



# *In Silico* Prediction Analysis of Idiotope-Driven T–B Cell Collaboration in Multiple Sclerosis

Rune A. Høglund<sup>1,2\*</sup>, Andreas Lossius<sup>1,3</sup>, Jorunn N. Johansen<sup>3</sup>, Jane Homan<sup>4</sup>, Jūratė Šaltytė Benth<sup>2,5</sup>, Harlan Robins<sup>6</sup>, Bjarne Bogen<sup>2,3,7</sup>, Robert D. Bremel<sup>4†</sup> and Trygve Holmøy<sup>1,2†</sup>

<sup>1</sup>Department of Neurology, Akershus University Hospital, Lorenskog, Norway, <sup>2</sup>Institute of Clinical Medicine, University of Oslo, Oslo, Norway, <sup>3</sup>Faculty of Medicine, Department of Immunology and Transfusion Medicine, University of Oslo and Oslo University Hospital Rikshospitalet, Oslo, Norway, <sup>4</sup>EigenBio LLC, Madison, WI, United States, <sup>5</sup>Health Services Research Unit, Akershus University Hospital, Lorenskog, Norway, <sup>6</sup>Adaptive Biotechnologies, Seattle, WA, United States, <sup>7</sup>Centre for Immune Regulation, University of Oslo, Oslo, Norway

## OPEN ACCESS

### Edited by:

Zsolt Illes,  
University of Southern Denmark  
Odense, Denmark

### Reviewed by:

Lindsay B. Nicholson,  
University of Bristol,  
United Kingdom  
Chandrasegaran Massilamany,  
National Institutes  
of Health (NIH), United States  
ADI Vaknin-Dembinsky,  
Hadassah Medical Center, Israel

### \*Correspondence:

Rune A. Høglund  
r.a.hoglund@medisin.uio.no

<sup>†</sup>Joint senior authorship.

### Specialty section:

This article was submitted to Multiple Sclerosis and Neuroimmunology, a section of the journal *Frontiers in Immunology*

Received: 26 May 2017

Accepted: 20 September 2017

Published: 02 October 2017

### Citation:

Høglund RA, Lossius A, Johansen JN, Homan J, Benth JS, Robins H, Bogen B, Bremel RD and Holmøy T (2017) *In Silico* Prediction Analysis of Idiotope-Driven T–B Cell Collaboration in Multiple Sclerosis. *Front. Immunol.* 8:1255. doi: 10.3389/fimmu.2017.01255

Memory B cells acting as antigen-presenting cells are believed to be important in multiple sclerosis (MS), but the antigen they present remains unknown. We hypothesized that B cells may activate CD4<sup>+</sup> T cells in the central nervous system of MS patients by presenting idiotopes from their own immunoglobulin variable regions on human leukocyte antigen (HLA) class II molecules. Here, we use bioinformatics prediction analysis of B cell immunoglobulin variable regions from 11 MS patients and 6 controls with other inflammatory neurological disorders (OINDs), to assess whether the prerequisites for such idiotope-driven T–B cell collaboration are present. Our findings indicate that idiotopes from the complementarity determining region (CDR) 3 of MS patients on average have high predicted affinities for disease associated HLA-DRB1\*15:01 molecules and are predicted to be endosomally processed by cathepsin S and L in positions that allows such HLA binding to occur. Additionally, complementarity determining region 3 sequences from cerebrospinal fluid (CSF) B cells from MS patients contain on average more rare T cell-exposed motifs that could potentially escape tolerance and stimulate CD4<sup>+</sup> T cells than CSF B cells from OIND patients. Many of these features were associated with preferential use of the IGHV4 gene family by CSF B cells from MS patients. This is the first study to combine high-throughput sequencing of patient immune repertoires with large-scale prediction analysis and provides key indicators for future *in vitro* and *in vivo* analyses.

**Keywords:** multiple sclerosis, idiotope, B cell, T cell, bioinformatics, immunoglobulin heavy variable, immunosequencing, immunoglobulin

## INTRODUCTION

Multiple sclerosis (MS) is a chronic inflammatory, demyelinating, and neurodegenerative disease of the central nervous system (CNS), thought to be mainly mediated by the immune system (1). Although T cells as mediators of disease have been investigated thoroughly over the years, recent trials of B cell targeted therapies (i.e., rituximab and ocrelizumab) point to these cells as equally

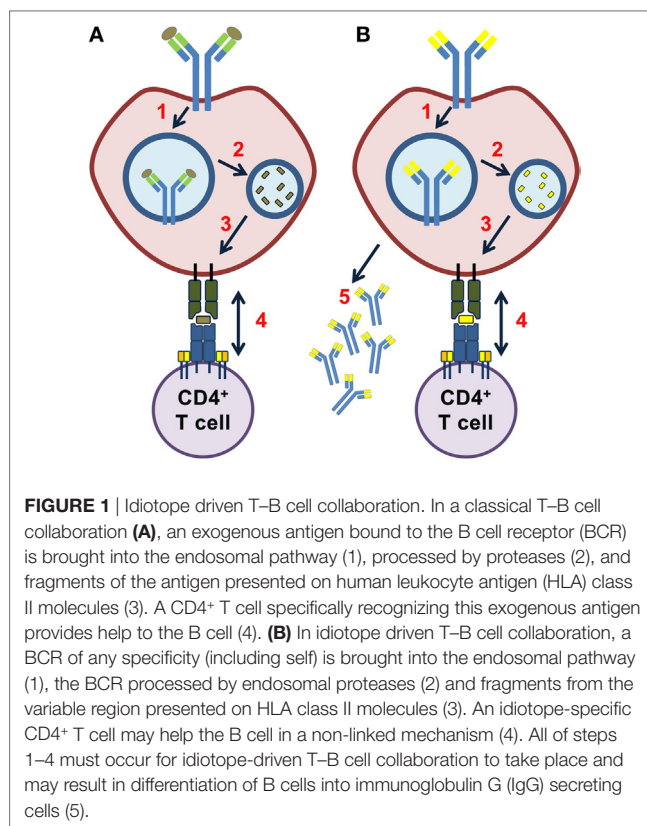
**Abbreviations:** APC, antigen-presenting cell; BCR, B cell receptor; CDR, complementarity determining region; mAb, monoclonal antibody; c/mTEC, cortical/medullary thymic epithelial cell; CI, confidence interval; CNS, central nervous system; CSF, cerebrospinal fluid; FC, Frequency class; FW, framework; HLA, human leukocyte antigen; IC<sub>50</sub>, the half maximal inhibitory concentration; IgG, immunoglobulin G; IGHV, immunoglobulin heavy variable; MHC, major histocompatibility complex; MS, multiple sclerosis; OIND, other inflammatory neurological disease; TCEM, T cell-exposed motif.

important contributors (2, 3). Notably, depleting B cells in the periphery has a substantial effect within the CNS (4). It is also possible that other approved therapies for MS act by depleting or prohibiting CD19<sup>+</sup>, CD27<sup>+</sup> memory B cells from invading the CNS (5). In MS, B cell immunoglobulin heavy chain variable (IGHV) repertoires suggest that clonally expanded plasma cells in the brain and cerebrospinal fluid (CSF) are derived from peripheral B cells that have matured in cervical lymph nodes (6–8). Hence, it seems that peripheral memory B cells play an important role in MS immunopathology.

The mechanisms by which memory B cells induce pathology could involve antibody production or secretion of cytokines (9). However, B cell-depleting therapies targeting CD20 ameliorate disease before reducing immunoglobulin G (IgG) production (10), which is therefore not likely their main mechanism. Whereas the discovery of intrathecal Ig production in the CNS is an old one (11), a common antigenic determinant has yet to be discovered. However, B cells in MS lesions (12) and CSF (13–17) have evidently undergone somatic hypermutation indicating T cell help, suggestive of a possible antigen being involved with B cells as antigen-presenting cells (APC) (18). We have proposed an alternative hypothesis to explain how T–B cell collaboration in absence of a common antigen can result in intrathecal IgG production (19). It was shown that B cells present endogenously processed variable region fragments (idiotopes) on major histocompatibility complex (MHC) class II molecules (20, 21). T cells can specifically recognize this idiotope–MHC complex, resulting in a T cell response (22, 23). Such an interaction between idiotope<sup>+</sup> B cells that present idiotope–MHCII and idiotope-specific CD4<sup>+</sup> T cells is named idiotope-driven T–B collaboration (21, 24, 25) (**Figure 1**). An important feature of idiotope-driven T–B collaboration is that unlike conventional antigen-linked T–B collaboration (26, 27), it is unlinked in the sense that while the B cell can recognize any (self) antigen, the T cell recognizes a different antigen (idiotope–MHCII). Thus, B cells of theoretically any specificity, including self-specificity, can be helped by idiotope-specific CD4<sup>+</sup> T cells to develop into IgG producing plasma cells (25). Consistent with this idea, endogenous idiotopes were eluted from MHC II molecules on B cells (28, 29), and idiotope-driven T–B cell collaboration has been shown to drive the development of autoimmune disease in transgenic mice (30).

Extending idiotope driven T–B collaboration to humans, we have previously demonstrated that human leukocyte antigen (HLA)-DR restricted CD4<sup>+</sup> T cells from blood and CSF of MS patients can recognize multiple idiotopes within the complementarity determining region 3 (CDR3) and mutated framework (FW) regions on autologous CSF IgG (31–33), showing that MS patients have a repertoire idiotope-matched T–B cell pairs. Idiotope-specific CD4<sup>+</sup> T cells specifically recognized idiotopes presented by autologous Epstein Barr virus transformed CSF B cells, suggesting that B cells can process and present their endogenous idiotopes on HLA class II molecules (33), and they are also induced to kill oligodendrocytes upon activation (34).

Further large-scale investigations into this mechanism in MS have been hampered by overwhelming numbers of possible IGHV region idiotopes. High-throughput sequencing now offers



a possibility to characterize the immune repertoire in unprecedented depth and detail (35) and has triggered a rapid growth of bioinformatic approaches for diagnostic and research purposes (36), including methods to assess possible immunogenicity of B cell variable region sequences (37).

There are several prerequisites for idiotope-driven T–B cell collaboration: the idiotopes would need to undergo endosomal processing; the processed idiotope fragments must have sufficient affinity for HLA class II molecules; and they must be sufficiently rare to avoid T cell tolerance (**Figure 1**). In this article, we combined high-throughput sequencing of the B cell receptor (BCR) transcriptome with *in silico* prediction analysis to assess whether these prerequisites exist in the intrathecal compartment of MS patients.

## MATERIALS AND METHODS

### Patients

In this study, we included 11 relapsing-remitting MS patients and six patients with other inflammatory neurological disorders (OINDs), recruited at Akershus University Hospital and Oslo University Hospital. All patients (except MS-11) and the procedures for PBMC isolation, RNA extraction and cDNA preparation have been described previously (7). The cDNA sequencing was performed by Adaptive Biotechnologies using the immunoSEQ level assay (38), with resulting 130bp sequences spanning the IGH-VDJ region.

Multiple sclerosis patients were either treatment naive or treated with first-line therapies (MS-2, MS-3, and MS-4), while OIND patients were untreated at the time of lumbar puncture. All MS patients and one OIND patient had oligoclonal IgG bands in CSF. A summary of patient and sample characteristics are shown in S1 in Supplementary Material.

All participants provided written informed consent for participation. The study was approved by the Committee for Research Ethics in the South-Eastern Norwegian Healthy Authority (REK Sør-Øst S-04143a), the Norwegian Social Science Data Services (no. 11069) and the review boards at AHUS and OUS.

Genotyping for HLA-A, HLA-B, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1, and HLA-DPB1 was performed with four-digit resolution at the Department of Immunology and Transfusion Medicine at Oslo University Hospital, by utilizing a combination of sequence-specific primer- and sequencing based typing technologies. For some patients (MS-1, 3, 5, 8, OIND-1, 5, and 6), we used the strong linkage disequilibrium with HLA-DPB1 to deduce their likely DPA1 alleles (39). HLA types are shown in **Table 1**.

## Preparation of Datasets

After removing non-productive sequences, IGHV amino acid sequences were deduced using the ImMunoGeneTics (IMGT) database and the IMGT/High-V-Quest analysis tool (version 3.3.4) (40, 41). This analysis identified additional non-productive sequences that were removed. IGHV transcripts comprising more than 0.5% of total reads within each compartment were designated “highly transcribed.” Finally a single FASTA file containing all the IGHV sequences with tagged information consisting of patient code, compartment, frequency rank and tag describing whether it was highly transcribed was prepared. These sequences are deposited online at <http://doi.org/10.6084/m9.figshare.5035703>.

An extensive public dataset of IGHV nucleotide sequences from three healthy individuals (42, 43) was obtained online (<http://datadryad.org/resource/doi:10.5061/dryad.35ks2>). The corresponding IGHV amino acid sequences were deduced according to IMGT standards, and used for further analysis.

The compiled patient dataset of IGHV amino acid sequences was submitted to EigenBio (WI, USA) for processing and prediction analysis. Each sequence was given a unique general identifier (gi), and sequences with exact matching amino acid sequence within each patient were identified and given a clonal identifier for statistical purposes. Every possible 15-mer and 9-mer (denoted as IGHV fragments) were derived from each IGHV amino acid sequence, and indexed according to their N-terminus CDR3-relative position as determined by IMGT standards (44), designating cysteine 104 at the start of CDR3 as position 0. This indexing process resulted in extensive databases of overlapping IGHV fragments offset by one single amino acid, and provided a basis for systematic comparison of fragments in the FW3 and CDR3 regions across MS and OIND patients, and healthy individuals.

## Cathepsin Cleavage Probabilities

Peptidase cleavage by cathepsins S, L, and B was predicted with neural network models developed using datasets from Biniossek et al. (45), by methods described previously (46). In short, all IGHV sequences were converted into sequential octamers using the P4P3P3P1-P1' P2' P3' P4' convention with the scissile bond designated as the bond between amino acid 4 and amino acid 5 designated P1P1'. The neural network ensembles each produce a probability of cleavage of the scissile bond ranging from 0 (uncleaved) to 1 (cleaved). An ensemble median cleavage probability of >0.8 was set as the prediction threshold for the analyses.

Endosomal enzymes digest proteins into peptides of varying lengths and a peptide 15-mer is commonly presented on HLA

**TABLE 1** | Patients HLA types.

Patient	HLA class I alleles		HLA class II alleles				
	HLA-A	HLA-B	DRB1	DQA1	DQB1	DPA1	DPB1
MS-1	02:01 + 02:01	15:01 + 44:02	13:02 + 15:01	01:02 + 01:02	06:02 + 06:04	01:03/05 + 01:03/05 <sup>a</sup>	04:01 + 04:01
MS-2	02:01 + 02:01	07:02 + 27:05	01:01 + 15:01	01:01 + 01:02	05:01 + 06:02	01:03 + 01:03	04:01 + 04:02
MS-3	03:01 + 25:01	07:02 + 18:01	15:01 + 15:01	01:02 + 01:02	06:02 + 06:02	01:03/05 + 01:03/05 <sup>a</sup>	04:01 + 04:01
MS-4	03:01 + 24:02	07:02 + 35:01	01:01 + 15:01	01:01 + 01:02	05:01 + 06:02	01:03 + 01:03	02:01 + 20:01
MS-5	02:01 + 30:01	13:02 + 44:03	07:01 + 15:01	01:02 + 02:01	02:02 + 06:02	01:03/05 + 01:03/05 <sup>a</sup>	04:01 + 04:01
MS-6	03:01 + 32:01	07:02 + 08:01	03:01 + 13:01	01:03 + 05:01	02:01 + 06:03	01:03 + 02:01	01:01 + 04:01
MS-7	24:02 + 69:01	35:01 + 44:02	11:04 + 15:01	01:02 + 05:05	03:01 + 06:02	01:03 + 01:03	04:01 + 04:02
MS-8	01:01 + 31:01	07:02 + 15:01	15:01 + 15:01	01:02 + 01:02	06:02 + 06:02	01:03 + 01:03 <sup>a</sup>	02:01 + 02:01
MS-9	25:01 + 31:01	18:01 + 44:03	07:01 + 08:01	02:01 + 04:01	02:02 + 04:02	01:03 + 01:03	02:01 + 04:01
MS-10	03:01 + 29:02	35:01 + 45:01	01:01 + 04:05	01:01 + 03:03	03:02 + 05:01	01:03 + 01:03	04:01 + 04:02
MS-11	02:01 + 02:01	15:01 + 40:01	04:04 + 11:01	03:01 + 05:05	03:01 + 03:02	01:03 + 02:02	02:01 + 05:01
OIND-1	01:01 + 68:01	08:01 + 35:03	03:01 + 15:01	01:02 + 05:01	02:01 + 06:02	02:01/02/03:02 + 01:03 <sup>a</sup>	01:01 + 03:01
OIND-2	01:01 + 02:05	08:01 + 50:01	03:01 + 07:01	02:01 + 05:01	02:01 + 02:02	01:03 + 01:03	04:02 + 104:01
OIND-3	01:01 + 01:01	08:01 + 08:01	03:01 + 03:01	05:01 + 05:01	02:01 + 02:01	02:01 + 02:01	01:01 + 14:01
OIND-4	02:01 + 02:01	15:01 + 27:05	01:01 + 04:01	01:01 + 03:01	03:02 + 05:01	01:03 + 01:03	04:01 + 04:02
OIND-5	03:01 + 68:01	07:02 + 44:02	07:01 + 15:01	01:02 + 02:01	02:02 + 06:02	01:03/03:01/04:01 + 02:01 <sup>a</sup>	04:02 + 11:01
OIND-6	01:01 + 01:01	08:01 + 08:01	03:01 + 03:01	05:01 + 05:01	02:01 + 02:01	02:01/02/03:02 + 1:03/03:01/04:01 <sup>a</sup>	01:01 + 04:02

<sup>a</sup>Deduced by DPB1 linkage disequilibrium.

class II (47). However, the HLA class II molecules display peptides of widely varying lengths. In an effort to simulate this process, a “fuzzy logic” system was devised where peptide excision probability was determined by examining the simultaneous cleavage probabilities from the N-terminus minus three amino acids to the C-terminus plus three amino acids. Thus, the process generates predicted excised peptides ranging from 15 to 21 amino acids. The cutoffs for this “fuzzy logic” excision prediction were intentionally lowered to avoid under-prediction. If the maximum probability of cathepsin cleavage at either terminal was  $\geq 0.5$  and simultaneously had a probability of  $\geq 0.25$  on the other end, that would lead to an “Excision” call.

## HLA Affinity Predictions

For each CSF-derived 9- and 15-mer peptide, we predicted the affinity for 37 HLA class I and 28 class II molecules using previously described models (37, 48, 49). The neural network ensembles used for affinity predictions were developed using public datasets of allelic affinities [half-maximal inhibitory concentration ( $IC_{50}$ ) units] from <http://www.iedb.org> (downloaded June 2012). The alleles for which affinities were predicted are shown in S2 in Supplementary Material. Predicted affinities are either presented as the natural logarithm of  $IC_{50}$  [ $\ln(IC_{50})$ ] or a Johnson SI standardized value of  $\ln(IC_{50})$  within patient and compartment. Standardization was performed to bring the predicted values onto an equal scale for comparative and/or illustrative purposes. In previous publications using such predictions, standardizations were performed within protein, as all peptides within a protein may compete with each other in terms of HLA affinity (37, 46). For the current publication, the IGHV sequences were shorter, and there is also a possibility of HLA affinity competition between different IgG molecules, hence the overall within patient standardization.

## T Cell-Exposed Motifs (TCEMs)

In a peptide-HLA (pHLA) complex, some amino acids of the peptide will be exposed to T cells (TCEM), and others will be oriented inwards toward the HLA molecule (groove-exposed motif). For a 15-mer peptide in a pHLA complex, we numbered the amino acid residues from  $-3$  to 12. Utilizing the work of Rudolph et al. (50) and Calis et al. (51), as previously described (37) three different types of non-continuous TCEM were deduced and designated TCEM I (amino acid residues 4, 5, 6, 7, 8 of a 9-mer), TCEM IIa (2, 3, 5, 7, 8 of a 15-mer with a core 9-mer), and TCEM IIb ( $-1$ , 3, 5, 7, 8 of a 15-mer with a core 9-mer). The latter two are relevant for HLA class II predictions, the former for HLA class I predictions. Analysis of the models of Rudolph et al. (50) indicate that TCEM IIa or TCEM IIb motifs occur in approximately equal proportions of TCR:MHC class II structures. All possible TCEM patterns were identified for all IGHV fragments.

Rare IGHV sequences may escape tolerance and be stimulatory under the right circumstances (52). TCEM frequencies are unequally distributed in the IGHV region and also elsewhere in the proteome as some motifs occur far more frequently than others (37). Their frequency in the IGHV region can be assessed by assigning a frequency class (FC), which is a reverse log<sub>2</sub> scale where FC 0 ( $1/2^0$ ) corresponds to “occurring in every IGHV

sequence” and FC 21 ( $1/2^{21}$ ) to “occurring once every approx. 2 million sequence” (37). In this study, we considered a TCEM with FC above 16 (once every 65,536 sequence) as rare.

To compare the mean FC of our patients TCEMs, we compiled a FC classification system using a public database of 37 million unique BCRs spanning the FW3 and CDR3 from memory and naive B cells from three healthy donors published by Dewitt et al. (42, 43), consistent with a previously published classification based on 56,000 sequences from Genbank (37). The IGVH sequences of this database are of the same length and were established with similar technology as those from our patients, thereby minimizing technical or disease-related bias. All TCEMs occurring in the dataset were assigned a FC class based on their mean patient and compartment-specific  $-\log_2$  frequencies.

The FC as a measure of presumed likelihood for IGVH sequences to escape tolerance does not take into account the possible occurrence of similar TCEMs elsewhere in the human proteome or in the gut microbiome. While the human proteome is common for patients, and can to some degree be accounted for, the gut microbiome displays variations across populations and ages (53). For this publication, we used databases of TCEM occurrences in the human proteome (assembled from UniProt, with removal of Ig variable regions) (54) and microbiome (from NIH Human Microbiome Project Reference Genomes database) (55) as described previously (56), by searching for all 3.2 million possible variations of each TCEM.

Each TCEM occurs at a characteristic frequency in proteomes. These frequencies were then normalized to a zero mean unit variance scale using Johnson SI scale transformation of log<sub>2</sub> frequency values.

## Validation of Prediction Analyzes

We have previously derived two monoclonal antibodies from CSF B cells of two MS patients (CSF mAbs), and demonstrated that one idiotope from each of these mAbs (pMS1 and pMS2) was both processed, presented on HLA class II molecules and recognized by cloned CD4<sup>+</sup> T cells *in vitro* (32, 33). These were therefore suitable for validation of the prediction analyzes. Cathepsin cleavage, HLA affinity, and TCEM occurrence were predicted as for the main dataset. The FC was calculated using a previously described dataset comprising the complete IGVH region (37, 56). Idiotope pMS1-VH1 was presented on DRB1\*13:02 and pMS2-VH3 on DRB1\*13:01 encoded HLA molecules. As these have identical amino acid sequences, the affinity for DRB1\*13:02 was predicted for both peptides.

## Statistics

All predictive models were built by EigenBio using JMP® software version 12.1/13.0 (SAS Institute, Cary, NC, USA), by script processing. STATA v 14.1 (StataCorp LLC, TX, USA) and JMP® 12.1 were used for statistical analyses. Plots were created in JMP® 12.1. All plots displaying CDR3 relative positions are cropped to include ~99% of the IGHV fragments.

For bioinformatics processing purposes, to avoid end-effects in various algorithms, we added a standard immunoglobulin signal peptide and three amino acid sequence (“DTR”) to the

beginning, and a 26 amino acid sequence derived from the IgG1 constant region (“GTLVTVSSASTKGPSVFPLAPSSKST”) at the end of each IGHV sequence. Parts of these were retained as described below for the comparative statistical analyses. Our IGHV fragments were indexed by the position of the first cysteine determining the start of the CDR3 region, and changes due to mutations in the CDR3 region could influence both TCEM and affinity predictions at indexed positions even prior to position 0.

For statistical testing, we created three subsets of the IGHV sequences from our patients. For comparison of differences in mean FC and mean HLA affinities between MS and OIND patients within the CDR3, we used a subset limited to fragments with approximately half of the amino acids of the IGHV-fragment within the CDR3. This subset contained fragments starting at indexed position  $-7$ , and ended in the position where the fragment would contain eight amino acids of the added constant region. A second subset containing FW3 (spanning approximately amino acid 73–104 by IMGT numbering) and CDR3 regions were compared within MS and OIND patients. For this purpose, we used the patients dataset, with the addition of the “DTR” amino acid sequence at the start and “GTL” amino acid sequence at the end, and defined fragments as being influenced by CDR3 changes similarly as above, with a cutoff at position  $-7$ . In the third subset, we used a similar approach as for the second subset, and included the blood-derived sequences for comparisons between blood and CSF. No extra amino acid sequences were attached to the sequences of this dataset.

In the first subset, the differences between patients and controls, between low and high FC within patients and controls, and between IGHV4 and the other IGHV families were assessed by estimating the multilevel mixed effects model for each outcome variable. A multilevel approach was chosen because the data exhibit a three-level hierarchical structure, with levels for patient, relative CDR3 position, and clone. The intraclass correlation coefficient was calculated to assess the cluster effect on each level. The cluster effect was highest at patient-level for all variables. Clone-level demonstrated negligible or no cluster effect and was therefore not taken into account. Adjustment for cluster effect on CDR3 relative position-level caused convergence problems. Therefore, the differences between the categories were assessed by estimating a linear mixed model with fixed effects for factor variable and random intercepts for patients at each relative CDR3 position separately. Benjamini-Hochberg adjustment for multiple testing was applied within each outcome variable with acceptable false discovery rate (FDR) set to 20% (57).

In the second subset, the cluster effect on protein clone-level was also negligible or zero, and hence ignored. The cluster effect on patient- or relative position-level or both were present for most variables. A linear mixed model with fixed effect for factor defining the region was estimated. Random effects for either patients or relative position were included in the model. For variables with cluster effect on both levels, random effects for relative position nested within the patient were assessed, but as these were negligible only the models with single random effect were estimated.

No cluster effect on protein clone-level was found in the third subset. Inpatient correlations were close to zero or not present.

Consequently, comparisons between blood and CSF within patients and controls as well as within IGHV families for patients and controls separately were performed by independent samples *t*-test at each relative CDR3 position, and *p*-values were further adjusted for multiple testing by Benjamini-Hochberg procedure. As this was an exploratory study also aiming to guide further studies on the proposed mechanism, the FDR was set at 20%.

Finally, generalized linear models with random effects for patients were estimated to assess the differences in number of IGHV sequences containing fragments meeting a set of idiotope criteria between MS and OIND groups, as well as between highly transcribed IGHV sequences and other sequences.

The results are reported as mean differences between the groups with the corresponding 95% confidence intervals (CIs) and *p*-values.

## RESULTS

### IGHV Transcripts

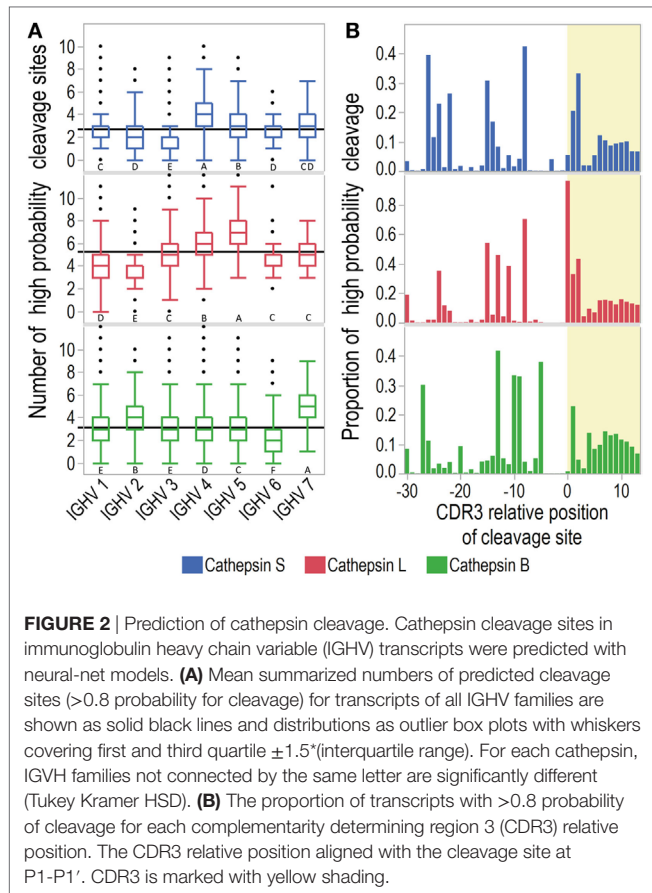
From all patients a total of 1,812,920 IGHV amino acid sequences were deduced after removing non-productive transcripts, with a mean of 3,552 (95% CI 1,399–5,706) sequences obtained from CSF and 125,180 (95% CI 78,262–172,100) from blood. The mean CDR3 lengths of CSF sequences were 15.55 amino acids (95% 15.5–15.6,  $n = 25,556$ ) for MS patients and 15.25 amino acids (95% CI 15.20–15.29,  $n = 34,833$ ) for controls ( $p < 0.001$ , independent samples *t*-test). For blood, the corresponding lengths were 15.48 (95% CI 15.48–15.49,  $n = 1,551,154$ ) for MS and 16.03 (95% CI 16.01–16.04,  $n = 201,377$ ) for controls ( $p < 0.001$ , independent samples *t*-test). The IGHV gene family usage is shown in Table S3 in Supplementary Material. As reported previously by us and others (7, 58), there was a preferential use of IGHV4 in CSF from MS patients.

Most IGHV sequences and transcripts used in this project were previously published (7). However, in the present study we also included exceedingly rare sequences for the purpose of creating a thorough database of TCEM, resulting in a higher total number than previously described (7).

### Cathepsin Cleavage

Due to the IGHV4 bias in CSF among the MS patients, we first investigated whether the predicted pattern of cathepsin cleavage differed across IGHV families. An analysis of variance by IGHV family yielded significant variations (Welch ANOVA  $F(6, 951.32/951.49/949.68)$  for cathepsins S, L, and B, respectively,  $p < 0.0001$  for all). Cathepsin S was predicted to preferentially cleave IGHV4 derived sequences, cathepsin L was predicted preferentially to cleave those from IGHV5, and cathepsin B to preferentially cleave those from IGHV7. Interestingly, cathepsin S was also predicted to cleave IGHV3 sequences least efficiently (Figure 2; Table S4 in Supplementary Material).

As CDR3 is most diverse and therefore hypothesized to be the main source of immunogenic idiotopes, we further investigated whether the cathepsins were likely to release CDR3 fragments (Figure 2B). All three cathepsins displayed a similar overall pattern of cleavage sites in FW3 and at the start of CDR3.

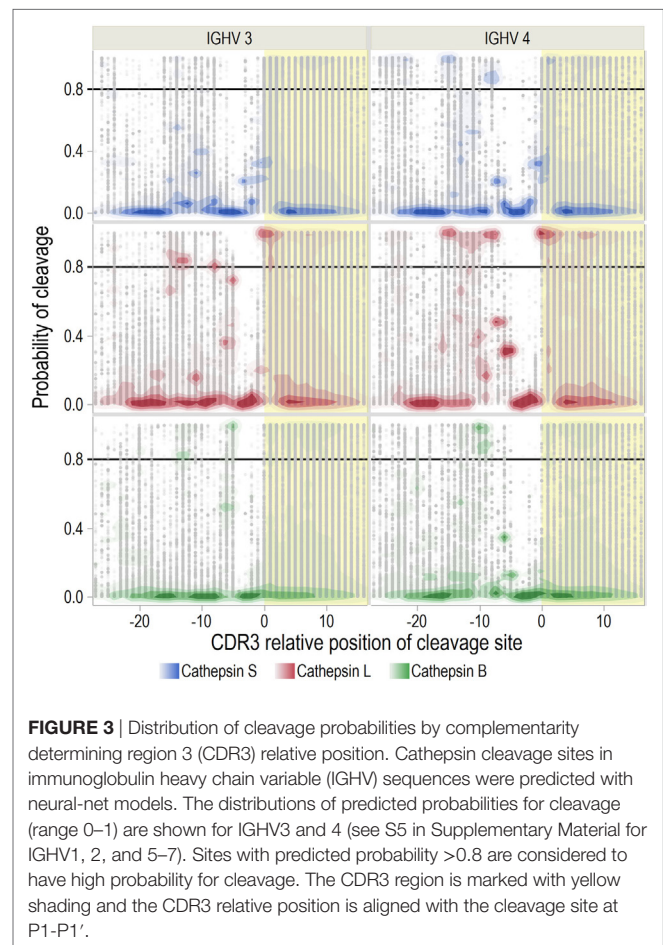


Notably, cathepsin L was predicted to cleave almost all IGHV sequences before or after the cysteine marking the start of CDR3. Cathepsins S and B showed less pronounced peaks for cleavage at the same positions.

We next investigated whether these patterns differ across IGHV families (Figure 3; S5 in Supplementary Material). The previously identified hotspot for predicted cathepsin L activity at the CDR3 start was consistently found for all IGHV families. For cleavage of IGHV4 by cathepsin L and S, there were also two hotspots in FW3. No hotspot was identified for cathepsin S for IGHV3. As cathepsins S and L are endopeptidases that recognize octamers within the peptide, cleavage at these hotspots would effectively block other cleavages in the immediate vicinity. Therefore, these hotspots probably represent the most likely cleavage sites.

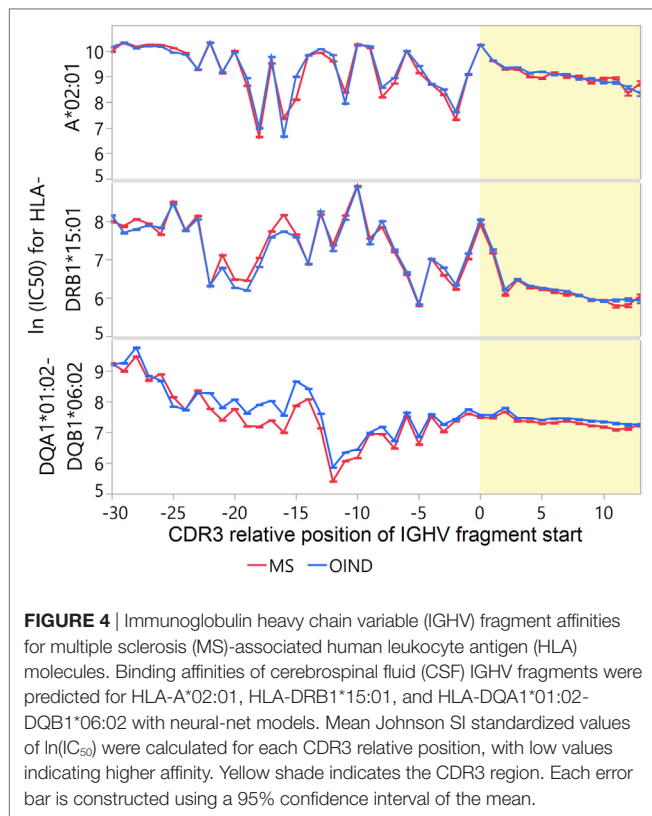
## HLA Affinities

Different HLA molecules may display different binding affinities for IGHV fragments. Analyzing all CSF IGHV fragments together, fragments from CDR3 had consistently higher predicted affinities for HLA-DR and -DP than fragments from FW3 (Figure S6 in Supplementary Material). While DRB1\*15:01 was among the DR molecules with the highest mean standardized affinities for CDR3-derived fragments, the same was not true for the linked DQA1\*01:02-DQB1\*06:02 among DQ molecules. In general, the predicted patterns of affinities for different HLA class II molecules were similar between MS and OIND patients (not shown).



We next investigated the IGHV sequences from CSF for binding affinity to MS-associated HLA molecules (Figure 4). CDR3 fragments from MS patients had higher predicted affinity compared to FW3 fragments for DRB1\*15:01 and to a lesser extent for DQA1\*01:02-DQB1\*06:02, but lower affinity for A\*02:01 ( $p < 0.001$  for all, S7 in Supplementary Material). After correcting for intraclass correlations at patient-level, there were no significant differences between the MS and OIND patients in predicted affinity for these HLA molecules for any IGHV fragment in vicinity of the CDR3. Similarly, no significant differences were detected for MS or OIND patients when comparing highly transcribed to other IGHV sequences (data not shown). However, IGHV4 family fragments from MS and OIND patients had higher predicted affinity for DRB1\*15:01 than other IGHV fragments at almost every position within the CDR3 (Table S8B in Supplementary Material). For DQA1\*01:02-DQB1\*06:02 and A\*02:01, the results were similar to those for DRB1\*15:01, except for IGHV4 where both higher and lower mean affinities were predicted within the CDR3 (Tables S8C,D in Supplementary Material).

Because each patient did not carry all HLA alleles, we extracted the affinity predictions for those carried by each individual. For heterozygous patients, we used the allele with the

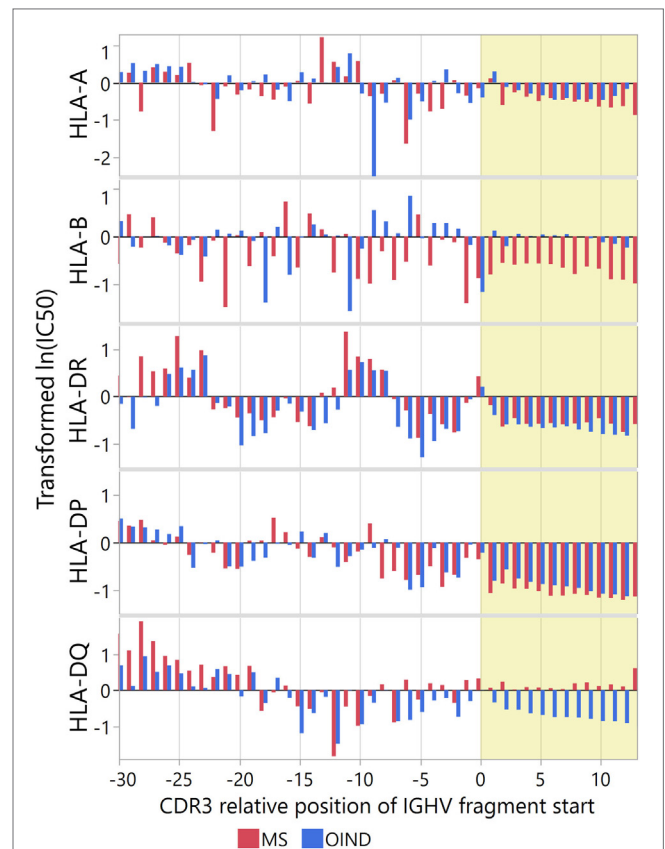


highest predicted affinity (**Figure 5**). In general, the standardized predicted affinities were very similar for DR and DP, with high predicted affinity in CDR3 for both MS and OIND patients. For DQ on the other hand, only OIND patients followed this pattern. This most likely reflects that MS patients more frequently carry DQA1\*01:02-DQB1\*06:02, which we previously showed did not display the highest affinities within the CDR3.

### TCEM Patterns

From all possible IGHV fragments ( $n = 52,566,906$ ) we identified TCEM I, IIa and IIb and created databases of TCEM occurrences. We identified approximately 1.5 million unique TCEM of each pattern from a theoretical maximum of 3.2 million. These overlapped >98% with the TCEMs in the dataset derived from healthy individuals by DeWitt et al. (42), which could therefore be used for frequency classification in our dataset (**Table 2**).

Different occurrences of TCEM between populations and compartments could point to a selection process. The results of cluster analysis performed on pairwise correlations for the summarized occurrences of TCEM IIa in blood and CSF for each patient group are shown in **Figure 6**. The TCEM patterns of IGHV sequences from CSF of MS patients differed from those in blood and also from those of the OIND patients. There were two notable exceptions; TCEM from CSF of OIND-4 clustered consistently with that in CSF of MS patients. This patient was the only OIND patient with oligoclonal IgG bands (OCB) and IGHV4 dominance in the CSF. The other exception was MS-6, from whom the TCEM in CSF clustered with that in blood.



**FIGURE 5** | Immunoglobulin heavy chain variable (IGHV) fragment affinities for patient-specific human leukocyte antigen (HLA) molecules. Binding affinities of IGHV fragments were predicted for patient-specific HLA A, B, DP, DR, and DQ molecules (listed in **Table 1**). Results are presented as mean Johnson SI standardized values of  $\ln(I_{C50})$  for each complementarity determining region 3 (CDR3) relative position, with low values indicating high affinity. For heterozygous patients with two sets of predictions for one HLA molecule, we used the lowest standardized  $\ln(I_{C50})$ . Yellow shade indicates the CDR3 region.

**TABLE 2** | Number of unique TCEM identified in each dataset.

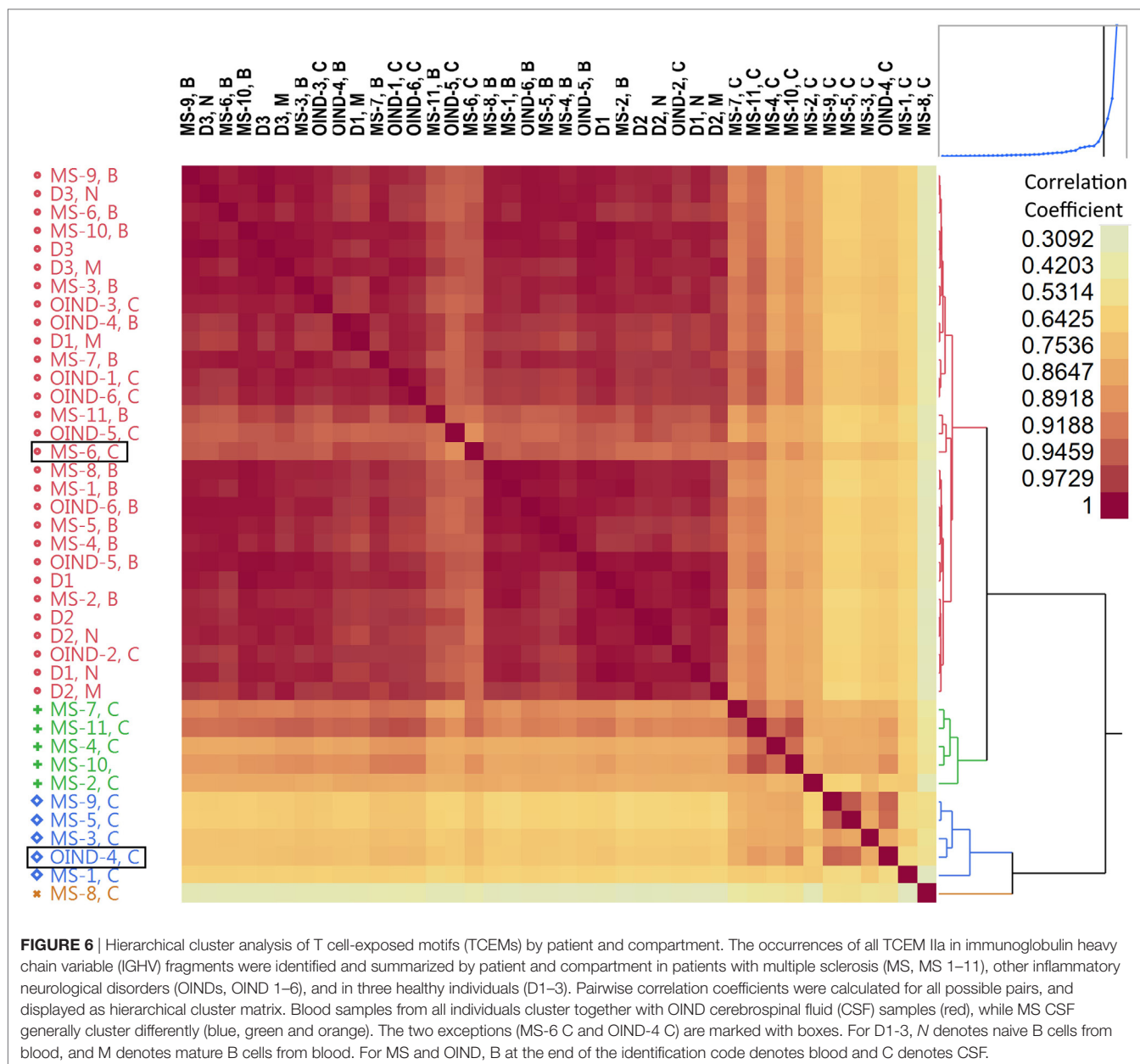
	Patient dataset <sup>a</sup>	Healthy BCR dataset	Overlap (%) <sup>b</sup>
TCEM I	1,448,203	2,747,840	98.64
TCEM IIa	1,439,194	2,736,223	98.62
TCEM IIb	1,403,685	2,710,199	98.64

<sup>a</sup>Combination of blood and CSF derived IGHV fragments from both MS and OIND patients.

<sup>b</sup>Percentage of TCEM identified in the patients, also identified in the DeWitt et al. healthy BCR set.

Notably, MS-6 was one of two MS patients with dominant IGHV3 use in CSF. TCEM I and IIb displayed similar patterns as TCEM IIa (data not shown). Corresponding results were found using standardized ( $z$  transformed) TCEM occurrences (S9 in Supplementary Material).

We next performed cluster analysis on pairwise correlations of TCEM occurrences by compartment and IGHV family.



**FIGURE 6 |** Hierarchical cluster analysis of T cell-exposed motifs (TCEMs) by patient and compartment. The occurrences of all TCEM Ila in immunoglobulin heavy chain variable (IGHV) fragments were identified and summarized by patient and compartment in patients with multiple sclerosis (MS, MS 1–11), other inflammatory neurological disorders (OINDs, OIND 1–6), and in three healthy individuals (D1–3). Pairwise correlation coefficients were calculated for all possible pairs, and displayed as hierarchical cluster matrix. Blood samples from all individuals cluster together with OIND cerebrospinal fluid (CSF) samples (red), while MS CSF generally cluster differently (blue, green and orange). The two exceptions (MS-6 C and OIND-4 C) are marked with boxes. For D1–3, N denotes naive B cells from blood, and M denotes mature B cells from blood. For MS and OIND, B at the end of the identification code denotes blood and C denotes CSF.

This revealed clustering on IGHV family level (S10 in Supplementary Material), implying that either dominance of IGHV4 or relative lack of IGHV3 could be driving the differences in TCEM patterns. This also suggests that differences in number of IGHV sequences were not driving the TCEM clustering, as could be suspected because CSF samples generally clustered together. Similar results were obtained when analyzing samples by IGHV family, patient-ID and compartment. Again the samples clustered mainly by IGHV family and secondly by compartment (data not shown).

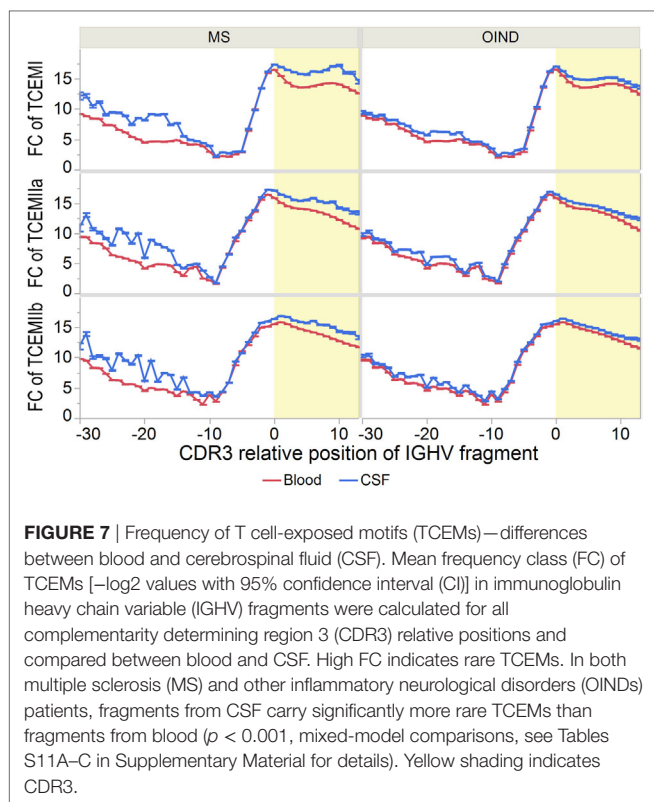
## Frequency Classifications

As CSF IgG is more mutated than IgG from blood (7, 12, 15, 59), the TCEMs of IGHV sequences from CSF could be rarer

than those from blood. We therefore compared FC distributions in blood vs. CSF (Figure 7; Tables S11A–F in Supplementary Material). Although most pronounced among the MS patients, the mean FC of TCEMs from the CDR3 was significantly higher in CSF at nearly every CDR3 relative positions for all three TCEM for both patient groups. The mean FC was also significantly higher in CSF for each IGHV family analyzed separately (data not shown).

Our analyzes predicted that CDR3 fragments were both likely to be released by cathepsins and to bind HLA class II molecules with higher affinities than FW3 fragments. Hence, we analyzed whether the CDR3 region generated more rare TCEMs. The CDR3 regions of CSF from both MS and OIND patients generated on average more rare TCEM than FW3 ( $p < 0.001$  for





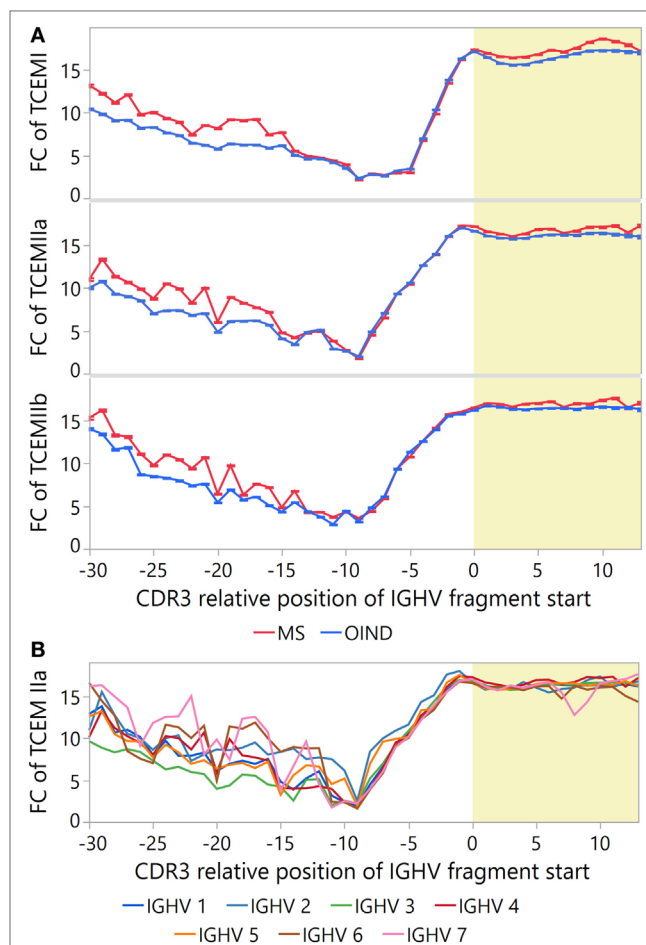
all TCEMs), probably driven by greater diversity in the CDR3 (Table S12 in Supplementary Material).

Because rare motifs are expected to be most likely to escape tolerance (52), we also tested whether CDR3 fragments from MS patients contain on average rarer motifs than those from OIND patients. After correcting for intra-class correlations in a multilevel hierarchical mixed model and multiple testing, the MS patients had significantly higher FC than the OIND patients at several CDR3 positions (**Figure 8A**; Tables S13B–D in Supplementary Material). This was most evident for TCEM I and IIa.

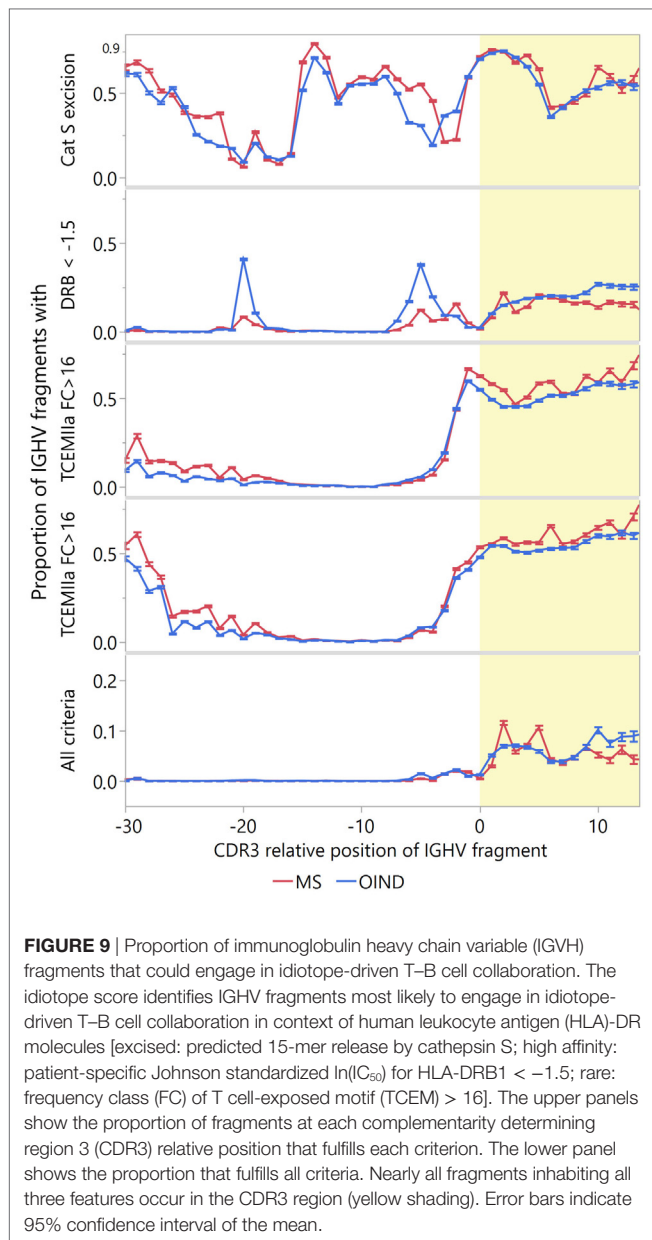
As IGHV4 family bias was found in MS CSF samples (Table S3 in Supplementary Material), we compared the FC of sequences carrying IGHV4 to other IGHV families (**Figure 8B**). Differences in FC between IGHV families were most pronounced prior to position  $-7$ . For CDR3, we found statistically higher FC at nearly all positions for sequences carrying IGHV4 than for sequences carrying any other IGHV family (Tables S13E–G in Supplementary Material). Thus, it seems that IGHV4 possess all the assumed prerequisites for T cell stimulation.

## Summarized Attributes

To identify IGHV sequences most likely to engage in idiotope-driven T–B cell collaboration, we devised an “idiotope score” identifying IGHV fragments with both high patient-specific HLA-DRB1 standardized affinity [ $\ln(\text{IC}_{50}) < -1.5$ ]; TCEM II (a or b) FC  $> 16$ ; and predicted fuzzy cut “Excision” by cathepsin S (**Figure 9**). Although many IGHV fragments fulfill one of these

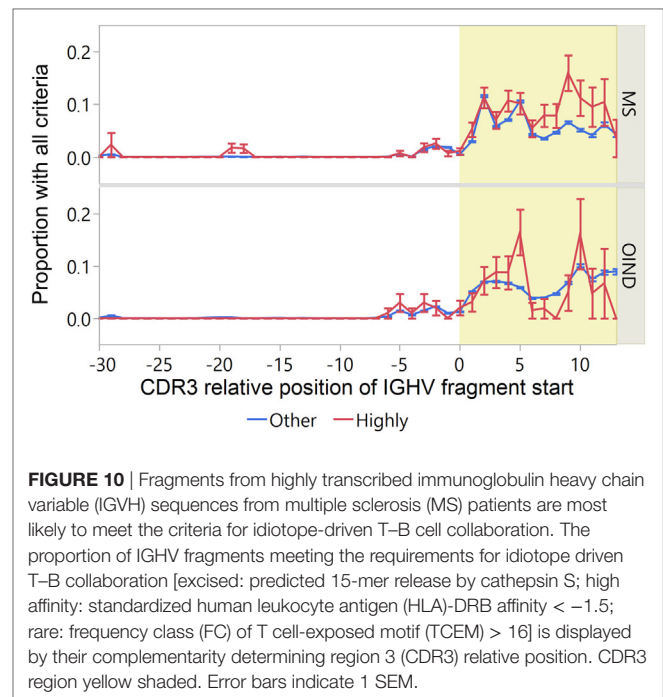


criteria relatively few fulfill all, and these almost exclusively reside in the CDR3 region. In a few peaks in the CDR3, almost 10% of the IGHV fragments fulfilled all criteria. Among the MS patients, the highly transcribed CDR3 fragments in CSF were generally most likely to carry all these traits (**Figure 10**). Moreover, in MS patients highly transcribed IGHV sequences were more likely than others to carry at least one IGHV fragment with these attributes (OR = 1.39,  $p = 0.01$  unadjusted model). However, after adjusting for cluster effect on patient-id-level the difference was no longer significant (OR = 1.24,  $p = 0.11$ ). For OIND patients, we found no significant difference in unadjusted (OR = 1.20,  $p = 0.38$ ) or adjusted models (OR 1.25,  $p = 0.29$ ). Among highly transcribed IGHV sequences, 42% from the MS patients and 34% from the OIND patients carried at least one fragment fulfilling all criteria (OR = 1.41,  $p = 0.16$ ; adjusted on patient-id-level: OR = 1.40,  $p = 0.27$ ).



## TCEM in the Human Proteome and Gut Microbiome

By plotting the mean Johnson standardized frequencies of TCEM found in the human proteome and gut microbiome in a similar fashion as the immunoglobulin FC scale, we found that the CDR3 regions generate TCEM that generally occur rarely in both the gut microbiome and the human proteome (S14 and S15 in Supplementary Material). There was a significantly lower standardized mean occurrence of TCEM in the CDR3 than FW3 for both gut microbiome ( $p < 0.001$ ) and the human proteome ( $p < 0.001$ ). Hence it seems that somatic hypermutation and recombination in the CDR3 is capable of generating TCEM that are rare, both in the healthy IGHV repertoire and in the human proteome and gut microbiome.

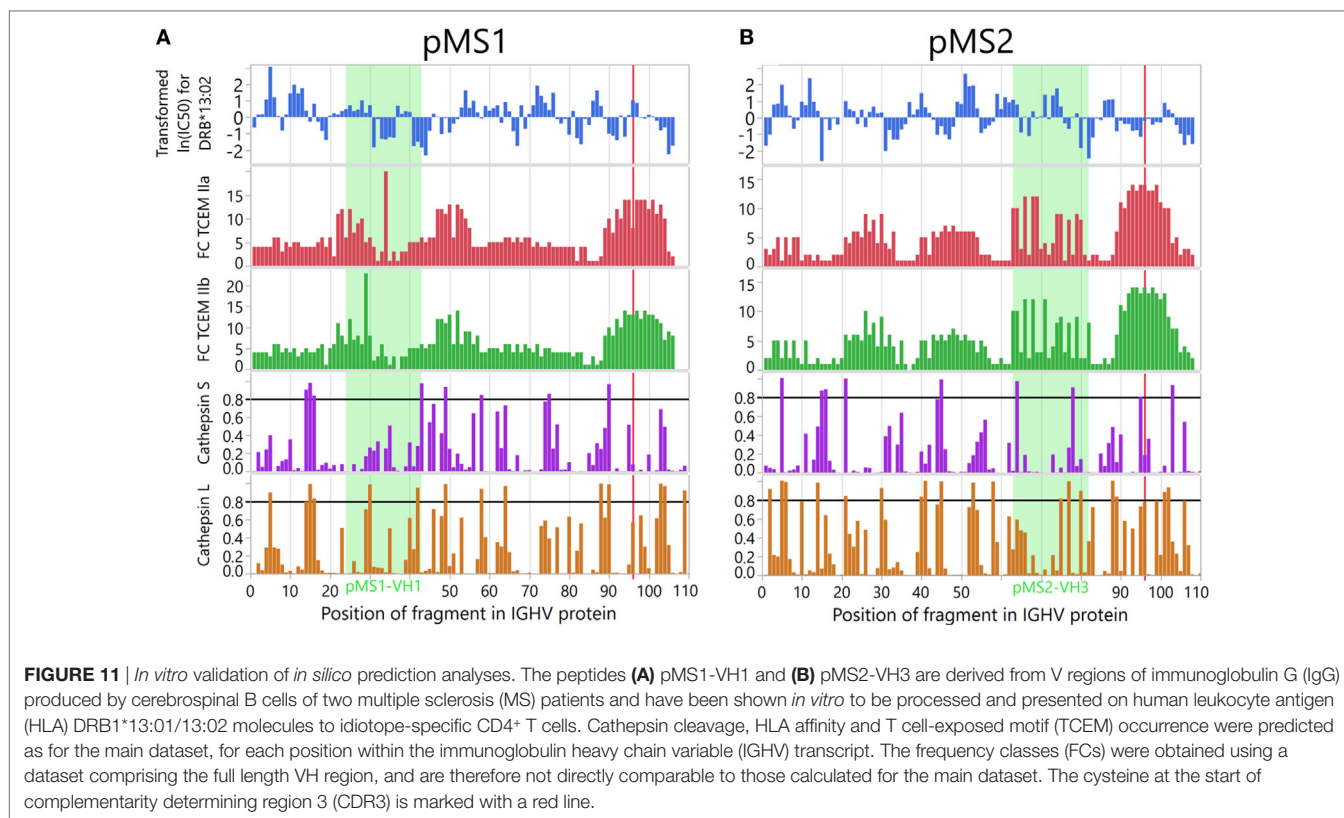


## Validation

To validate the epitope prediction, we performed prediction analysis of the V regions of two CSF mAbs previously derived from CSF B cells of two MS patients (32). We have previously demonstrated *in vitro* that each VH region of these CSF mAbs carry an idiotope (pMS1-VH1 and pMS2-VH3) that was processed by APCs, and presented on DRB1\*13:01/13:02 molecules to CD4<sup>+</sup> T cells which specifically recognized the particular idiotope (32, 33). As shown in **Figure 11** prediction analysis anticipated that both peptides were likely to be cleaved at positions allowing for presentation by the relevant HLA-DR molecule, and that they would bind their restriction element (DRB1\*13:01/13:02) with high affinity. The FCs had to be calculated for the complete VH region and are therefore not directly comparable to those used for the complete dataset, but nevertheless shows that the TCEM associated with these idiotopes are rare.

## DISCUSSION

We have hypothesized that idiotope-driven T–B cell collaboration may drive the intrathecal immune response in MS (19, 33). Although proof of concept studies have provided some evidence compatible with this hypothesis (31–33), the immense diversity of the immune repertoire have previously precluded further analyses. Here we combined high-throughput sequencing of IGHV transcripts with *in silico* prediction analyses to assess whether the requirements for such T–B cell collaboration exist. Our findings indicate that idiotopes from the CDR3 regions of MS patients on average have high affinities for disease associated HLA-DRB1\*15:01 molecules and are predicted to be endosomally processed by cathepsin S and L in positions that allows such HLA binding to occur. Additionally, CDR3 sequences from



CSF B cells from MS patients contain on average rarer TCEM that could potentially stimulate non-tolerant CD4<sup>+</sup> T cells, than corresponding sequences from OIND patients. Many of these features were associated with the previously described IGHV4 gene family bias (7, 58, 59) of CSF B cells in MS, indicating a possible explanation for this previously unexplained predominance. The IGHV gene family distribution with IGHV3-gene family dominance in blood also correlated with previously published results from healthy individuals (37, 42, 60).

Cathepsins S and L are essential for antigen presentation on MHC class II molecules (61, 62), including both processing of invariant chain (Ii) and preparation of protein fragments expressed on MHC class II molecules (63). Cathepsin S was shown to be particularly important for endosomal processing in B cells, dendritic cells and macrophages, while cathepsin L was essential for the Ii chain processing in cortical thymic epithelial cells and in macrophages (64, 65). Both these cathepsins are therefore likely relevant for endosomal processing of idiotopes. Cathepsin S was also implicated in MS by early expression studies suggesting a possible disease association (66), but this has not been replicated in more recent GWAS studies (67). The cathepsin S (*CTSS*) gene was further reported to be associated with treatment responses of both glatiramer acetate and IFN-beta (68). It was shown that cathepsin S could cleave myelin basic protein as a possible mechanism of action (69), and a cathepsin S-like helminthic protease was efficient in cleaving IgG (70). Even earlier publications showed that human lysosomal proteases can cleave IgG at acidic and to a lesser extent at neutral pH (71).

Cathepsin S and L have similar cleavage patterns in general (72). Our cleavage predictions are in line with these findings and indicate that cathepsin S especially may be important for endosomal processing of BCRs in a way that allows idiotopes of the CDR3 region to be released.

Three observations point to CDR3 as crucial for generation of idiotopes capable of stimulating CD4<sup>+</sup> T cells: First, cathepsins S and L both displayed high probability of cleavage around the CDR3 start; second, these cleavage spots were immediately followed by regions with high predicted affinity for HLA class II molecules; and third, the same region was associated with relatively high mean FC values for all TCEM, implying high likelihood for T cell stimulation.

It has been shown that IGHV fragments from endogenous IgG are indeed processed and presented on MHC II molecules (20). However, while the effects of cathepsin S and L on foreign antigens and Ii-chain have been well characterized (73), there have been no specific studies to our knowledge on how human cathepsins S and L act on Ig variable regions in endosomal conditions. The predictive models were built on data from proteomic identification of cleavage sites assays (45), where cleavage sites are readily available in preprocessed polypeptide cocktails (45, 46). Native IgG molecules, on the other hand, contain disulfide bonds that could interfere with cathepsin activity (74). Future studies addressing these questions would be important for validating our predictions.

B cells are likely to play a role as APCs in MS (18), but which antigens they present are unclear. B cells are capable of processing

and presenting their own Ig, as demonstrated almost 30 years ago in mice (20, 22), and recently shown to occur in large scale on HLA class II molecules in humans with mantle cell lymphoma (75). Such presentation would normally be enhanced by BCR stimulation, which activates B cells and induces proper antigen presentation potential (76). An alternative way for the B cell to upregulate their HLA class II expression and antigen presenting potential, was shown in mice to occur through CD40 stimulation in the thymus (77), as thymus B cells upregulated AIRE, HLA class II and CD80 in a CD40 dependent (and BCR independent) mechanism. It is not known whether B cells from the CNS or CSF of MS patients express AIRE.

Our results of predicted high affinity for HLA class II molecules for CDR3-derived fragments are consistent with findings in the Genbank IGHV dataset from healthy donors, including the different patterns observed for HLA-DQ (37). Notably peptides derived from the CDR3/FW3 region studied here were recently shown to be extensively presented on HLA class II in human mantle cell lymphoma B cells, and also recognized by idiotope-specific CD4<sup>+</sup> T cells (75). Interestingly, for the single lymphoma patient with available HLA data and T cell specificity (75), our predictions confirmed high affinity of the eluted VH peptide to HLA-DRB1\*04:01 molecules [predicted  $\ln(\text{IC}_{50}) = 3.69$ ,  $\text{SD} = 0.69$ ]. Along with our results, this implies that the diversity of this region either generates on average higher affinity peptides, or that some idiotopes are selected for their affinity to HLA class II. As our sequences did not span the whole Ig variable region, we were unable to compare affinities for peptides derived from CDR3 and FW3 with those derived from CDR1, -2 and other FW regions. However, the recent results from human mantle cell lymphoma suggest that the most relevant part of the Ig molecule for idiotope-driven T-B cell collaboration were included in this study (75).

The concept of idiotope-driven T-B cell collaboration is founded on an idea of lack of tolerance for the IGHV region, but it is not fully known to which extent such tolerance occurs (52). Central T cell tolerance is mediated through positive and negative selection within the thymus, with help of cortical and medullary thymic epithelial cells (mTECs) as well as dendritic cells (78). During negative selection T cells are exposed to self-peptides by mTECs with help of promiscuous gene expression regulated by the autoimmune regulator AIRE protein, leading to either clonal deletion or induction of regulatory T cells (78). However, V (D) J recombinations of the IGHV genes only occur in B cells, and mTECs presumably are unable to present the huge number of idiotopes resulting from this process. Yet it was shown that T cells are likely tolerant to germline-encoded (non-mutated) IGHV regions (23, 79). It is possible that this could be mediated through circulating Ig, as very high concentrations of monoclonal Ig can induce tolerance through clonal deletion in the thymus (80, 81). Recent studies have found that both naive and class-switched B cells in the thymus of mice are of peripheral origin and capable of AIRE-induced antigen presentation (77, 82), even without BCR stimulation. This could provide another explanation as to how B cells can generate tolerance in the thymus, but the studies did not investigate to what extent such B cells present their own IGHV regions. Also, only a few B cells are present in

the thymus at any time (77), providing a relatively small pool of BCRs to generate tolerance.

Our study utilizes TCEM as a model for how TCR interact with pHLA. Rudolph et al. described how only a few amino acid residues of the peptide in a pHLA complex interact with TCRs (50). These observations were then used to deduce the atomic contacts of motifs exposed to T cells (37). As HLA class II TCEMs are non-linear, matching TCEM can appear in context of both high and low affinity peptides. The TCEM model is applicable to any protein in a pHLA complex, and we expect TCEM occurring in the human proteome to be associated with tolerance if they are presented in context of HLA (56). A logarithmic scale of TCEM frequency classification (FC system) was developed and described for an IGHV repertoire from healthy individuals, and it was shown that each TCEM has a characteristic frequency of use. Some TCEM occur very frequently in IGHV regions (low FC), while others are incredibly rare (high FC) (37). The observation that some TCEM are present in every single, second, fourth, etc., IGHV sequence could possibly explain why relatively few thymic B cells may induce central tolerance for a substantial proportion of the IGHV repertoire (37). In agreement with previous findings of high diversity in transcribed CDR3 regions from CSF B cells (7), we found here that the CDR3 sequences from CSF contained on average quite rare TCEM, compatible with high likelihood of escaping tolerance. Moreover, the finding that CSF B cells have higher mean FC (more rare) TCEM than blood B cells is compatible with the notion that B cells are selected into the intrathecal compartment for their ability to stimulate idiotope-specific T cells. Finally, our comparison of TCEM in the IGHV sequences and the human proteome and gut microbiome found that CDR3-derived IGHV fragments more frequently carried TCEM that were rare in both these compartments, again suggesting that they would be more likely to escape tolerance. It was previously shown that neither the gut microbiome nor the human proteome cover the entirety of TCEM diversity (56), allowing for occurrence of many TCEM unique to the IGHV regions.

In agreement with the recently discovered lymph drainage of the CNS (83), B cells in the brain of MS patients seem to mature in cervical lymph nodes (8), and B cells in the CSF are clonally related to those in blood (7, 59). B cells may also proliferate within ectopic lymphoid follicles within the CNS of MS patients with long-standing disease (84). Hence, maturation necessary for idiotope-driven T-B cell collaboration could occur both in the periphery and within the CNS (8, 59). Such maturation may increase CDR3 variability and influence any of the parameters investigated in this study, for instance mutations that generate rare TCEM could also influence HLA affinity or cathepsin cleavage.

There are several limitations to this study. The number of included patients was quite low. Most importantly the findings *in silico* study needs to be further validated *in vitro*. In this study, the controls were restricted to OIND patients. To address whether the proposed mechanism drives the intrathecal synthesis of oligoclonal IgG in MS, MS patients without evidence of this phenomenon could be informative. Such patients are however rare, and can be hard to identify as the absence of two or more OCB bands by routine methods not necessarily rules out intrathecal synthesis of oligoclonal Ig (85–88). Moreover, the

proposed hypothesis does not exclude that individuals without intrathecal synthesis of oligoclonal IgG have a repertoire of idiotope-matched T–B cell pairs, but rather that a break of immune tolerance against self-IgG leads to dysregulated idiotope-driven T–B cell collaboration (19). This corresponds to other candidate autoantigens in MS, as myelin-specific T cells in the blood of healthy individuals are a frequent finding (89). Moreover, T cell responses against self-IgG is not unique for MS, but have previously been shown in patients with other inflammatory diseases such as systemic lupus erythematosus (90–92), granulomatosis with polyangiitis (Wegener’s granulomatous) (93), and rheumatoid arthritis (94).

## CONCLUSION

The overwhelming complexity of the immune repertoire calls for novel approaches to chart the interactions between immune receptors. This is the first work combining high-throughput sequencing of the IGHV transcriptome with *in silico* predictions analysis for T cell activation in a human disease. We predict that the three proposed prerequisites (successful endosomal processing, high HLA class II affinity and sufficiently rare TCEM) for idiotope-driven T–B cell collaboration are likely to occur in the CDR3 region in the CSF of MS patients, with as many as 42% of the highly transcribed IGHV sequences possess at least one segment with these features.

## ETHICS STATEMENT

All participants provided written informed consent for participation. The study was approved by the Committee for Research Ethics in the South-Eastern Norwegian Healthy Authority (REK Sør-Øst S-04143a), the Norwegian Social Science Data Services (no. 11069) and the review boards at AHUS and OUS.

## AUTHOR CONTRIBUTIONS

RH contributed with dataset preparations, statistical designs, analysis and interpretation of the data, and drafting the

manuscript. AL contributed by sample acquisition, interpreting the data, and writing the manuscript. JJ contributed with collection and preparation of samples and provided the immunosequence patient database and revising the manuscript. JH contributed by interpreting the data and revising the manuscript. JB contributed by designing and performing statistical experiments, the interpretation of these and revising the manuscript. BB contributed by developing the concept and design of the study, as well as revising the manuscript. HR contributed by design of immunosequencing techniques as well as revising the manuscript. RB contributed by designing the bioinformatics algorithms, preparing datasets, interpreting the data and revising the manuscript. TH contributed by designing the study, interpreting the data, and writing the manuscript.

## ACKNOWLEDGMENTS

We would like to thank all patients that have participated in this study. We also thank Marte K. Viken at the Department of Immunology and Transfusion Medicine, Faculty of Medicine, University of Oslo and Oslo University Hospital, Rikshospitalet, Oslo, Norway, for the kind help with HLA typing. The authors also would like to thank Alleen Hager for assistance in computational processing. Some of the data appearing in this manuscript was first published as an abstract and poster (P435) at ECTRIMS 2016, London.

## FUNDING

The study was supported with a grant from the Norwegian Research Council (FriMedBio) (grant/project number 250864/F20) and Akershus University Hospital internal strategic funds.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fimmu.2017.01255/full#supplementary-material>.

## REFERENCES

- Compston A, Coles A. Multiple sclerosis. *Lancet* (2008) 372(9648):1502–17. doi:10.1016/s0140-6736(08)61620-7
- Hauser SL. The Charcot Lecture | beating MS: a story of B cells, with twists and turns. *Mult Scler* (2015) 21(1):8–21. doi:10.1177/1352458514561911
- Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung HP, Hemmer B, et al. Ocrelizumab versus interferon beta-1a in relapsing multiple sclerosis. *N Engl J Med* (2017) 376(3):221–34. doi:10.1056/NEJMoa1601277
- Hohlfeld R, Meinl E. Ocrelizumab in multiple sclerosis: markers and mechanisms. *Lancet Neurol* (2017) 16(4):259–61. doi:10.1016/S1474-4422(17)30048-0
- Baker D, Marta M, Pryce G, Giovannoni G, Schmierer K. Memory B cells are major targets for effective immunotherapy in relapsing multiple sclerosis. *EBioMedicine* (2017) 16:41–50. doi:10.1016/j.ebiom.2017.01.042
- Obermeier B, Mentele R, Malotka J, Kellermann J, Kumpfel T, Wekerle H, et al. Matching of oligoclonal immunoglobulin transcriptomes and proteomes of cerebrospinal fluid in multiple sclerosis. *Nat Med* (2008) 14(6):688–93. doi:10.1038/nm1714
- Johansen JN, Vartdal F, Desmarais C, Tuttunen AE, de Souza GA, Lossius A, et al. Intrathecal BCR transcriptome in multiple sclerosis versus other neuroinflammation: equally diverse and compartmentalized, but more mutated, biased and overlapping with the proteome. *Clin Immunol* (2015) 160(2): 211–25. doi:10.1016/j.clim.2015.06.001
- Stern JN, Yaari G, Vander Heiden JA, Church G, Donahue WF, Hintzen RQ, et al. B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci Transl Med* (2014) 6(248):248ra107. doi:10.1126/scitranslmed.3008879
- Kinzel S, Weber MS. B cell-directed therapeutics in multiple sclerosis: rationale and clinical evidence. *CNS Drugs* (2016) 30(12):1137–48. doi:10.1007/s40263-016-0396-6
- Cross AH, Stark JL, Lauber J, Ramsbottom MJ, Lyons JA. Rituximab reduces B cells and T cells in cerebrospinal fluid of multiple sclerosis patients. *J Neuroimmunol* (2006) 180(1–2):63–70. doi:10.1016/j.jneuroim.2006.06.029
- Kabat EA, Moore DH, Landow H. An electrophoretic study of the protein components in cerebrospinal fluid and their relationship to the serum proteins. *J Clin Invest* (1942) 21(5):571–7. doi:10.1172/JCI101335

12. Baranzini SE, Jeong MC, Butunoi C, Murray RS, Bernard CCA, Oksenberg JR. B cell repertoire diversity and clonal expansion in multiple sclerosis brain lesions. *J Immunol* (1999) 163(9):5133–44.
13. Owens GP, Bennett JL, Lassmann H, O'Connor KC, Ritchie AM, Shearer A, et al. Antibodies produced by clonally expanded plasma cells in multiple sclerosis cerebrospinal fluid. *Ann Neurol* (2009) 65(6):639–49. doi:10.1002/ana.21641
14. Wings KM, Gilden DH, Bennett JL, Yu X, Ritchie AM, Owens GP. Analysis of multiple sclerosis cerebrospinal fluid reveals a continuum of clonally related antibody-secreting cells that are predominantly plasma blasts. *J Neuroimmunol* (2007) 192(1–2):226–34. doi:10.1016/j.jneuroim.2007.10.009
15. 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(5):2725. doi:10.4049/jimmunol.171.5.2725
16. Monson NL, Brezinschek H-P, Brezinschek RI, Mobley A, Vaughan GK, Frohman EM, et al. Receptor revision and atypical mutational characteristics in clonally expanded B cells from the cerebrospinal fluid of recently diagnosed multiple sclerosis patients. *J Neuroimmunol* (2005) 158(1–2):170–81. doi:10.1016/j.jneuroim.2004.04.022
17. Obermeier B, Lovato L, Mentele R, Bruck W, Forne I, Imhof A, et al. Related B cell clones that populate the CSF and CNS of patients with multiple sclerosis produce CSF immunoglobulin. *J Neuroimmunol* (2011) 233(1–2):245–8. doi:10.1016/j.jneuroim.2011.01.010
18. Blauth K, Owens GP, Bennett JL. The ins and outs of B cells in multiple sclerosis. *Front Immunol* (2015) 6:565. doi:10.3389/fimmu.2015.00565
19. Holmoy T, Vartdal F, Hestvik AL, Munthe L, Bogen B. The idiotope connection: linking infection and multiple sclerosis. *Trends Immunol* (2010) 31(2):56–62. doi:10.1016/j.it.2009.11.001
20. Weiss S, Bogen B. MHC class II-restricted presentation of intracellular antigen. *Cell* (1991) 64(4):767–76. doi:10.1016/0092-8674(91)90506-T
21. Weiss S, Bogen B. B-lymphoma cells process and present their endogenous immunoglobulin to major histocompatibility complex-restricted T cells. *Proc Natl Acad Sci U S A* (1989) 86(1):282–6. doi:10.1073/pnas.86.1.282
22. Bogen B, Malissen B, Haas W. Idiotope-specific T cell clones that recognize synergic immunoglobulin fragments in the context of class II molecules. *Eur J Immunol* (1986) 16(11):1373–8. doi:10.1002/eji.1830161110
23. Eyerman MC, Zhang X, Wysocki LJ. T cell recognition and tolerance of antibody diversity. *J Immunol* (1996) 157(3):1037–46.
24. Bogen B, Weiss S. Processing and presentation of idiotypes to MHC-restricted T cells. *Int Rev Immunol* (1993) 10(4):337–55. doi:10.3109/08830189309061709
25. Munthe LA, Os A, Zangani M, Bogen B. MHC-restricted Ig V region-driven T-B lymphocyte collaboration: B cell receptor ligation facilitates switch to IgG production. *J Immunol* (2004) 172(12):7476–84. doi:10.4049/jimmunol.172.12.7476
26. Mitchison NA. The carrier effect in the secondary response to hapten-protein conjugates. II. Cellular cooperation. *Eur J Immunol* (1971) 1(1):18–27. doi:10.1002/eji.1830010104
27. Lanzavecchia A. Antigen-specific interaction between T and B cells. *Nature* (1985) 314(6011):537–9. doi:10.1038/314537a0
28. Rudensky A, Preston-Hurlburt P, al-Ramadi BK, Rothbard J, Janeway CA Jr. Truncation variants of peptides isolated from MHC class II molecules suggest sequence motifs. *Nature* (1992) 359(6394):429–31. doi:10.1038/359429a0
29. Chicz RM, Urban RG, Gorga JC, Vignali DA, Lane WS, Strominger JL. Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J Exp Med* (1993) 178(1):27–47. doi:10.1084/jem.178.1.27
30. Munthe LA, Corthay A, Os A, Zangani M, Bogen B. Systemic autoimmune disease caused by autoreactive B cells that receive chronic help from Ig V region-specific T cells. *J Immunol* (2005) 175(4):2391–400. doi:10.4049/jimmunol.175.4.2391
31. Holmoy T, Vandvik B, Vartdal F. T cells from multiple sclerosis patients recognize immunoglobulin G from cerebrospinal fluid. *Mult Scler* (2003) 9(3):228–34. doi:10.1191/1352458503ms9060a
32. Holmoy T, Fredriksen AB, Thompson KM, Hestvik AL, Bogen B, Vartdal F. Cerebrospinal fluid T cell clones from patients with multiple sclerosis: recognition of idiotopes on monoclonal IgG secreted by autologous cerebrospinal fluid B cells. *Eur J Immunol* (2005) 35(6):1786–94. doi:10.1002/eji.200425417
33. Hestvik AL, Vartdal F, Fredriksen AB, Thompson KM, Kvale EO, Skorstad G, et al. T cells from multiple sclerosis patients recognize multiple epitopes on self-IgG. *Scand J Immunol* (2007) 66(4):393–401. doi:10.1111/j.1365-3083.2007.01955.x
34. Hestvik AL, Skorstad G, Vartdal F, Holmoy T. Idiotope-specific CD4(+) T cells induce apoptosis of human oligodendrocytes. *J Autoimmun* (2009) 32(2):125–32. doi:10.1016/j.jaut.2009.01.004
35. Boyd SD, Crowe JE Jr. Deep sequencing and human antibody repertoire analysis. *Curr Opin Immunol* (2016) 40:103–9. doi:10.1016/j.coi.2016.03.008
36. Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and statistical analysis of adaptive immune repertoires. *Trends Immunol* (2015) 36(11):738–49. doi:10.1016/j.it.2015.09.006
37. Bremel RD, Homan EJ. Frequency patterns of T-cell exposed amino acid motifs in immunoglobulin heavy chain peptides presented by MHCs. *Front Immunol* (2014) 5:541. doi:10.3389/fimmu.2014.00541
38. Larimore K, McCormick MW, Robins HS, Greenberg PD. Shaping of human germline IgH repertoires revealed by deep sequencing. *J Immunol* (2012) 189(6):3221–30. doi:10.4049/jimmunol.1201303
39. Hollenbach JA, Madbouly A, Gragert L, Vierra-Green C, Flesch S, Spellman S, et al. A combined DPA1~DPB1 amino acid epitope is the primary unit of selection on the HLA-DP heterodimer. *Immunogenetics* (2012) 64(8):559–69. doi:10.1007/s00251-012-0615-3
40. Alamyar E, Duroux P, Lefranc MP, Giudicelli V. IMGT(R) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol Biol* (2012) 882:569–604. doi:10.1007/978-1-61779-842-9\_32
41. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res* (2008) 36(Web Server issue):W503–8. doi:10.1093/nar/gkn316
42. DeWitt WS, Lindau P, Snyder TM, Sherwood AM, Vignali M, Carlson CS, et al. A public database of memory and naive B-cell receptor sequences. *PLoS One* (2016) 11(8):e0160853. doi:10.1371/journal.pone.0160853
43. DeWitt WS, Lindau P, Snyder TM, Sherwood AM, Vignali M, Carlson CS, et al. Data from: A Public Database of Memory and Naive B-Cell Receptor Sequences. Dryad Data Repository (2016). doi:10.5061/dryad.35ks2
44. Lefranc MP, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, et al. IMGT, the international ImmunoGeneTics information system. *Nucleic Acids Res* (2009) 37(Database issue):D1006–12. doi:10.1093/nar/gkn838
45. Biniossek ML, Nagler DK, Becker-Pauly C, Schilling O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. *J Proteome Res* (2011) 10(12):5363–73. doi:10.1021/pr200621z
46. Bremel RD, Homan EJ. Recognition of higher order patterns in proteins: immunological kernels. *PLoS One* (2013) 8(7):e70115. doi:10.1371/journal.pone.0070115
47. Chicz RM, Urban RG, Lane WS, Gorga JC, Stern LJ, Vignali DAA, et al. Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature* (1992) 358(6389):764–8. doi:10.1038/358764a0
48. Bremel RD, Homan EJ. An integrated approach to epitope analysis II: a system for proteomic-scale prediction of immunological characteristics. *Immunome Res* (2010) 6:8. doi:10.1186/1745-7580-6-8
49. Bremel RD, Homan EJ. An integrated approach to epitope analysis I: dimensional reduction, visualization and prediction of MHC binding using amino acid principal components and regression approaches. *Immunome Res* (2010) 6:7. doi:10.1186/1745-7580-6-7
50. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol* (2006) 24:419–66. doi:10.1146/annurev.immunol.23.021704.115658
51. Calis JJ, de Boer RJ, Kesmir C. Degenerate T-cell recognition of peptides on MHC molecules creates large holes in the T-cell repertoire. *PLoS Comput Biol* (2012) 8(3):e1002412. doi:10.1371/journal.pcbi.1002412
52. Bogen B, Ruffini P. Review: to what extent are T cells tolerant to immunoglobulin variable regions? *Scand J Immunol* (2009) 70(6):526–30. doi:10.1111/j.1365-3083.2009.02340.x

53. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature* (2012) 486(7402):222–7. doi:10.1038/nature11053
54. Consortium U. UniProt: a hub for protein information. *Nucleic Acids Res* (2015) 43(D1):D204–12. doi:10.1093/nar/gku989
55. The NIH HMPWG, Peterson J, Garges S, Giovanni M, McInnes P, Wang L, et al. The NIH human microbiome project. *Genome Res* (2009) 19(12):2317–23. doi:10.1101/gr.096651.109
56. Bremel RD, Homan EJ. Extensive T-cell epitope repertoire sharing among human proteome, gastrointestinal microbiome, and pathogenic bacteria: implications for the definition of self. *Front Immunol* (2015) 6:538. doi:10.3389/fimmu.2015.00538
57. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol* (1995) 57(1):289–300. doi:10.2307/2346101
58. Owens GP, Kannus 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(2):236–43. doi:10.1002/ana.410430214
59. von Budingen HC, Kuo TC, Sirota M, van Belle CJ, Apeltsin L, Glanville J, et al. B cell exchange across the blood-brain barrier in multiple sclerosis. *J Clin Invest* (2012) 122(12):4533–43. doi:10.1172/jci63842
60. Boyd SD, Gaeta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, et al. Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J Immunol* (2010) 184(12):6986–92. doi:10.4049/jimmunol.1000445
61. Nakagawa TY, Rudensky AY. The role of lysosomal proteinases in MHC class II-mediated antigen processing and presentation. *Immunol Rev* (1999) 172:121–9. doi:10.1111/j.1600-065X.1999.tb01361.x
62. Villadangos JA, Bryant RA, Deussing J, Driessen C, Lennon-Dumenil AM, Riese RJ, et al. Proteases involved in MHC class II antigen presentation. *Immunol Rev* (1999) 172:109–20. doi:10.1111/j.1600-065X.1999.tb01360.x
63. Roche PA, Furuta K. The ins and outs of MHC class II-mediated antigen processing and presentation. *Nat Rev Immunol* (2015) 15(4):203–16. doi:10.1038/nri3818
64. Honey K, Rudensky AY. Lysosomal cysteine proteases regulate antigen presentation. *Nat Rev Immunol* (2003) 3(6):472–82. doi:10.1038/nri1110
65. Hsing LC, Rudensky AY. The lysosomal cysteine proteases in MHC class II antigen presentation. *Immunol Rev* (2005) 207:229–41. doi:10.1111/j.0105-2896.2005.00310.x
66. Haves-Zburuf D, Paperna T, Gour-Lavie A, Mandel I, Glass-Marmor L, Miller A. Cathepsins and their endogenous inhibitors cystatins: expression and modulation in multiple sclerosis. *J Cell Mol Med* (2011) 15(11):2421–9. doi:10.1111/j.1582-4934.2010.01229.x
67. Mahurkar S, Suppiah V, O'Doherty C. Pharmacogenomics of interferon beta and glatiramer acetate response: a review of the literature. *Autoimmun Rev* (2014) 13(2):178–86. doi:10.1016/j.autrev.2013.10.012
68. Foti Cuzzola V, Palella E, Celi D, Barresi M, Giacoppo S, Bramanti P, et al. Pharmacogenomic update on multiple sclerosis: a focus on actual and new therapeutic strategies. *Pharmacogenomics J* (2012) 12(6):453–61. doi:10.1038/tpj.2012.41
69. Beck H, Schwarz G, Schroter CJ, Deeg M, Baier D, Stevanovic S, et al. Cathepsin S and an asparagine-specific endoprotease dominate the proteolytic processing of human myelin basic protein in vitro. *Eur J Immunol* (2001) 31(12):3726–36. doi:10.1002/1521-4141(200112)31:12<3726::AID-IMMU3726>3.0.CO;2-O
70. Kong Y, Chung YB, Cho SY, Kang SY. Cleavage of immunoglobulin G by excretory-secretory cathepsin S-like protease of *Spirometra mansoni* plerocercoid. *Parasitology* (1994) 109(Pt 5):611–21. doi:10.1017/S0031182000076496
71. Fehr K, LoSpalluto J, Ziff M. Degradation of immunoglobulin G by lysosomal acid proteases. *J Immunol* (1970) 105(4):973–83.
72. Kirschke H. Cathepsin S A2. In: Salvesen G, Rawlings ND, editors. *Handbook of Proteolytic Enzymes*, Chap. 413. London: Academic Press (2013). p. 1824–30.
73. van Kasteren SI, Overkleef HS. Endo-lysosomal proteases in antigen presentation. *Curr Opin Chem Biol* (2014) 23:8–15. doi:10.1016/j.cbpa.2014.08.011
74. Liu H, May K. Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function. *MAbs* (2012) 4(1):17–23. doi:10.4161/mabs.4.1.18347
75. Khodadoust MS, Olsson N, Wagar LE, Haabeth OAW, Chen B, Swaminathan K, et al. Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. *Nature* (2017) 543(7647):723–7. doi:10.1038/nature21433
76. Lanzavecchia A. Receptor-mediated antigen uptake and its effect on antigen presentation to class II-restricted T lymphocytes. *Annu Rev Immunol* (1990) 8:773–93. doi:10.1146/annurev.iy.08.040190.004013
77. Yamano T, Nedjic J, Hinterberger M, Steinert M, Koser S, Pinto S, et al. Thymic B cells are licensed to present self antigens for central T cell tolerance induction. *Immunity* (2015) 42(6):1048–61. doi:10.1016/j.immuni.2015.05.013
78. Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol* (2014) 14(6):377–91. doi:10.1038/nri3667
79. Bogen B, Jorgensen T, Hannestad K. T helper cell recognition of idiotopes on lambda 2 light chains of M315 and T952: evidence for dependence on somatic mutations in the third hypervariable region. *Eur J Immunol* (1985) 15(3):278–81. doi:10.1002/eji.1830150313
80. Bogen B, Dembic Z, Weiss S. Clonal deletion of specific thymocytes by an immunoglobulin idiotype. *EMBO J* (1993) 12(1):357–63.
81. Lauritszen GF, Hofgaard PO, Schenck K, Bogen B. Clonal deletion of thymocytes as a tumor escape mechanism. *Int J Cancer* (1998) 78(2):216–22. doi:10.1002/(SICI)1097-0215(19981005)78:2<216:AID-IJC16>3.0.CO;2-8
82. Yamano T, Steinert M, Klein L. Thymic B cells and central T cell tolerance. *Front Immunol* (2015) 6:376. doi:10.3389/fimmu.2015.00376
83. Louveau A, Smirnov I, Keyes TJ, Eccles JD, Rouhani SJ, Peske JD, et al. Structural and functional features of central nervous system lymphatic vessels. *Nature* (2015) 523(7560):337–41. doi:10.1038/nature14432
84. Serafini B, Rosicarelli B, Magliozzi R, Stigliano E, Aloisi F. Detection of ectopic B-cell follicles with germinal centers in the meninges of patients with secondary progressive multiple sclerosis. *Brain Pathol* (2004) 14(2):164–74. doi:10.1111/j.1750-3639.2004.tb00049.x
85. Jarius S, Eichhorn P, Franciotta D, Peteret HF, Akman-Demir G, Wick M, et al. The MRZ reaction as a highly specific marker of multiple sclerosis: re-evaluation and structured review of the literature. *J Neurol* (2017) 264(3):453–66. doi:10.1007/s00415-016-8360-4
86. Hassan-Smith G, Durant L, Tsentemeidou A, Assi LK, Faint JM, Kalra S, et al. High sensitivity and specificity of elevated cerebrospinal fluid kappa free light chains in suspected multiple sclerosis. *J Neuroimmunol* (2014) 276(1–2):175–9. doi:10.1016/j.jneuroim.2014.08.003
87. Poyraz T, Kaya D, Idiman E, Cevik S, Karabay N, Arslan D, et al. What does an isolated cerebrospinal fluid monoclonal band mean: a tertiary centre experience. *Neurology* (2015) 84(14 Suppl):P5.247.
88. Sandberg-Wollheim M, Vandvik B, Nadj K, Norrby E. The intrathecal immune response in the early stage of multiple sclerosis. *J Neurol Sci* (1987) 81(1):45–53. doi:10.1016/0022-510X(87)90182-1
89. Zhang J, Markovic-Plese S, Lacet B, Raus J, Weiner HL, Hafler DA. Increased frequency of interleukin 2-responsive T cells specific for myelin basic protein and proteolipid protein in peripheral blood and cerebrospinal fluid of patients with multiple sclerosis. *J Exp Med* (1994) 179(3):973–84. doi:10.1084/jem.179.3.973
90. Williams WM, Staines NA, Muller S, Isenberg DA. Human T cell responses to autoantibody variable region peptides. *Lupus* (1995) 4(6):464–71. doi:10.1177/096120339500400608
91. Blank M, Shoenfeld Y. The story of the 16/6 idiotype and systemic lupus erythematosus. *Isr Med Assoc J* (2008) 10(1):37–9.
92. Dayan M, Segal R, Stoeber Z, Waisman A, Brosh N, Elkayam O, et al. Immune response of SLE patients to peptides based on the complementarity determining regions of a pathogenic anti-DNA monoclonal antibody. *J Clin Immunol* (2000) 20(3):187–94. doi:10.1023/A:1006685413157
93. Peen E, Malone C, Myers C, Williams RC Jr, Peck AB, Csernok E, et al. Amphipathic variable region heavy chain peptides derived from monoclonal human Wegener's anti-PR3 antibodies stimulate lymphocytes from patients with Wegener's granulomatosis and microscopic polyangiitis. *Clin Exp Immunol* (2001) 125(2):323–31. doi:10.1046/j.1365-2249.2001.01482.x

94. van Schooten WC, Devereux D, Ho CH, Quan J, Aguilar BA, Rust CJ. Joint-derived T cells in rheumatoid arthritis react with self-immunoglobulin heavy chains or immunoglobulin-binding proteins that copurify with immunoglobulin. *Eur J Immunol* (1994) 24(1):93–8. doi:10.1002/eji.1830240115

**Conflict of Interest Statement:** RB and JH hold equity in EigenBio, the company responsible for designing the bioinformatics models used in this project. HR is an equity holder in Adaptive Biotechnologies. The two companies are independent. All other authors declare that the research was conducted in the absence of any

commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Høglund, Lossius, Johansen, Homan, Benth, Robins, Bogen, Bremel and Holmøy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*