Title

MSPrecise: A molecular diagnostic test for multiple sclerosis using next generation sequencing

Journal

Gene, 572(2)

ISSN

0378-1119

Authors

Rounds, William H
Salinas, Edward A
Wilks, Tom B
et al.

Publication Date

2015-11-01

DOI

10.1016/j.gene.2015.07.011

Peer reviewed

HHS Public Access

Author manuscript

Gene. Author manuscript; available in PMC 2016 November 10.

Published in final edited form as:

Gene. 2015 November 10; 572(2): 191–197. doi:10.1016/j.gene.2015.07.011.

MSPrecise: a molecular diagnostic test for multiple sclerosis using next generation sequencing

William H. Rounds[1], **Edward A. Salinas**[2], **Tom B. Wilks II**[6], **Mikhail K. Levin**[2], **Ann J. Ligocki**[1], **Carolina Ionete**[4], **Carlos Pardo-Villamizar**[5], **Steven Vernino**[1], **Benjamin M. Greenberg**[1], **Douglas W. Bigwood**[6], **Eric M. Eastman**[6], **Lindsay G. Cowell**[2], and **Nancy L. Monson**[1,3,*]

[1]Department of Neurology and Neurotherapeutics, UT Southwestern Medical Center, Dallas, TX, USA

[2]Department of Clinical Sciences, UT Southwestern Medical Center, Dallas, TX, USA

[3]Department of Immunology, UT Southwestern Medical Center, Dallas, TX, USA

[4]Department of Neurology, UMass Memorial Medical Center, Worcester, MA, USA

[5]Department of Neurology and Neurosurgery, John Hopkins University, Baltimore, MD, USA

[6]DioGenix Inc., Bethesda, MD, USA

Abstract

Background—We have previously demonstrated that cerebrospinal fluid-derived B cells from early relapsing-remitting multiple sclerosis (RRMS) patients that express a *VH4* gene accumulate specific replacement mutations that can be quantified as a score that identifies such patients as having or likely to convert to RRMS. Furthermore, we showed that next generation sequencing is an efficient method for obtaining the sequencing information required by this mutation scoring tool, originally developed using the less clinically viable single-cell Sanger sequencing.

Objective—To determine the accuracy of **MS***Precise*, the diagnostic test that identifies the presence of the RRMS-enriched mutation pattern from patient cerebrospinal fluid B cells.

Methods—Cerebrospinal fluid cell pellets were obtained from RRMS and other neurological disease (OND) patient cohorts. *VH4* gene segments were amplified, sequenced by next generation sequencing and analyzed for mutation score.

*Corresponding author: Dr. Nancy L. Monson, UT Southwestern Medical Center, Dept. of Neurology and Neurotherapeutics, 6000 Harry Hines Blvd., NL9.110E, Dallas, TX, 75390-8813, USA, nancy.monson@utsouthwestern.edu.

Conflict of interest statement: WHR, EAS, MKL, AJL, CI, CPV, SV, LGC and NLM have no competing interests.
DWB and EME are employees of DioGenix and, as such, they each own equity shares in the company and are inventors on pending patents relating to the work described in this publication.
BMG owns equity shares in DioGenix.

Results—The diagnostic test showed a sensitivity of 75% on the RRMS cohort and a specificity of 88% on the OND cohort. The accuracy of the test in identifying RRMS patients or patients that will develop RRMS is 84%.

Conclusion—MS*Precise* exhibits good performance in identifying patients with RRMS irrespective of time with RRMS.

Keywords

Multiple sclerosis; neurological disease; Biomarker; B cell; Genetics; High-throughput nucleotide sequencing

1. Introduction

Multiple sclerosis (MS) is a demyelinating autoimmune disease of the central nervous system (CNS). Several studies have underscored the impact of T and B cells in this disease and have broadened the community's search for more effective immunomodulatory therapies for the treatment of relapsing-remitting MS (RRMS). For example, early evidence for a role of B cells in the pathoetiology of MS, including oligoclonal bands,[1, 2] altered antibody genetics [3-5] and B cell responses to neuroantigens in vitro [6, 7] provided the basis for use of Rituximab, a B cell depleting antibody for the efficacious treatment of RRMS.[8,9]

A number of reports consistently demonstrate that B cells in the CNS of RRMS patients undergo extensive clonal expansion,[2,10-12] and in some cases, recognize neuroantigens. Our laboratory hypothesized that since antigen-driven B cell selection is dependent on somatic hypermutation (SHM) accumulation in antibody genes, the cerebrospinal fluid (CSF)-derived B cell pool of RRMS patients would be enriched for a unique pattern of SHM reflecting their potential to recognize neuroantigens. Since variable heavy chain family 4 (*VH4*) genes are enriched in RRMS patient CNS,[3, 4, 13-17] this gene family was examined for patterns of SHM. Indeed, we have demonstrated and confirmed that CSF-derived B cells from RRMS patients expressing rearranged variable heavy chain family 4 (*VH4*) genes have an exaggerated accumulation of replacement mutations at 6 codon positions.[18, 19]

Earlier studies of this SHM pattern used a pool of memory B cells isolated from healthy donor peripheral blood (N=2) to establish baseline SHM accumulation at each codon position. Our next goal was to compare the SHM pattern identified in MS patients with CSF B cell antibody repertoires from patients with other neurological diseases (OND). However, these early studies included comparison to only 3 OND patients. Thus, further confirmation is required regarding the specificity of SHM accumulation at these codon positions in B cells from RRMS patients and a larger OND cohort. In addition, the majority of patients analyzed in the previous two studies were patients who were very early in their disease (N=17/19). Thus, it is unclear whether established RRMS patients who meet the McDonald criteria for RRMS[20] have the same exaggerated accumulation of SHM at these codon positions.

To address these issues, we analyzed the *VH4* antibody gene repertoires in CSF cell pellets from 26 patients with OND and 13 patients with confirmed RRMS using next generation sequencing (NGS). Our results indicate that RRMS patients exhibited the expected pattern

of SHM at these codon positions. In addition, 23/26 OND patients did not appreciably accumulate SHM at these codon positions or displayed insufficient sequence data indicative of low B cell abundance in the CSF.

2. Material and methods

2.1 Patient description and CSF sample preparation

CSF cell pellets were collected from 26 OND patients and 13 patients with confirmed or possible RRMS (Supplementary Tables 1&2). All CSF samples were collected by lumbar puncture in accordance with IRB-approved protocols at UT Southwestern Medical Center, the University of Massachusetts Memorial Medical Center (UMass), John Hopkins University (JHU), or purchased from a commercial biorepository (PrecisionMed, Solana Beach, CA). See Supplementary Method 1.2 for additional sample processing information.

2.2 PCR and next generation sequencing of antibody genes from CSF-derived B cell pools

All PCR reactions and sequencing were performed as previously published with modifications made to account for usage of gDNA (see Supplementary Method l.l).[19] Of note, only *VH4* amplifications were performed for this analysis since the unique SHM accumulation was identified only in this family.

2.3 NGS 454 data processing

Each raw sequence was analyzed using the VDJserver online repertoire analysis tool (https://vdjserver.org/). Unique reads were identified and filtered as detailed in Supplementary Method 1.2.

2.4 Mutation analyses

Mutation analyses were performed as previously published[19] and as detailed in Supplementary Method 1.3.

2.5 Statistical analyses

Statistical analyses were done using GraphPad Software 6.00 (San Diego, California, USA, www.graphpad.com). Specific tests for each comparison are detailed in Supplementary Method 1.4.

3. Results

For this study, we generated *VH4* antibody repertoires using NGS of CSF cell pellets isolated from 39 patients (Table 1). Of the 39 patient-derived CSF cell pellets, 13 were from patients with confirmed or possible RRMS, and 26 were from patients with OND. 14 patient samples (1 RRMS and 13 OND) were excluded due to recovery of insufficient sequence reads after sequence filtering (Tables 2&3). A pool of purified CD19+CD27- naïve B cells from peripheral blood of one healthy donor (run in 10 replicates) was included as a sequencing control for 454 error rates and as a control for random *VH4* gene usage in the naïve B cell pool.

We first determined how a series of process and analytical modifications made since previous analyses affected sequence coverage (Supplementary Methods 1.1 and 1.2).[19] One modification was to include only unique sequences that had two or more copies after sequence filtering (redundancy 1) in an attempt to increase our confidence that the sequences being analyzed were representative of the B cell pool and not a result of sequence errors generated during either PCR amplification or NGS. We compared the sequence coverage obtained with redundancy filter (Rl) and without (R0) (Table 4). The previously published dataset had an average of 2,426 unique sequences per RRMS sample at R0 and an average of 583 sequences per RRMS sample at Rl. The current dataset had an average of 751 sequences for the RRMS samples and an average of 632 sequences for the OND samples at Rl (Table 4). This resulted in a 1.3-fold increase per RRMS patient in the number unique sequences in CSF-derived antibody repertoires using our current method. The healthy control naïve (HCN) cohort had an average of 1,363 sequences per sample, which resulted in 2.5-fold more coverage in the peripheral HCN B cell pools in comparison to all CSF B cell pools, which likely relates to a larger initial pool of purified B cells.

Next, we sought to determine if the distributions of variable heavy chain family 4 (*VH4*) gene segments in each cohort were comparable (Figure la). The *VH4* gene distributions differed significantly between all pairs of cohorts with some pairs being more divergent than others. The RRMS *VH4* gene distribution was most distinct relative to the other two cohorts (Chi-squared value = 5652 for RRMS versus HCN; 3741 for RRMS versus OND), while the OND and HCN distributions were more similar (Chi-squared value = 2114). As expected,[21] the usage frequency of *VH4* genes in the HCN B cell pool was comparable to a uniform distribution of 12.5% for each individual gene (Chi-squared value = 4665), with an underrepresentation of *VH4-4* (percentage deviation = -81%) and an overrepresentation of *VH4-b* (percentage deviation = 119%) contributing most to the overall Chi-squared value. Similarly, for the OND cohort, deviation from a uniform distribution of gene usage is primarily due to one or two genes, with underrepresentation of *VH4-31* showing the largest deviation (percent deviation = -96%). In contrast, the RRMS cohort was very different from a uniform distribution (Chi-squared value = 7804) and utilized *VH4-39* (percentage deviation = 190%) and *VH4-59* (percentage deviation = 105%) more frequently than expected, which others have previously observed for *VH4-39*.[13, 14]

The distribution of joining heavy chain (JH) gene segments in naïve B cells is heavily skewed towards *JH4* usage.[21] Indeed, the healthy donor peripheral naïve B cell pools in the current dataset demonstrated skewing towards *JH4* usage (Figure lb). However, the RRMS cohort for this dataset had a JH usage rank of 5>6>4>2>1=3. The high usage of *JH5* and *JH6* gene segments was unexpected and contrasted with the previous dataset where *JH4* was maintained as the most frequently used JH gene segment in the RRMS cohort.[19] Further investigation confirmed that 8 of the 12 RRMS patients had unusually high skewing towards *JH5* or *JH6* usage, which resulted in an unexpected JH usage rank in the cohort. Thus, the overall distribution of JH gene segments in the RRMS cohort was significantly different from that of the HCN cohort (Chi-squared value = 2416). The OND cohort had a JH gene segment usage rank of 4>5>6>1=3>2, which more closely followed the JH rank of the HCN B cell repertoire (Chi-squared value = 1791).

We next determined whether the RRMS and OND cohorts from this dataset had accumulated SHMs into the variable regions of their antibody genes as established in the literature by calculating both the overall mutation frequency (MF), which considers all nucleotide substitutions, and the replacement mutation frequency (RMF), which considers only amino acid substitutions (Figure 2). Whereas the HCN B cell pools had very low MFs (median 1.9%) as expected from a naïve B cell population with low background sequencing error, the RRMS and OND cohorts had very high MFs (medians 6.7% for RRMS and 3.4% for OND), demonstrating that CSF B cells accumulate SHMs at a high frequency as previously published.[22] Interestingly, the MF of the RRMS and OND cohorts were not significantly different (p=0.50). The RMF calculations demonstrate a similar result (i.e. high and comparable RMF in the RRMS and OND CSF cohorts compared to the peripheral HCN). No correlation was found between patient age and RMF for either cohort (RRMS p=0.8; OND p=0.2). Proper targeting of these mutations to the hypervariable regions within the complementarity determining regions (CDRs) was also confirmed (Figure 2b).

Next, we compared the RMF at each codon position in the 6 codons that we originally used to calculate antibody gene signature (AGS) scores (31B, 40, 56, 57, 81, 89)[18]. The RMF at codons 31B, 40, 56, and 57 were all statistically greater in the RRMS cohort compared to the OND cohort (Table 5). However, the RMF at codons 81 and 89 were statistically greater in the OND cohort compared to the RRMS cohort. In fact, codon 89 had the lowest RMF of all 6 AGS codons in the RRMS cohort (9.3%), and thus contributed the least to scores for the RRMS cohort combined as well as for individual patients.

Finally, we calculated **MS***Precise* scores for all 25 patient CSF samples (Figure 3), excluding codon 89 in the calculations due to its low impact on scores for the RRMS cohort. As expected, the RRMS samples had a median **MS***Precise* score of 10.6 and IQR of 5.7 to 17.7. The OND samples had a median **MS***Precise* score of 4.5 and IQR of -3.3 to 11.7. Thus, the **MS***Precise* scores of the RRMS cohort were statistically higher than the **MS***Precise* scores of the OND cohort (p=0.05). The HCN cohort had very consistent and low **MS***Precise* scores as expected for a sequencing control that demonstrates non-targeted background sequence error, with a median score of -0.6 and an interquartile range (IQR) of -1.1 to 0.6.

As expected, 10 of 13 OND patients had **MS***Precise* scores below the previously established threshold of 6.8. However, the 6.8 threshold was based on Sanger sequencing data and NGS sequences have a low level of background RMs which tends to lower **MS***Precise* scores. Therefore, we identified an alternative threshold of 5.8 where we would expect to find some NGS samples with **MS***Precise* scores above but close to the threshold by Sanger sequencing. This new threshold did not affect the number of OND patients that had **MS***Precise* scores low enough to be properly identified. Four of the OND patients had **MS***Precise* scores just below the 5.8 **MS***Precise* threshold, and six of them had negative **MS***Precise* scores. There was no correlation between diagnoses of the OND patients and their **MS***Precise* scores (Supplementary Tables 1&2).

MS*Precise* scores for 9 of the 12 RRMS patients were above the **MS***Precise* threshold of 5.8 and included 2 patients who were on interferon beta-1a (one for 9 months, MS05, and

one for 2 years, MS07), one patient who was on glatiramer acetate for 5 years (MS04) and one patient who was on mycophenolic acid for 7 years (MS06). All four of the patients diagnosed with RRMS who were oligoclonal banding (OCB) negative had **MS***Precise* scores above the threshold (scores = 33.2; 10.0; 26.8; 7.5), two of which were on disease-modifying therapies (DMT) (MS06, MS07). Of the three RRMS patients who were OCB positive, two had **MS***Precise* scores below the threshold (scores = -3.5 and 6.6), but had been sampled while on DMTs (MS02, steroids; MS05, interferon beta-1a). One OCB positive RRMS patient who was not on DMT at the time of sampling had an **MS***Precise* score above the threshold (score = 15.2).

No correlations were found between **MS***Precise* score and age or mutation frequency (Supplementary Figure1). There was a trend towards higher diversity in *VH4* gene usage (termed “diversity index”) for RRMS patients with low **MS***Precise* scores (Figure 4a), which did not correlate with sequence read count (Supplementary Figure2). The two RRMS patients that had high diversity indices and low **MS***Precise* scores were MS08 (diversity index = 1.10; score = 5.37) and MS10 (diversity index = 1.22; score = -1.51). The OND cohort did not display any correlation of **MS***Precise* score with the diversity index (Figure 4b), even though the diversity index for the RRMS and OND cohorts were not statistically different (Figure 4c; p=0.6). The HCN cohort displayed a high diversity index that was statistically different from both the RRMS and OND cohorts ($p<0.0001$ for both) as expected from a large peripheral B cell pool compared to CSF B cells (Figure 4c).

4. Discussion

The application of antibody genetics to human disease has begun to emerge rapidly, particularly since NGS became readily available. Indeed, the power of this technology has been applied to monitoring minimal residual disease in cases of B cell lymphomas[23], and establishing that CSF-derived B cell clones matriculate from the periphery.[12] Our application of NGS has been to develop a new approach to identify patients with clinically isolated syndrome (CIS) who are at high risk of converting to fulminant MS. Indeed, our early work using Sanger DNA sequencing methods demonstrated that AGS scoring identified CIS patients who later converted to definite RRMS with 91% accuracy.[18]

However, four questions remained. First, was the accumulation of SHM in these codons specific to MS patients? Second, would established RRMS patients that meet the revised McDonald criteria[20] have a similar pattern of SHM as early-stage patients? Third, does OCB status affect the score? Fourth, does treatment with immunomodulatory drugs affect the score? To address these issues, we generated antibody gene repertoires from CSF-derived B cells of ONDs, OCB^+ and OCB^- RRMS patients as well as treatment-naïve RRMS patients and RRMS patients who had been on DMTs for more than a year.

We obtained CSF cell pellets from 26 OND patients with a variety of diagnoses including headache (n=6), paraneoplastic disease (n=4) and others (Supplementary Tables 1&2). Of the 26 OND patients, 13 were excluded from analysis due to a very low number of sequence reads. Since this primarily occurred in the OND cohort, we concluded that those 13 OND patients either did not display an expanded B cell mediated CNS immune response that we

could detect, or that the response was negligible. In either case, the inability to recover antibody sequences from such samples is likely indicative of a lack of B cell recruitment and confirms why the literature is limited in the area of antibody genetics in patients with non-inflammatory neurological diseases. In fact, there was one RRMS patient with insufficient reads that we did not include in the present cohort because this patient had been on natalizumab for more than 4 years, a well-known drug that prevents B cells and other lymphocytes from entering the CNS.[24]

Our ability to detect antibody genes of rare B cells by PCR might provide OND samples an advantage and result in an **MS***Precise* score that might not properly reflect their OND status. In addition, low antibody sequence reads might be indicative of their OND status. Indeed, of the 14 samples we removed based on recovery of an insufficient number of unique sequence reads, 13 of them were within the OND cohort. If we assigned such samples the lowest **MS***Precise* score possible (**MS***Precise* score = -8.9), and inserted them back into the OND cohort, the median **MS***Precise* score of the OND group decreases to -8.9 (Supplementary Figure 3).

In those 13 OND cases where we were able to recover a sufficient number of unique antibody sequences from CSF-derived cells, we observed that the accumulation of replacement mutations was slightly lower than in the RRMS patients, but not significantly different (OND, median RMF 6.5; RRMS, median RMF 9.9; p=0.5). In addition, the distribution of *VH4* gene segments in the OND cohort did not differ significantly from the expected random frequency. JH gene segment usage was also no different from the expected frequency established in naïve B cell pools. This suggests that in the OND cases for which CSF B cells can be detected, antigen-driven selection is not as prominent as it is in RRMS patients.

There is very little available information regarding the impact of DMTs on numbers or types of B cells found in the CSF of RRMS patients. Even in the case of B cell-depleting monoclonal antibodies, such as Rituximab, our understanding of B cell dynamics in the CSF is limited.[25, 26] Nevertheless, the RRMS cohort used for this study included 4 patients on DMTs for an extended period of time, most of which had high **MS***Precise* scores regardless of OCB status. The one RRMS patient who had been on steroids for 7 days at the time of sampling had a negative **MS***Precise* score. It is difficult to make conclusions based on these small samples, but these data suggest that the clinical benefit of many immunomodulatory drugs used to treat RRMS, including the beta-interferons and glatiramer acetate, is independent of the CSF B cell pool. Further study is warranted to determine if particular DMTs impact the CSF B cell pool and **MS***Precise* scores.

Finally, there is an increasing need for new methods to determine whether a patient has MS or not.[27] **MS***Precise* scoring may be one supportive approach to aid clinicians in this task. Indeed, if we include the OND samples with insufficient reads, the specificity of identifying patients with OND based on **MS***Precise* scoring is 88%. The sensitivity of this test in identifying RRMS patients is 75%, although the impact of DMTs and steroids on the **MS***Precise* scoring system for our RRMS cohort remains unclear. This puts the overall accuracy of **MS***Precise* scoring in this study at 84% if samples with insufficient reads are

included and 76% if they are omitted. Previously, we presented data generated using Sanger DNA sequencing suggesting that **MS***Precise* scoring is able to identify CIS patients who will convert to RRMS but who are not yet on immunomodulatory therapy with 91% accuracy. [18] Determining whether **MS***Precise* scoring using NGS performs as well to identify CIS patients who will convert to RRMS will be the subject of future investigations. More work also needs to be done to determine whether the codons we used to calculate **MS***Precise* scores are still appropriate on the NGS platform which will require a larger patient cohort with preferably several sub-cohorts of RRMS patients on particular DMTs and OND patients of a particular diagnosis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank the patients who provided samples for this study. We thank Carolyn Griffin for coordinating sample collection at the University of Massachusetts. We also thank Dr. Lee Szkotnicki and SeqWright Genomic Services (a GE company, Houston, TX) for performing PCR amplification and sequencing to generate the NGS database. We also want to thank Dr. Mark Erlander (Trovagene, San Diego, CA) for his critical reading of the manuscript.

Funding acknowledgements: This study was supported by grants from the National Multiple Sclerosis Society to NLM [RG3267, RG4653); DioGenix, Inc. to CI, CPV and NLM; the NIH to AJL and WHR [NIH NRSA5 T32 A1 005284-28 from NIAID]; NIAID to LGC [R01AI097403] and the Burroughs Wellcome Fund to LGC [Career Award].

This research was supported in part by a Sponsored Research Agreement that DioGenix has with UTSWMC.

References

1. Andersson M, Alvarez-Cermeno J, Bernardi G, et al. Cerebrospinal fluid in the diagnosis of multiple sclerosis: a consensus report. Journal of neurology, neurosurgery, and psychiatry. 1994; 57:897–902.
2. Krumbholz M, Derfuss T, Hohlfeld R, Meinl E. B cells and antibodies in multiple sclerosis pathogenesis and therapy. Nature reviews Neurology. 2012; 8:613–23. [PubMed: 23045237]
3. Harp C, Lee J, Lambracht-Washington D, et al. Cerebrospinal fluid B cells from multiple sclerosis patients are subject to normal germinal center selection. Journal of neuroimmunology. 2007; 183:189–99. [PubMed: 17169437]
4. Owens GP, Winges KM, Ritchie AM, et al. VH4 gene segments dominate the intrathecal humoral immune response in multiple sclerosis. J Immunol. 2007; 179:6343–51. [PubMed: 17947712]
5. Bennett JL, Haubold K, Ritchie AM, et al. CSF IgG heavy-chain bias in patients at the time of a clinically isolated syndrome. Journal of neuroimmunology. 2008; 199:126–32. [PubMed: 18547652]
6. Antel J, Bar-Or A. Roles of immunoglobulins and B cells in multiple sclerosis: From pathogenesis to treatment. Journal of neuroimmunology. 2006; 180:3–8. [PubMed: 16934338]
7. Lambracht-Washington D, O'Connor KC, Cameron EM, et al. Antigen specificity of clonally expanded and receptor edited cerebrospinal fluid B cells from patients with relapsing remitting MS. Journal of neuroimmunology. 2007; 186:164–76. [PubMed: 17451814]
8. Hauser SL, Waubant E, Arnold DL, et al. B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. N Engl J Med. 2008; 358:676–88. [PubMed: 18272891]
9. Kappos L, Li D, Calabresi PA, et al. Ocrelizumab in relapsing-remitting multiple sclerosis: a phase 2, randomised, placebo-controlled, multicentre trial. Lancet. 2011; 378:1779–87. [PubMed: 22047971]

10. Qin Y, Duquette P, Zhang Y, Talbot P, Poole R, Antel J. Clonal expansion and somatic hypermutation of V(H) genes of B cells from cerebrospinal fluid in multiple sclerosis. J Clin Invest. 1998; 102:1045–50. [PubMed: 9727074]
11. von Budingen HC, Harrer MD, Kuenzle S, Meier M, Goebels N. Clonally expanded plasma cells in the cerebrospinal fluid of MS patients produce myelin-specific antibodies. Eur J Immunol. 2008; 38:2014–23. [PubMed: 18521957]
12. von Budingen HC, Kuo TC, Sirota M, et al. B cell exchange across the blood-brain barrier in multiple sclerosis. J Clin Invest. 2012; 122:4533–43. [PubMed: 23160197]
13. Owens GP, Kraus H, Burgoon MP, Smith-Jensen T, Devlin ME, Gilden DH. Restricted use of VH4 germline segments in an acute multiple sclerosis brain. Ann Neurol. 1998; 43:236–43. [PubMed: 9485065]
14. Baranzini SE, Jeong MC, Butunoi C, Murray RS, Bernard CC, Oksenberg JR. B cell repertoire diversity and clonal expansion in multiple sclerosis brain lesions. J Immunol. 1999; 163:5133–44. [PubMed: 10528220]
15. Colombo M, Dono M, Gazzola P, et al. Accumulation of clonally related B lymphocytes in the cerebrospinal fluid of multiple sclerosis patients. J Immunol. 2000; 164:2782–9. [PubMed: 10679121]
16. Owens GP, Ritchie AM, Burgoon MP, Williamson RA, Corboy JR, Gilden DH. Single-cell repertoire analysis demonstrates that clonal expansion is a prominent feature of the B cell response in multiple sclerosis cerebrospinal fluid. J Immunol. 2003; 171:2725–33. [PubMed: 12928426]
17. Monson NL, Hans-Peter Brezinschek, Ruth I Brezinschek, et al. Receptor revision and atypical mutational characteristics in clonally expanded B cells from the cerebrospinal fluid of recently diagnosed mutiple sclerosis patients. Journal of neuroimmunology. 2005; 158:170–81. [PubMed: 15589051]
18. Cameron EM, Spencer S, Lazarini J, et al. Potential of a unique antibody gene signature to predict conversion to clinically definite multiple sclerosis. Journal of neuroimmunology. 2009; 213:123–30. [PubMed: 19631394]
19. Rounds WH, Ligocki AJ, Levin MK, et al. The antibody genetics of multiple sclerosis: comparing next-generation sequencing to Sanger sequencing. Frontiers in neurology. 2014; 5:166. [PubMed: 25278930]
20. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Annals of neurology. 2011; 69:292–302. [PubMed: 21387374]
21. Brezinschek HP, Brezinschek RI, Lipsky PE. Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction. J Immunol. 1995; 155:190–202. [PubMed: 7602095]
22. Monson NL, Cravens PD, Frohman EM, Hawker K, Racke MK. Effect of Rituximab on the peripheral blood and cerebrospinal fluid b cells in patients with primary progressive multiple sclerosis. Archives of Neurology. 2005; 62:258–64. [PubMed: 15710854]
23. Boyd SD, Marshall EL, Merker JD, et al. Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. Sci Transl Med. 2009; 1:12ra23.
24. Stuve O, Bennett JL. Pharmacological properties, toxicology and scientific rationale for the use of natalizumab (Tysabri) in inflammatory diseases. CNS Drug Rev. 2007; 13:79–95. [PubMed: 17461891]
25. Monson NL, Cravens PD, Frohman EM, Hawker K, Racke MK. Effect of rituximab on the peripheral blood and cerebrospinal fluid B cells in patients with primary progressive multiple sclerosis. Archives of neurology. 2005; 62:258–64. [PubMed: 15710854]
26. Evdoshenko E, Maslyanskiy A, Lapin S, et al. Dynamics of B-Cell Populations in CSF and Blood in Patients Treated with a Combination of Rituximab and Mitoxantrone. ISRN neurology. 2013; 2013:748127. [PubMed: 24109519]
27. Kroksveen AC, Opsahl JA, Guldbrandsen A, et al. Cerebrospinal fluid proteomics in multiple sclerosis. Biochimica et biophysica acta. 2014

Abbreviations

RRMS	relapsing-remitting multiple sclerosis
OND	other neurological disease
MS	multiple sclerosis
CNS	central nervous system
SHM	somatic hypermutation
CSF	cerebrospinal fluid
VH4	variable heavy chain family 4
NGS	next generation sequencing
HCN	healthy control naïve
JH	joining heavy chain
MF	mutation frequency
RMF	replacement mutation frequency
CDR	complementarity determining region
AGS	antibody gene signature
IQR	interquartile range
OCB	oligoclonal banding
DMT	disease-modifying therapy
CIS	clinically isolated syndrome

Highlights

- The diagnostic MSPrecise supports identification of multiple sclerosis patients.
- MSPrecise uses B cell antibody sequences from patient cerebrospinal fluid.
- MSPrecise performs well in identifying MS among a broad cohort of neurological diseases.

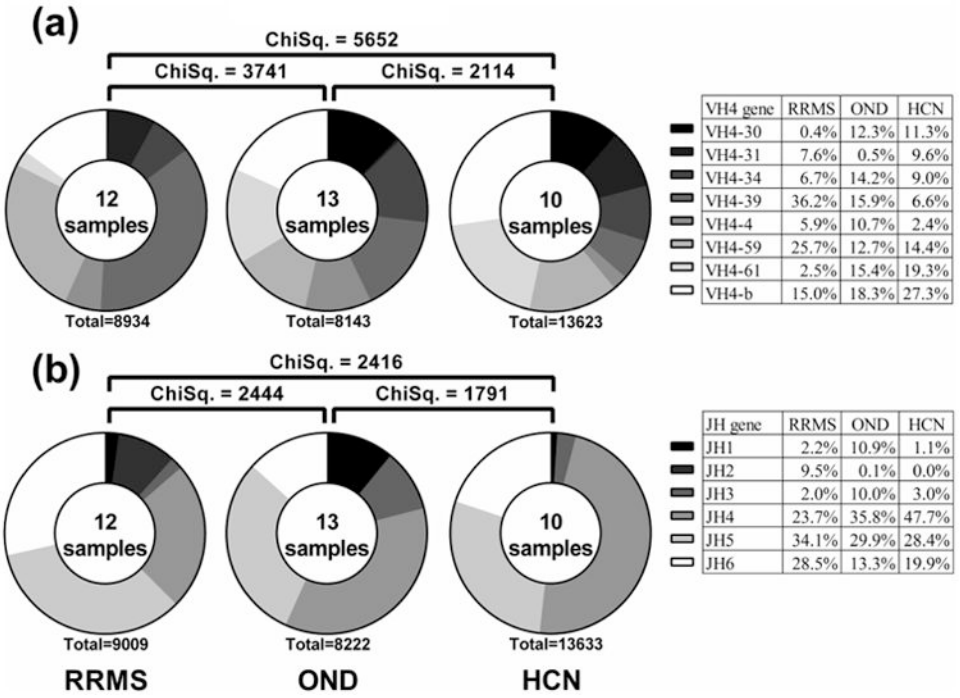

Figure 1. *VH4* and JH gene distributions of CSF B cells from RRMS patients are more divergent from healthy control naïve peripheral B cell repertoires than those from OND patients

VH4 (a) and JH (b) gene calls were obtained by IgBlast alignment (see methods). Total unique sequences used in cohort databases are indicated inside the pie charts. Chi-squared analysis values between cohort gene distributions are shown above the bars. Gene frequencies are shown in the table. Abbreviations: RRMS, relapsing-remitting MS; OND, other neurological disorder; HCN, healthy control naïve peripheral B cells. HCN samples are all replicates from a single patient.

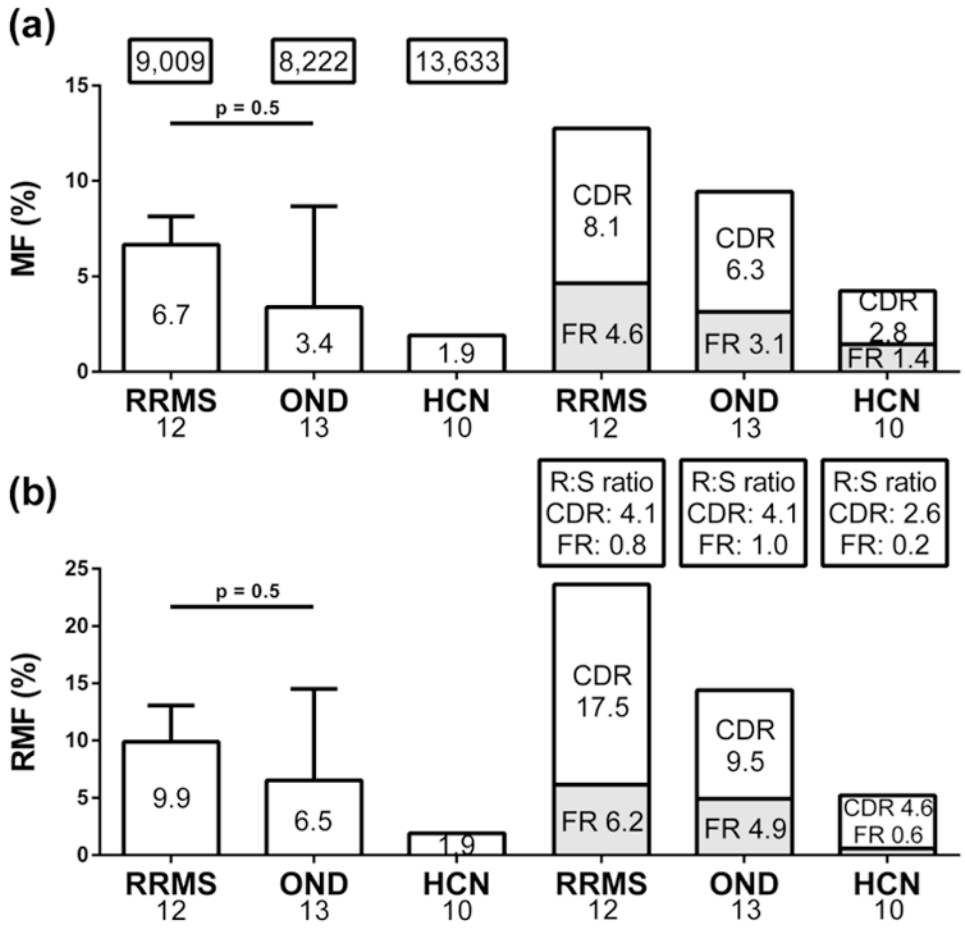

Figure 2. Mutation characteristics of *VH4* sequences in RRMS and OND patients

(a) Mutation frequency (MF) analysis was done by nucleotide; boxes indicate total unique sequences in each cohort and sample numbers are marked under cohort names. (b) Replacement mutation frequency (RMF) analysis was done by codon. RRMS sequence data includes 119,483 total point mutations and 62,749 total replacement mutations (RM); OND sequence data includes 74,769 total point mutations and 39,324 total replacement mutations (RM); RRMS sequence data includes 51,238 total point mutations and 17,375 total replacement mutations (RM). MF and RMF were calculated by sample and bar graphs show median (indicated on the bar graphs) and interquartile range (statistical significance of the difference between RRMS and OND was tested by Mann Whitney test). MF, RMF and R:S ratios for CDR and FR regions were calculated independently by region for each sample and are shown as cohort medians. HCN samples are all replicates from a single patient.

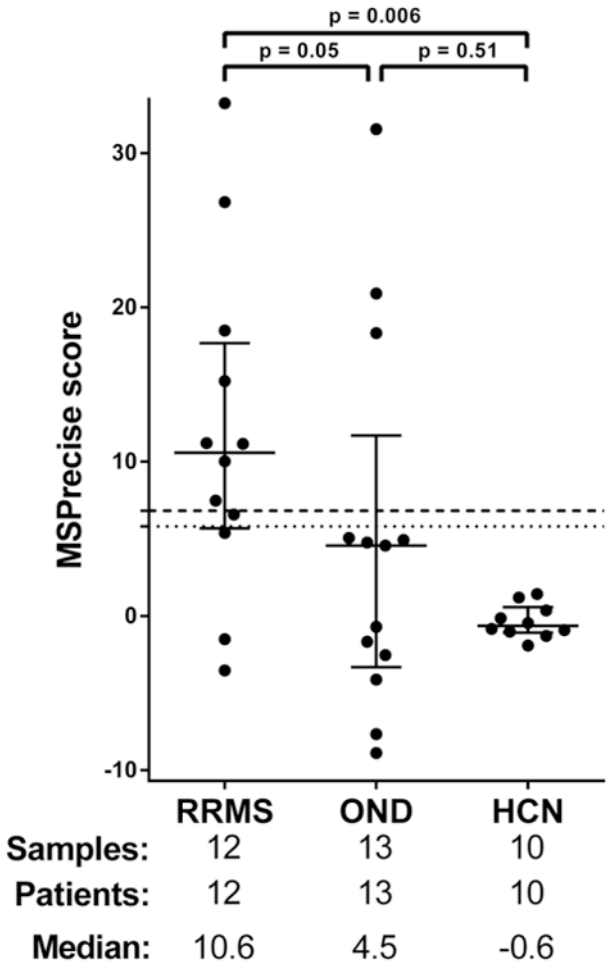

Figure 3. MS*Precise* scores in RRMS and OND patients

Each data point represents a single sample sequence pool (median and interquartile range are marked on the figure). The dashed line represents the **MS***Precise* cut-off point of 6.8 above which patients are expected to have or convert to relapsing-remitting MS (RRMS). The dotted line delineate an indeterminate range (-1) below the 6.8 cut-off where the results of the **MS***Precise* score test are less clear cut. Samples are grouped by most current diagnosis as RRMS, other neurological diseases (OND), and healthy control naïve (HCN). Only samples that pass our filtering criteria are displayed with their calculated **MS***Precise* scores. Statistical significance of the difference between cohorts was calculated by Mann Whitney test. HCN samples are all replicates from a single patient.

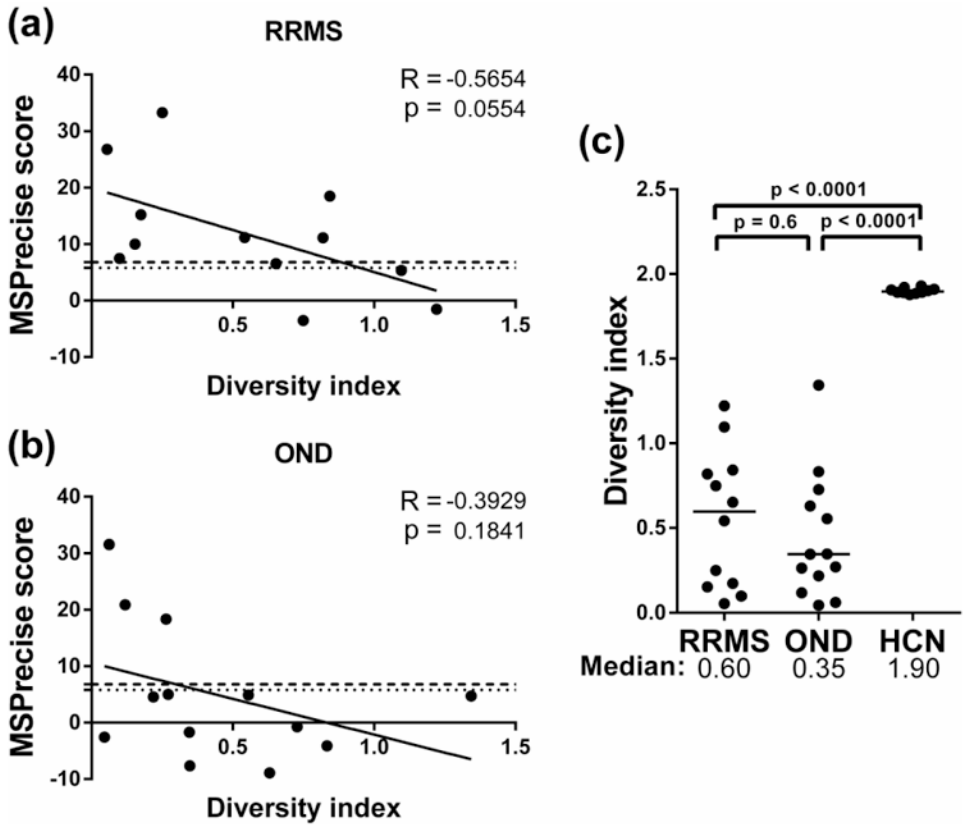

Figure 4. Low diversity correlates with high MS*Precise* score in the RRMS cohort but not in the OND cohort

Each data point represents a single sample sequence pool from (a) the RRMS cohort or (b) the OND cohort. The diversity index was calculated as described in the methods section and high values indicate a more even distribution across the *VH4* genes. Pearson's correlation coefficient (R) indicates the linear correlation between **MS***Precise* and the diversity index, and the two-tailed p-value of the correlation is also indicated. The dashed line represents the **MS***Precise* cut-off point of 6.8 above which patients are expected to have or convert to relapsing-remitting MS (RRMS). The dotted lines delineate an indeterminate range (-1) below the 6.8 cut-off where the results of the **MS***Precise* score test are less clear cut. (c) Distribution of the diversity index is shown here with the median marked on the graph. HCN samples are all replicates from a single patient. Statistical significance of the difference between cohorts was tested by Mann Whitney test.

Table 1

Filtering of samples by cohort.

Cohort name[a]	RRMS	OND	HCN[b]	TOTAL
Initial sample number	13	26	10	40
Samples with insufficient reads (<10 unique reads after filtering)	1[c]	13	0	14
Total analyzed	**12**	**13**	**10**	**26**

[a] Samples were grouped into patient cohorts by final diagnosis.

[b] Replicates from a single patient

[c] Patient on natalizumab at time of sampling

Abbreviations: RRMS, relapsing-remitting MS; OND, other neurologica disorder; HCN, healthy control naïve peripheral B cells.

Table 2

RRMS sample summary.

Patient ID	Age[a]	Gender	Diagnosis at tap	Diagnosis[b]	Time with MS[c]	Treatment[d]	OCB status[a]
MS01	35	F	RRMS	RRMS	NR	None	NR
MS02	27	F	RRMS	RRMS	1	steroids (7 days)	POS
MS03	39	F	RRMS	RRMS	0	None	NEG
MS04	26	F	RRMS	RRMS	76	glatiramer acetate (5 years)	NR
MS05	31	F	RRMS	RRMS	25	IFN-Bla (9 months)	POS
MS06	42	F	RRMS	RRMS	92	mycophenolic acid (7 years)	NEG
MS07	35	M	RRMS	RRMS	24	IFN-Bla (2 years)	NEG
MS08	36	F	Possible MS	RRMS	0	None	NR
MS09	31	M	RRMS	RRMS	0	None	NR
MS10	23	F	Possible MS	RRMS	0	None	NR
MS11	32	F	RRMS	RRMS	0	None	POS
MS12	58	F	Possible MS	RRMS	0	None	NEG

[a] At time of sampling (yrs)

[b] Most up-to-date available

[c] At time of sampling (months)

[d] If immunomodulatory and at time of sampling

Abbreviations: NR, not reported; OCB, oligoclonal bands; RRMS, relapsing-remitting MS.

Table 3

Non-RRMS sample summary.

Patient ID	Age[a]	Gender	Diagnosis[b]
OND01	37	M	OND
OND02	61	M	Dementia
OND03	54	M	Stroke
OND04	65	F	Dementia
OND05	52	F	Headache
OND06	48	F	Neurosarcoidosis
OND07	NR	F	Headache
OND08	57	F	PND
OND09	44	F	Encephalitis
OND10	67	F	PND
OND11	49	F	Urge incontinence
OND12	52	M	Alzheimer's
OND13	22	F	Headache
HCN	NR	NR	NA

[a] At time of sampling (yrs)

[b] Most up-to-date available

Abbreviations: NR, not reported; OND, other neurological disorder; PND, paraneoplastic neurologic disorder; HCN, healthy control naïve peripheral B cells; NA, not applicable.

Gene. Author manuscript; available in PMC 2016 November 10.

Table 4

Sequence yield per cohort.

		Redundancy 0[a]		Redundancy 1[b]			
Cohort name	**N**	**Total unique sequences[c]**	**Avg. unique sequences per sample**	**Total unique sequences**	**Avg. unique sequences per sample**	**Avg. RM Per sample**	**Avg. RM per sequence**
RRMS	12	28,489	2,374	9,009	751	5,229	7.0
OND	13	20,201	1,554	8,222	632	3,025	4.8
HCN	10[d]	93,204	9,320	13,633	1,363	1,738	1.3
Previous MS	7	16,984	2,426	4,082	583	5,466	9.4

[a] Unique sequences with any number of reads are included in the sequence database as with the previous study (labelled “Previous MS” in the table).[19]

[b] Filter used for this study. Unique sequences with at least two reads are included in the analysis database.

[c] After sequence filtering

[d] Replicates from a single patient

Abbreviations: RM, replacement mutation; RRMS, relapsing-remitting MS; OND, other neurological disorder; HCN, healthy control naïve peripheral B cells.

Table 5

AGS codon replacement mutation frequency relative to germline in RRMS and OND patients.

AGS codon	Location	RRMS RMF[a]	OND RMF	Fold higher in RRMS	p-value[b]
31B	CDR1	53.8%	38.5%	1.40	<0.001
40	FR2	16.4%	13.0%	1.26	<0.001
56	CDR2	33.6%	15.7%	2.14	<0.001
57	CDR2	21.1%	4.8%	4.42	<0.001
81	FR3	20.1%	27.0%	0.75	<0.001
89	FR3	9.3%	13.7%	0.68	<0.001

[a] Calculated relative to the total possible replacement mutations for each cohort (i.e. the number of reads that have a specific numbered codon in the germline)

[b] Calculated by Chi-squared test

Abbreviations: RMF, replacement mutation frequency; RRMS, relapsing-remitting MS; OND, other neurological disorder.