Genetic subtypes associated with multiple sclerosis severity and response to treatment.

Karim L Kreft, MD, PhD, MSc(Hons) [1,2], Nienke J. Mekkes MSc [3,4], Emeka Uzochukwu PhD [2], Sam Loveless PhD [2], Ray Wynford-Thomas MD, PhD [2], Katharine E. Harding MD, PhD [5], Mark Wardle MD [1], Peter Holmans PhD [2], J William L Brown MD, PhD [6], Michael Lawton PhD [7], Emma C Tallantyre MD, PhD [1,2], Inge R. Holtman PhD [3,4], Neil P Robertson, MBBS, MD [1,2]

Affiliations

[1] Department of Neurology, University Hospital of Wales, Cardiff, UK.

[2] Division of Psychological Medicine and Clinical Neuroscience, Cardiff University, Cardiff, UK.

[3] Department of Biomedical Sciences, Section Molecular Neurobiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[4] Machine Learning Lab, Data Science Center in Health, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

[5] Department of Neurology, Royal Gwent Hospital, Newport, United Kingdom

[6] Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom

[7] Bristol Population Health Science Institute, Bristol Medical School, University of Bristol, Bristol, United Kingdom

Running head:

Multiple sclerosis genetic subtypes and outcomes

Corresponding authors:

Karim L. Kreft, karim.kreft@nottingham.ac.uk

Neil P. Robertson, robertsonnp@cardiff.ac.uk

Number of characters in title: 87

Number of characters in running head: 48

Word count article: 3495

Word count abstract: 248

Abstract

Background: Predicting response to treatment and long-term disability in multiple sclerosis (MS) remains challenging. In other complex diseases, combining genetic risk variants has enabled detection of relevant clinical endophenotypes associated with important outcomes, but this strategy has never been applied to MS.

Methods: We applied unsupervised hierarchical clustering to genomic risk scores in a prospective Welsh MS cohort (n=1,455) and replicated findings in the post-mortem Netherlands Brain Bank MS cohort (NBB-MS, n=272). Disease progression was assessed using survival analysis for time to Expanded Disability Status Scale (EDSS) milestones.

Results: Three genomic clusters were identified, each with similar genetic profiles. Baseline demographics did not differ between clusters. Welsh patients in cluster 1 attained EDSS6 and EDSS8 significantly later than cluster 2 and 3 (by respectively 6 years, $p=3x10^{-3}$ and 13 years, p=0.02 respectively). These findings were replicated in the NBB-MS cohort (6-year delay to EDSS6 for cluster 1 vs. 2, p=0.04). Genomic clustering independently predicted disease progression (hazard ratios 1.3–2.0, all $p<0.05$), beyond established risk factors. Clusters 2 and 3 showed greater annual T2 lesion load increase on serial MR imaging (p=0.04). In cluster 2, patients receiving disease-modifying treatments (DMTs) had delayed progression to EDSS6 ($p=3\times10^{-3}$), whilst no such benefit was observed in clusters 1 or 3. Cluster 2 patients also had earlier onset of symptoms including dysphagia (p=0.02) and spasticity ($p=8\times10^{-4}$) in the NBB-MS cohort.

Conclusion: Genetic clustering reveals clinically meaningful MS subtypes with distinct prognoses and treatment responses, highlighting its potential role in precision medicine for MS management.

Key messages

What is already known on this topic

Multiple sclerosis (MS) susceptibility and severity is a partly heritable. Currently, only one genome wide significant single nucleotide variant (SNV) is associated with MS severity. In other disease areas, combinations of SNVs have been shown to improve identification of subgroups of patients with distinct clinical profiles.

What this study adds

In this study, we have identified three subgroups of people living with MS with significantly different prognosis based on their genomics. The most severely affected group become wheelchair dependent 13 years earlier than the group with the best prognosis. This group also accumulates more T2 lesions on serial MR imaging, have more cognitive and mental health problems, and more swallowing difficulties.

How this study might affect research, practice or policy

This study could support precision medicine by enabling early identification of individuals at risk of a more aggressive disease course, potentially guiding timely initiation of high-efficacy disease modifying treatments for those who require them most.

Introduction

Multiple sclerosis (MS) is a neuro-inflammatory and degenerative disorder with heterogenous outcomes. Over 15 disease modifying treatments (DMT) are currently approved, with a range of efficacy and safety. Accurate risk prediction of inflammation and development of sustained disability is important for personalised management, but remains difficult at disease onset[1].

233 single nucleotide variants (SNVs) have been linked to MS susceptibility[2], along with one genome-wide significant locus and several suggestive loci associated with increased risk of a more severe disease course[3,4]. Genetic studies assessing disease outcomes have mainly focused at the effects of variants in isolation with conflicting results[5,6].

In other disease areas, clustering of patients based on a combination of genetic variants has increased accuracy of predicting outcomes[7,8], although this approach has never been employed in MS. We hypothesize that a combination of genomic scores will also improve the accuracy of predicting long-term disability in MS. Therefore, we integrated results of three susceptibility and progression genome wide association studies (GWAS)[2–4] and applied genetic clustering. We associated the resultant patient clusters with clinical outcomes in a cohort of people with MS (pwMS, n=1455) for whom detailed longitudinal phenotypic data was available. Results were then replicated and expanded within the Netherlands Brain Bank– MS cohort (NBB-MS, n=272, Fig. 1).

Materials and methods

Participants

1455 patients were recruited from the South Wales MS Registry established in 1985. Consecutive patients undergo annual assessment and during each visit, demographic and clinical data (including DMT, MRI and Expanded Disability Severity Score, EDSS) are collected by trained physicians recorded in a standardised web-based form. Inclusion was limited to Caucasian pwMS with an established diagnosis according to 2017 McDonald Criteria[9]. This study was approved by the Wales Research Ethics Committee.

The replication cohort consisted of Caucasian donors with post-mortem confirmed MS (n=272) from the Netherlands Brain Bank (NBB). The study was approved by the Free University Medical Center Ethics Committee. Clinical, neuropathological, and genetic information from donors were obtained, cleaned, and linked to ontology structures[10].

All participants provided written informed consent.

Genotyping

Welsh cohort genotyping methodology has been described elsewhere[5]. Briefly, DNA from blood was stored at -80°C. Genotyping was performed on either Infinium CoreExome-24 v2 or v3 or Immunochip (all Illumina). Stringent quality criteria were applied, and genotypic data was imputed.

For the NBB-MS cohort, genotyping was performed on blood or brain tissue using Infinium Global Screening Array v.3.0 according to manufacturers' instructions. Related and non-European donors were excluded after applying quality control and imputation as previously described[11].

Genomic risk scores and unsupervised hierarchical clustering

We constructed genomic risk scores (wGRS) from the largest MS susceptibility GWAS (significant associations only)[2] and the two most recent progression GWAS (both significant and suggestive loci as per definition of the original GWAS ($p=5*10^{-6}$) were used due to limited power)[3,4] to capture all genomic loci contributing to MS. All available SNVs were used (Supplementary table 1). The number of risk alleles per patient and wGRS (the total genetic risk for MS per individual patient) was calculated using $wGRS = \sum_{i=1}^{k} \beta i Ni$, where N is the number of risk alleles and β the logit of the OR of SNV i[12]. Human leukocyte antigen genomic burden (HLAGB) score was calculated based on the MS susceptibility GWAS[2]. Individual SNVs were uniquely assigned to a single genomic risk score, without overlap across scores. We compared different unsupervised hierarchical clustering algorithms using agglomerative coefficient and selected the algorithm with the highest coefficient (Euclidean distance with Ward D2 clustering algorithm) according to a pre-specified data analysis plan. Only results of this algorithm were taken forward to compare clinical characteristics based on patient's assigned genomic cluster (Fig. 1). To control for population structure, we computed the first 10 principal components (PC) on the genomic data using PLINK2[13], removing SNP with Hardy-Weinberg disequilibrium $<1*10^{-10}$, a minor allele frequency <0.05 and samples with >1% missing SNP data. PC1 and PC2 were plotted with color-coding for the assigned genomic cluster. We did not observe evidence that our clustering was due to population structure (supplementary figure 1). Collider bias (a biased association between a genetic variant associated with susceptibility having an effect on severity)[14] was excluded using multiple approaches (Suppl. methods). To evaluate the robustness of our clustering approach and assess the potential influence of methodological artefacts, we conducted a permutation-based sensitivity analysis (Suppl methods). For the NBB cohort, we used the Welsh cluster genomic data as a classifier and applied the k-nearest neighbours' algorithm to assign clusters to the NBB-MS patients.

Clinical parameters

Annualised relapse rate (ARR) was calculated by dividing the total number of relapses by follow-up in years. Individuals with primary progressive disease were excluded from analysis. EDSS obtained during relapse were excluded to prevent overestimation of disability due to temporary worsening in neurological function.

MRIs were collected as part of routine clinical care. The number of T2-hyperintense lesions (WML) was assessed longitudinally through visual inspection by experienced physicians, using a standardised web-based data collection form. Baseline number of WML were dichotomised into < 9 or ≥9. We calculated annual increase in WML for patients who underwent MRIs at least every 2 years. Due to variability in imaging acquisition protocols and different scanners, enlarging lesions were excluded.

Cerebrospinal fluid (CSF) analysis, including evaluation for oligoclonal bands (OCBs), was performed when clinically indicated. OCB testing followed established local diagnostic laboratory protocols.

DMTs may delay time to develop sustained disability and patients presenting with more unfavourable prognostic factors are more likely to receive high-efficacy DMT. Therefore, DMT use may result in confounding-by-indication. To control for this, patients were categorised into groups who were commenced on moderate efficacy (all interferons, teriflunomide, fumarates and glatiramer acetate), high efficacy (cladribine, S1P inhibitors, monoclonal antibodies), or patients switching from moderate to high efficacy DMT or never exposed to DMTs[15] and stratified survival analysis on DMT use was performed.

Multi-level modelling

Covariates included sex, age of onset, and genotype clusters. Time since onset was used as a time metric, DMTs are time-varying covariates as they are administered at different time points. Factional polynomials were utilised to account for non-linear disease progression. To avoid modelling EDSS during relapse, scores within 6-months post-relapse were excluded. To reduce autocorrelation caused by short-interval observations or prolonged unchanged periods, we summarised observations within quarter-year intervals using the median score before model fitting. Variation within individuals may depend on time, allowing nonconstant residual variance. We accounted for this using a complex level 1 variance model, modelling level 1 residuals as a function of time, previously described[16].

Survival analysis

Primary endpoint was time to sustained disability, defined by reaching EDSS milestones of 4, 6, or 8, with no subsequent EDSS falling below the attained milestone or time to develop secondary progressive MS (SPMS, determined based on an increase in EDSS as previously described, in relapsing onset patients only[5]), adjusted for sex, age at onset and DMT use. Patients were assigned to a genomic cluster which was used as an independent predictor. We constructed multivariate Cox Proportional Hazard models with genomic clusters and well-known clinical risk factors for disease progression (cerebellar disease onset, number of relapses in first 5 years, and optic neuritis as protective factor)[1]. We computed three hazard ratios (HR) to assess if genomic clustering adds predictive value for the risk to develop EDSS milestones, the reference groups were (i) patient in genetic cluster 1 presenting without a cerebellar syndrome, (ii) patients presenting with an optic neuritis and genomic cluster 1 and (iii) patients in genomic cluster 1 with ≤2 relapses in the first five years after onset.

Symptomatology, comorbidities and EDSS in the NBB-MS cohort

Survival analysis for time to EDSS 6 and other clinical outcomes were performed in the Welsh cohort. Differences in disease progression might also manifest as differences in experienced symptomatology. For the NBB-MS cohort, neuro(-psychiatric) symptoms were scored in medical record summaries through natural language processing as described previously[10]. We compared the number of observations of a symptom between the genetic clusters, assessed difference in temporal distribution (Kruskal-Wallis test), and performed survival analysis for each symptom. Comorbidities are potential confounders of disease progression and were grouped into higher-order categories of the Human Disease Ontology. For each cluster, the proportion of donors with at least one member of each higher-order category was compared (χ^2-test).

Statistical analysis

Comparison of baseline characteristics was undertaken using non-parametric analysis (Kruskal-Wallis test). Correlation between the genomic scores was calculated (Spearman's rho). Statistical analyses were performed with RStudio (Supplementary materials). For all secondary analyses, we applied false discovery rate (FDR) adjustment for multiple testing and p-value <0.05 were considered statistically significant.

Results

Individual genomic risk scores are not associated with time to disability

Both the South Wales and NBB MS cohort had a mean age at onset of MS around 33 years of age and were predominantly female (approximately 70%). However the NBB-MS had relatively more patients presenting with primary progressive multiple sclerosis (22.4% in NBB-MS versus 10.4% in the South Wales MS cohort (Table 1). We assessed whether different wGRS for MS susceptibility and severity were predictive for development of sustained disability. We did not observe a consistent association between wGRS and the HR to develop EDSS 4, 6, and 8 or SPMS (Supplementary tables 2-5), validating another study using MS susceptibility SNPs to predict disease severity[17], indicating that single wGRS are unable to predict disease worsening.

Unsupervised hierarchical clustering on genomic risk scores identified three clusters of patients with similar genomic risk profiles.

Unsupervised hierarchical clustering was applied only to genomic risk scores revealing a Ward agglomerative coefficient of 0.97, indicating clear clustering patterns based on genomics alone (Fig. 1). Collider bias was excluded by applying several different methods (supplementary figure 2). No significant differences in wGRS were found between the Welsh and NBB-MS cohort (all $p>0.05$), indicating robust and generalisable results between the two cohorts. Next, we validated that wGRS were significantly different between different clusters and that all scores independently contributed to the clustering approach (Figure 2A-H), all p-values $<7.4*10^{-8}$).

Severity of MS at disease onset is similar between genomic clusters

No significant differences in baseline characteristics were identified between clusters, indicating that severity at disease onset is independent of genetics (Table 1-2).

Patients in genomic cluster 2 have significantly increased risk of fixed disability
We compared long-term disease outcomes and found no significant differences in time to impaired ambulation (EDSS 4) between different genetic clusters (Supplementary figure 3). Interestingly, patients in cluster 2 and 3 had a significantly shorter time to requiring a walking aid (EDSS 6, 5.7 and 4.9 years earlier respectively, Fig 3A) and time to wheelchair dependency (EDSS 8, 2.8 and 16.2 years earlier respectively Fig. 3B) compared to cluster 1. Time to develop SPMS was significantly shorter for patients in genomic cluster 2 and 3 (3.9 and 5.1 years respectively Fig. 3C). Patients in genomic cluster 2 and 3 also had a significantly higher risk of developing EDSS 6 (HR 1.32, p=0.009 and 1.46, $p=4*10^{-5}$, Supplementary table 3), EDSS 8 (HR 1.45, p=0.038 and 1.49, p=0.0087, Supplementary table 4) and SPMS (HR 1.27, p=0.030 and 1.30, p=0.0063, Fig. 3E and Supplementary table 5). No significant differences were found between genomic clusters in ARR in RRMS (Cohen's d max 0.078, p=0.23, data not shown), indicating that the genetic variants are likely related to accelerated development of sustained disability independent of relapse activity. Next, we compared EDSS trajectories using a multi-level model to compare the annual increase in EDSS. We included 8,941 non-relapse EDSS from 1,029 relapsing pwMS. Nine years post-onset, EDSS trajectories started to diverge, patients in genomic cluster 2 had a 0.38 (95% CI 0.00075-0.75, Fig 4F and Supplementary table 6) higher increase in EDSS compared with cluster 1. Next, we replicated time to EDSS 6 in the independent NBB-MS cohort; patients in genomic clusters 2 and 3 had an accelerated time to development of sustained disability (p=0.044, Fig. 3D) independent of comorbidities (Fig. 3G). Our longitudinal cohort included patients diagnosed under different revisions of the diagnostic criteria. With each successive revision, the interval between symptom onset and diagnosis shortened. This may have allowed earlier treatment initiation, potentially influencing long-term outcomes. To account for this, we adjusted our survival

analyses for year of diagnosis. This adjustment did not alter our findings (supplementary tables 4 and 5).

Sensitivity analysis of clustering

To account for the potential influence of random artefacts inherent to unsupervised clustering, we performed sensitivity analysis by introducing SNVs not associated with MS into the wGRS. Incorporating four SNVs (equivalent of one per wGRS) into the clustering analysis abolished the previously observed associations with long-term outcomes. Additional permutations, involving different combinations and distributions of random SNVs across wGRS, consistently failed to yield significant associations with outcomes (supplementary table 7). These results reinforce the specificity of our clustering approach and suggest that the identified clusters reflect truly underlying biological signals rather than random statistical noise.

Genomic clustering is an independent predictor of disease progression compared to well-validated risk factors

We assessed performance of genomic clustering to predict risk of disease progression compared to well-validated clinical predictors[1]. We found that patients presenting with a cerebellar syndrome at disease onset have a HR of 1.44 (95% CI 1.15-1.81, p=0.002, fig. 3H) of developing EDSS 6. Interestingly, patients without cerebellar syndrome at onset (favourable outcome based on current risk prediction) within genomic cluster 2 and 3 had a similar risk of developing EDSS 6 as patients presenting with cerebellar signs (Fig. 3H). We found a similar pattern that genomic clustering improves risk prediction for patients with two or less relapses within the first 5 years after onset (Supplementary figure 4). Patients with optic neuritis at onset overall had a lower risk of developing EDSS 6 (HR 0.81). However, optic neuritis patients within cluster 2 and 3 had a significantly increased risk of developing EDSS 6 compared to cluster 1 (HR 1.3, Fig. 3I). Therefore, genetic testing at onset provides additional predictive information for the risk of sustained disability (Fig. 3J).

Cluster 1 patients have significantly lower annualised increase in T2 lesions

539 pwMS had at least bi-annual brain MR imaging. At baseline, no differences were found in the percentage of patients having 9 or more lesions between the different genomic clusters (p=0.61). Comparing annualised increase in WML in patients having active MRI revealed that patients in genomic cluster 2 and 3 had a significantly higher number of new lesions compared to cluster 1 patients (p=0.04, Table 2).

Cluster 2 patient have best response to treatment

Time to develop disability milestones may be delayed by DMTs. Therefore, we assessed effect of DMT within our genetic clusters comparing patients who had never been treated versus patients ever receiving a DMT. Cluster 2 patients who received a DMT had a significantly longer time to develop EDSS 6 compared to untreated patients in cluster 2 (11.7 years, p=0.003, Fig. 5A) and similar effects were found for time to EDSS 8 (p=0.005, Fig. 4B) with a trend towards significance for SPMS (p=0.06, Fig. 4C). In contrast, we only observed an effect of a DMT in cluster 1 patients for time to EDSS 8 (Fig. 4B) and no effect of a DMT in cluster 3 patients (Fig. 4A-C). These effects were replicated in the multi-level model (Supplementary table 6). We did not observe a difference in time to EDSS 4 between treated and untreated patients between the different genomic clusters (Supplementary figure 5). Genomic clustering may help identify good responders to DMTs.

Patients in genomic cluster 1 have a more favourable phenotype

Cluster 1 patients had a more favourable disease course on a broad range of symptoms, including cognition, bladder involvement and fatigue (Fig. 5A)[10]. Patients in cluster 2 were younger when they developed motor and non-motor symptoms (Fig. 5B-D) and did so more frequently (Fig. 5E-F), indicating that genetics is associated with disease phenotype.

Discussion

We have identified three clusters of patients with similar genetic characteristics and baseline characteristics but different disease trajectories, in two independent cohorts. Patients in cluster 2 and 3 had a greater increase in WML on serial MRI and a higher risk of developing early fixed disability, independent of relapses. Cluster 2 patients showed the best response to DMT. It may seem counterintuitive that cluster 2 patients who carry relatively fewer inflammation-associated susceptibility variants compared to other clusters, exhibit the strongest response to immunomodulatory therapies. These patients still possess some inflammation-related variants and targeting these with DMTs may unmask the influence of progression-related variants. Alternative, specific combinations of genetic variants may play a critical role in long-term outcomes. This hypothesis warrants further investigation in future studies. We confirmed that patients in genomic cluster 2 have a worse phenotype compared to cluster 1 in a second independent cohort of patients. While one could consider this second cohort small when performing unbiased genome-wide analysis, we could replicate our findings, because of the large effect sizes compared to GWAS. To prevent a self-fulling prophecy, we could not include patients from previous GWASs[2–4], significantly limiting replication cohorts. Large scale biobanks (UK Biobank among others) lack the required in-depth phenotypes and longitudinal outcome measurements.

To our knowledge, this is the first study applying clustering analysis integrating all available genetic information into one model. Previously, an Australian multi-centre study applied Ensemble machine learning to a cohort of 279 patients with a first demyelinating event and found that 7 out of 208 MS susceptibility loci are associated with the risk of sustained disability[18]. Although MS susceptibility variants are mainly associated with immune function, a recent meta-analysis provided further support that MS susceptibility is also linked to neuronal and glial dysfunction[19]. In other disease areas, genomic clustering has been shown to identify relevant disease phenotypes. In patients with insulin resistance, genomic clustering was able to identify subgroups who had a higher risk of developing hypertension and coronary artery disease, whereas another cluster was associated with lipid metabolism[8]. In type 2 diabetes, genomic clustering was associated with a subgroup with insulin resistance, and another cluster with reduced insulin secretion[7]. In a large study involving Alzheimer's (AD), Parkinson's disease (PD), frontotemporal dementia and amyotrophic lateral sclerosis, genomic clustering revealed novel insights in underlying pathologies[20]. By integrating clinical and genomic data from PD, AD risk loci were shown to be related to clinical phenotypes in PD[21] and also clustering on clinical data can identify subgroups of patients in primary lateral sclerosis with novel genetic mutations[22].

In the current study, MS patients in genomic cluster 2 and 3 had a significantly shorter time to developing sustained disability. These two patient groups also had the highest annual increase in WML. The development of new WML is a well-known risk factor for developing sustained disability[23–25] and could, at least partially, be driven by genetic factors. Another interesting observation was that using newly developed multi-level modelling[16], within 9 years of disease onset, the rate of increase in EDSS is similar between genomic clusters. However, after 10 years, particularly in cluster 2, there was accelerated disability accumulation, validating the non-linear disease trajectory of MS and underlining the importance of genetics in disability progression independent of relapse activity.

Our study has limitations. We observed differences in response to treatment, but we were unable to stratify this effect to specific DMTs because of sample size. Larger studies, preferably nested within a clinical trial, will need to address this further. Secondly, the inter-rater variability of EDSS is relatively high[26] and could lead to some ascertainment bias. Collection of EDSS measurements from 1985-

2022, used different physicians. However, considering the high number of EDSS per patients, we do not expect any systemic bias. In addition, accumulation of WML and non-EDSS outcomes all showed similar patterns with patients in cluster 2 having a more severe phenotype compared to cluster 1. Thirdly, because we included a historical cohort, not all pwMS underwent regular MRIs. The imaging-based outcomes could therefore only be assessed in a subgroup of patients over-represented by those receiving a DMT. However, this would be more likely to lead to an underestimation of annualised increase in WML, since untreated patients would be more likely to develop new lesions compared to treated patients. Fourthly, patients in cluster 2 had a significantly shorter time to EDSS milestones, but the best response to treatment. This could reflect some degree of residual confounding-by-indication or collider bias and the effects of DMT on progression are not causal effects as we have not adjusted for all confounding. Some patients were diagnosed before the licensing of DMT and therefore the lag time between disease onset and DMT start could contribute to a diminished response to treatment as those patients had an increased risk of disease progression due to a higher ARR[27]. Future research could explore methods that account for time-varying treatments and confounders (such as marginal structural models)[28] to robustly assess whether these genetic clusters are associated with the effect of DMTs.

Finally, the use of unsupervised artificial intelligence models has advantages and disadvantages. The main disadvantage is the risk of overfitting of the model and thus reduced generalizability, however the external validation of MS prognosis in our validation cohort showed that overfitting is unlikely in our study. Moreover, our sensitivity analysis further ascertained that our results reflect true biological effects rather than statistical artefacts. Supervised models require an unbiased and well-defined clinical outcome to train the model on, and due to the relative lack of consensus in outcomes in MS (e.g. what we consider a good and worse outcome due to the heterogeneity of the disease), results of supervised models will be more variable due to differences in primary endpoint chosen. Unsupervised clustering on genomics overcomes this limitation, because the clustering is performed on genetics rather than clinical grounds, and subsequently the clusters are used to compare clinical outcomes between genetic subgroups.

In summary, we have identified three genetic clusters associated with important long-term outcomes. Our study suggest that genomic clustering provides a valuable tool to detect endophenotypes of MS and to increase accuracy of prognostication in MS. Further studies are necessary to understand the underlying biology (which genetic loci are driving the effect, and which pathways are affected) and to develop more targeted therapeutic approaches.

Acknowledgement

We are grateful to Prof. Davey Smith, Karen Ho BSc and Dr. Tom Clark (University of Bristol, UK) for assistance with the genotyping and all people with MS who have participated in this study. We would like to acknowledge the staff from the Netherlands Brain Bank, as well as the brain donors and their families, and all participating pwMS from the South Wales MS Registry for their invaluable contributions to science.

Authors contribution

KLK, ECT, NPR contributed to the conception and design of the study; KLK, NJM, EU, SL, RWT, KEH, PH, JWJB, ECT, IRH, NPR contributed to the acquisition and analysis of data; KLK, NJM, EU, SL, RWT, KEH, PH, JWJB, ECT, IRH, NPR contributed to drafting the text or preparing the figures.

Potential Conflicts of Interest

None of the authors report a conflict of interest.

Funding

Not applicable

References

1 Rotstein D, Montalban X. Reaching an evidence-based prognosis for personalized treatment of multiple sclerosis. *Nat Rev Neurol*. 2019;15:287–300. doi: 10.1038/s41582-019-0170-8

2 International Multiple Sclerosis Genetics Consortium, Patsopoulos NA, Baranzini SE, *et al.* Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*. 2019;365:eaav7188. doi: 10.1126/science.aav7188

3 International Multiple Sclerosis Genetics Consortium, MultipleMS Consortium. Locus for severity implicates CNS resilience in progression of multiple sclerosis. *Nature*. 2023;619:323–31. doi: 10.1038/s41586-023-06250-x

4 Jokubaitis VG, Campagna MP, Ibrahim O, *et al.* Not all roads lead to the immune system: the genetic basis of multiple sclerosis severity. *Brain*. 2023;146:2316–31. doi: 10.1093/brain/awac449

5 Kreft KL, Uzochukwu E, Loveless S, *et al.* Relevance of Multiple Sclerosis Severity Genotype in Predicting Disease Course: A Real-World Cohort. *Ann Neurol*. 2024;95:459–70. doi: 10.1002/ana.26831

6 Campagna MP, Havrdova EK, Horakova D, *et al.* No evidence for association between rs10191329 severity locus and longitudinal disease severity in 1813 relapse-onset multiple sclerosis patients from the MSBase registry. *Mult Scler*. 2024;13524585241240406. doi: 10.1177/13524585241240406

7 Dimas AS, Lagou V, Barker A, *et al.* Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes*. 2014;63:2158–71. doi: 10.2337/db13-0949

8 Yaghootkar H, Scott RA, White CC, *et al.* Genetic evidence for a normal-weight "metabolically obese" phenotype linking insulin resistance, hypertension, coronary artery disease, and type 2 diabetes. *Diabetes*. 2014;63:4369–77. doi: 10.2337/db14-0318

9 Thompson AJ, Banwell BL, Barkhof F, *et al.* Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*. 2018;17:162–73. doi: 10.1016/S1474-4422(17)30470-2

10 Mekkes NJ, Groot M, Hoekstra E, *et al.* Identification of clinical disease trajectories in neurodegenerative disorders with natural language processing. *Nat Med*. 2024;30:1143–53. doi: 10.1038/s41591-024-02843-9

11 de Boer A, van den Bosch AMR, Mekkes NJ, *et al.* Disentangling the heterogeneity of multiple sclerosis through identification of independent neuropathological dimensions. *Acta Neuropathol*. 2024;147:90. doi: 10.1007/s00401-024-02742-w

12 De Jager PL, Chibnik LB, Cui J, *et al.* Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol*. 2009;8:1111–9. doi: 10.1016/S1474-4422(09)70275-3

13 Chang CC, Chow CC, Tellier LC, *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. doi: 10.1186/s13742-015-0047-8

14 Mahmoud O, Dudbridge F, Davey Smith G, *et al.* A robust method for collider bias correction in conditional genome-wide association studies. *Nat Commun*. 2022;13:619. doi: 10.1038/s41467-022-28119-9

15 Filippi M, Amato MP, Centonze D, *et al.* Early use of high-efficacy disease-modifying therapies makes the difference in people with multiple sclerosis: an expert opinion. *J Neurol*. 2022;269:5382–94. doi: 10.1007/s00415-022-11193-w

16 Uzochukwu EC, Harding KE, Hrastelj J, *et al.* Modelling Disease Progression of Multiple Sclerosis in a South Wales Cohort. *Neuroepidemiology*. 2024;58:218–26. doi: 10.1159/000536427

17 George MF, Briggs FBS, Shao X, *et al.* Multiple sclerosis risk loci and disease severity in 7,125 individuals from 10 studies. *Neurol Genet*. 2016;2:e87. doi: 10.1212/NXG.0000000000000087

18 Fuh-Ngwa V, Zhou Y, Melton PE, *et al.* Ensemble machine learning identifies genetic loci associated with future worsening of disability in people with multiple sclerosis. *Sci Rep*. 2022;12:19291. doi: 10.1038/s41598-022-23685-w

19 Zeng L, Atlas K, Lama T, *et al.* GWAS highlights the neuronal contribution to multiple sclerosis susceptibility. 2024, *pre-print*. doi: 10.1101/2024.12.04.24318500.

20 Koretsky MJ, Alvarado C, Makarious MB, *et al.* Genetic risk factor clustering within and across neurodegenerative diseases. *Brain*. 2023;146:4486–94. doi: 10.1093/brain/awad161

21 Sandor C, Millin S, Dahl A, *et al.* Universal clinical Parkinson's disease axes identify a major influence of neuroinflammation. *Genome Med*. 2022;14:129. doi: 10.1186/s13073-022-01132-9

22 Mitsumoto H, Nagy PL, Gennings C, *et al.* Phenotypic and molecular analyses of primary lateral sclerosis. *Neurol Genet*. 2015;1:e3. doi: 10.1212/01.NXG.0000464294.88607.dd

23 Fisniku LK, Brex PA, Altmann DR, *et al.* Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain*. 2008;131:808–17. doi: 10.1093/brain/awm329

24 Tur C, Carbonell-Mirabent P, Cobo-Calvo Á, *et al.* Association of Early Progression Independent of Relapse Activity With Long-term Disability After a First Demyelinating Event in Multiple Sclerosis. *JAMA Neurol*. 2023;80:151. doi: 10.1001/jamaneurol.2022.4655

25 Fahrbach K, Huelin R, Martin AL, *et al.* Relating relapse and T2 lesion changes to disability progression in multiple sclerosis: a systematic literature review and regression analysis. *BMC Neurol*. 2013;13:180. doi: 10.1186/1471-2377-13-180

26 Cohen M, Bresch S, Thommel Rocchi O, *et al.* Should we still only rely on EDSS to evaluate disability in multiple sclerosis patients? A study of inter and intra rater reliability. *Mult Scler Relat Disord*. 2021;54:103144. doi: 10.1016/j.msard.2021.103144

27 Scalfari A, Neuhaus A, Degenhardt A, *et al.* The natural history of multiple sclerosis, a geographically based study 10: relapses and long-term disability. *Brain*. 2010;133:1914–29. doi: 10.1093/brain/awq118

28 Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–60. doi: 10.1097/00001648-200009000-00011

Figure legends

Figure 1: overview of the study

1,455 Welsh MS patients were included, and we calculated 7 genomic risk scores based on 3 pivotal MS GWAS. Unsupervised hierarchical clustering (Ward D2) was applied on the genomic risk scores. 3 Genomic clusters of pwMS were identified and patients within those clusters were compared with regards to relevant clinical outcomes. The clustering was subsequently applied to the Dutch NBB-MS cohort and clinical outcomes were compared between the clusters.

Figure 2: genomic clustering differentiates genetic risk scores and shows comparable patterns between Welsh and NBB-MS cohort

Comparison of genomic risk scores between the genetic clusters, in every figure left boxplot represent Welsh MS patients (darker colour) and right boxplot Dutch MS patients (lighter colour). **A)** sum of risk alleles of susceptibility variants, **B)** sum of number of risk alleles of severity variants of IMSGC GWAS, **C)** sum of risk alleles of MSBase GWAS, **D)** weighted genomic risk score for susceptibility variants, **E)** weighted genomic risk scores for IMSGC progression GWAS, **F)** weighted genomic risk scores for MSBase GWAS and **G)** Human leukocyte antigen (HLA) genetic burden score. **H)** Heatmap of FDR adjusted p-values, left panel comparing within a genetic cluster the genomic scores between the Welsh and NBB-MS cohort and right panel ANOVA results comparing the genomic risk scores between the three genetic clusters for the pooled Welsh and NBB-MS cohort.

Figure 3: patients in genomic cluster 2 have significantly worse disease outcomes

Kaplan-Meier survival curves for time to **A)** EDSS 6, **B)** EDSS 8 and **C)** SPMS in the Welsh cohort. **D)** Kaplan Meier curve for time to EDSS 6 in the NBB-MS cohort. For all survival curves, the table below the graph shows the number of patients at risk and between brackets the number of patients censored at each time point and p-values represent log-rank p. **E)** Hazard ratios to develop several disability milestones, adjusted for gender, age at onset and use of DMT in the Welsh in MS cohort. **F)** Multi-level modelling to compare the changes in EDSS over time for Welsh pwMS between the different genetic clusters. **G)** Heatmap for proportion of donors with comorbidities comparing the genetic clusters in the Dutch cohort. Hazard ratios of genetic clusters compared to well-validated clinical risk factors; **H)** cerebellar syndrome or **I)** optic neuritis at disease onset. **J)** Multivariate Cox Proportional Hazard model for hazard ratio to develop EDSS 6, adjusted for sex, age at onset and use of DMT in the Welsh MS cohort.

Figure 4: patients in genetic cluster 2 have the best response to disease modifying treatments

Within each genomic cluster, we compared time to disability milestones between patients who ever received a DMT versus patients who have never been treated, and calculated hazard ratios (HR) for time to **A)** EDSS 6, **B)** EDSS 8 and **C)** SPMS. All reported p-values are log-rank p.

Figure 5: genetic cluster 1 patients have the most favourable disease course

Natural language processing was applied to medical records of the Dutch cohort. **A)** We compared the number of observations of a specific symptom (first column of the heatmap), the age occurrence of the symptoms (second column) and differences in event free survival (third column), all p-values are FDR corrected for multiple testing. * Denotes a statistically significant difference, + denotes a borderline non-significant finding after FDR adjustment. Differences in age at onset were compared between the three different clusters for **B)** changes in mood, **C)** development of muscle spasticity, **D)** and swallowing difficulties. All p-values are ANOVA p-values adjusted for multiple testing. Survival analysis for time **E)** difficulties with concentration and **F)** reduced hearing. The reported p-values are FDR adjusted log-rank p-values. The table below the survival curves indicated number of patients at risk and number of censored observations per time point.

Table 1: Characteristics at disease onset similar between genomic clusters

	South Wales cohort (n=1455)			p-value	Dutch cohort (n=272)			p-value
	Cluster 1 (n=750)	Cluster 2 (n=277)	Cluster 3 (n=446)		Cluster 1 (n=145)	Cluster 2 (n=38)	Cluster 3 (n=84)	
Age at onset (mean, SD)	33.0 (10.9)	33.5 (10.8)	33.1 (10.6)	0.76	32 (15.5)	32 (20)	32.5 (14)	0.80
Percentage female	70.6%	71.1%	67.9%	0.53	67.6%	68.4%	63.1%	0.37
Months between first and second relapse[1] (median, IQR)	22 (60)	23.3 (59.9)	61.6 (47)	0.25	-	-	-	-
Oligoclonal bands positive/number tested (positive %)	367/449 (81.7%)	142/167 (85.0%)	233/282 (82.6%)	0.63	-	-	-	-
Percentage PPMS	10.9%	10.3%	9.2%	0.67	22.1%	31.6%	20.2%	0.37
First EDSS after onset (median, IQR)	3.5 (3)	3.5 (3.5)	3.5 (3.5)	0.97	-	-	-	-
Disease modifying treatments (number, % of total)					-	-	-	-
Never	546 (72.8%)	208 (75.1%)	323 (72.4%)	0.95				
Only moderate efficacy	126 16.8%)	47 (17.0%)	80 (17.9%)					
Only high efficacy	35 (4.7%)	10 (3.6%)	20 (4.5%)					
Switch between moderate and high efficacy	43 (5.7%)	12 (4.3%)	23 (5.2%)					
Age at death [2] (mean, SD)	n=76 63.3 (12.4)	n=35 59.5 (13.0)	n=55 63.0 (9.9)	0.38	64.5(12.9)	62.1 (13.9)	61.8 (11.2)	0.12

[1] Relapsing onset patients only, Cohen's d comparing different clusters negligible (max 0.17); [2] ANOVA

Table 2: Welsh MS patients in genomic cluster 2 and 3 have a significantly higher annual increase in T2 lesion load on serial MR imaging

	Cluster 1	Cluster 2	Cluster 3	p-value
≥9 T2 lesions at baseline (%)	57.8	29.8	53.3	0.61
Patients developing new T2 lesions	21.7%	16.9%	22.5%	0.062
Annual increase in T2 lesions mean (SD)	**1.29 (6.45)**	**1.71 (19.9)**	**2.43 (34.2)**	**0.04**