

Genetic loci for Epstein-Barr virus nuclear antigen-1 are associated with risk of multiple sclerosis

Yuan Zhou, Gu Zhu, Jac C Charlesworth, Steve Simpson Jr., Rohina Rubicz, Harald HH Göring, Nikolaos A Patsopoulos, Caroline Laverty, Feitong Wu, Anjali Henders, Jonathan J Ellis, Ingrid van der Mei, Grant W Montgomery, John Blangero, Joanne E Curran, Matthew P Johnson, Nicholas G Martin, Dale R Nyholt and Bruce V Taylor; ANZgene consortium

Abstract

Background: Infection with the Epstein-Barr virus (EBV) is associated with an increased risk of multiple sclerosis (MS).

Objective: We sought genetic loci influencing EBV nuclear antigen-1 (EBNA-1) IgG titers and hypothesized that they may play a role in MS risk.

Methods: We performed a genome-wide association study (GWAS) of anti-EBNA-1 IgG titers in 3599 individuals from an unselected twin family cohort, followed by a meta-analysis with data from an independent EBNA-1 GWAS. We then examined the shared polygenic risk between the EBNA-1 GWAS (effective sample size (N_{eff})=5555) and a large MS GWAS (N_{eff} =15,231).

Results: We identified one locus of strong association within the human leukocyte antigen (HLA) region, of which the most significantly associated genotyped single nucleotide polymorphism (SNP) was rs2516049 ($p=4.11 \times 10^{-9}$). A meta-analysis including data from another EBNA-1 GWAS in a cohort of Mexican-American families confirmed that rs2516049 remained the most significantly associated SNP ($p=3.32 \times 10^{-20}$). By examining the shared polygenic risk, we show that the genetic risk for elevated anti-EBNA-1 titers is positively correlated with the development of MS, and that elevated EBNA-1 titers are not an epiphenomena secondary to MS. In the joint meta-analysis of EBNA-1 titers and MS, loci at 1p22.1, 3p24.1, 3q13.33, and 10p15.1 reached genome-wide significance ($p < 5 \times 10^{-8}$).

Conclusions: Our results suggest that apart from the confirmed HLA region, the association of anti-EBNA-1 IgG titer with MS risk is also mediated through non-HLA genes, and that studies aimed at identifying genetic loci influencing EBNA immune response provides a novel opportunity to identify new and characterize existing genetic risk factors for MS.

Keywords: Epstein-Barr virus, EBNA-1, multiple sclerosis, GWAS, polygenic risk, non-HLA

Date received: 17 July 2015; revised: 9 November 2015; accepted: 2 December 2015

Introduction

Epstein-Barr virus (EBV) is a double-stranded DNA human herpesvirus also known as HHV-4, which infects more than 90% of the world's adults and persists as a lifelong latent infection.¹ It primarily infects B-lymphocytes and permanently transforms them into latently infected lymphoblastoid cell lines,

which constantly produce a group of viral proteins, including EBV nuclear antigens (EBNAs) and latent membrane proteins.² Infection with EBV has been associated with the development of autoimmune diseases including multiple sclerosis (MS), systemic lupus erythematosus, and rheumatoid arthritis.³ Despite the high prevalence of EBV infection, only

Multiple Sclerosis Journal

1–10

DOI: 10.1177/
1352458515626598

© The Author(s), 2016.
Reprints and permissions:
[http://www.sagepub.co.uk/
journalsPermissions.nav](http://www.sagepub.co.uk/journalsPermissions.nav)

Correspondence to:
BV Taylor
Menzies Institute for Medical
Research, University of
Tasmania, 17 Liverpool
Street, Hobart, TAS 7000,
Australia.

Bruce.Taylor@utas.edu.au
DR Nyholt and BV Taylor
jointly directed this research.

Yuan Zhou
Jac C Charlesworth
Steve Simpson Jr.
Feitong Wu
Ingrid van der Mei
Bruce V Taylor
Menzies Institute for Medical
Research, University of
Tasmania, Hobart, TAS,
Australia

Gu Zhu
Anjali Henders
Jonathan J Ellis
Grant W Montgomery
Nicholas G Martin
Department of Genetics and
Computational Biology,
QIMR Berghofer Medical
Research Institute, Brisbane,
QLD, Australia

Rohina Rubicz
Division of Public Health
Sciences, Fred Hutchinson
Cancer Research Center,
Seattle, WA, USA

Harald HH Göring
John Blangero
Joanne E Curran
Matthew P Johnson
South Texas Diabetes and
Obesity Institute, University
of Texas Rio Grande
Valley School of Medicine,
Brownsville, TX, USA

Nikolaos A Patsopoulos
Program in Translational
Neuropsychiatric
Genomics, Institute for the
Neurosciences, Department
of Neurology, Brigham and
Women's Hospital, Boston,
MA, USA/Harvard Medical
School, Boston, MA, USA/
Broad Institute, Cambridge,
MA, USA

Caroline Laverty
Monash Antibody
Technologies Facility,
Monash University,
Melbourne, VIC, Australia

Dale R Nyholt

Statistical and Genomic Epidemiology Laboratory, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, Australia/Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

a small number of infected individuals will develop one of these diseases.⁴

EBV nuclear antigen-1 (EBNA-1) may act as an important viral antigen in MS pathogenesis.^{5,6} Recent work has suggested that human leukocyte antigen (HLA)-dependent and independent immune responses to EBNA-1 are important drivers of MS pathogenesis and may modulate effects on MS risk of protective and deleterious HLA antigens.⁷ Additionally, EBNA-1-specific T-cells that cross-react with myelin have been demonstrated,⁸ and anti-EBNA-1 antibodies have been described in cerebrospinal fluid oligoclonal bands in MS cases.⁹ Higher anti-EBNA-1 titers have been associated with a worse outcome in MS, both clinically¹⁰ and radiologically.¹¹

A number of studies have observed an association between anti-EBNA-1 IgG titers and MS risk.^{6,12,13} In a study of US military personnel who were anti-EBNA-1 IgG positive at study entry, the relative risk of MS over an average follow-up of 5 years was 36 times higher among those with high anti-EBNA titers compared with those with low titers. Importantly, in this study, in blood samples collected before the age of 20 years, the mean anti-EBNA-1 IgG titers were identical between participants who later developed MS and age- and sex-matched controls that remained healthy. However, after the age of 20 years, the anti-EBNA-1 IgG titers in those that later developed MS increased four-fold, while that of controls remained constant.¹² Similarly, in 10 cases who were seronegative for anti-EBNA-1 IgG at their first blood sample who subsequently developed MS, all became seropositive generally within a few months of MS onset.¹⁴ These findings suggest that the elevation in anti-EBNA-1 IgG titers may either directly predispose to the development of MS or may be a biomarker of immune dysregulation that results in an increased risk of MS.¹⁵

Recent work in Mexican-American families¹⁶ suggested that particular HLA variants are associated with elevated anti-EBNA-1 IgG titers. Thus, we hypothesized that an altered immune response to EBV infection—manifested by increased anti-EBNA-1 IgG titers or a potentiated inflammatory response to EBV infection—may be influenced by genetic factors, and that these genetic variants may also predispose to MS.

Materials and methods

A flow chart of the study design is shown in Figure 1.

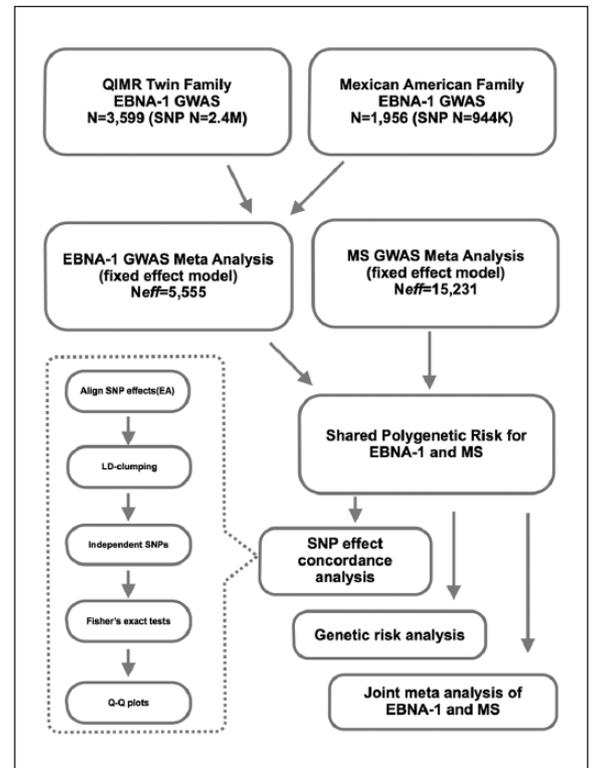


Figure 1. Study flowchart detailing cohorts used.

Genome-wide association analysis

Queensland Institute of Medical Research (QIMR) twin families EBNA-1 genome-wide association study (QTFEGWAS). In all, 3760 individuals from 1020 families in the QIMR Twin Family Cohort Study were studied. Details of the clinical protocol for blood collection and processing have been described elsewhere.^{17,18} For a more detailed description of the sample, please see Table S1. The ImmunoWELL EBNA IgG Test kit (GenBio, San Diego, CA, USA) was used for detection of anti-EBNA-1 IgG in human serum ($N=3760$) at the Monash Antibody Technologies Facility, Monash University. The following values were used to determine the status for seropositive/seronegative according to the manufacturer's protocol: negative (<200 units/mL); equivocal (200–300 units/mL); and positive (>300 units/mL). To check which data transformation should be used for analysis, the distribution of standardized residuals was examined by treating anti-EBNA-1 IgG titers as both continuous and categorical values (four groups: <250 , 250–1001, 1001–2000, and >2000 units/mL). The residuals were derived from generalized linear model (GLM) analysis using box effect, row effect, and column effect as fixed factors while also controlling for sex, age, age \times sex, age², and age² \times sex. Since

the residuals were normally distributed (Figure S1), we utilized these modeled residuals as continuous outcomes and predictors in place of as-measured anti-EBNA-1 IgG.

Genome-wide genotyping was performed using the Illumina 610-Quad BeadChip and imputed to extend the genomic coverage to 2.4 million single nucleotide polymorphisms (SNPs).¹⁹ As part of quality control, individuals from families with pedigree errors and a call rate <0.95 were excluded. SNPs with a call rate <0.95 or with a low imputation quality score ($R^2 < 0.3$), minor allele frequency <0.01 in the population, or significant deviation from Hardy–Weinberg equilibrium ($p_{\text{HWE}} < 1.0 \times 10^{-6}$), were excluded. In total, the data from 3599 individuals from 992 families and 2,428,106 SNPs passed that quality control were used for further analysis.

The association analysis of imputed dosage scores was performed using a family-based score test implemented in Merlin-offline²⁰ which corrects for relatedness of twins and family members. To detect whether there were independent effects at significant loci, we carried out conditional analysis by including the most significant SNP as a covariate.

Mexican-American families EBNA-1 GWAS (MAFEGWAS). We used summary results from the MAFEGWAS.¹⁶ In brief, the SNPs were typed on several versions of Illumina's microarrays and quality controls were done according to standard requirements. A total of 944,565 genotyped SNPs from 1956 individuals were available for analysis. The anti-EBNA-1 quantitative antibody titer level was used for genome-wide joint linkage and association analysis in SOLAR.²¹

GWAS meta-analysis

EBNA-1 GWAS meta-analysis (EGMA). Summary results from the two EBNA-1 GWAS (QTPEGWAS ($N=3599$) and MAFEGWAS ($N=1956$)) were used to carry out a meta-analysis using an inverse variance fixed-effect model in META.²² To facilitate comparison with p values for all 110 established non-HLA risk loci recently published by the International Multiple Sclerosis Genetics Consortium,²³ we used the Direct Imputation of summary Statistics (DIST)²⁴ program, to directly impute EGMA summary statistics based on the 1000 genome project reference data. The DIST is achieved by employing a conditional expectation formula for multivariate normal variants and using the correlation structure from a relevant

reference population. Default and recommended parameters of DIST were utilized in this study.

MS GWAS meta-analysis (MGMA). The MGMA was comprised of data from seven datasets of non-overlapping case and control subjects of European descent. A total of 5545 cases and 12,153 controls with 2,529,394 SNPs were used for analysis. For a detailed description, see Patsopoulos *et al.*²⁵

Shared polygenic risk between EBNA-1 GWAS and MS GWAS

To test whether the SNPs associated with anti-EBNA-1 IgG titers may also contribute to MS risk, we first matched the genome-wide significant SNPs ($p < 5 \times 10^{-8}$) in the EGMA with MGMA results. We also matched EGMA p values for all 110 established non-HLA risk loci recently published.²³ We then carried out SNP effect concordant analysis (SECA) and genetic risk score (GRS) analysis to further examine the shared polygenic risk between the EBNA-1 GWAS and the MS GWAS, including and excluding the extended HLA region (chr6: 25–35 Mb).

SNP effect concordant analysis. SECA was undertaken using the EGMA and MGMA, utilizing one study as a discovery sample while the other acted as a target sample, and vice-versa. Subsets of independent autosomal SNPs were extracted via linkage disequilibrium (LD) clumping. The approach iterated from the first to last SNP on each chromosome sorted from smallest (most significant) to largest p value (less significant) in the discovery sample that had not already been clumped (denoting this as the index SNP) and formed clumps of all other SNPs that were within 10 Mb of and not in LD with the index SNP ($r^2 > 0.1$) based on HapMap2 CEU genotype data. In total, for autosomal SNPs including the HLA region, there were 24,722 SNPs clumped using the EGMA as a discovery sample, and 24,642 SNPs clumped using the MGMA as a discovery sample. For autosomal SNPs excluding the HLA region, there were 24,688 SNPs clumped using the EGMA as a discovery sample, and 24,621 SNPs clumped using the MGMA as a discovery sample. To test for concordant SNP effects between the two datasets, a total of 144 key SNP subsets were examined, each generated using 12 p value thresholds ($p=0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, \text{ and } 1.0$). Fisher's exact test was applied to test whether the overlapping effects were in the same direction after conditioning on their GWAS p value, where a Fisher's exact test p value ($p_{\text{FT}} < 0.05$) means the SNP effects are nominally correlated. In

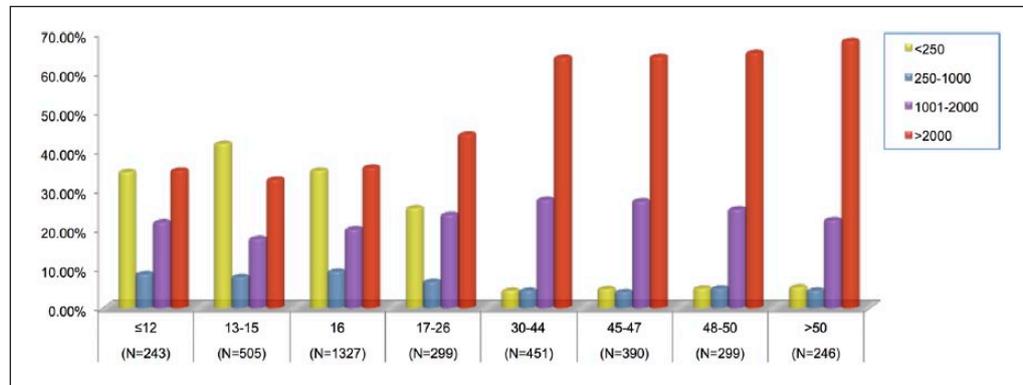


Figure 2. Anti-EBNA-1 IgG titer distribution in different age ranges for twins and their parents. For each age range, anti-EBNA-1 IgG titers are divided into four groups (from left to right: <250, 250–1000, 1001–2000, and > 2000).

all, 1000 permutations were performed to examine the significance of observing a specific proportion of subsets with correlated SNP effects—adjusted for testing all 144 subsets ($p_{F\text{Tsig-permuted}}$).²⁶

GRS analysis. To further validate the SECA results, GRSs²⁷ were constructed using the independent SNPs separated by <10Mb that are not in LD ($r^2 < 0.1$) and selected as having the smallest p value in the discovery sample with a nominally significant $p < 0.05$, and in the target sample with $p < 1.0$. GRS analysis requires individual-level GWAS SNP genotype data for the target sample. As only part of the genotype data for the EGMA and MGMA could be obtained, the first GRS analysis (EGMA as a discovery sample) used 5862 clumped SNPs to generate EGMA-based GRSs in the MS GWAS obtained from dbGaP (phs000275, phs000171, and phs000139) and ANZgene (cases=3252; controls=5725). For details on data cleaning, please refer to our previous paper.²⁸ The second GRS analysis (MGMA as a discovery sample) used 5770 clumped SNPs to generate MGMA-based GRSs in the QTFEWAS. For binary disease traits (cases and controls), the GRS was used as a predictor to test whether higher mean GRS were observed for cases (MS) than for controls in the target samples. For quantitative traits, the GRS was used to test the correlation between GRS and the target phenotype (anti-EBNA-1 IgG titers).

Joint EBNA-1 and MS meta-analysis. To identify genetic risk factors contributing to both anti-EBNA-1 IgG titer and MS risk outside the HLA region, a joint meta-analysis of EGMA results and MGMA results was carried out using a sample size-weighted method based on p values in METAL.²⁹

Gene–gene interaction analysis. To check whether interactions existed between the loci that were found

to contribute to both anti-EBNA-1 IgG titer and MS risk, we defined two logistic regression models, one taking into account only the marginal effects of both SNPs while the other was a full logistic regression model including a pairwise interaction between the two SNPs. A likelihood ratio test with four degrees of freedom was then used to test whether there was any statistical difference between the two models.³⁰ The datasets used for the analysis were the MS GWAS obtained from dbGaP and ANZgene listed above (cases=3252; controls=5725).

Results

HLA region as major quantitative trait loci influencing anti-EBNA-1 IgG titer

The distributions of anti-EBNA-1 IgG titers using four groups (<250, 250–1000, 1001–2000, and >2000) in different age ranges are shown in Figure 2. As age increased, the percentage of individuals with higher anti-EBNA-1 IgG titers increased dramatically. Heritability for anti-EBNA-1 IgG titers was 42%. The HLA region on chromosome 6 contained multiple SNPs exceeding the threshold for a significant association with anti-EBNA-1 IgG titers ($p < 5 \times 10^{-8}$) (Figure S2(a)). A Q-Q plot is shown in Figure S3(a). Of these HLA SNPs, the most significant imputed SNP was rs9268923 ($p = 1.22 \times 10^{-11}$), located between the genes *HLA-DRA* and *HLA-DRB5*. The most significant genotyped SNP was rs2516049 ($p = 4.11 \times 10^{-9}$; Table 1), located between the genes *HLA-DRB1* and *HLA-DQA1*. As the HLA region is genetically complex, with a low rate of recombination and long-range LD, we carried out two conditional analyses to check whether there were independent peaks, first by conditioning on rs9268923, and then on both rs9268923 and rs2267647 (Figure 3). However, after conditional

Table 1. Genome-wide association analysis results of QIMR twin families.

SNP	CHR	POS	E(A)	O(A)	SNP type	Score	Parents + children (N=3599)		Parents (N=1243)		Children ^a (N=2356)		Children ^b (N=1488)		
							Effect	Freq effect <i>p</i>	Effect	Effect <i>p</i>	Effect	Effect <i>p</i>	Effect	Effect <i>p</i>	
rs9268923	6	32540813	C	T	Imputed	0.905	0.552	0.180	1.22 × 10 ⁻¹¹	0.252	1.37 × 10 ⁻¹³	0.148	2.93 × 10 ⁻⁰⁵	0.225	9.72 × 10 ⁻¹²
rs9268969	6	32542327	C	T	Imputed	0.905	0.552	0.180	1.22 × 10 ⁻¹¹	0.252	1.37 × 10 ⁻¹³	0.148	2.93 × 10 ⁻⁰⁵	0.225	9.72 × 10 ⁻¹²
rs6916742	6	32561169	C	T	Imputed	0.879	0.569	-0.176	3.12 × 10 ⁻¹¹	-0.200	2.40 × 10 ⁻⁰⁹	-0.160	5.95 × 10 ⁻⁰⁶	-0.215	5.05 × 10 ⁻¹¹
rs2516049	6	32678378	T	C	Genotyped	0.999	0.666	0.162	4.11 × 10 ⁻⁰⁹	0.210	1.31 × 10 ⁻⁰⁹	0.142	1.36 × 10 ⁻⁰⁴	0.180	1.68 × 10 ⁻⁰⁷
rs522308	6	32689900	C	T	Imputed	0.989	0.664	0.159	8.56 × 10 ⁻⁰⁹	0.207	1.89 × 10 ⁻⁰⁹	0.138	2.03 × 10 ⁻⁰⁴	0.178	2.36 × 10 ⁻⁰⁷
rs11757159	6	32628250	C	T	Imputed	0.972	0.636	0.154	1.24 × 10 ⁻⁰⁸	0.206	1.44 × 10 ⁻⁰⁹	0.130	3.59 × 10 ⁻⁰⁴	0.170	4.07 × 10 ⁻⁰⁷
rs9405112	6	32553578	G	A	Imputed	0.977	0.662	0.156	1.57 × 10 ⁻⁰⁸	0.215	8.08 × 10 ⁻¹⁰	0.128	5.30 × 10 ⁻⁰⁴	0.176	2.38 × 10 ⁻⁰⁷
rs9268853	6	32537621	T	C	Imputed	0.989	0.657	0.152	3.30 × 10 ⁻⁰⁸	0.206	3.36 × 10 ⁻⁰⁹	0.125	6.96 × 10 ⁻⁰⁴	0.172	4.64 × 10 ⁻⁰⁷
rs9268858	6	32537736	T	C	Imputed	0.989	0.657	0.152	3.30 × 10 ⁻⁰⁸	0.206	3.36 × 10 ⁻⁰⁹	0.125	6.96 × 10 ⁻⁰⁴	0.172	4.64 × 10 ⁻⁰⁷
rs9268877	6	32539125	G	A	Imputed	0.977	0.585	-0.146	3.35 × 10 ⁻⁰⁸	-0.184	3.46 × 10 ⁻⁰⁸	-0.130	2.33 × 10 ⁻⁰⁴	-0.176	7.00 × 10 ⁻⁰⁸
rs7739203	6	32549533	G	A	Imputed	0.973	0.585	-0.145	3.90 × 10 ⁻⁰⁸	-0.187	1.99 × 10 ⁻⁰⁸	-0.129	2.65 × 10 ⁻⁰⁴	-0.175	8.13 × 10 ⁻⁰⁸
rs7739357	6	32549619	G	A	Imputed	0.973	0.585	-0.145	3.90 × 10 ⁻⁰⁸	-0.187	1.99 × 10 ⁻⁰⁸	-0.129	2.65 × 10 ⁻⁰⁴	-0.175	8.13 × 10 ⁻⁰⁸

E(A) refers to the effect allele; O(A) refers to the opposite allele; score refers to the imputed or genotype score of the SNP; Freq refers to the E(A) frequency.

Effect corresponds to standard deviation units for the transformed phenotype.

^aChildren refers to using all the children samples.

^bChildren refers to only for anti-EBNA-1 IgG titers positive children.

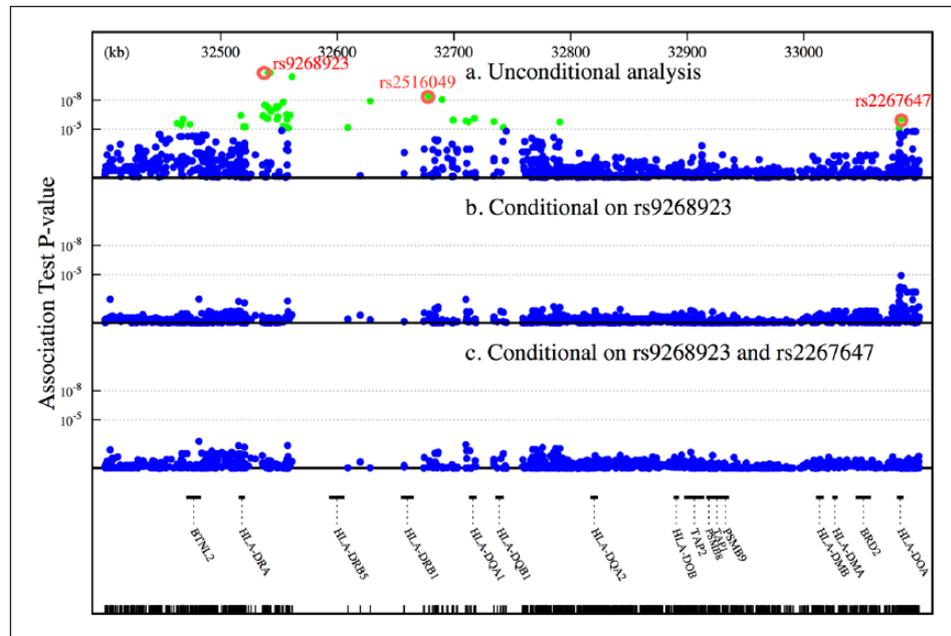


Figure 3. Association analysis of HLA region and EBNA-1 titers in QIMR Twin Families EBNA-1 GWAS ($N=3599$). Results are shown (a) unconditional analysis, (b) conditional analysis on rs9268923, and (c) conditional analysis on rs9268923 and rs2267647. The top significant genotyped SNP rs2516049 was also highlighted in red. SNPs with a $p < 1.0 \times 10^{-5}$ are highlighted in green.

analysis, no other SNPs showed a significant association, suggesting these top SNPs are within one haplotype block that contains all the variants modulating anti-EBNA-1 IgG titers.

Association analyses were also conducted separately for parents and children (Table S1). Analysis of parents only ($N=1243$), implicated the same SNPs in the HLA region (Figure S2(b)), that were found using all samples: imputed SNP rs9268923 ($p=1.37 \times 10^{-13}$) and genotyped SNP rs2516049 ($p=1.31 \times 10^{-9}$; Table 1). In contrast, analysis of all children ($N=2356$), did not identify any significant SNPs in the HLA region (Figure S2(c)). However, once we restricted analyses to only EBNA-1 positive (>300 units/mL) children ($N=1488$), we again observed the same significant associations in the HLA region (Figure S2(d)). Q-Q plots are shown in Figure S3(b)–(d), respectively.

In the EGMA (inflation factor, $\lambda=1.017$; Figure S3(e)), 45 SNPs that reached genome-wide significance were found to be located in the HLA region ($p < 5.0 \times 10^{-8}$; Table S2). Of these, rs2516407 remained the most significant ($p=3.32 \times 10^{-20}$). No other loci were genome-wide significant outside the HLA region (Figure S4). The inflation factors (λ) for the analyses were close to 1, suggesting population stratification had minimal influence on test statistical distribution; therefore, we did not adjust for genomic control in our analysis.

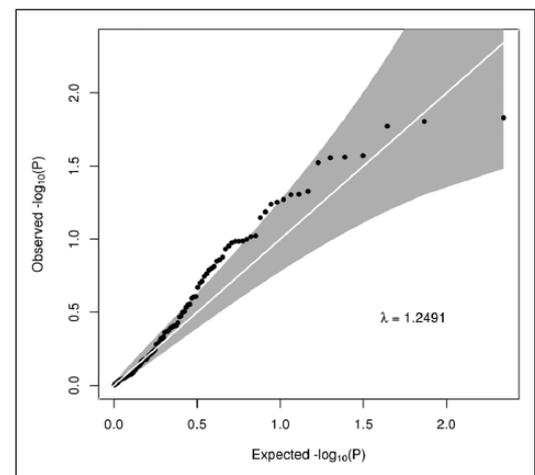


Figure 4. Q-Q plot for 110 non-HLA MS SNPs in the EBNA-1 GWAS meta-Analysis. The 95% confidence envelope (shaded grey) was formed by calculating, for each order statistic (p values ranked from smallest to largest), the 2.5th and 97.5th centiles of the distribution of the order statistic under random sampling and the null hypothesis.

Anti-EBNA-1 IgG titers as possible risk factors for the development of MS: the effect of HLA and non-HLA region genes

The Q-Q plot (Figure 4) shows a significant excess of smaller EGMA p values for the 110 non-HLA MS

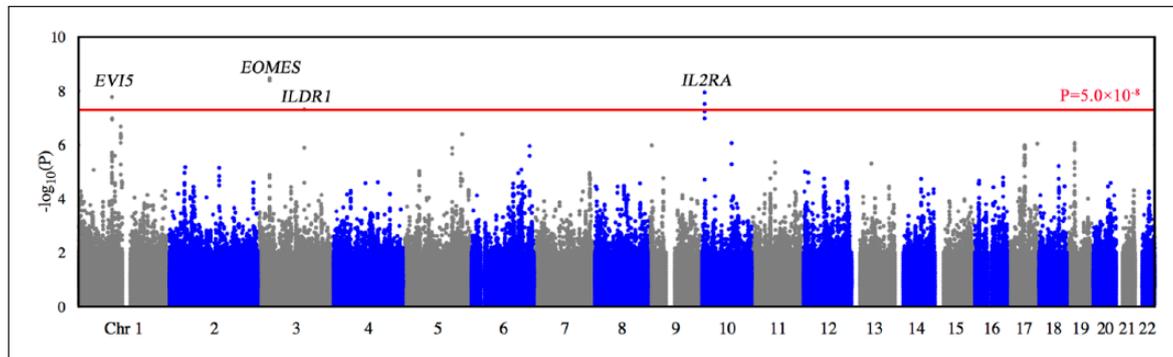


Figure 5. Manhattan plot for the pooled meta-analysis of EGMA and MGMA (excluding the HLA region). The 22 autosomal chromosomes separated by two different colors are displayed on the X-axis. The $-\log_{10}(p)$ values are displayed on the Y-axis. Genome-wide association significance level is shown as the red line ($p < 5.0 \times 10^{-8}$).

SNPs recently published by the International Multiple Sclerosis Genetics Consortium²³ than expected by chance ($\lambda = 1.25$). In SECA, using the EGMA as a discovery sample, the clumped SNP effects were positively correlated with MGMA effects, either including ($p_{\text{FTsig-permuted}} = 0.026$; 95% confidence interval (CI): 0.018–0.038) or excluding ($p_{\text{FTsig-permuted}} = 0.049$; 95% CI: 0.037–0.064) the HLA region (Figure S5). However, using the MGMA as a discovery sample, the clumped SNP effects were not positively correlated with the EGMA SNP effects, either including ($p_{\text{FTsig-permuted}} = 0.173$; 95% CI: 0.151–0.197) or excluding ($p_{\text{FTsig-permuted}} = 0.157$; 95% CI: 0.136–0.181) the HLA region (Figure S6).

The SECA results (excluding the HLA) were further validated by GRS analysis. A significantly higher GRS, constructed using the EGMA as a discovery sample, was observed in MS cases than controls in the target dbGaP and ANZgene GWAS ($p = 0.007$; Figure S7). In contrast, a GRS constructed using MGMA as a discovery sample, did not correlate with anti-EBNA-1 IgG titers in the target QTfEGWAS samples ($p = 0.270$, correlation coefficient = -0.030 ; Figure S8). Altogether, these results indicate that the genetic risk for elevated anti-EBNA-1 IgG titer is positively correlated with increased MS susceptibility, whereas the reverse does not hold. That is, the genetic risk for MS is not positively correlated with elevated anti-EBNA-1 IgG titer level.

The combined meta-analysis of the EGMA and MGMA results (excluding the HLA region) ($\lambda = 1.046$; Figure S3(f)) showed several SNP associations that reached genome-wide significance ($p < 5.0 \times 10^{-8}$; Table S3). The approximate chromosomal location of the significant SNPs is shown in the Manhattan plot in Figure 5 and labeled with the nearest gene. These

genes include *EVI5*(1p22.1), *EOMES*(3p24.1), *ILDR1*(3q13.33), and *IL2RA*(10p15.1). For detailed regional plots, see Figure S9. Of these, the most interesting gene is *Ecotropic viral integration site 5 (EVI5)* on chromosome 1, in which a missense mutation SNP rs11808092 ($p = 1.71 \times 10^{-8}$) results in an amino acid change from glutamine to histidine. The impact of this amino acid substitution on the structure and function of the protein product was predicted as benign using PolyPhen-2,³¹ with a score of 0.003 (sensitivity: 0.98; specificity: 0.44). Sorting Intolerant From Tolerant (SIFT)³² gave similar results (tolerated), while MutationAssessor³³ scored it as “medium” (2.255).

In the interaction analysis, tagged SNPs within each genome-wide significant peak were selected ($r^2 > 0.65$). In total, we studied five SNPs: rs11808092 (1p22.1), rs427221 (3p24.1), rs2255214 (3q13.33), rs2516049 (6p21.32), and rs12722561 (10p15.1). We observed a suggestive synergistic interaction between rs2516049 and rs11808092 ($p = 0.006$; Table S4).

Discussion

We have undertaken a large GWAS of anti-EBNA-1 IgG titers and provide strong evidence that the HLA region serves as the main quantitative trait locus for anti-EBNA-1 IgG titers. Our results strongly suggest that there are shared genetic risk factors that influence both anti-EBNA-1 IgG titers and MS risk. Our results further indicate that the genetic risk for elevated anti-EBNA-1 IgG titer is positively correlated with increased MS susceptibility but the reverse importantly is not supported, that is, MS risk genetic differences do not increase anti-EBNA-1 titers. In our study, the heritability of anti-EBNA-1 IgG titers is 42%, clearly suggesting that other factors besides genetic differences are also important in MS pathogenesis and

that there is still a clear strong association with prior EBV infection in MS risk modulated by the individual's genetic make-up.

In our results, the strongest non-HLA EBNA-1 association was found for *EVI5*. The major function of *EVI5* is the modulation of cell cycle progression, cytokinesis, and cellular membrane traffic.³⁴ Several studies have described the involvement of *EVI5* in the risk and severity of MS using case-control and cross-sectional study designs. Using the candidate gene approach, Hoppenbrouwers et al. found that SNPs within *EVI5* showed a moderate association with MS risk (odds ratio (OR) = 1.90–2.01; $p = 0.01$).³⁵ The ANZgene consortium found a significant interaction between SNPs in *EVI5* and *HLA-DR15* in predicting MS risk ($p = 0.001$).³⁶ This finding is supported by recent work demonstrating that *EVI5* genotype is associated with a greater odds of having a more severe relapse among individuals who carry the risk allele of *HLA-DRB1*1501*.³⁷ These studies showed associations with MS parameters, but did not consider genetic effects in concert with anti-EBNA-1 IgG levels. Direct evidence of a link between EBV infection and gene expression changes has been provided by the work of van Aalderen et al.,³⁸ who found that *EOMES* expression was altered after EBV infection. Also, in another study by Parnell et al.³⁹ analyzing the expression of MS-associated transcription factors using whole blood, *EOMES* expression was significantly lower in MS cases compared with healthy controls.

Further to the correlation in genetic risk between anti-EBNA-1 IgG titer and MS, the observed excess of smaller EGMA p values for the recently reported 110 non-HLA MS risk SNPs²³ (Figure 3) suggests that other genetic factors likely influence the interaction between MS risk and EBV infection. Hence, analysis of larger EBNA-1 GWAS datasets and/or analysis of other markers of EBNA immune responses (e.g. anti-EBNA-2 or EBNA-3 titers) should provide a novel opportunity to identify new and characterize existing genetic risk factors for MS. Our findings indicate that these studies offer a powerful approach to identify the specific immune responses components that contribute most to MS susceptibility. Indeed, considering the high cost and difficulty in amassing larger samples of MS cases, we suggest focus could switch to performing large studies, such as these, aimed at identifying genetic loci influencing anti-EBNA immune response. Such studies should also provide insight into other EBV-associated autoimmune diseases such as Burkitt's lymphoma and nasopharyngeal carcinoma.

In conclusion, our results suggest that genetic factors influencing differential immune system reactivity to EBV infection increase the risk of MS in genetically predisposed individuals exposed to EBV, most notably in adolescence or adult life.

Acknowledgements

The members of the ANZgene Consortium are as follows: Alan Baxter (School of Pharmacy and Molecular Sciences, James Cook University, Townsville, Australia), Allan Kermode (Department of Neurology, Sir Charles Gairdner Hospital, Nedlands, Australia), Melanie Bahlo (Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Australia), William Carroll (Department of Neurology, Sir Charles Gairdner Hospital, Perth, Australia), Jeanette Lechner-Scott (University of Newcastle, Newcastle, Australia), Jac C. Charlesworth, Bruce V Taylor (Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia), Helmut Butzkueven (Department of Medicine, Royal Melbourne Hospital, Melbourne, Australia), David Booth (Westmead Millennium Institute, University of Sydney, Sydney, Australia), Graeme Stewart (Westmead Millennium Institute, University of Sydney, Sydney, Australia), James Wiley (Howard Florey Institute, University of Melbourne, Melbourne, Australia), Judith Field (Howard Florey Institute, University of Melbourne, Melbourne, Australia), Lotti Tajouri (Genomics Research Centre, Griffith University, Brisbane, Australia), Lyn Griffiths (Genomics Research Centre, Griffith University, Brisbane, Australia), Michael Barnett (Brain and Mind Research Institute, University of Sydney, Sydney, Australia), Pablo Moscato (Hunter Medical Research Institute, Newcastle, Australia), Robert Heard (Westmead Millennium Institute, University of Sydney, Sydney, Australia), Rodney Scott (School of Biomedical Sciences, University of Newcastle, Newcastle, Australia), Shaun McColl (School of Molecular & Biomedical Science, University of Adelaide, Adelaide, Australia), Simon Foote (Australian School of Advanced Medicine, Macquarie University, Sydney, Australia), Simon Broadley (School of Medicine, Griffith University, Gold Coast Campus, Australia), Mark Slee (School of Medicine, Flinders University of South Australia, Adelaide, Australia), Steve Vucic (Western Clinical School, University of Sydney, Sydney, Australia), Trevor Kilpatrick (Centre for Neurosciences, Department of Anatomy and Neuroscience, University of Melbourne, Melbourne, Australia). Authors Dale R Nyholt and Bruce V Taylor jointly directed the research. YZ, GZ, JCC, SSJ, RR, NAP, FW, JE BT and DRN were involved in the analysis and interpretation of data. CL

and AH did EBNA-1 ELISA assay analysis. HHHG, IVDM, GWM, JB, JEC, MPJ, NGM, DRN and BVT were involved in study design and concept. YZ drafted the paper. All authors read and approved the final manuscript.

Conflicts of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

QIMR Twin Family Cohort Study: We thank the Brisbane twins and their families for their participation. Funding: Australia National Health and Medical Research Council (NHMRC; grants 486682, 1009064, 389875). *Mexican-American families EBNA-1 GWAS:* We thank the participants of the San Antonio Family Heart Study and San Antonio Family Diabetes/Gallbladder Study. Funding: NIH R01 grant HL080149, HL045522, HL080149, DK053889, and DK047482. *MS GWAS Meta-Analysis:* We thank all the participants involved in the MS study and thank those who have contributed datasets to dbGaP and the ANZgene consortium for providing access to genotyped data. SSJ was supported by the MS Research Australia Postdoctoral Research Fellowship. DRN was supported by the Australian Research Council (FT0991022) and National Health and Medical Research Council (APP0613674). IvdM is supported by an Australian Research Council Future Fellowship. NAP is supported by a Career Independence Award from the National Multiple Sclerosis Society.

References

- Lucas RM, Hughes AM, Lay ML, et al. Epstein-Barr virus and multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2011; 82: 1142–1148.
- Young LS and Rickinson AB. Epstein-Barr virus: 40 years on. *Nat Rev Cancer* 2004; 4: 757–768.
- Pender MP. CD8+ T-cell deficiency, Epstein-Barr virus infection, vitamin D deficiency, and steps to autoimmunity: A unifying hypothesis. *Autoimmun Dis* 2012; 2012: 189096.
- Simpson S Jr, Blizzard L, Othahal P, et al. Latitude is significantly associated with the prevalence of multiple sclerosis: A meta-analysis. *J Neurol Neurosurg Psychiatry* 2011; 82: 1132–1141.
- Pakpoor J, Disanto G, Gerber JE, et al. The risk of developing multiple sclerosis in individuals seronegative for Epstein-Barr virus: A meta-analysis. *Mult Scler* 2013; 19: 162–166.
- Munger KL, Levin LI, O'Reilly EJ, et al. Anti-Epstein-Barr virus antibodies as serological markers of multiple sclerosis: A prospective study among United States military personnel. *Mult Scler* 2011; 17: 1185–1193.
- Strautins K, Tschochner M, James I, et al. Combining HLA-DR risk alleles and anti-Epstein-Barr virus antibody profiles to stratify multiple sclerosis risk. *Mult Scler* 2014; 20: 286–294.
- Lunemann JD, Jelcic I, Roberts S, et al. EBNA1-specific T cells from patients with multiple sclerosis cross react with myelin antigens and co-produce IFN-gamma and IL-2. *J Exp Med* 2008; 205: 1763–1773.
- Cepok S, Zhou D, Srivastava R, et al. Identification of Epstein-Barr virus proteins as putative targets of the immune response in multiple sclerosis. *J Clin Invest* 2005; 115: 1352–1360.
- Lunemann JD, Tintore M, Messmer B, et al. Elevated Epstein-Barr virus-encoded nuclear antigen-1 immune responses predict conversion to multiple sclerosis. *Ann Neurol* 2010; 67: 159–169.
- Farrell RA, Antony D, Wall GR, et al. Humoral immune response to EBV in multiple sclerosis is associated with disease activity on MRI. *Neurology* 2009; 73: 32–38.
- Levin LI, Munger KL, Rubertone MV, et al. Temporal relationship between elevation of Epstein-Barr virus antibody titers and initial onset of neurological symptoms in multiple sclerosis. *JAMA* 2005; 293: 2496–2500.
- Lucas RM, Ponsonby AL, Dear K, et al. Current and past Epstein-Barr virus infection in risk of initial CNS demyelination. *Neurology* 2011; 77: 371–379.
- Levin LI, Munger KL, O'Reilly EJ, et al. Primary infection with the Epstein-Barr virus and risk of multiple sclerosis. *Ann Neurol* 2010; 67: 824–830.
- Ascherio A, Munger KL and Lunemann JD. The initiation and prevention of multiple sclerosis. *Nat Rev Neurol* 2012; 8: 602–612.
- Rubicz R, Yolken R, Drigalenko E, et al. A genome-wide integrative genomic study localizes genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1). *PLoS Genet* 2013; 9: e1003147.
- Zhu G, Montgomery GW, James MR, et al. A genome-wide scan for naevus count: Linkage to CDKN2A and to other chromosome regions. *Eur J Hum Genet* 2007; 15: 94–102.
- Wright MJ and Martin NG. Brisbane Adolescent Twin Study: Outline of study methods and research projects. *Aust J Psychol* 2004; 56: 65–78.
- Ferreira MA, Mangino M, Brumme CJ, et al. Quantitative trait loci for CD4:CD8 lymphocyte ratio

- are associated with risk of type 1 diabetes and HIV-1 immune control. *Am J Hum Genet* 2010; 86: 88–92.
20. Abecasis GR, Cherny SS, Cookson WO, et al. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; 30: 97–101.
 21. Almasy L and Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; 62: 1198–1211.
 22. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* 2010; 42: 436–440.
 23. International Multiple Sclerosis Genetics Consortium; Beecham AH, Patsopoulos NA, Xifara DK, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* 2013; 45: 1353–1360.
 24. Lee D, Bigdeli TB, Riley BP, et al. DIST: Direct imputation of summary statistics for unmeasured SNPs. *Bioinformatics* 2013; 29: 2925–2927.
 25. Patsopoulos NA, Bayer Pharma MS Genetics Working Group, Steering Committees of Studies Evaluating IFN β -1b and a CCR1-Antagonist, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann Neurol* 2011; 70: 897–912.
 26. Nyholt DR. SECA: SNP effect concordance analysis using genome-wide association summary results. *Bioinformatics* 2014; 30: 2086–2088.
 27. International Schizophrenia Consortium; Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; 460: 748–752.
 28. Lin R, Charlesworth J, Stankovich J, et al. Identity-by-descent mapping to detect rare variants conferring susceptibility to multiple sclerosis. *PLoS ONE* 2013; 8: e56379.
 29. Willer CJ, Li Y and Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; 26: 2190–2191.
 30. Emily M, Mailund T, Hein J, et al. Using biological networks to search for interacting loci in genome-wide association studies. *Eur J Hum Genet* 2009; 17: 1231–1240.
 31. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; 7: 248–249.
 32. Kumar P, Henikoff S and Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; 4: 1073–1081.
 33. Reva B, Antipin Y and Sander C. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res* 2011; 39: e118.
 34. Lim YS and Tang BL. The Evi5 family in cellular physiology and pathology. *FEBS Lett* 2013; 587: 1703–1710.
 35. Hoppenbrouwers IA, Aulchenko YS, Ebers GC, et al. EVI5 is a risk gene for multiple sclerosis. *Genes Immun* 2008; 9: 334–337.
 36. Australia New Zealand Multiple Sclerosis Genetics Consortium. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet* 2009; 41: 824–828.
 37. Mowry EM, Carey RF, Blasco MR, et al. Multiple sclerosis susceptibility genes: Associations with relapse severity and recovery. *PLoS ONE* 2013; 8: e75416.
 38. Van Aalderen MC, Remmerswaal EB, Verstegen NJ, et al. Infection history determines the differentiation state of human CD8⁺ T cells. *J Virol* 2015; 89: 5110–5123.
 39. Parnell GP, Gatt PN, Krupa M, et al. The autoimmune disease-associated transcription factors EOMES and TBX21 are dysregulated in multiple sclerosis and define a molecular subtype of disease. *Clin Immunol* 2014; 151: 16–24.