J, Castilló J, Vidal-Jordana A, Galán I, Nos C, *et al* (2018) The value of oligoclonal bands in the multiple sclerosis diagnostic criteria. *Brain* 141: 1075–1084

Ascherio A, Munger KL, White R, Köchert K, Simon KC, Polman CH, Freedman MS, Hartung H-P, Miller DH, Montalbán X, *et al* (2014) Vitamin D as an early predictor of multiple sclerosis activity and progression. *JAMA Neurol* 71: 306–14

Avasarala JR, Cross AH & Trotter JL (2001) Oligoclonal band number as a marker for prognosis in multiple sclerosis. *Arch Neurol* 58: 2044–2045

Bach J-F (2002) The Effect of Infections on Susceptibility to Autoimmune and Allergic Diseases. *N Engl J Med* 347: 911–920

Baranzini SE, Srinivasan R, Khankhanian P, Okuda DT, Nelson SJ, Matthews PM, Hauser SL, Oksenberg JR & Pelletier D (2010) Genetic variation influences glutamate concentrations in brains of patients with multiple sclerosis. *Brain* 133: 2603–2611

Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, Radue EW, Lindberg RLP, Uitdehaag BMG, Johnson MR, *et al* (2009) Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet* 18: 767–778

Barcellos LF, Oksenberg JR, Begovich AB, Martin ER, Schmidt S, Vittinghoff E, Goodin DS, Pelletier D, Lincoln RR, Bucher P, *et al* (2003) HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course. *Am J Hum Genet* 72: 710–716

Barkhof F, Scheltens P, Frequin ST, Nauta JJ, Tas MW, Valk J & Hommes OR (1992) Relapsing-remitting multiple sclerosis: sequential enhanced MR imaging vs clinical findings in determining disease activity. *AJR Am J Roentgenol* 159: 1041–7

Barro C, Benkert P, Disanto G, Tsagkas C, Amann M, Naegelin Y, Leppert D, Gobbi C, Granziera C, Yaldizli Ö, *et al* (2018) Serum neurofilament as a predictor of disease worsening and brain and spinal cord atrophy in multiple sclerosis. *Brain* 141: 2382–2391

Beck RW, Cleary PA, Anderson MM, Keltner JL, Shults WT, Kaufman DI, Buckley EG, Corbett JJ, Kupersmith MJ, Miller NR, *et al* (1992) A Randomized,

Controlled Trial of Corticosteroids in the Treatment of Acute Optic Neuritis. *N Engl J Med* 326: 581–588

Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kemppinen A, Cotsapas C, Shah TS, Spencer C, Booth D, Goris A, *et al* (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* 45: 1353–1360

Beilin O, Karussis DM, Korczyn AD, Gurwitz D, Aronovich R, Hantai D, Grigoriadis N, Mizrachi-Kol R & Chapman J (2005) Increased thrombin inhibition in experimental autoimmune encephalomyelitis. *J Neurosci Res* 79: 351–359

Bejarano B, Bianco M, Gonzalez-Moron D, Sepulcre J, Goñi J, Arcocha J, Soto O, Carro UD, Comi G, Leocani L, *et al* (2011) Computational classifiers for predicting the short-term course of Multiple sclerosis. *BMC Neurol* 11

Berer K, Mues M, Koutrolos M, AlRasbi Z, Boziki M, Johner C, Wekerle H & Krishnamoorthy G (2011) Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. *Nature* 479: 538–541

Bevan CJ & Cree BAC (2014) Disease Activity Free Status. *JAMA Neurol* 71: 269

Bi W, Fritsche LG, Mukherjee B, Kim S & Lee S (2020) A Fast and Accurate Method for Genome-Wide Time-to-Event Data Analysis and Its Application to UK Biobank. *Am J Hum Genet* 107

Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, Elledge SJ, Niebuhr DW, Scher AI, Munger KL, *et al* (2022) Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* 375: 296–301

Briggs FBS, Shao X, Goldstein BA, Oksenberg JR, Barcellos LF & De Jager PL (2011) Genome-wide association study of severity in multiple sclerosis. *Genes Immun* 12: 615–625

Britanova O V., Putintseva E V., Shugay M, Merzlyak EM, Turchaninova MA, Staroverov DB, Bolotin DA, Lukyanov S, Bogdanova EA, Mamedov IZ, *et al* (2014) Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. *J Immunol* 192: 2689–2698

Britanova O V., Shugay M, Merzlyak EM, Staroverov DB, Putintseva E V., Turchaninova MA, Mamedov IZ, Pogorelyy M V., Bolotin DA, Izraelson M, *et al*

(2016) Dynamics of Individual T Cell Repertoires: From Cord Blood to Centenarians. *J Immunol* 196: 5005–5013

Brodsky M, Nazarian S, Orengo-Nania S, Hutton GJ, Buckley EG, Massey EW, Bhatti MT, Greer M, Goodwin J, Wall M, *et al* (2008) Multiple sclerosis risk after optic neuritis: final optic neuritis treatment trial follow-up. *Arch Neurol* 65: 727–732

De Brouwer E, Becker T, Moreau Y, Havrdova EK, Trojano M, Eichau S, Ozakbas S, Onofrj M, Grammond P, Kuhle J, *et al* (2022) Corrigendum to Longitudinal machine learning modeling of MS patient trajectories improves predictions of disability progression: [Computer Methods and Programs in Biomedicine, Volume 208, (September 2021) 106180]. *Comput Methods Programs Biomed* 213: 106479

Brownlee WJ, Altmann DR, Prados F, Miszkiel KA, Eshaghi A, Gandini Wheeler-Kingshott CAM, Barkhof F & Ciccarelli O (2019) Early imaging predictors of long-term outcomes in relapse-onset multiple sclerosis. *Brain* 142: 2276–2287

Cadavid D, Wolansky LJ, Skurnick J, Lincoln J, Cheriyan J, Szczepanowski K, Kamin SS, Pachner AR, Halper J & Cook SD (2009) Efficacy of treatment of MS with IFN -1b or glatiramer acetate by monthly brain MRI in the BECOME study. *Neurology* 72: 1976–1983

Calabrese M, Poretto V, Favaretto A, Alessio S, Bernardi V, Romualdi C, Rinaldi F, Perini P & Gallo P (2012) Cortical lesion load associates with progression of disability in multiple sclerosis. *Brain* 135: 2952–2961

Calabresi PA, Radue E-W, Goodin D, Jeffery D, Rammohan KW, Reder AT, Vollmer T, Agius MA, Kappos L, Stites T, *et al* (2014) Safety and efficacy of fingolimod in patients with relapsing-remitting multiple sclerosis (FREEDOMS II): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Neurol* 13: 545–556

Campbell GR, Ziabreva I, Reeve AK, Krishnan KJ, Reynolds R, Howell O, Lassmann H, Turnbull DM & Mahad DJ (2011) Mitochondrial DNA deletions and neurodegeneration in multiple sclerosis. *Ann Neurol* 69: 481–492

Cannella B & Raine CS (1995) The adhesion molecule and cytokine profile of multiple sclerosis lesions. *Ann Neurol* 37: 424–35

Carton H, Vlietinck R, Debruyne J, De Keyser J, D'Hooghe MB, Loos R, Medaer R, Truyen L, Yee IM & Sadovnick AD (1997) Risks of multiple sclerosis in relatives of patients in Flanders, Belgium. *J Neurol Neurosurg Psychiatry* 62: 329–33

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM & Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4: 7

Charcot JM (1880) Leçons sur les maladies du système nerveux faites à la Salpêtrière. 4th ed. Paris: Delahaye

Chelmicka Szorc E & Arnason BGW (1972) Partial suppression of experimental allergic encephalomyelitis with heparin. *Arch Neurol* 27: 153–158

Chen J, Chia N, Kalari KR, Yao JZ, Novotna M, Soldan MMP, Luckey DH, Marietta E V., Jeraldo PR, Chen X, *et al* (2016) Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls. *Sci Rep* 6

Chiarini M, Sottini A, Bertoli D, Serana F, Caimi L, Rasia S, Capra R & Imberti L (2015) Newly produced T and B lymphocytes and T-cell receptor repertoire diversity are reduced in peripheral blood of fingolimod-treated multiple sclerosis patients. *Mult Scler* 21: 726–734

Claussen MC & Korn T (2012) Immune mechanisms of new therapeutic strategies in MS — Teriflunomide. *Clin Immunol* 142: 49–56

Cohen JA, Barkhof F, Comi G, Hartung H-P, Khatri BO, Montalban X, Pelletier J, Capra R, Gallo P, Izquierdo G, *et al* (2010) Oral Fingolimod or Intramuscular Interferon for Relapsing Multiple Sclerosis. *N Engl J Med* 362: 402–415

Cohen JA & Chun J (2011) Mechanisms of fingolimod's efficacy and adverse effects in multiple sclerosis. *Ann Neurol* 69: 759–77

Cohen JA, Coles AJ, Arnold DL, Confavreux C, Fox EJ, Hartung H-P, Havrdova E, Selmaj KW, Weiner HL, Fisher E, *et al* (2012) Alemtuzumab versus interferon beta 1a as first-line treatment for patients with relapsing-remitting multiple sclerosis: a randomised controlled phase 3 trial. *Lancet (London, England)* 380: 1819–28

Cohen JA, Comi G, Selmaj KW, Bar-Or A, Arnold DL, Steinman L, Hartung HP, Montalban X, Kubala Havrdová E, Cree BAC, *et al* (2019) Safety and efficacy of ozanimod versus interferon beta-1a in relapsing multiple sclerosis (RADIANCE): a multicentre, randomised, 24-month, phase 3 trial. *Lancet Neurol* 18: 1021–1033

Coles AJ, Compston DAS, Selmaj KW, Lake SL, Moran S, Margolin DH, Norris K, Tandon PK & Investigators CT (2008) Alemtuzumab vs. interferon beta-1a in

early multiple sclerosis. *N Engl J Med* 359: 1786–801

Coles AJ, Twyman CL, Arnold DL, Cohen JA, Confavreux C, Fox EJ, Hartung H-P, Havrdova E, Selmaj KW, Weiner HL, *et al* (2012) Alemtuzumab for patients with relapsing multiple sclerosis after disease-modifying therapy: a randomised controlled phase 3 trial. *Lancet (London, England)* 380: 1829–39

Comi G, Kappos L, Selmaj KW, Bar-Or A, Arnold DL, Steinman L, Hartung HP, Montalban X, Kubala Havrdová E, Cree BAC, *et al* (2019) Safety and efficacy of ozanimod versus interferon beta-1a in relapsing multiple sclerosis (SUNBEAM): a multicentre, randomised, minimum 12-month, phase 3 trial. *Lancet Neurol* 18: 1009–1020

Comi G, De Stefano N, Freedman MS, Barkhof F, Polman CH, Uitdehaag BMJ, Casset-Semanaz F, Hennessy B, Moraga MS, Rocak S, *et al* (2012) Comparison of two dosing frequencies of subcutaneous interferon beta-1a in patients with a first clinical demyelinating event suggestive of multiple sclerosis (REFLEX): a phase 3 randomised controlled trial. *Lancet Neurol* 11: 33–41

Confavreux C, O'Connor P, Comi G, Freedman MS, Miller AE, Olsson TP, Wolinsky JS, Bagulho T, Delhay J-L, Dukovic D, *et al* (2014) Oral teriflunomide for patients with relapsing multiple sclerosis (TOWER): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Neurol* 13: 247–56

Confavreux C, Vukusic S & Adeleine P (2003) Early clinical predictors and progression of irreversible disability in multiple sclerosis: an amnesic process. *Brain* 126: 770–82

Cook S, Vermersch P, Comi G, Giovannoni G, Rammohan K, Rieckmann P, Sørensen PS, Hamlett A, Miret M, Weiner J, *et al* (2011) Safety and tolerability of cladribine tablets in multiple sclerosis: the CLARITY (CLAdRIbine Tablets treating multiple sclerosis orallY) study. *Mult Scler J* 17: 578–593

Danelakis A, Theoharis T & Verganelakis DA (2018) Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. *Comput Med Imaging Graph* 70: 83–100

Dargahi N, Katsara M, Tselios T, Androutsou M-E, de Courten M, Matsoukas J & Apostolopoulos V (2017) Multiple Sclerosis: Immunopathology and Treatment Update. *Brain Sci* 7: 78

Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, *et al* (2016) Next-generation genotype imputation service and methods. *Nat Genet* 48: 1284–1287

Dean G & Kurtzke JF (1971) On the risk of multiple sclerosis according to age at immigration to South Africa. *Br Med J* 3: 725–9

Degenhardt A, Ramagopalan S V., Scalfari A & Ebers GC (2009) Clinical prognostic factors in multiple sclerosis: a natural history review. *Nat Rev Neurol* 5: 672–682

Detels R, Visscher BR, Malmgren RM, Coulson AH, Lucia M V & Dudley JP (1977) Evidence for lower susceptibility to multiple sclerosis in Japanese-Americans. *Am J Epidemiol* 105: 303–10

Devarajan A, Bourquard N, Hama S, Navab M, Grijalva VR, Morvardi S, Clarke CF, Vergnes L, Reue K, Teiber JF, *et al* (2011) Paraoxonase 2 deficiency alters mitochondrial function and exacerbates the development of atherosclerosis. *Antioxid Redox Signal* 14: 341–351

Devlin B & Roeder K (1999) Genomic control for association studies. *Biometrics* 55: 997–1004

Dhib-Jalbut S (2002) Mechanisms of action of interferons and glatiramer acetate in multiple sclerosis. *Neurology* 58: S3-9

Dilokthornsakul P, Valuck RJ, Nair K V., Corboy JR, Allen RR & Campbell JD (2016) Multiple sclerosis prevalence in the United States commercially insured population. *Neurology* 86: 1014–1021

Ding Z, Dai C, Zhong L, Liu R, Gao W, Zhang H & Yin Z (2021) Neuregulin-1 converts reactive astrocytes toward oligodendrocyte lineage cells via upregulating the PI3K-AKT-mTOR pathway to repair spinal cord injury. *Biomed Pharmacother* 134: 111168

Ebers GC, Sadovnick AD & Risch NJ (1995) A genetic basis for familial aggregation in multiple sclerosis. *Nature* 377: 150–151

Elian M, Nightingale S & Dean G (1990) Multiple sclerosis among United Kingdom-born children of immigrants from the Indian subcontinent, Africa and the West Indies. *J Neurol Neurosurg Psychiatry* 53: 906–11

Eriksson M, Andersen O & Runmarker B (2003) Long-term follow up of patients with clinically isolated syndromes, relapsing-remitting and secondary progressive

multiple sclerosis. *Mult Scler* 9: 260–74

Erlich PM, Lunetta KL, Cupples LA, Huyck M, Green RC, Baldwin CT, Farrer LA, Auerbach S, Akomolafe A, Griffith P, *et al* (2006) Polymorphisms in the PON gene cluster are associated with Alzheimer disease. *Hum Mol Genet* 15: 77–85

Eshaghi A, Riyahi-Alam S, Saeedi R, Roostaei T, Nazeri A, Aghsaei A, Doosti R, Ganjgahi H, Bodini B, Shakourirad A, *et al* (2015) Classification algorithms with multi-modal data fusion could accurately distinguish neuromyelitis optica from multiple sclerosis. *NeuroImage Clin* 7: 306–314

Eshaghi A, Wottschel V, Cortese R, Calabrese M, Sahraian MA, Thompson AJ, Alexander DC & Ciccarelli O (2016) Gray matter MRI differentiates neuromyelitis optica from multiple sclerosis using random forest. *Neurology* 87: 2463–2470

Eshaghi A, Young AL, Wijeratne PA, Prados F, Arnold DL, Narayanan S, Guttmann CRG, Barkhof F, Alexander DC, Thompson AJ, *et al* (2021) Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat Commun* 12

Esposito F, Sorosina M, Ottoboni L, Lim ET, Replogle JM, Raj T, Brambilla P, Liberatore G, Guaschino C, Romeo M, *et al* (2015) A pharmacogenetic study implicates SLC9a9 in multiple sclerosis disease activity. *Ann Neurol* 78: 115–27

Farez MF, Fiol MP, Gaitán MI, Quintana FJ & Correale J (2015a) Sodium intake is associated with increased disease activity in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 86: 26–31

Farez MF, Mascanfroni ID, Méndez-Huergo SP, Yeste A, Murugaiyan G, Garo LP, Balbuena Aguirre ME, Patel B, Ysrraelit MC, Zhu C, *et al* (2015b) Melatonin Contributes to the Seasonality of Multiple Sclerosis Relapses. *Cell* 162: 1338–1352

Feigin VL, Abajobir AA, Abate KH, Abd-Allah F, Abdulle AM, Abera SF, Abyu GY, Ahmed MB, Aichour AN, Aichour I, *et al* (2017) Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Neurol* 16: 877–897

Filippi M, Horsfield MA, Morrissey SP, MacManus DG, Rudge P, McDonald WI & Miller DH (1994) Quantitative brain MRI lesion load predicts the course of clinically isolated syndromes suggestive of multiple sclerosis. *Neurology* 44: 635–41

Filippi M, Horsfield MA, Tofts PS, Barkhof F, Thompson AJ & Miller DH (1995) Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis. *Brain* 118 (Pt 6): 1601–12

Filippi M & Rocca MA (2011) MR Imaging of Multiple Sclerosis. *Radiology* 259: 659–681

Fischer MT, Sharma R, Lim JL, Haider L, Frischer JM, Drexhage J, Mahad D, Bradl M, van Horssen J & Lassmann H (2012) NADPH oxidase expression in active multiple sclerosis lesions in relation to oxidative tissue damage and mitochondrial injury. *Brain* 135: 886–99

Fisher E, Rudick RA, Cutter G, Baier M, Miller D, Weinstock-Guttman B, Mass MK, Dougherty DS & Simonian NA (2000) Relationship between brain atrophy and disability: an 8-year follow-up study of multiple sclerosis patients. *Mult Scler* 6: 373–7

Fisniku LK, Brex PA, Altmann DR, Miszkiel KA, Benton CE, Lanyon R, Thompson AJ & Miller DH (2008) Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain* 131: 808–817

Fitzgerald KC, Munger KL, Köchert K, Arnason BGW, Comi G, Cook S, Goodin DS, Filippi M, Hartung H-P, Jeffery DR, *et al* (2015) Association of Vitamin D Levels With Multiple Sclerosis Activity and Progression in Patients Receiving Interferon Beta-1b. *JAMA Neurol* 72: 1458

Fogdell-Hahn A, Ligers A, Grønning M, Hillert J & Olerup O (2000) Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. *Tissue Antigens* 55: 140–8

Fox RJ, Miller DH, Phillips JT, Hutchinson M, Havrdova E, Kita M, Yang M, Raghupathi K, Novas M, Sweetser MT, *et al* (2012) Placebo-Controlled Phase 3 Study of Oral BG-12 or Glatiramer in Multiple Sclerosis. *N Engl J Med* 367: 1087–1097

Frohman EM, Racke MK & Raine CS (2006) Multiple sclerosis--the plaque and its pathogenesis. *N Engl J Med* 354: 942–55

Fromont A, Debouverie M, Le Teuff G, Quantin C, Binquet C & Moreau T (2008) Clinical Parameters to Predict Response to Interferon in Relapsing Multiple Sclerosis. *Neuroepidemiology* 31: 150–156

Gabbert L, Dilling C, Meybohm P & Burek M (2020) Deletion of Protocadherin Gamma C3 Induces Phenotypic and Functional Changes in Brain Microvascular Endothelial Cells In Vitro. *Front Pharmacol* 11

Gabr RE, Coronado I, Robinson M, Sujit SJ, Datta S, Sun X, Allen WJ, Lublin FD, Wolinsky JS & Narayana PA (2020) Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: A large-scale study. *Mult Scler* 26: 1217–1226

Gale CR & Martyn CN Migrant studies in multiple sclerosis. *Prog Neurobiol* 47: 425–48

Gedvilaite G, Vilkeviciute A, Kriauciuniene L, Asmoniene V & Liutkeviciene R (2019) Does CETP rs5882, rs708272, SIRT1 rs12778366, FGFR2 rs2981582, STAT3 rs744166, VEGFA rs833068, IL6 rs1800795 polymorphisms play a role in optic neuritis development? *Ophthalmic Genet* 40: 219–226

Geraldes R, Ciccarelli O, Barkhof F, De Stefano N, Enzinger C, Filippi M, Hofer M, Paul F, Preziosa P, Rovira A, *et al* (2018) The current role of MRI in differentiating multiple sclerosis from its imaging mimics. *Nat Rev Neurol* 14: 213–213

Gianfrancesco MA, Acuna B, Shen L, Briggs FBS, Quach H, Bellesis KH, Bernstein A, Hedstrom AK, Kockum I, Alfredsson L, *et al* (2014) Obesity during childhood and adolescence increases susceptibility to multiple sclerosis after accounting for established genetic and environmental risk factors. *Obes Res Clin Pract* 8: e435-47

Gilmore CP, Bo L, Owens T, Lowe J, Esiri MM & Evangelou N (2006) Spinal Cord Gray Matter Demyelination in Multiple Sclerosis—A Novel Pattern of Residual Plaque Morphology. *Brain Pathol* 16: 202–208

Giovannoni G, Comi G, Cook S, Rammohan K, Rieckmann P, Sørensen PS, Vermersch P, Chang P, Hamlett A, Musch B, *et al* (2010) A Placebo-Controlled Trial of Oral Cladribine for Relapsing Multiple Sclerosis. *N Engl J Med* 362: 416–426

Giovannoni G, Soelberg Sorensen P, Cook S, Rammohan K, Rieckmann P, Comi G, Dangond F, Adeniji AK & Vermersch P (2017) Safety and efficacy of cladribine tablets in patients with relapsing–remitting multiple sclerosis: Results from the randomized extension trial of the CLARITY study. *Mult Scler J*: 135245851772760

Giovannoni G, Southam E & Waubant E (2012) Systematic review of disease-modifying therapies to assess unmet needs in multiple sclerosis: tolerability and adherence. *Mult Scler* 18: 932–46

Göbel K, Kraft P, Pankratz S, Gross CC, Korsukewitz C, Kwiecien R, Mesters R, Kehrel BE, Wiendl H, Kleinschnitz C, *et al* (2016) Prothrombin and factor X are elevated in multiple sclerosis patients. *Ann Neurol* 80: 946–951

Gold R, Kappos L, Arnold DL, Bar-Or A, Giovannoni G, Selmaj K, Tornatore C, Sweetser MT, Yang M, Sheikh SI, *et al* (2012) Placebo-Controlled Phase 3 Study of Oral BG-12 for Relapsing Multiple Sclerosis. *N Engl J Med* 367: 1098–1107

Goodin DS, Reder AT, Traboulsee AL, Li DK, Langdon D, Cutter G, Cook S, O'Donnell T, Kremenchutzky M, Oger J, *et al* (2018) Predictive validity of NEDA in the 16- and 21-year follow-up from the pivotal trial of interferon beta-1b. *Mult Scler J*: 135245851877351

Goronzy JJ, Qi Q, Olshen RA & Weyand CM (2015) High-throughput sequencing insights into T-cell receptor repertoire diversity in aging. *Genome Med* 7

Graetz C, Gröger A, Luessi F, Salmen A, Zöller D, Schultz J, Siller N, Fleischer V, Bellenberg B, Berthele A, *et al* (2019) Association of smoking but not HLA-DRB1*15:01, APOE or body mass index with brain atrophy in early multiple sclerosis. *Mult Scler* 25: 661–668

Gregory SG, Schmidt S, Seth P, Oksenberg JR, Hart J, Prokop A, Caillier SJ, Ban M, Goris A, Barcellos LF, *et al* (2007) Interleukin 7 receptor α chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat Genet* 39: 1083–1091

Guaschino C, Esposito F, Liberatore G, Colombo B, Annovazzi P, D'Amico E, Cavalla P, Capello E, Capra R, Galimberti D, *et al* (2014) Familial clustering in Italian progressive-onset and bout-onset multiple sclerosis. *Neurol Sci* 35: 789–791

Haahr S, Plesner AM, Vestergaard BF & Höllsberg P (2004) A role of late Epstein-Barr virus infection in multiple sclerosis. *Acta Neurol Scand* 109: 270–5

Hafler D, Compston A, Sawcer S, Lander E, Daly M, De Jager P, de Bakker P, Gabriel S, Mirel D, Ivinson A, *et al* (2007) Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 357: 851–862

Haider L, Fischer MT, Frischer JM, Bauer J, Höftberger R, Botond G, Esterbauer H, Binder CJ, Witztum JL & Lassmann H (2011) Oxidative damage in multiple

sclerosis lesions. *Brain* 134: 1914–1924

Handel AE, Williamson AJ, Disanto G, Dobson R, Giovannoni G & Ramagopalan S V (2011) Smoking and multiple sclerosis: an updated meta-analysis. *PLoS One* 6: e16149

Hansen T, Skytthe A, Stenager E, Petersen HC, Brønnum-Hansen H & Kyvik KO (2005) Concordance for multiple sclerosis in Danish twins: an update of a nationwide study. *Mult Scler* 11: 504–10

Harris KM, Lim N, Lindau P, Robins H, Griffith LM, Nash RA, Turka LA & Muraro PA (2020) Extensive intrathecal T cell renewal following hematopoietic transplantation for multiple sclerosis. *JCI Insight* 5

Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung H-P, Hemmer B, Lublin F, Montalban X, Rammohan KW, Selmaj K, *et al* (2017) Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. *N Engl J Med* 376: 221–234

Hauser SL, Fleischnick E, Weiner HL, Marcus D, Awdeh Z, Yunis EJ & Alper CA (1989) Extended major histocompatibility complex haplotypes in patients with multiple sclerosis. *Neurology* 39: 275–7

Hauser SL, Oksenberg JR, Lincoln R, Garovoy J, Beck RW, Cole SR, Moke PS, Kip KE, Gal RL & Long DT (2000) Interaction between HLA-DR2 and abnormal brain MRI in optic neuritis and early MS. *Neurology* 54: 1859–1861

Havrdova E, Horakova D & Kovarova I (2015) Alemtuzumab in the treatment of multiple sclerosis: key clinical trial results and considerations for use. *Ther Adv Neurol Disord* 8: 31–45

Hayashi F, Isobe N, Glanville J, Matsushita T, Maimaitijiang G, Fukumoto S, Watanabe M, Masaki K & Kira J ichi (2021) A new clustering method identifies multiple sclerosis-specific T-cell receptors. *Ann Clin Transl Neurol* 8: 163

Hayes CE, Hubler SL, Moore JR, Barta LE, Praska CE & Nashold FE (2015) Vitamin D Actions on CD4(+) T Cells in Autoimmune Disease. *Front Immunol* 6: 100

He H, Hu Z, Xiao H, Zhou F & Yang B (2018) The tale of histone modifications and its role in multiple sclerosis. *Hum Genomics* 12

Healy BC, Ali EN, Guttmann CRG, Chitnis T, Glanz BI, Buckle G, Houtchens M, Stazzone L, Moodie J, Berger AM, *et al* (2009) Smoking and Disease Progression in Multiple Sclerosis. *Arch Neurol* 66: 858–64

Hedström AK, Olsson T & Alfredsson L (2012) High body mass index before age 20 is associated with increased risk for multiple sclerosis in both men and women. *Mult Scler J* 18: 1334–1336

Henderson APD, Barnett MH, Parratt JDE & Prineas JW (2009) Multiple sclerosis: Distribution of inflammatory cells in newly forming lesions. *Ann Neurol* 66: 739–753

Hernán MA, Jick SS, Logroscino G, Olek MJ, Ascherio A & Jick H (2005) Cigarette smoking and the progression of multiple sclerosis. *Brain* 128: 1461–5

Hindson J (2017) Multiple sclerosis: A possible link between multiple sclerosis and gut microbiota. *Nat Rev Neurol* 13: 705–705

Hofman FM, Hinton DR, Johnson K & Merrill JE (1989) Tumor necrosis factor identified in multiple sclerosis brain. *J Exp Med* 170: 607–12

Holman DW, Klein RS & Ransohoff RM (2011) The blood–brain barrier, chemokines and multiple sclerosis. *Biochim Biophys Acta - Mol Basis Dis* 1812: 220–230

Houenou LJ, Turner PL, Li L, Oppenheim RW & Festoff BW (1995) A serine protease inhibitor, protease nexin I, rescues motoneurons from naturally occurring and axotomy-induced cell death. *Proc Natl Acad Sci U S A* 92: 895–899

Hu X, Xiao G, He L, Niu X, Li H, Lou T, Hu Q, Yang Y, Xu Q, Wei Z, *et al* (2021) Sustained ErbB Activation Causes Demyelination and Hypomyelination by Driving Necroptosis of Mature Oligodendrocytes and Apoptosis of Oligodendrocyte Precursor Cells. *J Neurosci* 41: 9872

Hughey JJ, Rhoades SD, Fu DY, Bastarache L, Denny JC & Chen Q (2019) Cox regression increases power to detect genotype-phenotype associations in genomic studies using the electronic health record. *BMC Genomics* 20: 1–7

Hupperts R, Smolders J, Vieth R, Holmøy T, Marhardt K, Schluep M, Killestein J, Barkhof F, Beelke M & Grimaldi LME (2019) Randomized trial of daily high-dose vitamin D 3 in patients with RRMS receiving subcutaneous interferon β-1a. *Neurology* 93: E1906–E1916

Inaba Y, Ichikawa M, Koh CS, Inoue A, Itoh M, Kyogashima M & Komiyama A (1999) Suppression of experimental autoimmune encephalomyelitis by dermatan sulfate. *Cell Immunol* 198: 96–102

Isobe N, Keshavan A, Gourraud PA, Zhu AH, Datta E, Schlaeger R, Caillier SJ,

Santaniello A, Lizée A, Himmelstein DS, *et al* (2016) Association of HLA Genetic Risk Burden With Disease Phenotypes in Multiple Sclerosis. *JAMA Neurol* 73: 795

Jackson KC, Sun K, Barbour C, Hernandez D, Kosa P, Tanigawa M, Weideman AM & Bielekova B (2020) Genetic model of MS severity predicts future accumulation of disability. *Ann Hum Genet* 84: 1–10

Jacobs LD, Beck RW, Simon JH, Kinkel RP, Brownscheidle CM, Murray TJ, Simonian NA, Slasor PJ & Sandrock AW (2000) Intramuscular Interferon Beta-1A Therapy Initiated during a First Demyelinating Event in Multiple Sclerosis. *N Engl J Med* 343: 898–904

Jacobs LD, Cookfair DL, Rudick RA, Herndon RM, Richert JR, Salazar AM, Fischer JS, Goodkin DE, Granger C V., Simon JH, *et al* (1996) Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. *Ann Neurol* 39: 285–294

Janeway CAJ, Travers P, Walport M & Shlomchik MJ (2001) Immunobiology,The Immune System in Health and Disease. *Garl Sci*

Jangi S, Gandhi R, Cox LM, Li N, Von Glehn F, Yan R, Patel B, Mazzola MA, Liu S, Glanz BL, *et al* (2016) Alterations of the human gut microbiome in multiple sclerosis. *Nat Commun* 7

Jannot A-S, Ehret G & Perneger T (2015) P < 5 × 10−8 has emerged as a standard of statistical significance for genome-wide association studies. *J Clin Epidemiol* 68: 460–465

Jensen CJ, Stankovich J, Van der Walt A, Bahlo M, Taylor B V., van der Mei IAF, Foote SJ, Kilpatrick TJ, Johnson LJ, Wilkins E, *et al* (2010) Multiple sclerosis susceptibility-associated SNPs do not influence disease severity measures in a cohort of Australian MS patients. *PLoS One* 5

Jersild C, Fog T, Hansen GS, Thomsen M, Svejgaard A & Dupont B (1973) Histocompatibility determinants in multiple sclerosis, with special reference to clinical course. *Lancet (London, England)* 2: 1221–5

Jersild C, Svejgaard A & Fog T (1972) HL-A antigens and multiple sclerosis. *Lancet (London, England)* 1: 1240–1

Jokubaitis VG, Li V, Kalincik T, Izquierdo G, Hodgkinson S, Alroughani R, Lechner-Scott J, Lugaresi A, Duquette P, Girard M, *et al* (2014) Fingolimod after

natalizumab and the risk of short-term relapse. *Neurology* 82: 1204–11

Jones JL, Thompson SAJ, Loh P, Davies JL, Tuohy OC, Curry AJ, Azzopardi L, Hill-Cawthorne G, Fahey MT, Compston A, *et al* (2013) Human autoimmunity after lymphocyte depletion is caused by homeostatic T-cell proliferation. *Proc Natl Acad Sci U S A* 110: 20200–20205

Jordan KR, Parra-Izquierdo I, Gruber A, Shatzel JJ, Pham P, Sherman LS, McCarty OJT & Verbout NG (2021) Thrombin generation and activity in multiple sclerosis. *Metab Brain Dis* 36: 407–420

Kalincik T, Vivek V, Jokubaitis V, Lechner-Scott J, Trojano M, Izquierdo G, Lugaresi A, Grand'Maison F, Hupperts R, Oreja-Guevara C, *et al* (2013) Sex as a determinant of relapse incidence and progressive course of multiple sclerosis. *Brain* 136: 3609–3617

Kappos L, Bar-Or A, Cree BAC, Fox RJ, Giovannoni G, Gold R, Vermersch P, Arnold DL, Arnould S, Scherz T, *et al* (2018) Siponimod versus placebo in secondary progressive multiple sclerosis (EXPAND): a double-blind, randomised, phase 3 study. *Lancet (London, England)* 391: 1263–1273

Kappos L & European Study Group on interferon beta-1b in secondary progressive MS . (1998) Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. *Lancet (London, England)* 352: 1491–7

Kappos L, O'Connor P, Radue E-W, Polman C, Hohlfeld R, Selmaj K, Ritter S, Schlosshauer R, von Rosenstiel P, Zhang-Auberson L, *et al* (2015) Long-term effects of fingolimod in multiple sclerosis: the randomized FREEDOMS extension trial. *Neurology* 84: 1582–91

Kappos L, Polman C, Pozzilli C, Thompson A, Beckmann K, Dahlke F & European Study Group in Interferon beta-1b in Secondary-Progressive MS (2001) Final analysis of the European multicenter trial on IFNbeta-1b in secondary-progressive MS. *Neurology* 57: 1969–75

Kappos L, Radue E-W, O'Connor P, Polman C, Hohlfeld R, Calabresi P, Selmaj K, Agoropoulou C, Leyk M, Zhang-Auberson L, *et al* (2010) A Placebo-Controlled Trial of Oral Fingolimod in Relapsing Multiple Sclerosis. *N Engl J Med* 362: 387–401

Kataria H, Hart CG, Alizadeh A, Cossoy M, Kaushik DK, Bernstein CN, Marrie RA, Yong VW & Karimi-Abdolrezaee S (2021) Neuregulin-1 beta 1 is implicated in pathogenesis of multiple sclerosis. *Brain* 144: 162–185

Kidd D, Barkhof F, McConnell R, Algra PR, Allen I V & Revesz T (1999) Cortical lesions in multiple sclerosis. *Brain* 122 (Pt 1): 17–26

Kim H, Lee Y, Kim YH, Lim YM, Lee JS, Woo J, Jang SK, Oh YJ, Kim HW, Lee EJ, *et al* (2020) Deep Learning-Based Method to Differentiate Neuromyelitis Optica Spectrum Disorder From Multiple Sclerosis. *Front Neurol* 11

Klaver R, Popescu V, Voorn P, Galis-de Graaf Y, van der Valk P, de Vries HE, Schenk GJ & Geurts JJG (2015) Neuronal and Axonal Loss in Normal-Appearing Gray Matter and Subpial Lesions in Multiple Sclerosis. *J Neuropathol Exp Neurol* 74: 453–458

Klistorner A, Chai Y, Leocani L, Albrecht P, Aktas O, Butzkueven H, Ziemssen T, Ziemssen F, Frederiksen J, Xu L, *et al* (2018) Assessment of Opicinumab in Acute Optic Neuritis Using Multifocal Visual Evoked Potential. *CNS Drugs* 32: 1159–1171

Koch-Henriksen N & Sørensen PS (2010) The changing demographic pattern of multiple sclerosis epidemiology. *Lancet Neurol* 9: 520–532

Kuhle J, Barro C, Disanto G, Mathias A, Soneson C, Bonnier G, Yaldizli Ö, Regeniter A, Derfuss T, Canales M, *et al* (2016) Serum neurofilament light chain in early relapsing remitting MS is increased and correlates with CSF levels and with MRI measures of disease severity. *Mult Scler* 22: 1550–1559

Kurtzke JF, Beebe GW & Norman JE (1979) Epidemiology of multiple sclerosis in U.S. veterans: 1. Race, sex, and geographic distribution. *Neurology* 29: 1228–35

Kurtzke JF, Beebe GW & Norman JE (1985) Epidemiology of multiple sclerosis in US veterans: III. Migration and the risk of MS. *Neurology* 35: 672–8

Kutzelnigg A, Faber-Rod JC, Bauer J, Lucchinetti CF, Sorensen PS, Laursen H, Stadelmann C, Brück W, Rauschka H, Schmidbauer M, *et al* (2007) Widespread Demyelination in the Cerebellar Cortex in Multiple Sclerosis. *Brain Pathol* 17: 38–44

Kutzelnigg A & Lassmann H (2014) Pathology of multiple sclerosis and related inflammatory demyelinating diseases. In *Handbook of clinical neurology* pp 15–

58.

Lassmann H (2018) Multiple Sclerosis Pathology. *Cold Spring Harb Perspect Med* 8: a028936

Lassmann H, Brück W & Lucchinetti CF (2007) The immunopathology of multiple sclerosis: an overview. *Brain Pathol* 17: 210–8

Lassmann H, van Horssen J & Mahad D (2012) Progressive multiple sclerosis: pathology and pathogenesis. *Nat Rev Neurol* 8: 647–656

Lassmann H, Niedobitek G, Aloisi F, Middeldorp JM & NeuroproMiSe EBV Working Group (2011) Epstein-Barr virus in the multiple sclerosis brain: a controversial issue--report on a focused workshop held in the Centre for Brain Research of the Medical University of Vienna, Austria. *Brain* 134: 2772–86

Lavorgna L, Bonavita S, Ippolito D, Lanzillo R, Salemi G, Patti F, Valentino P, Coniglio G, Buccafusca M, Paolicelli D, *et al* (2014) Clinical and magnetic resonance imaging predictors of disease progression in multiple sclerosis: a nine-year follow-up study. *Mult Scler* 20: 220–226

Lebrun-Frénay C, Rollot F, Mondot L, Zephir H, Louapre C, Le Page E, Durand-Dubief F, Labauge P, Bensa C, Thouvenot E, *et al* (2021) Risk Factors and Time to Clinical Symptoms of Multiple Sclerosis Among Patients With Radiologically Isolated Syndrome. *JAMA Netw Open* 4

Lee X, Shao Z, Sheng G, Pepinsky B & Mi S (2014) LINGO-1 regulates oligodendrocyte differentiation by inhibiting ErbB2 translocation and activation in lipid rafts. *Mol Cell Neurosci* 60: 36–42

Lehmann-Horn K, Penkert H, Grein P, Leppmeier U, Teuber-Hanselmann S, Hemmer B & Berthele A (2016) PML during dimethyl fumarate treatment of multiple sclerosis: How does lymphopenia matter? *Neurology* 87: 440–1

Leist TP, Comi G, Cree BAC, Coyle PK, Freedman MS, Hartung H-P, Vermersch P, Casset-Semanaz F, Scaramozza M & oral cladribine for early MS (ORACLE MS) Study Group (2014) Effect of oral cladribine on time to conversion to clinically definite multiple sclerosis in patients with a first demyelinating event (ORACLE MS): a phase 3 randomised trial. *Lancet Neurol* 13: 257–267

Leocani L & Comi G (2014) Clinical neurophysiology of multiple sclerosis. In *Handbook of clinical neurology* pp 671–679.

Li DK & Paty DW (1999) Magnetic resonance imaging results of the PRISMS trial: a randomized, double-blind, placebo-controlled study of interferon-beta1a in relapsing-remitting multiple sclerosis. Prevention of Relapses and Disability by Interferon-beta1a Subcutaneously in Multiple Sclerosis. *Ann Neurol* 46: 197–206

Lill CM (2014) Recent advances and future challenges in the genetics of multiple sclerosis. *Front Neurol* 5: 130

Link H & Huang Y-M (2006) Oligoclonal bands in multiple sclerosis cerebrospinal fluid: an update on methodology and clinical usefulness. *J Neuroimmunol* 180: 17–28

Linker RA, Lee D-H, Ryan S, van Dam AM, Conrad R, Bista P, Zeng W, Hronowsky X, Buko A, Chollate S, *et al* (2011) Fumaric acid esters exert neuroprotective effects in neuroinflammation via activation of the Nrf2 antioxidant pathway. *Brain* 134: 678–92

Liu C & Blumhardt LD (2000) Disability outcome measures in therapeutic trials of relapsing-remitting multiple sclerosis: effects of heterogeneity of disease course in placebo cohorts. *J Neurol Neurosurg Psychiatry* 68: 450–7

Liu X, Finucane HK, Gusev A, Bhatia G, Gazal S, O'Connor L, Bulik-Sullivan B, Wright FA, Sullivan PF, Neale BM, *et al* (2017) Functional Architectures of Local and Distal Regulation of Gene Expression in Multiple Human Tissues. *Am J Hum Genet* 100: 605–616

Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, *et al* (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45: 580–585

Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550

Lublin FD & Reingold SC (1996) Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology* 46: 907–11

Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sorensen PS, Thompson AJ, Wolinsky JS, Balcer LJ, Banwell B, Barkhof F, *et al* (2014) Defining the clinical course of multiple sclerosis: The 2013 revisions. *Neurology* 83: 278–286

Lucchinetti C, Brück W, Parisi J, Scheithauer B, Rodriguez M & Lassmann H (2000) Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Ann Neurol* 47: 707–17

Lundmark F, Duvefelt K, Iacobaeus E, Kockum I, Wallström E, Khademi M, Oturai A, Ryder LP, Saarela J, Harbo HF, *et al* (2007) Variation in interleukin 7 receptor α chain (IL7R) influences risk of multiple sclerosis. *Nat Genet* 39: 1108–1113

Magliozzi R & Cross AH (2020) Can CSF biomarkers predict future MS disease activity and severity? *Mult Scler* 26: 582–590

Magliozzi R, Howell O, Vora A, Serafini B, Nicholas R, Puopolo M, Reynolds R & Aloisi F (2006) Meningeal B-cell follicles in secondary progressive multiple sclerosis associate with early onset of disease and severe cortical pathology. *Brain* 130: 1089–1104

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, *et al* (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747–53

Martinelli Boneschi F, Boneschi FM, Rovaris M, Comi G & Filippi M (2004) The use of magnetic resonance imaging in multiple sclerosis: lessons learned from clinical trials. *Mult Scler* 10: 341–7

Martinelli V, Cocco E, Capra R, Salemi G, Gallo P, Capobianco M, Pesci I, Ghezzi A, Pozzilli C, Lugaresi A, *et al* (2011) Acute myeloid leukemia in Italian patients with multiple sclerosis treated with mitoxantrone. *Neurology* 77: 1887–1895

Martinelli V, Rocca MA, Annovazzi P, Pulizzi A, Rodegher M, Martinelli Boneschi F, Scotti R, Falini A, Sormani MP, Comi G, *et al* (2009) A short-term randomized MRI study of high-dose oral vs intravenous methylprednisolone in MS. *Neurology* 73: 1842–8

Martinez-Lapiscina EH, Arnow S, Wilson JA, Saidha S, Preiningerova JL, Oberwahrenbrock T, Brandt AU, Pablo LE, Guerrieri S, Gonzalez I, *et al* (2016) Retinal thickness measured with optical coherence tomography and risk of disability worsening in multiple sclerosis: a cohort study. *Lancet Neurol* 15: 574–584

Martínez C, García-Martín E, Benito-León J, Calleja P, Díaz-Sánchez M, Pisa D, Alonso-Navarro H, Ayuso-Peralta L, Torrecilla D, Agúndez JAG, *et al* (2010)

Paraoxonase 1 polymorphisms are not related with the risk for multiple sclerosis. *Neuromolecular Med* 12: 217–223

Matarese G, Carrieri PB, Montella S, De Rosa V & La Cava A (2010) Leptin as a metabolic link to multiple sclerosis. *Nat Rev Neurol* 6: 455–461

McGuigan C, Craner M, Guadagno J, Kapoor R, Mazibrada G, Molyneux P, Nicholas R, Palace J, Pearson OR, Rog D, *et al* (2015) Stratification and monitoring of natalizumab-associated progressive multifocal leukoencephalopathy risk: recommendations from an expert group. *J Neurol Neurosurg Psychiatry* 87: jnnp-2015-311100

Meins M, Piosik P, Schaeren-Wiemers N, Franzoni S, Troncoso E, Kiss JZ, Brösamle C, Schwab ME, Molnár Z & Monard D (2001) Progressive neuronal and motor dysfunction in mice overexpressing the serine protease inhibitor protease nexin-1 in postmitotic neurons. *J Neurosci* 21: 8830–8841

Miclea A, Leussink VI, Hartung HP, Gold R & Hoepner R (2016) Safety and efficacy of dimethyl fumarate in multiple sclerosis: a multi-center observational study. *J Neurol* 263: 1626–1632

Miller AE, Wolinsky JS, Kappos L, Comi G, Freedman MS, Olsson TP, Bauer D, Benamor M, Truffinet P & O'Connor PW (2014) Oral teriflunomide for patients with a first clinical episode suggestive of multiple sclerosis (TOPIC): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Neurol* 13: 977–986

Minneboo A, Barkhof F, Polman CH, Uitdehaag BMJ, Knol DL & Castelijns JA (2004) Infratentorial lesions predict long-term disability in patients with initial findings suggestive of multiple sclerosis. *Arch Neurol* 61: 217–221

Mishra A & Macgregor S (2015) VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res Hum Genet* 18: 86–91

Mokry LE, Ross S, Ahmad OS, Forgetta V, Smith GD, Leong A, Greenwood CMT, Thanassoulis G, Richards JB & Richards JB (2015) Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. *PLOS Med* 12: e1001866

Montalban X, Hauser SL, Kappos L, Arnold DL, Bar-Or A, Comi G, de Seze J, Giovannoni G, Hartung H-P, Hemmer B, *et al* (2017) Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. *N Engl J Med* 376: 209–220

Munger KL, Bentzen J, Laursen B, Stenager E, Koch-Henriksen N, Sørensen TIA & Baker JL (2013) Childhood body mass index and multiple sclerosis risk: a long-term cohort study. *Mult Scler* 19: 1323–9

Munger KL, Levin LI, Hollis BW, Howard NS & Ascherio A (2006) Serum 25-Hydroxyvitamin D Levels and Risk of Multiple Sclerosis. *JAMA* 296: 2832

Munger KL, Levin LI, O'Reilly EJ, Falk KI & Ascherio A (2011) Anti-Epstein-Barr virus antibodies as serological markers of multiple sclerosis: a prospective study among United States military personnel. *Mult Scler* 17: 1185–93

Muraro PA, Robins H, Malhotra S, Howell M, Phippard D, Desmarais C, de Paula Alves Sousa A, Griffith LM, Lim N, Nash RA, *et al* (2014) T cell repertoire following autologous stem cell transplantation for multiple sclerosis. *J Clin Invest* 124: 1168–1172

Nait-Oumesmar B, Picard-Riera N, Kerninon C, Decker L, Seilhean D, Höglinger GU, Hirsch EC, Reynolds R & Baron-Van Evercooren A (2007) Activation of the subventricular zone in multiple sclerosis: evidence for early glial progenitors. *Proc Natl Acad Sci U S A* 104: 4694–9

Naito S, Namerow N, Mickey MR & Terasaki PI (1972) Multiple sclerosis: association with HL-A3. *Tissue Antigens* 2: 1–4

Nicoletti F, Mazzon E, Fagone P, Mangano K, Mammana S, Cavalli E, Basile MS, Bramanti P, Scalabrino G, Lange A, *et al* (2019) Prevention of clinical and histological signs of MOG-induced experimental allergic encephalomyelitis by prolonged treatment with recombinant human EGF. *J Neuroimmunol* 332: 224–232

O'Connor P, Wolinsky JS, Confavreux C, Comi G, Kappos L, Olsson TP, Benzerdjeb H, Truffinet P, Wang L, Miller A, *et al* (2011a) Randomized Trial of Oral Teriflunomide for Relapsing Multiple Sclerosis. *N Engl J Med* 365: 1293–1303

O'Connor PW, Goodman A, Kappos L, Lublin FD, Miller DH, Polman C, Rudick RA, Aschenbach W & Lucas N (2011b) Disease activity return during natalizumab treatment interruption in patients with multiple sclerosis. *Neurology* 76: 1858–1865

Oksenberg JR, Barcellos LF, Cree BAC, Baranzini SE, Bugawan TL, Khan O, Lincoln RR, Swerdlin A, Mignot E, Lin L, *et al* (2004) Mapping Multiple Sclerosis

Susceptibility to the HLA-DR Locus in African Americans. *Am J Hum Genet* 74: 160–167

Oksenberg JR, Seboun E & Hauser SL (1996) Genetics of demyelinating diseases. *Brain Pathol* 6: 289–302

Orton S-M, Herrera BM, Yee IM, Valdar W, Ramagopalan S V, Sadovnick AD, Ebers GC & Canadian Collaborative Study Group (2006) Sex ratio of multiple sclerosis in Canada: a longitudinal study. *Lancet Neurol* 5: 932–936

Ostmeyer J, Christley S, Rounds WH, Toby I, Greenberg BM, Monson NL & Cowell LG (2017) Statistical classifiers for diagnosing disease from immune repertoires: A case study using multiple sclerosis. *BMC Bioinformatics* 18: 1–10

Ota M, Nagafuchi Y, Hatano H, Ishigaki K, Terao C, Takeshima Y, Yanaoka H, Kobayashi S, Okubo M, Shirai H, *et al* (2021) Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* 184: 3006-3021.e17

Palazuelos J, Klingener M, Raines EW, Crawford HC & Aguirre A (2015) Oligodendrocyte regeneration and CNS remyelination require TACE/ADAM17. *J Neurosci* 35: 12241–12247

Papeix C, Vukusic S, Casey R, Debard N, Stankoff B, Mrejen S, Uhry Z, Van Ganse E, Castot A, Clanet M, *et al* (2016) Risk of relapse after natalizumab withdrawal: Results from the French TYSEDMUS cohort. *Neurol Neuroimmunol neuroinflammation* 3: e297

Parsons ME, O'Connell K, Allen S, Egan K, Szklanna PB, McGuigan C, Ní Áinle F & Maguire PB (2017) Thrombin generation correlates with disease duration in multiple sclerosis (MS): Novel insights into the MS-associated prothrombotic state. *Mult Scler J - Exp Transl Clin* 3: 205521731774762

Patsopoulos NA, Baranzini SE, Santaniello A, Shoostari P, Cotsapas C, Wong G, Beecham AH, James T, Replogle J, Vlachos IS, *et al* (2019) Multiple Sclerosis Genomic Map implicates peripheral immune cells & microglia in susceptibility. *Science* 365: 50

Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, Raj T, Gourraud P-A, Stranger BE, Oksenberg J, *et al* (2013) Fine-Mapping the Genetic Association of the Major Histocompatibility Complex in Multiple Sclerosis: HLA and Non-HLA Effects. *PLoS Genet* 9: e1003926

Paty DW & Li DK (1993) Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group. *Neurology* 43: 662–7

Pereira-Santos M, Costa PRF, Assis AMO, Santos CAST & Santos DB (2015) Obesity and vitamin D deficiency: a systematic review and meta-analysis. *Obes Rev* 16: 341–349

Pérez-Miralles F, Sastre-Garriga J, Tintoré M, Arrambide G, Nos C, Perkal H, Río J, Edo MC, Horga A, Castilló J, *et al* (2013) Clinical impact of early brain atrophy in clinically isolated syndromes. *Mult Scler* 19: 1878–1886

Pinto MF, Oliveira H, Batista S, Cruz L, Pinto M, Correia I, Martins P & Teixeira C (2020) Prediction of disease progression and outcomes in multiple sclerosis with machine learning. *Sci Rep* 10

Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH, Phillips JT, Lublin FD, Giovannoni G, Wajgt A, *et al* (2006) A Randomized, Placebo-Controlled Trial of Natalizumab for Relapsing Multiple Sclerosis. *N Engl J Med* 354: 899–910

Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, Fujihara K, Havrdova E, Hutchinson M, Kappos L, *et al* (2011) Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol* 69: 292–302

Prineas JW, Kwon EE, Cho ES, Sharer LR, Barnett MH, Oleszak EL, Hoffman B & Morgan BP (2001) Immunopathology of secondary-progressive multiple sclerosis. *Ann Neurol* 50: 646–57

Prokopenko I, Montomoli C, Ferrai R, Musu L, Piras ML, Ticca A, Murgia BS & Bernardinelli L (2003) Risk for Relatives of Patients with Multiple Sclerosis in Central Sardinia, Italy. *Neuroepidemiology* 22: 290–296

Qi Q, Liu Y, Cheng Y, Glanville J, Zhang D, Lee JY, Olshen RA, Weyand CM, Boyd SD & Goronzy JJ (2014) Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci U S A* 111: 13139–13144

Ramanujam R, Hedström AK, Manouchehrinia A, Alfredsson L, Olsson T, Bottai M & Hillert J (2015) Effect of Smoking Cessation on Multiple Sclerosis Prognosis. *JAMA Neurol* 72: 1117–1123

Reinhard E, Suidan HS, Pavlik A & Monard D (1994) Glia-derived nexin/protease nexin-1 is expressed by a subset of neurons in the rat brain. *J Neurosci Res* 37: 256–270

Richard-Miceli C & Criswell LA (2012) Emerging patterns of genetic overlap across autoimmune disorders. *Genome Med* 4: 6

Río J, Castilló J, Rovira A, Tintoré M, Sastre-Garriga J, Horga A, Nos C, Comabella M, Aymerich X & Montalbán X (2009) Measures in the first year of therapy predict the response to interferon beta in MS. *Mult Scler* 15: 848–53

Río J, Nos C, Tintoré M, Téllez N, Galán I, Pelayo R, Comabella M & Montalban X (2006) Defining the response to interferon-β in relapsing-remitting multiple sclerosis patients. *Ann Neurol* 59: 344–352

Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 46: 222–8

Robertson NP, Fraser M, Deans J, Clayton D, Walker N & Compston DA (1996) Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. *Brain* 119 (Pt 2): 449–55

Rodgers J, Friede T, Vonberg FW, Constantinescu CS, Coles A, Chataway J, Duddy M, Emsley H, Ford H, Fisniku L, *et al* (2021) The impact of smoking cessation on multiple sclerosis disease progression. *Brain*

Ross CJ, Towfic F, Shankar J, Laifenfeld D, Thoma M, Davis M, Weiner B, Kusko R, Zeskind B, Knappertz V, *et al* (2017) A pharmacogenetic signature of high response to Copaxone in late-phase clinical-trial cohorts of multiple sclerosis. *Genome Med* 9: 50

Rotstein D & Montalban X (2019) Reaching an evidence-based prognosis for personalized treatment of multiple sclerosis. *Nat Rev Neurol* 15: 287–300

Rotstein DL, Healy BC, Malik MT, Chitnis T & Weiner HL (2015) Evaluation of No Evidence of Disease Activity in a 7-Year Longitudinal Multiple Sclerosis Cohort. *JAMA Neurol* 72: 152

Ruck T, Pfeuffer S, Schulte-Mecklenbeck A, Gross CC, Lindner M, Metze D, Ehrchen J, Sondermann W, Pul R, Kleinschnitz C, *et al* (2018) Vitiligo after alemtuzumab treatment: Secondary autoimmunity is not all about B cells. *Neurology* 91: E2233–E2237

Rudick RA, Lee J-C, Simon J, Ransohoff RM & Fisher E (2004) Defining interferon ? response status in multiple sclerosis patients. *Ann Neurol* 56: 548–555

Rudick RA, Stuart WH, Calabresi PA, Confavreux C, Galetta SL, Radue E-W, Lublin FD, Weinstock-Guttman B, Wynn DR, Lynn F, *et al* (2006) Natalizumab plus Interferon Beta-1a for Relapsing Multiple Sclerosis. *N Engl J Med* 354: 911–923

Runia TF, Hop WCJ, de Rijke YB, Buljevac D & Hintzen RQ (2012) Lower serum vitamin D levels are associated with a higher relapse risk in multiple sclerosis. *Neurology* 79: 261–266

Runmarker B & Andersen O (1993) Prognostic factors in a multiple sclerosis incidence cohort with twenty-five years of follow-up. *Brain* 116 (Pt 1): 117–34

Sadovnick AD & Baird PA (1988) The familial nature of multiple sclerosis: age-corrected empiric recurrence risks for children and siblings of patients. *Neurology* 38: 990–1

Sadovnick AD, Yee IM, Guimond C, Reis J, Dyment DA & Ebers GC (2009) Age of onset in concordant twins and other relative pairs with multiple sclerosis. *Am J Epidemiol* 170: 289–96

Salzer J, Hallmans G, Nystrom M, Stenlund H, Wadell G & Sundstrom P (2012) Vitamin D as a protective factor in multiple sclerosis. *Neurology* 79: 2140–2145

Sangalli F, Moiola L, Ferrè L, Radaelli M, Barcella V, Rodegher M, Colombo B, Martinelli Boneschi F, Martinelli V & Comi G (2014) Long-term management of natalizumab discontinuation in a large monocentric cohort of multiple sclerosis patients. *Mult Scler Relat Disord* 3: 520–526

Sawcer S, Hellenthal G, Pirinen M, Spencer CCA, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, *et al* (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476: 214–219

Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, Ha B, Altay G, Greenbaum JA, McVicker G, *et al* (2018) Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* 175: 1701-1715.e16

Schumacher G, Beebe G, Kibler R, Kurland L, Kurtzke J, Mcdowell F, Nagler B, Sibley W, Tourtellotte W & Willmon T (1965) Problems of experimental trials of therapy in Multiple Sclerosis: report by the panel on the evaluation of experimental trials of

therapy in Multiple Sclerosis. *Ann N Y Acad Sci* 122: 552–68

Seccia R, Gammelli D, Dominici F, Romano S, Landi AC, Salvetti M, Tacchella A, Zaccaria A, Crisanti A, Grassi F, *et al* (2020) Considering patient clinical history impacts performance of machine learning models in predicting course of multiple sclerosis. *PLoS One* 15

Seccia R, Romano S, Salvetti M, Crisanti A, Palagi L & Grassi F (2021) Machine Learning Use for Prognostic Purposes in Multiple Sclerosis. *Life 2021, Vol 11, Page 122* 11: 122

Shi Y, Ding Y, Li G, Wang L, Osman RA, Sun J, Qian L, Zheng G & Zhang G (2021) Discovery of Novel Biomarkers for Diagnosing and Predicting the Progression of Multiple Sclerosis Using TMT-Based Quantitative Proteomics. *Front Immunol* 12: 1

Simon JH, Jacobs LD, Campion M, Wende K, Simonian N, Cookfair DL, Rudick RA, Herndon RM, Richert JR, Salazar AM, *et al* (1998) Magnetic resonance studies of intramuscular interferon ?-1a for relapsing multiple sclerosis. *Ann Neurol* 43: 79–87

Simpson S, Blizzard L, Otahal P, Van der Mei I & Taylor B (2011) Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis. *J Neurol Neurosurg Psychiatry* 82: 1132–1141

Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars N, Benner C, Aguirre M, Venkataraman GR, Wainberg M, Ollila HM, Kiiskinen T, *et al* (2021) Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat Genet* 53: 185–194

Slowik A, Wloch D, Szermer P, Wolkow P, Malecki M, Pera J, Turaj W, Dziedzic T, Klimkowicz-Mrowiec A, Kopec G, *et al* (2007) Paraoxonase 2 gene C311S polymorphism is associated with a risk of large vessel disease stroke in a Polish population. *Cerebrovasc Dis* 23: 395–400

Sombekke MH, Wattjes MP, Balk LJ, Nielsen JM, Vrenken H, Uitdehaag BMJ, Polman CH & Barkhof F (2013) Spinal cord lesions in patients with clinically isolated syndrome: a powerful tool in diagnosis and prognosis. *Neurology* 80: 69–75

Sormani M, Signori A, Siri P & De Stefano N (2013) Time to first relapse as an endpoint in multiple sclerosis clinical trials. *Mult Scler J* 19: 466–474

Sweeney MD, Sagare AP & Zlokovic B V. (2018) Blood–brain barrier breakdown in

Alzheimer's disease and other neurodegenerative disorders. *Nat Rev Neurol* 14: 133

Teunissen CE, Malekzadeh A, Leurs C, Bridel C & Killestein J (2015) Body fluid biomarkers for multiple sclerosis--the long road to clinical application. *Nat Rev Neurol* 11: 585–96

Thebault S, Reaume M, Marrie RA, Marriott JJ, Furlan R, Laroni A, Booth RA, Uccelli A & Freedman MS (2022) High or increasing serum NfL is predictive of impending multiple sclerosis relapses. *Mult Scler Relat Disord* 59

Therneau TM (2020) A Package for Survival Analysis in R.

Therneau TM & Grambsch PM (2000) Modeling survival data : extending the Cox model Springer

Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, Correale J, Fazekas F, Filippi M, Freedman MS, *et al* (2018) Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol* 17: 162–173

Thompson AJ, Polman CH, Miller DH, McDonald WI, Brochet B, Filippi M Montalban X & De Sá J (1997) Primary progressive multiple sclerosis. *Brain* 120 (Pt 6): 1085–96

Tintore M, Rovira À, Río J, Otero-Romero S, Arrambide G, Tur C, Comabella M, Nos C, Arévalo MJ, Negrotto L, *et al* (2015) Defining high, medium and low impact prognostic factors for developing multiple sclerosis. *Brain* 138: 1863–1874

Toledo J, Sepulcre J, Salinas-Alaman A, García-Layana A, Murie-Fernandez M, Bejarano B & Villoslada P (2008) Retinal nerve fiber layer atrophy is associated with physical and cognitive disability in multiple sclerosis. *Mult Scler* 14: 906–912

Tran JQ, Rana J, Barkhof F, Melamed I, Gevorkyan H, Wattjes MP, De Jong R, Brosofsky K, Ray S, Xu L, *et al* (2014) Randomized phase I trials of the safety/tolerability of anti-LINGO-1 monoclonal antibody BIIB033. *Neurol Neuroimmunol neuroinflammation* 1: e18

Trapp BD, Peterson J, Ransohoff RM, Rudick R, Mörk S & Bö L (1998) Axonal Transection in the Lesions of Multiple Sclerosis. *N Engl J Med* 338: 278–285

Tremlett H & Waubant E (2018) Gut microbiome and pediatric multiple sclerosis. *Mult Scler* 24: 64–68

Vermersch P, Czlonkowska A, Grimaldi LME, Confavreux C, Comi G, Kappos L,

Olsson TP, Benamor M, Bauer D, Truffinet P, *et al* (2014) Teriflunomide versus subcutaneous interferon beta-1a in patients with relapsing multiple sclerosis: a randomised, controlled phase 3 trial. *Mult Scler* 20: 705–16

Viechtbauer W (2010) Conducting Meta-Analyses in *R* with the **metafor** Package. *J Stat Softw* 36: 1–48

Viehover A, Miller RH, Park SK, Fischbach G & Vartanian T (2001) Neuregulin: an oligodendrocyte growth factor absent in active multiple sclerosis lesions. *Dev Neurosci* 23: 377–386

Villoslada P & Steinman L (2020) New targets and therapeutics for neuroprotection, remyelination and repair in multiple sclerosis. *Expert Opin Investig Drugs* 29: 443–459

Vukusic S, Van Bockstael V, Gosselin S & Confavreux C (2007) Regional variations in the prevalence of multiple sclerosis in French farmers. *J Neurol Neurosurg Psychiatry* 78: 707–9

Vyshkina T, Banisor I, Shugart YY, Leist TP & Kalman B (2005) Genetic variants of Complex I in multiple sclerosis. *J Neurol Sci* 228: 55–64

Walsh K, Shah R, Armstrong JK, Moore ES & Oliver BJ (2022) Comparing traditional modeling approaches versus predictive analytics methods for predicting multiple sclerosis relapse. *Mult Scler Relat Disord* 57

Wang J, Vasaikar S, Shi Z, Greer M & Zhang B (2017) WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res* 45: W130–W137

Warnke C, Mausberg AK, Stettner M, Dehmel T, Nekrich L, Horste GMZ, Hartung HP, Fogdell-Hahn A, Adams O & Kieseier BC (2013) Natalizumab affects the T-cell receptor repertoire in patients with multiple sclerosis. *Neurology* 81: 1400–1408

Waubant E, Vukusic S, Gignoux L, Dubief FD, Achiti I, Blanc S, Renoux C & Confavreux C (2003) Clinical characteristics of responders to interferon therapy for relapsing MS. *Neurology* 61: 184–9

Weinshenker BG, Bass B, Rice GP, Noseworthy J, Carriere W, Baskerville J & Ebers GC (1989) The natural history of multiple sclerosis: a geographically based study. 2. Predictive value of the early clinical course. *Brain* 112 (Pt 6): 1419–28

Weinshenker BG, O'Brien PC, Petterson TM, Noseworthy JH, Lucchinetti CF, Dodick

DW, Pineda AA, Stevens LN & Rodriguez M (1999) A randomized trial of plasma exchange in acute central nervous system inflammatory demyelinating disease. *Ann Neurol* 46: 878–86

Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC & Canadian Collaborative Study Group (2003) Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc Natl Acad Sci* 100: 12877–12882

Windhagen A, Newcombe J, Dangond F, Strand C, Woodroofe MN, Cuzner ML & Hafler DA (1995) Expression of costimulatory molecules B7-1 (CD80), B7-2 (CD86), and interleukin 12 cytokine in multiple sclerosis lesions. *J Exp Med* 182: 1985–96

Wottschel V, Alexander DC, Kwok PP, Chard DT, Stromillo ML, De Stefano N, Thompson AJ, Miller DH & Ciccarelli O (2014) Predicting outcome in clinically isolated syndrome using machine learning. *NeuroImage Clin* 7: 281–287

Wu M, Zhao M, Wu H & Lu Q (2021) Immune repertoire: Revealing the “real-time” adaptive immune response in autoimmune diseases. *https://doi.org/101080/0891693420211887149* 54: 61–75

Yoo Y, Tang LYW, Li DKB, Metz L, Kolind S, Traboulsee AL & Tam RC (2017) Deep learning of brain lesion patterns and user-defined clinical and MRI features for predicting conversion to multiple sclerosis from clinically isolated syndrome. *https://doi.org/101080/2168116320171356750* 7: 250–259