Genetic and environmental factors influencing susceptibility to the complex disease multiple sclerosis

DPhil thesis

Giulio Disanto
Merton College
Nuffield Department of Clinical Neurosciences
University of Oxford

Index

List of publications:

1) **Disanto G**, Handel AE, Damoiseaux J, Hupperts R, Giovannoni G, Smolders J, Ramagopalan SV (2013) Vitamin D supplementation and antibodies against the Epstein-Barr virus in multiple sclerosis patients. Mult Scler (in press)
2) **Disanto G**, Vcelakova J, Pakpoor J, Elangovan RI, Sumnik Z, Ulmannova T, Ebers GC, Ramagopalan SV, Štechová K (2013) DNA methylation in a monozygotic quadruplet affected by type 1 diabetes. Diabetologia (in press).
3) **Disanto G**, Watson CT, Meier UC, Ebers GC, Giovannoni G, Ramagopalan SV (2013) Month of birth and thymic output. JAMA neurology 70: 527-528
4) **Disanto G**, Hall C, Lucas R, et al. (2013) Assessing interactions between HLA-DRB1*15 and infectious mononucleosis on the risk of multiple sclerosis. Mult Scler
5) Handel AE, **Disanto G**, Ramagopalan SV (2013) Next-generation sequencing in understanding complex neurological disease. Expert review of neurotherapeutics 13: 215-227
6) **Disanto G**, Ramagopalan SV (2013) On the sex ratio of multiple sclerosis. Mult Scler 19: 3-4
7) Watson CT, **Disanto G**, Breden F, Giovannoni G, Ramagopalan SV (2012) Estimating the proportion of variation in susceptibility to multiple sclerosis captured by common SNPs. Scientific reports 2: 770
8) Ragnedda G, **Disanto G**, Giovannoni G, Ebers GC, Sotgiu S, Ramagopalan SV (2012) Protein-protein interaction analysis highlights additional loci of interest for multiple sclerosis. PloS one 7: e46730
9) Watson CT, **Disanto G**, Sandve GK, Breden F, Giovannoni G, Ramagopalan SV (2012) Age-associated hyper-methylated regions in the human brain overlap with bivalent chromatin domains. PloS one 7: e43840
10) **Disanto G**, Ramagopalan SV (2012) Multiple sclerosis and co-morbid autoimmune disease: the final nail in the coffin? Mult Scler 18: 1370-1371
11) **Disanto G**, Pakpoor J, Morahan JM, et al. (2013) Epstein-Barr virus, latitude and multiple sclerosis. Mult Scler 19: 362-365
12) **Disanto G**, Chaplin G, Morahan JM, et al. (2012) Month of birth, vitamin D and risk of immune mediated disease: a case control study. BMC medicine 10: 69
13) Dyment DA, Cader MZ, Chao MJ, et al. (2012) Exome sequencing identifies a novel multiple sclerosis susceptibility variant in the TYK2 gene. Neurology 79: 406-411
14) Pakpoor J, **Disanto G**, Gerber JE, et al. (2013) The risk of developing multiple sclerosis in individuals seronegative for Epstein-Barr virus: a meta-analysis. Mult Scler 19: 162-166
15) Pakpoor J, **Disanto G**, Lacey MV, Hellwig K, Giovannoni G, Ramagopalan SV (2012) Breastfeeding and multiple sclerosis relapses: a meta-analysis. Journal of neurology 259: 2246-2248
16) **Disanto G**, Sandve GK, Berlanga-Taylor AJ, et al. (2012) Vitamin D receptor binding, chromatin states and association with multiple sclerosis. Human molecular genetics 21: 3575-3586
17) **Disanto G**, Morahan JM, Ramagopalan SV (2012) Multiple sclerosis: risk factors and their interactions. CNS & neurological disorders drug targets 11: 545-555
18) **Disanto G**, Morahan JM, Lacey MV, et al. (2012) Seasonal distribution of psychiatric births in England. PloS one 7: e34866
19) Berlanga-Taylor AJ, **Disanto G**, Ramagopalan SV (2012) Re: "Use of penicillin and other antibiotics and risk of multiple sclerosis: a population-based case-control study". American journal of epidemiology 175: 727-728; author reply 728-729
20) **Disanto G**, Morahan JM, Barnett MH, Giovannoni G, Ramagopalan SV (2012) The evidence for a role of B cells in multiple sclerosis. Neurology 78: 823-832

21) **Disanto G**, Sandve GK, Berlanga-Taylor AJ, et al. (2012) Genomic regions associated with multiple sclerosis are active in B cells. PloS one 7: e32281
22) Ramagopalan SV, Dyment DA, Cader MZ, et al. (2011) Rare variants in the CYP27B1 gene are associated with multiple sclerosis. Annals of neurology 70: 881-886
23) Berlanga-Taylor AJ, **Disanto G**, Ebers GC, Ramagopalan SV (2011) Vitamin D-gene interactions in multiple sclerosis. Journal of the neurological sciences 311: 32-36
24) **Disanto G**, Meier U, Giovannoni G, Ramagopalan SV (2011) Vitamin D: a link between Epstein-Barr virus and multiple sclerosis development? Expert review of neurotherapeutics 11: 1221-1224
25) Handel AE, **Disanto G**, Jarvis L, et al. (2011) Seasonality of admissions with multiple sclerosis in Scotland. European journal of neurology : the official journal of the European Federation of Neurological Societies 18: 1109-1111
26) **Disanto G**, Handel AE, Morahan JM, et al. (2011) Vitamin D and multiple sclerosis hospital admissions in Scotland. QJM : monthly journal of the Association of Physicians 104: 1001-1003
27) Banwell B, Bar-Or A, Arnold DL, et al. (2011) Clinical, environmental, and genetic determinants of multiple sclerosis in children with acute demyelination: a prospective national cohort study. Lancet neurology 10: 436-445
28) **Disanto G**, Handel AE, Morahan JM, Ramagopalan SV (2011) Comment on "Epigenetic reduction in invariant NKT cells following in utero vitamin D deficiency in mice". J Immunol 186: 3803-3804
29) **Disanto G**, Handel AE, Para AE, Ramagopalan SV, Handunnetthi L (2011) Season of birth and anorexia nervosa. The British journal of psychiatry : the journal of mental science 198: 404-405
30) **Disanto G**, Magalhaes S, Handel AE, et al. (2011) HLA-DRB1 confers increased risk of pediatric-onset MS in children with acquired demyelination. Neurology 76: 781-786
31) Handel AE, Williamson AJ, **Disanto G**, Dobson R, Giovannoni G, Ramagopalan SV (2011) Smoking and multiple sclerosis: an updated meta-analysis. PloS one 6: e16149
32) **Disanto G**, Berlanga AJ, Handel AE, et al. (2010) Heterogeneity in multiple sclerosis: scratching the surface of a complex disease. Autoimmune diseases 2011: 932351
33) **Disanto G**, Handel AE, Ramagopalan SV (2011) Estrogen-vitamin D interaction in multiple sclerosis. Fertility and sterility 95: e3; author reply e4
34) **Disanto G**, Ramagopalan SV, Para AE, Handunnetthi L (2011) The emerging role of vitamin D binding protein in multiple sclerosis. Journal of neurology 258: 353-358
35) Handel AE, **Disanto G**, Ramagopalan SV (2010) Visceral obesity and brain volume. Annals of neurology 68: 770-771; author reply 771-772
36) Handel AE, Williamson AJ, **Disanto G**, Handunnetthi L, Giovannoni G, Ramagopalan SV (2010) An updated meta-analysis of risk of multiple sclerosis following infectious mononucleosis. PloS one 5

ABSTRACT

Genetic and environmental factors influencing susceptibility to the complex disease multiple sclerosis. *Giulio Disanto, Merton College, DPhil in Clinical Neurology, Nuffield Department of Clinical Neurosciences, University of Oxford, Trinity term 2013.*

Multiple sclerosis is a complex immune mediated condition of the central nervous system characterized by myelin loss and progressive neurodegeneration. The risk of developing MS is influenced by both genetic and environmental agents and, among them, several lines of evidence support a role for vitamin D deficiency, Epstein-Barr virus (EBV) infection and smoking in the aetiology of this disease. The aim of this work was to further elucidate how nature and nurture act in the causal cascade leading to MS. In chapter 1, I show that the main genetic factor in adult MS (the *HLA-DRB1*1501* allele) plays an equally important role in paediatric cases of MS (PMS) and that EBV negative PMS patients represent a separate entity characterized by lower age at disease onset, lower female to male ratio and a trend towards a lower frequency of the *HLA-DRB1*1501* allele. In chapter 2, I provide evidence in support of month of birth having a role on MS risk and T cell production and that vitamin D may underlie this effect. In chapter 3 I demonstrate the presence of a link between vitamin D deficiency and the immune response against EBV, whereby the proportion of EBV seropositive MS patients and controls increases with increasing latitude and high dose vitamin D supplementation appears to reduce the level of antibodies against this virus. In chapter 4, I show that MS associated genetic variants are located in genomic regions that exert a regulatory function and are active in immune cell types. In chapter 5, I illustrate how vitamin D receptor binding is also located within active regulatory regions in immune cells and that this is particularly evident near MS associated genes. Finally, in chapter 6, I use chromatin data on more than 100 different cell types and conclude that MS associated genetic variants are particularly active in T helper, T cytotoxic and B cells. Further work is needed to elucidate how genetic and environmental agents play a role in the cause of MS and to develop effective strategies for disease treatment and prevention.

INTRODUCTION TO MULTIPLE SCLEROSIS

Multiple sclerosis (MS) is a highly debilitating immune mediated disorder of the central nervous system (CNS) and represents a substantial burden to the developed world (1). The prevalence of MS in Western Europe and North America is approximately 1/1,000, while incidence estimates range between 5 and 10 new cases per 100,000 individuals every year (2). Between 85,000 and 100,000 individuals are affected by this often devastating condition in the United Kingdom and this makes MS the most common cause of acquired neurological disability in young adults in this country.

CLINICAL FEATURES AND DIAGNOSIS

The clinical course of MS is highly variable, ranging from individuals showing occasional sensory nuisance throughout their lifetime to patients with fulminant course and death within months after disease onset. Approximately 85% of MS patients present with a clinically isolated syndrome (CIS). This can be defined as a single demyelinating and clinical evident episode that lasts for at least 24 hours and occurs in the absence of fever or infection, with no clinical features of encephalopathy (3). Although polysymptomatic CIS cases are possible, the attack usually involves a single CNS site and can have a variety of manifestations including limb numbness (paraesthesia), optic neuritis (inflammation of the optic nerve leading to loss of vision) and motor symptoms such as ataxia and hemiparesis (3). Approximately 60% of CIS patients later develop the relapsing-remitting form (RRMS) within 20 years time, in which acute exacerbations are followed by periods of remission of symptoms (4). Few clinical features are disease-specific, but particularly characteristic are Lhermitte's symptom (an electrical sensation running down the spine or limbs on neck flexion) and the Uhthoff phenomenon (transient worsening of symptoms and signs when core body temperature increases, such as after exercise or a hot bath) (1, 5). With time, recovery from each episode is incomplete and approximately 60-70% of

RRMS cases enter the secondary progressive phase of MS (SPMS) in which neurological disability accumulates in the absence of remission (1). About 20% of MS patients instead develop the primary progressive form of MS (PPMS), in which the disease is progressive from disease onset (1).

The principle of diagnosis in RRMS is to establish from clinical evidence, supplemented by laboratory investigations, that disease activity (consistent with focal demyelination) has affected more than one part of the CNS (dissemination in space (DIS)) on different occasions (dissemination in time (DIT)) (1). The current diagnostic criteria (2010 revision of McDonald criteria) define a clinical attack or relapse as patient-reported symptoms or objectively observed signs typical of an acute inflammatory demyelinating event in the CNS, current or historical, with duration of at least 24 hours, in the absence of fever or infection (6). DIS and DIT can be clinically established in the presence of two demyelinating episodes affecting separate sites of the CNS and occurring at least 30 days apart. Magnetic resonance imaging (MRI) can be used as a substitute for one of two clinical episodes. DIS can be demonstrated by the presence of at least 1 T2 lesion in at least 2 of 4 locations considered characteristic for MS (juxtacortical, periventricular, infratentorial, and spinal cord). DIT is instead indicated by at least one of the two following conditions: 1) a new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, with reference to a baseline scan irrespective of the timing of the baseline MRI; 2) Simultaneous presence of asymptomatic gadolinium-enhancing and non-enhancing lesions at any time (6).

For PPMS to be diagnosed, the presence of both of the following criteria is necessary: 1) One year of disease progression (retrospectively or prospectively determined); 2) Two of the 3 following conditions: I- Evidence for DIS in the brain based on ≥1 T2 lesions in at least 1 area characteristic for MS (periventricular, juxtacortical, or infratentorial); II- Evidence for DIS in the spinal cord based on ≥2 T2 lesions in the cord; III- Positive cerebrospinal fluid (CSF) (isoelectric focusing evidence of oligoclonal bands and/or elevated IgG index) (6).

PATHOLOGICAL AND IMMUNOLOGICAL SIGNATURES OF MS

Although more than 100 years have passed since Charcot, Carswell, Cruveilhier, and others described the clinical and pathological characteristic of MS, both the etiology and the pathogenesis of this disease are not yet conclusively known. The hallmark of MS is the formation of a sclerotic plaque in the CNS, which represents the end stage of a process involving inflammation, demyelination, remyelination and axonal degeneration (1, 5). Partially demyelinated axons conducting action potentials at reduced velocity and neuronal loss lead to the characteristic delay in evoked potentials in MS patients and to neurological disability.

In the relapsing-remitting phase of MS inflammatory demyelination within active white matter plaques plays a central role. Myelin-laden macrophages and (to a lesser extent) CD8+ T cells dominate the lesions, while CD4+ T helper (Th) cells (both Th1 and Th17) are found primarily in the perivascular regions and relatively smaller numbers in the parenchyma (7-10). This immune-mediated demyelinating process is accompanied with a variable extent of acute axonal injury and loss and partial remyelination (8). However, cortical demyelinating lesions are also present in RRMS and have been shown to correlate with cortical atrophy, disease progression, physical disability, and cognitive impairment at later stages (11-13).

The pathology of progressive MS (both SPMS and PPMS) is considerably different. Focal demyelinating lesions are still present but appear less immunologically active than in RRMS. While moderate inflammatory infiltrates (mainly composed of T cells and microglial activation) are located at the margins of the lesions, neurodegeneration represents the main pathological finding with diffuse atrophy of both grey and white matter. Normal-appearing white matter (NAWM) also shows diffuse inflammation, generalized activation of microglia and considerable axonal injury and destruction (8). Also B and plasma cells become increasingly prominent. Inflammatory aggregates

organized as B-cell germinal center–like structures have been observed in the subarachnoid space of SPMS cases and associated with features indicative of a more severe disease course (14, 15).

However, in MS as well as in other complex immune mediated diseases, primary causal and secondary events are difficult to separate. Furthermore, lymphocytes may be absent in early MS lesions with extensive oligodendrocyte loss and demyelination (16, 17). Therefore, despite the amount of information that is today available regarding the pathological features of MS, which cell types drive the demyelinating process and how this happens is still largely unknown. T cells and in particular CD4+ Th cells have until recently been considered the primary immune drivers in MS. Evidence supporting this hypothesis includes the fact that the main MS genetic risk factor resides in the major histocompatibility complex (MHC) class II region, which plays a central role in the development of CD4+ T cell central tolerance and that the animal model of MS, experimental autoimmune encephalomyelitis (EAE), can be adoptively transferred to mice by the injection of encephalitogenic myelin-specific CD4+ Th cells (18, 19). However, phase II clinical trials testing the effect of monoclonal antibodies targeting the Th response such as Ustekinumab and cM-T412 reported no benefit in RRMS (20, 21). In clear contrast, analogous antibodies prevented the development of EAE (22, 23). These trials demonstrate why, although animal models can represent useful tools in modern research, one needs to be aware of their distance from the human disease.

It seems therefore unlikely that MS is driven by a single pathogenic cell type and the role of several components of both the adaptive and innate immune system is now widely acknowledged. For example, the importance of B cells in MS pathogenesis is suggested by the fact that the intrathecal synthesis of clonal IgG in the CSF is the most consistent immunological finding in MS and this indicates abnormal B cell activation within the CNS of MS patients (24). Furthermore, B cell abnormalities influence both conversion to clinically definite MS, MRI activity, onset of relapses and disease progression (14, 25-29). As mentioned above, in active MS lesions and unlike the lesions of

EAE, CD4+ T cells are outnumbered by MHC class I–restricted, clonally expanded CD8+ cytotoxic T cells (30). In addition, also MHC class I genes (important for CD8+ T cell central tolerance) are associated with the risk of MS (31). Finally, accumulating evidence supports the presence of functional abnormalities in both regulatory T and natural killer cell compartments in both EAE mice and human MS patients (32-34).

RISK FACTORS AND DISEASE AETIOLOGY

It is now established that genes, environmental factors and their interactions underlie the aetiology of MS (1, 35). However, which genes and environmental factors are implicated, to what extent they contribute to MS pathogenesis, and the precise meaning of the term "interaction" remain largely unknown and the subject of debate. Nowadays researchers are able to explore disease aetiology using tools that only a few years ago appeared unrealistic, including the investigation of DNA sequence and epigenetic modification across the whole genome at the single nucleotide level. Still, it is surprising to see how the most recent and relevant findings in MS were in fact suggested years earlier by ecological and epidemiological studies investigating the incidence, prevalence and geographical distribution of this complex disease. The identification of risk factors influencing MS development and their time points of action will aid our understanding of MS pathogenesis and ultimately in the development of disease treatment and prevention strategies.

Clues from epidemiology

That susceptibility to MS is influenced by the genes we inherit was clearly shown by large population-based studies investigating the familial aggregation of the disease. The recurrence risk for MS increases with degree of kinship, being higher in the nuclear than in the extended family and the highest among monozygotic (MZ) twin pairs (figure 1) (36). The same does not apply to non-biological

relatives such as those of adoptees suffering from MS and spouses of MS patients indicating that the shared familial microenvironment exerts little influence on susceptibility (37, 38).

Although these studies have provided the rationale behind genetic studies in MS, other epidemiological observations clearly suggested that genes were not deterministic and that genetic susceptibility can be shaped by unknown external agents that we are now starting to recognize and understand. This can be intuitively appreciated from the observation that MS concordance in MZ twins is far lower than 100% (39). The heritability estimates (i.e. the proportion of the phenotype that is due to genetic differences) which are calculated based on twin concordance rates, vary widely between studies (ranging from 0.25 to 0.76) (36, 39-44). While different assessment strategies and statistical methods represent potential factors of data variation (39), it appears that external variables such as latitude of birth can also influence MS twin concordance rates (44).

One of the most historically acknowledged features of MS is the positive correlation between its prevalence and distance from the equator (45-47). When incidence rates rather than prevalence are considered, the latitude gradient is absent in the northern hemisphere but striking in the southern hemisphere. Data from Australia and New Zealand suggest that when investigated in a genetically homogeneous background, latitude has a substantial influence on MS risk (48, 49).

Several lines of evidence point to pre-birth and early life factors in MS aetiology. For example the risk of MS in second and third generation Asian, African and West Indian immigrants to the UK appears relatively high, similar to that of the general English population and higher than first generation immigrants (50). In addition, MS patients' births follow a seasonal distribution being more and less frequent during the spring and the autumn respectively. This has now been shown in a number of regions and countries including Canada, Denmark, Sweden, Finland, Sardinia, Australia, France, England and Scotland (51-58). Furthermore, the observation that the recurrence risk for dizygotic (DZ) twins is almost twice that for normal full siblings and concordance rates among maternally

related half siblings is higher than among paternal ones, more specifically highlight the importance of gestational environment on MS risk (36, 59).

Studies investigating the incidence of MS among immigrants to other countries strongly suggest that factors acting during childhood and adolescence also influence the risk of MS. For example a nearly 70% decreased risk of MS was found among British individuals who migrated to South Africa before the age of 15 (60). Similarly MS risk was shown to be higher in Indian and Pakistani immigrants who entered the UK under 15 years of age (61). Perhaps the strongest evidence for environmental factors in MS is the observation that the female to male ratio of MS has been substantially increasing worldwide over the last century and this increase is particularly clear in large population-based studies from Denmark and Canada (2, 62). In the latter, the sex ratio of MS patients born in the 1930s was lower than 2 and then increased to more than 3 females for each male patient in the latest birth cohort analysed (1976–1980). A change occurring within a century is too short a time for a genetic cause and suggests that environmental factors are at work in a sex-specific manner.

Taken together, these findings indicate that MS is a complex disorder in which environmental exposures operating at different time points (pre-birth, childhood and adolescence) trigger disease onset in genetically susceptible individuals (63). Latitude related agents appear to play an important role. In addition to genes, three environmental risk factors now have overwhelming support from the scientific community: vitamin D deficiency, Epstein-Barr virus (EBV) infection, and smoking behaviour.

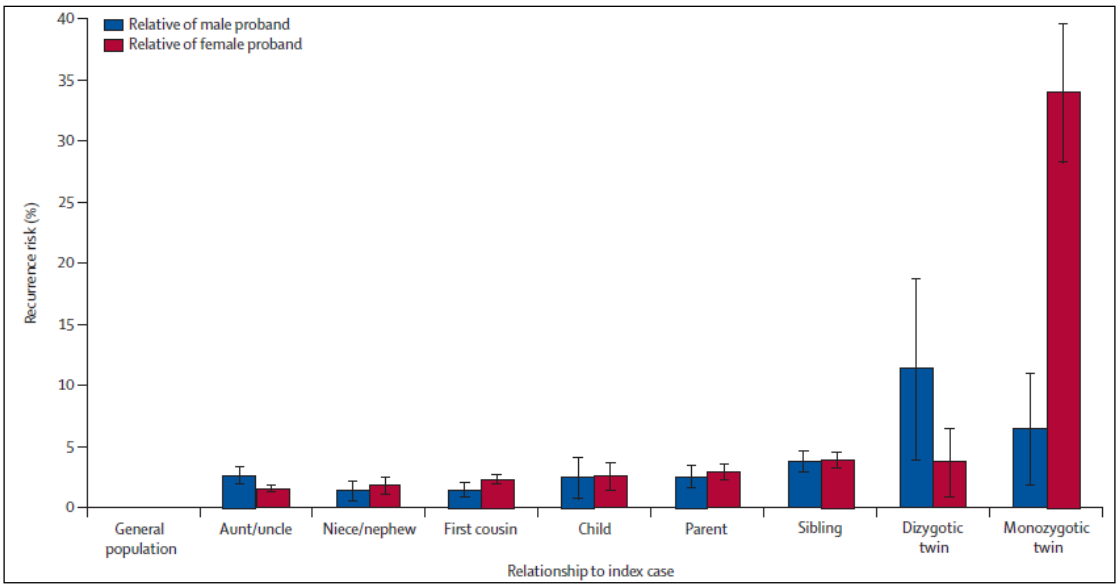

Figure 1: Age adjusted recurrence risk in biological relatives of MS patients (data from Willer et al. 2003)

The role of genes

The development of genome wide association studies (GWAS) has been a valuable tool for enabling the discovery of genetic variants contributing to complex traits. However, the main genetic locus for MS was discovered in the 1970s, long before we entered the GWAS era (64). Family-based and association studies more precisely identified the responsible variant for MS in the MHC region, specifically the *HLA-DRB1*1501* class II allele (18, 65-67). Notably, this allele has been found to increase the risk of MS in nearly all populations studied and its presence in homozygosity increases the risk of MS more than 6 fold (18, 35). However, almost every *HLA-DRB1* allele has been either positively or negatively associated with MS and substantial heterogeneity is observed between populations. For example, *DRB1*03, DRB1*04* and *DRB1*13* alleles increase the risk of MS in the Sardinian population, while protection is exerted by *DRB1*01, DRB1*10, DRB1*11* and *DRB1*14* in Canadians and *DRB1*09* in Japanese (18, 68, 69). This scenario is further complicated by the presence of *cis* and *trans* epistasis between different HLA-class II genes (65, 66, 70). But the role of genetics in

MS is not limited to the MHC and in recent years the development of GWAS has provided further insights into the genetic architecture of MS (31, 71-73). The International Multiple Sclerosis Genetics Consortium (IMSGC) and the Wellcome Trust Case Control Consortium 2 (WTCCC2) recently performed a GWAS including more than 9,000 MS patients and found evidence for 57 genomic regions in addition to the MHC. In the latter, independent signals were identified from both *HLA class II* (*DRB1*15:01, DRB1*03:01, DRB1*13:03*) and *class I* (*A*02*) alleles (31). A gene-ontology analysis of the genes located within MS-associated regions showed a substantial overrepresentation of immune-related processes which further confirmed the immunological nature of the disease (19, 31). Interestingly no association was found with either clinical course (relapsing remitting *vs* primary progressive) or rate of progression as measured by the Multiple Sclerosis Severity Scale (MSSS) (31).

Despite the large number of identified risk alleles, taken together these variants could explain only ~20% of MS genetic risk, most of which arose from the MHC variants. It thus appears the early expectations that GWAS would be definitive in identifying the genetic basis of complex diseases have now been supplanted with the reality that only a small fraction of heritability is explainable by GWAS findings (74). Further increasing the sample size will be unlikely to provide a solution. Several mechanisms have been proposed to explain such missing heritability including gene-gene and gene x environment interactions, epigenetics, and rare genetic variants (frequency <0.5% in the population) (74, 75).

It is not so uncommon to observe several individuals affected by MS within the same family. Studies have shown that multiplex families (families with several affected members) are particularly loaded with MS associated common variants, but this cannot entirely explain such aggregation of disease (76, 77). Therefore a role for rare and highly penetrant variants appears plausible. Such rare variants may also be involved in sporadic cases since even when odds ratios (ORs) are higher than 2, penetrance may be too low to create familial aggregation (78). Since the identity of these variants is

largely unknown, they cannot be included in current SNP panels and sequencing strategies are required for their discovery. The development of next generation sequencing technologies will facilitate the discovery of rare variants of large effects in MS (79).

Vitamin D

An important role for sunshine and ultraviolet (UV) light exposure and consequent vitamin D production in MS is now widely supported (80). Epidemiological evidence linking vitamin D and MS comes from the observation that the prevalence of MS in England and France inversely correlates with UV radiation and that sunshine exposure during childhood and adolescence protects against MS (81-85). This does not necessarily implicate a role for vitamin D since any other factor related to sunshine/UV exposure could be mediating these findings. However the involvement of vitamin D is further supported by studies showing how risk of MS is inversely associated with both vitamin D intake (RR=0.59, 95%CI=0.38-0.91 for intake > 400 IU/day) and 25-hydroxylvitamin D (25-OH-D) levels (OR=0.59; 95%CI=0.36-0.97 for a 50 nmol/L increase) before disease onset (86, 87).

Nonetheless even this type of study is susceptible to confounding by both social and behavioural factors, highlighted by the contradictory results of observational and randomized controlled investigations of the effect of antioxidant vitamins on cardiovascular disease and cancer where mortality is observed (88). Randomized controlled trials assessing the risk of MS after vitamin D supplementation are lacking as they would be substantial undertakings. If vitamin D exposure *in utero* or childhood is the key period of exposure for impacting MS then any prevention trial would have to last decades. Further, about 7 people per 100,000 are diagnosed with MS each year in Northern Europe and North America. Therefore, to show a significant change in the rate of MS, an adequately powered trial would need to enrol hundreds of thousands of participants. This would be the ‘gold standard’ way of providing evidence for a causal role for vitamin D by randomising all other

potential confounders. However, strong evidence for causation has been achieved by the observation of rare variants in the *CYP27B1* gene in MS patients, since this links a genetic change that is directly responsible for vitamin D deficiency with risk of MS (89). *CYP27B1* encodes the enzyme responsible for vitamin D activation converting 25-hydroxyl-vitamin D (25-OH-D) into its active form (1,25-OH-D) (90). Notably, the rare variants identified in the MS population have well described functional roles and are associated with particularly severe forms of rickets (vitamin D dependent rickets) and lower 1,25-OH-D levels (89). In addition to these rare variants, two common polymorphisms located in the proximity of *CYP27B1* and *CYP24A1* (the enzyme that initiates the degradation of 1,25-OH-D) have also been found to influence MS risk (31). These genetic findings add substantial support to the well documented epidemiological evidence for vitamin D in MS.

In addition to epidemiological and genetic association studies, several lines of evidence illustrate a functional link between MS and vitamin D. This exceptionally pleiotropic hormone has a profound and broad impact on the immune system. Both the vitamin D receptor (*VDR*) and *CYP27B1* are broadly expressed in immune cells such as monocytes, macrophages, neutrophils, natural killer cells, dendritic cells, as well as B and T cells (91). Moreover, lack of VDR and vitamin D exacerbates experimental models of MS while ablation of *CYP27B1* has also been shown to impair the immune system (80, 92, 93). Finally, recent findings strongly support the presence of a gene-environment interaction for vitamin D in MS. Vitamin D acts in the genome via its cognate nuclear receptor, the vitamin D receptor (VDR). Using chromatin immunoprecipitation followed by massively parallel DNA sequencing (ChIP-seq) in B cell lines, our group has shown the presence of 2,776 different vitamin D responsive elements throughout the genome bound by the VDR. Notably, genetic loci associated with MS (as well as other immune mediated conditions) were substantially enriched for VDR binding sites (figure 2)(94).

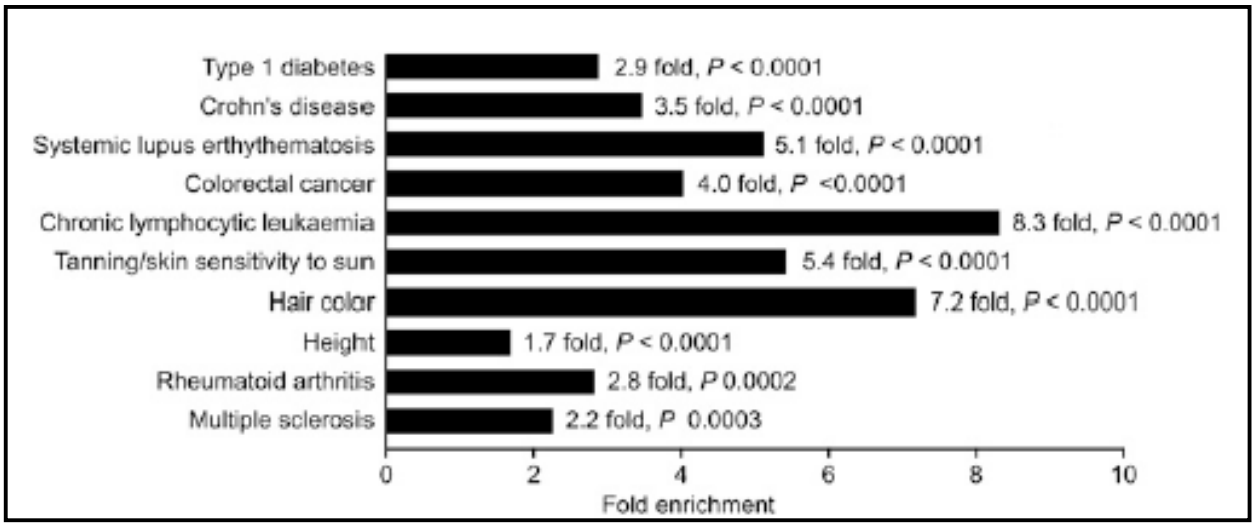

Figure 2: Genomic regions associated with variety of complex traits including MS are enriched for vitamin D responsive elements (VDREs). Genomic regions associated with MS overlap with VDREs more than expected by chance (p=0.0003) with a 2.2 fold enrichment (from Ramagopalan et al. 2010).

Epstein Barr virus (EBV)

EBV is a gamma human herpesvirus which primarily infects B cells and is able to immortalize them *in vitro* and induce lymphoproliferative disorders *in vivo* (95). The infection is generally asymptomatic but in about 30-50% of cases, especially when infection occurs during the second decade of life, EBV infection manifests as glandular fever, also known as infectious mononucleosis (IM) (96). Although approximately 90% of the adult population has encountered EBV at some point during life, several lines of evidence suggest a role for EBV in the pathogenesis of MS (97). Nearly all adult MS patients have been infected with EBV and the OR for developing MS in those who are EBV negative is extremely low (97, 98). Perhaps the strongest evidence for the involvement of EBV in MS comes from a single study investigating EBV status in a cohort of MS patients from the US military personnel on which several serum samples were collected before disease onset. The authors identified 10 cases of MS who were EBV sero-negative at baseline and found that all of them sero-converted before MS onset indicating how MS risk sharply increases after EBV infection (99). In addition, both high anti-EBV antibody titers and a history of IM also increase the risk of developing MS and this relation is

temporally consistent (97, 100-104). The risk of MS is 36-fold higher among individuals with anti-EBNA complex IgG titers ≥ 320 as compared to those with titers <20 and this appears independent from vitamin D status (101). In a meta-analysis collecting data from over 19,000 MS patients, the relative risk for MS after IM was 2.17 (95%CI=1.97-2.39) (105). In agreement with these findings, recent data from England demonstrate that IM and MS share the same geographical distribution (106). Interestingly MS patients are more likely to be infected with both strains of EBV (type 1 and type 2) perhaps suggesting that the effect of EBV on the development of MS may be enhanced in the presence of double infection (107).

Despite a lower frequency of infection, the association of EBV with MS is also present in the paediatric MS population. Although the potential for confounding remains, the paediatric findings are important since in this population less time has accrued between exposure and onset, making the identification of risk factors for MS potentially more tractable. In a longitudinal investigation of paediatric cases of either monofocal or polyfocal (with or without encephalopathy) acute demyelinating syndromes, those who were EBV positive and had high EBV antibody titres were more likely to convert to MS (108).

Whether this strong association is causal is still debated. For example MS associated genetic variants could also influence susceptibility to EBV. However this latter possibility appears unlikely since the *HLA-DRB1*1501* class II allele is not associated with the development of IM (109). Similarly, good hygiene during childhood may predispose both to MS and to a later contact with EBV and therefore IM, but this does not agree with the extremely low risk of MS in EBV negative individuals (likely to be exposed to the highest levels of hygiene) (97). Taken together, a causative role for EBV in MS is strongly supported but further studies are needed to elucidate if, how and when EBV plays a role in MS.

Smoking

Cigarette smoking is another environmental risk factor associated with MS. Interest in the link between smoking and disease began in 1965 with one study showing that previous smokers had a significantly higher risk of MS (110). These findings have then been confirmed by several case control and population based studies including some demonstrating a dose response effect with heavier smokers incurring higher risks (111-115) . The strongest evidence for smoking playing a role in MS comes from longitudinal studies. Two longitudinal studies in women from the UK showed higher risk of MS in smokers, although statistical significance was not reached (116, 117). However, the Nurses' Health Study and Nurses' Health Study II addressed this question in a larger cohort and found a dose response effect and increase in risk for current smokers (OR=1.6, 95% CI=1.2-2.1) (118). A meta-analysis carried out by our group showed that the overall evidence for smoking increasing the risk of MS is strong (RR=1.48, 95%CI=1.35–1.63)(119). Further increasing the strength of this association is the observation that parental smoking is also positively associated with the risk of paediatric MS (RR=2.12, 95% CI=1.43-3.15) (120). Studies have also shown the presence of an increasing female to male ratio in smoking behaviour and suggested that smoking may in part contribute to the increasing female to male ratio seen in MS (121).

The potential mechanism underlying the association of smoking with MS is not currently known. Smoking has a documented detrimental effect on the immune system, being able to influence T cells, NK cells, humoral and cell mediated immunity (122, 123). Interestingly snuff use does not carry an increased risk of MS, a finding suggesting that the injurious factor must be contained within cigarette smoke (112). The most likely candidate in cigarette smoke is the free radical nitric oxide, which could contribute to demyelination and axonal loss (124). Another mechanistic possibility involves post-translational modification in the brain. Peptidylarginine deiminase 2 (*PAD 2*), an enzyme which can

citrullinate myelin antigens such as myelin basic protein, is significantly upregulated in MS white matter (125, 126) and smoking has been shown to increase *PAD2* expression (127).

AIMS

The aim of my work was to further elucidate the role of genetic and environmental factors influencing susceptibility to MS. In chapter 1 I investigated the role of genes and environment in the susceptibility to paediatric cases of MS with a particular focus on the *HLA-DRB1*1501* MHC class II allele and EBV infection. In chapter 2 I investigated whether the reported season of birth effect is a specific feature of MS or is shared by other immune mediated diseases using a large cohort of patients from the United Kingdom. In addition I explored a potential mechanism by which month of birth can influence the development of the immune system and MS risk. In chapter 3 I investigated the association between latitude and EBV seroprevalence and the potential interaction between vitamin D and immune response to EBV in MS aetiology. In chapters 4 and 5 I aimed to study the relation between genetic variants influencing the risk of MS, chromatin states in immunological cell types and binding of the vitamin D receptor (VDR) across the whole genome. Finally, in chapter 6 I aimed to define the pathogenic cell types driving the onset of MS by investigating the overlap between MS associated genetic variants and regions of open chromatin across more than 100 different cell types.

CHAPTER 1: HLA STATUS AND ROLE OF EBV INFECTION IN PAEDIATRIC CASES OF MS

-***Disanto G**, Magalhaes S, Handel AE, Morrison KM, Sadovnick AD, Ebers GC, et al. HLA-DRB1 confers increased risk of pediatric-onset MS in children with acquired demyelination. Neurology. 2011 Mar 1;76(9):781-6.*

INTRODUCTION

MS is being increasingly recognised in individuals aged less than 16 years (128). In children, an initial attack of demyelination (acquired demyelinating syndrome [ADS] of the CNS) often remains a monophasic illness, but in at least 20% of cases it represents the first clinical attack of MS (129). This contrasts with adult-onset disease, where 60% of individuals presenting with acute demyelination are subsequently diagnosed with MS over a period of 20 years (4). Whether paediatric MS (PMS) shares the same biological underpinnings (including environmental and genetic risk factors) implicated in adult-onset MS is an area of active investigation. While a single study has reported a higher frequency of the *HLA-DRB1*15* allele in PMS (130), and another reported the absence of an association of HLA in a cohort of children with acute disseminated encephalomyelitis (ADEM) (131), little is known regarding the HLA status in an unselected paediatric population presenting with a first demyelinating attack who are at high risk for developing MS. In particular, it is not known whether certain HLA alleles confer risk of developing ADS in general (including children who remain with monophasic ADS), or more specifically confer the risk of recurrent disease (MS).

PMS also represents a potentially valuable study population for the identification of environmental risk factors. As a consequence of early disease onset, less time is available for potential confounders and thus any associated factor is more likely to be truly causative. In a recent investigation of a large cohort of PMS patients both EBV positivity and high EBV antibody titres predicted conversion to MS independently of HLA and vitamin D status (108). However in this as well as in other studies of PMS, the frequency of EBV positive patients ranges between approximately 80% to almost 100% (108, 132-

135), in contrast to the consistent presence of remote EBV infection in the adult MS population (>99%) (98) which would suggest perhaps a necessary role for EBV in MS. The inconsistency raises an important question on the pathogenic role of EBV in MS. The decreased frequency of EBV in PMS could mean that EBV is not a necessary factor for MS to develop. Alternatively, EBV negative (EBV-) PMS patients may represent a subgroup of individuals who does not require EBV infection to develop MS.

My first aim was to assess whether the presence of *HLA-DRB1*15* alleles influences the risk of conversion from ADS to MS. My next objective was to investigate whether any difference exists between EBV+ and EBV- PMS cases in order to shed more light on the role played by this virus in MS and PMS.

METHODS

Participants

Children age 16 years, 0 months, or younger presenting within 30 days of onset of ADS at 1 of 23 sites across Canada were recruited into the 8-year prospective Canadian National Paediatric Demyelinating Disease Study. Comprehensive structured clinical history and neurologic examinations were performed at study entry and at all study visits (baseline, 3, 6, and 12 months and annually, and at the time of recurrent demyelination). Presenting features at ADS were characterized by clinical history and examination as monofocal ADS (optic neuritis (ON), transverse myelitis (TM), other CNS sites) or polyfocal ADS (multiple CNS sites). ADEM was diagnosed using predefined criteria which require polyfocal deficits accompanied by encephalopathy (136). MS was diagnosed when either subsequent examination confirmed new neurologic deficits persisting for 24 hours or longer in the absence of fever or infection and occurring more than 30 days from ADS or if serial MRI scans revealed new lesions meeting criteria for dissemination in time and space (137). A total of 316

children meeting inclusion criteria were enrolled. Twenty children were excluded after subsequent investigations revealed an alternative diagnosis. In 12 children limited quantity or quality of DNA did not permit accurate genotyping and after the exclusion of those families refusing to participate in genetic studies, *HLA-DRB1* alleles were typed in 266 (91%) eligible children. A control group of 196 unrelated, healthy adults of Northern European ancestry was recruited through the longitudinal, population-based Canadian Collaborative Project on the Genetic Susceptibility to Multiple Sclerosis (CCPGSMS) (138).

Genotyping

I used total genomic DNA, extracted from whole blood, to type *HLA-DRB1* alleles by an allele-specific PCR amplification method (139). I performed seventy-two (high-resolution) PCR reactions to amplify allelotypes corresponding to alleles *HLA-DRB1*01, 03, 04, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16*. All the samples were genotyped blinded to clinical diagnosis.

EBV serology

Serum IgG antibodies directed against Epstein-Barr virus capsid antigens, nuclear antigens (EBNA1), and early antigens were detected using standardized ELISA kits (DiaSorin) at The Hospital for Sick Children (Toronto, ON, Canada) (108).

Statistical analyses

I compared frequencies of *HLA-DRB1* alleles (each allele contributing independently) between groups (MS vs monophasic ADS; MS vs healthy controls; monophasic ADS vs healthy controls) using the Chi square test. I also assessed the likelihood of MS diagnosis in children with ADS as a function of *HLA-DRB1*15* status (proportion of patients with one or both alleles, compared to those with no alleles).

Finally I compared age at onset (AO), sex ratio and HLA status of EBV+ and EBV- PMS cases using the Student's t and Chi square tests.

RESULTS

Of 266 children with ADS, 64 (24%) met criteria for the diagnosis of MS over an average follow-up duration of 3.2 ± 1.5 years. The average time from ADS presentation to MS diagnosis was 10.2 ± 9.0 months (median 6.3; range 1.2–37.3). Clinical and demographic characteristics of all children are presented in table 1.

Table 1: Clinical and demographic characteristics of paediatric patients

		Monophasic ADS (n=202)	**MS (n=64)**
Age at onset (mean ± SD)		9 ± 4.6	12.1 ± 3.6
Years of follow up (mean ± SD)		3.1 ± 1.5	3.5 ± 1.4
F/M (ratio)		96/106 (0.9)	42/22 (1.9)
Ancestry	**European**	137 (68%)	39 (61%)
	Non-European	65 (32%)	25 (39%)
Presentations	**ON**	65 (32%)	12 (19%)
	TM	42 (21%)	8 (13%)
	Monofocal other	51 (25%)	16 (25%)
	Polyfocal	18 (9%)	23 (36%)
	ADEM	26 (13%)	5 (8%)

DRB1*15 allele frequency is higher in MS than monophasic ADS and controls

For all children presenting with ADS (n=266), the allele frequency for *HLA-DRB1*15* was 0.2. *HLA-DRB1*15* allele frequencies were higher in children with ADS subsequently diagnosed with MS, compared to children with monophasic ADS (0.31 vs 0.15, p<0.0001) (table 2). It is plausible that this may reflect population stratification. However, I found that in the 176 ADS patients of European ancestry (monophasic ADS n=137; MS n=39) *HLA-DRB1*15* allele frequency was still higher in MS as

compared to monophasic ADS (p=0.001) or controls (p<0.001). I did not observe any difference comparing children with monophasic ADS to controls (p=0.90). Similarly, I did not see any significant difference in allele frequencies for other *HLA-DRB1* alleles.

Table 2: HLA-DRB1 allele frequencies in MS, monophasic ADS and controls (all individuals)

HLA-DRB1 allele	MS	ADS	Controls
1	0.09	0.11	0.11
3	0.13	0.12	0.12
4	0.17	0.15	0.16
7	0.06	0.11	0.14
8	0.03	0.04	0.02
9	0	0.02	0.02
10	0.01	0.008	0.01
11	0.1	0.13	0.08
12	0	0.01	0.03
13	0.1	0.11	0.13
14	0	0.02	0.04
15	0.31	0.15	0.14
16	0	0.03	0.02

DRB1*15 positive ADS cases are more likely to convert to MS

Of the 39 children of European ancestry with ADS diagnosed with MS, 21 (53.8%) had at least one *HLA-DRB1*15* allele, compared to 36 of 137 (26.3%) children with monophasic ADS (figure 3). I found that children presenting with ADS who harboured one or two *HLA-DRB1*15* alleles were more likely to be diagnosed with MS during the period of follow-up (37% [34/93 children]) as compared to children who did not harbour *HLA-DRB1*15* alleles (17% [30/173]) (p<0.001; RR=2.11, 95%CI=1.38-3.21). The increased MS risk conferred by *HLA-DRB1*15* appeared to be greater in children reporting European ancestry: 37% (21/57) of *HLA-DRB1*15*-positive children were diagnosed with MS, vs 15% (18/119) of *HLA-DRB1*15*-negative children (p=0.001, RR=2.43, 95%CI=1.41-4.20). In contrast, the risk of MS diagnosis in children with ADS of non-European ancestry did not appear to be influenced

by the presence of *HLA-DRB1*15* alleles (however the effect size was in a similar direction and non-significance may reflect a smaller sample size): 36% (13/36) of *HLA-DRB1*15*-positive children were diagnosed with MS, vs 22% (12/54) of *HLA-DRB1*15*-negative children (*p*=0.15, OR=1.62, 95%CI=0.84-3.15).

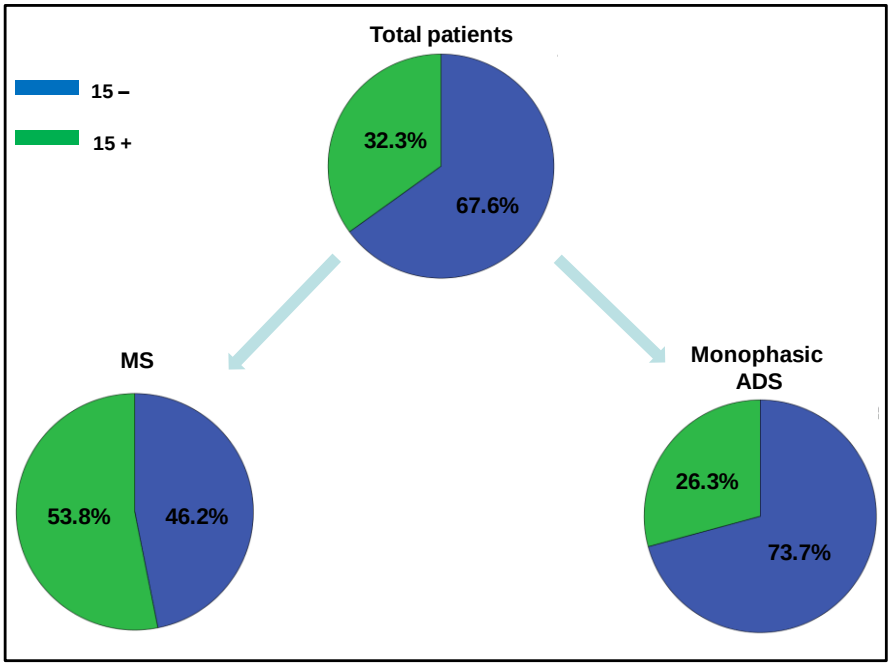

*Figure 3: Proportion of DRB1*15 positive and negative individuals in all patients, those who converted to MS and those who remained ADS.*

Characteristics of EBV- PMS

DNA and serum samples were available on a total of 51 PMS patients of which 43 (84.3%) and 8 (15.7%) were EBV+ and EBV- respectively. Demographic features and *HLA-DRB1*15* status are presented in table 3 and figure 4. I observed that both age at onset and female to male ratio were much lower in the EBV- than EBV+ groups. Furthermore the MS associated allele *HLA-DRB1*15* also appeared less frequent in EBV- than EBV+ PMS patients, but this did not reach statistical significance.

Table 3: Demographic and clinical features of EBV+ vs EBV- PMS patients

		EBV+	EBV-	*P*
N		43	8	
Age at onset	**Mean**	13.36	7.86	0.0065
	95%CI	12.6-14.1	4.4-11.3	
Sex	**F**	29	3	0.13
	M	14	5	
	F/M	2.07	0.6	
Genotypic freq	**15+**	21 (48.8%)	3 (37.5%)	0.7
	15-	22 (51.2%)	5 (62.5%)	
Allelic freq	**15**	26 (30.2%)	3 (18.7%)	0.39
	X (no15)	60 (69.8%)	13 (81.3%)	

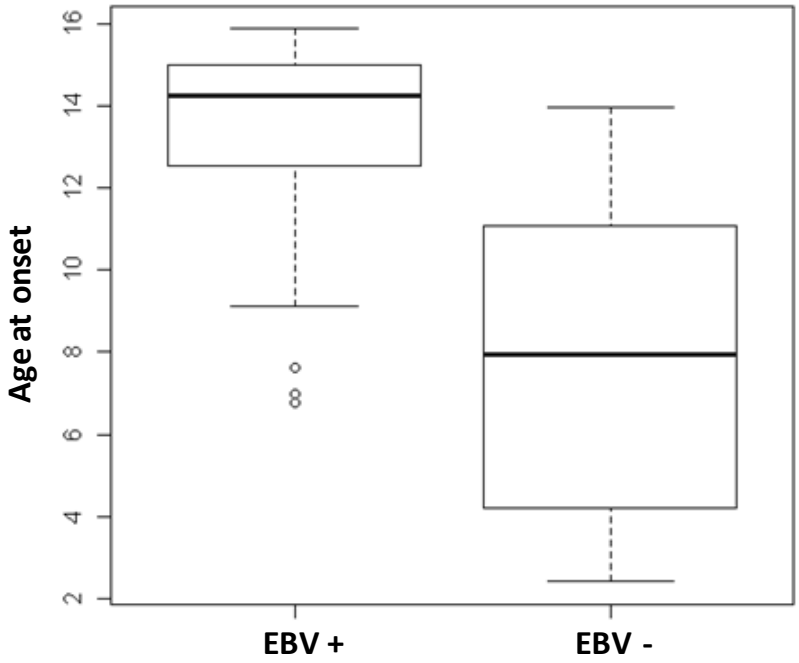

Figure 4: Boxplot of age at disease onset for EBV+ and EBV- PMS patients

DISCUSSION

Unique to this work is the evaluation of HLA alleles in a non selected prospectively followed population of children with ADS, representing a population at risk for the diagnosis of MS. Here I show that presence and frequencies of *HLA-DRB1*15* alleles in these children confers increased risk for a subsequent diagnosis of PMS. This was not the case for monophasic children with ADS, indicating that *HLA-DRB1*15* does not represent a risk factor for development of acquired demyelination in general. The association between *HLA-DRB1*15* and diagnosis of MS in this ADS population was mainly influenced by the increased frequency of this allele in children with European ancestry, while *HLA-DRB1*15* did not appear to increase MS risk in children of non-European ancestry (although this part of the study was less powered). In a single prior study examining HLA alleles in children with established MS, 44.6% of Russian paediatric patients with MS were found to have at least one *HLA-DRB1*15* allele (130), similar to the frequency in the children with ADS who subsequently developed MS. I can therefore conclude that the *HLA-DRB1*15* allele is associated with MS in the paediatric population to a similar extent that in the adult population and significantly increase the risk of conversion to MS after a first demyelinating event.

I found that EBV- PMS cases have a significantly lower age at onset as compared to EBV+ individuals. In addition, the EBV- group had a female to male ratio which was much lower than that of the EBV+ group (similar to that of adult MS), although this did not reach statistical significance. This is perhaps unsurprising since studies have shown that the female to male ratio in PMS increases with increasing age (128, 140). The interpretation of these findings is not easy. The youngest PMS patients have had less time to encounter EBV and this may explain their negative serology. However the role of exceptional factors (in particular rare genetic variants) capable of initiating MS regardless of EBV status and other risk factors typical of MS cannot be excluded. Although not statistically significant, the lower frequency of *HLA-DRB1*15* observed in EBV- PMS is supporting the latter hypothesis.

However, at the moment this study is clearly underpowered to reliably detect a genetic difference between EBV+ and EBV- patients. Assuming an effect size of approximately 3 (like that of *HLA-DRB1*15*), a risk allele frequency of 0.14 in the general population and a log-additive model, 42 EBV- and an equal number of EBV+ would be required to have 80% statistical power. New samples are currently being collected.

CHAPTER 2: MONTH OF BIRTH, VITAMIN D AND RISK OF IMMUNE MEDIATED DISEASES

***-Disanto G**, Chaplin G, Morahan JM, Giovannoni G, Hyppönen E, Ebers GC et al. Month of birth, vitamin D and risk of immune mediated disease: a case control study. BMC medicine (2011) 10: 69.*
***-Disanto G**, Watson CT, Meier UC, Ebers GC, Giovannoni G, Ramagopalan SV (2013) Month of birth and thymic output. JAMA neurology 70: 527-528*

INTRODUCTION

Immune mediated diseases (ID) affect approximately 5-10% of the developed world and the overall incidence seems to be increasing (141). This observation suggests that changes in environment and lifestyle play a central role in influencing prevalence.

Seasonality dominates the global environment and diet is closely related to seasonality by the effect of these environmental fluctuations on agriculture (142-144). Seasonal factors can potentially act even before birth, when according to the "fetal origin of adult disease hypothesis", environmental influences leading to changes in embryonic/fetal tissue structure and function can influence the risk of adult physiological and pathological conditions (145, 146). As a consequence, being born at a certain time of the year may influence susceptibility to disease later in life. Indeed, month of birth effects have already been documented in ID such as multiple sclerosis (MS) and type 1 diabetes (T1D) (51, 147, 148). In addition to MS and T1D, a few other studies have investigated the presence of a month of birth effect in other ID. However, poor sample sizes and inadequate statistical methods have significantly hampered these attempts and results are inconsistent (149-157).

The mechanisms involved in the pathogenesis of immune mediated disorders are variable and both adaptive and innate immune responses have been implicated in diseases such as MS, rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), Crohn's disease (CD) and ulcerative colitis (UC) (19, 158-160). For example, in MS and RA tolerance breakdown is thought to cause immune mediated demyelination of the central nervous system and cartilage and bone destruction respectively (1, 158). By contrast several lines of evidence suggest that CD and UC arise from an inappropriate immune

reaction to the intestinal microbiota in genetically predisposed hosts (160). Despite these differences, an abnormal activation of the immune system is a common thread linking these conditions and several observations indicate that similar genetic pathways and environmental agents such as vitamin D deficiency, smoking behaviour and various infections are involved in the pathogenesis of these disorders (35, 90, 97, 158-162). This led me to the hypothesis that a similar seasonality of birth may be present among different ID. I investigated whether the month of birth influences susceptibility to RA, SLE, CD and UC in addition to MS using the largest cohort to date to investigate these effects (n=115,172). Since all these conditions have been linked to vitamin D deficiency (90, 162), I also tested whether the risk of disease by month of birth follows the same seasonal distribution of predicted ultraviolet B (UVB) light radiation and 25-hydroxyvitamin D (25-OH-D) levels during gestation.

Finally, since thymic development and T cell central tolerance mainly occur *in utero,* are susceptible to intrauterine exposures (163) and are thought to play an important role in MS as well as in other ID, I sought to assess whether the month of birth has any effect on thymic function and whether cord blood vitamin D levels are associated with thymic output.

METHODS

Month of birth analysis

I obtained month of birth data for MS (n= 15,492), RA (n=39,666), SLE (n= 4,046), CD (n= 20,574) and UC (n= 23,892) patients seen by a doctor between 1997 and 2009 in Scotland and between 2003 and 2009 in England from the NHS National Service Scotland and the English Hospital Episode Statistics. For MS, an additional cohort of patients (n=11,502) and matched controls was collected as previously described (51), giving a total of 115,172 ID patients (26,162 English and 89,010 Scottish). I obtained general population controls from the General Register Office (http://www.gro-scotland.gov.uk/) and

the Office for National Statistics (http://www.ons.gov.uk/). Scottish controls (n=3,028,621) were based on month of birth registration between 1954 and 1973 and actual month of birth between 1974 and 1990. English controls (n=29,202,890) were based on actual month of birth between 1950 and 1990.

I compared cases and controls using the R package "Season" and the Cosinor test which is able to capture seasonal distributions and is particularly suitable for relatively simple and symmetric seasonal patterns. This test fits a generalized linear model under the Poisson distribution using sine and cosine terms that together describe the sinusoid. In addition to statistical significance, the model provides information on the amplitude (the height) and the phase (the peak point from 1 to 12 indicating months) of the predicted sinusoid (164). I also calculated monthly odds ratios (OR) by comparing frequencies of patients and controls born in a certain month vs the rest of the year.

I obtained average monthly UVB radiation data at the wavelength of 305 nm at noon (joules/square metre) in England and Scotland between 1979 and 1992 from the NASA's Total Ozone Mapping Program on the Nimbus 7 satellite (81). Average monthly 25-OH-D levels were collected from a large cohort of adult Scottish and English women (n=3,787) as previously described (165) and used as a proxy for the seasonal variation in gestational vitamin D status. I calculated the average predicted UVB exposure as well as vitamin D status during the first, second and third trimesters of gestation for each month of birth and I tested these for correlation with risk of ID (monthly OR) using the Spearman's correlation coefficient.

Table 4: Total number of patients used in the analysis

	MS	RA	SLE	CD	UC	All ID
England	13,075	4,747	1,622	2,463	4,255	26,162
Scotland	13,919	34,919	2,424	18,111	19,637	89,010
Total	26,994	39,666	4,046	20,574	23,892	115,172

Month of birth and thymic output

The basic mechanism for generating T-cell receptors (TCR) during T cell development provides an assay to measure thymic output (166). Signal sequences flanking the *TCR* gene segments in their germline configuration are cleaved during TCR production and join to form extrachromosomal DNA circles termed signal joint TCR excision circles (sjTRECs) (166). Since the risk of MS in England peaks in May and drops in November (51), I obtained cord blood from 50 healthy Caucasian individuals born in November and 50 born in May between 2009 and 2010 in London after ethical approval. CD4+ and CD8+ T cell populations were separated using magnetic microbeads. Quantification of sjTRECs in CD4+ and CD8+ populations was performed by real time quantitative PCR in duplicate as previously described (167). In brief, species-specific forward and reverse PCR primers and a fluorescently labelled species-specific real-time PCR probe were used for the amplification of a unique site on the T cell receptor δ sjTRECs of both collected samples and that of a calibrated progressively diluted sjTREC DNA plasmid standard for quantitation. Both samples and DNA standard were amplified for 45 PCR cycles using a real-time thermal cycler: forward primer: 5'-CAC ATC CCT TTC AAC CAT GCT-3'; reverse primer: 5'-GCC AGC TGC AGG GTT TAG G-3'; probe: 5'-/56-FAM/AC ACC TCT G/ZEN/G TTT TTG TAA AGG TGC CCA CT/3IABkFQ/-3'. Cycles used were: 1) step 1, 95°C for 10 min (1 time); 2) step 1, 95°C for 15 s; step 2, 60°C for 1 min (45 times); 3) step 1, 4°C HOLD (1 time). Ct values of experimental samples were then converted to number of sjTRECs molecules by comparing sample Ct to the curve

obtained from the standard DNA. Concentrations of 25-OH-D were measured by isotope-dilution liquid chromatography–tandem mass spectrometry with limit of detection=0.25 nmol/L. Concentrations were measured by a laboratory at Homerton Hospital (London) certified by DEQAS (Vitamin D External Quality Assessment Scheme: http://www.deqas.org/).

RESULTS

Month of birth bias in ID patients

In order to assess whether month of birth influences susceptibility to immune disorders I initially compared the distribution of all ID patients with that of the general population. Using the Cosinor test, I found that the birth distribution of ID patients follows a seasonal distribution as compared to the general population (p=5e-12, amplitude=0.033, phase=3.08, low point=9.08). When I calculated monthly ORs, I found a statistically significant peak in April (OR=1.045, 95%CI=1.024-1.067, p<0.0001) and a significant trough exactly six months later in October (OR=0.945, 95%CI=0.925-0.966, p<0.0001). I also detected a smaller deficit in August (OR=0.972, 95%CI=0.951-0.9927, p=0.008) (figure 5). The peak to trough ratio indicated the presence of a 6.5% increased risk for individuals born in April vs those born in October (OR=1.065, 95%CI=1.035- 1.096, p<0.0001).

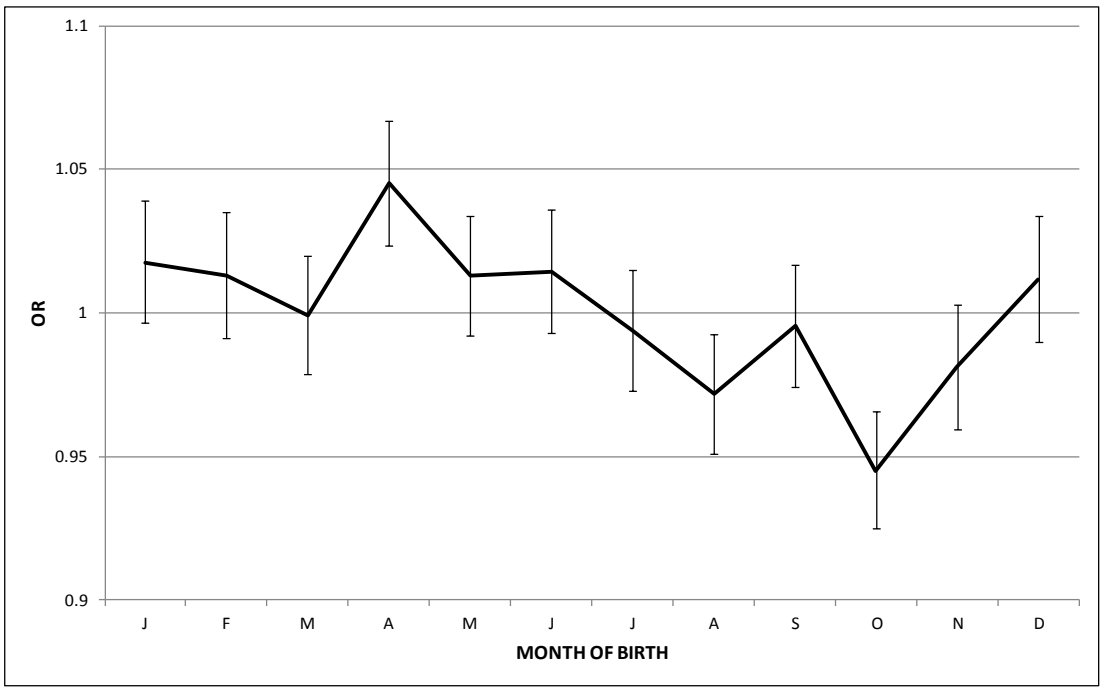

Figure 5: Odds ratio distribution with 95% CI based on month of birth in all ID (n=115,172) vs general population. April peak and October trough of risk can be observed.

When the analysis was performed according to country, the seasonal effect appeared to be present in both England and Scotland (Scotland *p*=5e-10, amplitude=0.034, phase=3.05, low point=9.05; England *p*=0.005, amplitude=0.032 phase=3.23, low point=9.23). The highest and lowest monthly ORs were present in the Scottish population (figure 6).

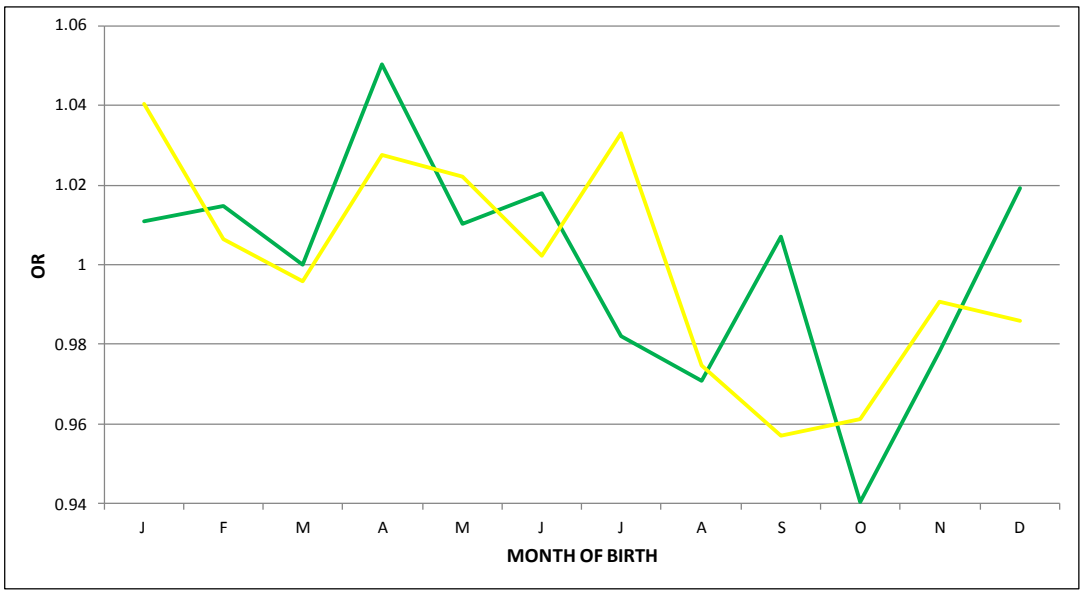

Figure 6: Odds ratio distribution based on month of birth in England (yellow) and Scotland (green). The seasonal distribution of ID births appears more striking in Scotland.

The seasonality of birth detected by grouping all ID patients could arise from a single disease such as MS, for which the presence of a month of birth effect has already been described. I therefore stratified the analysis by disease type. The Cosinor test indicated the presence of clear seasonality in all ID but CD: MS (p=5e-06; amplitude=0.041, phase=4.12, low point=10.12); RA (p=5e-04, amplitude=0.032, phase=2.69, low point=8.69); UC (p=5e-04, amplitude=0.04, phase=2.74, low point=8.74); SLE (p=0.025; amplitude=0.063; phase=2.89; low point=8.89); CD (p>0.05). When calculating monthly ORs, I observed peaks in spring and deficits in autumn in each ID apart from CD in which a January rather than spring peak was found. Birth percentages and monthly ORs with 95% confidence intervals are presented in table 5.

Table 5: Birth percentages and monthly ORs with 95% CI for each ID and all ID. The 2 months with the highest and lowest ORs are shown in red and green respectively.

Month	All ID			MS			RA		
	Birth %	OR	95% CI	Birth %	OR	95% CI	Birth %	OR	95% CI
Jan	8.63	1.02	0.99-1.04	8.51	1.01	0.97-1.05	8.44	0.99	0.95-1.03
Feb	7.90	1.01	0.99-1.04	7.76	0.99	0.95-1.03	7.94	1.02	0.98-1.06
Mar	8.88	1.00	0.98-1.02	8.67	0.97	0.93-1.01	8.99	1.01	0.98-1.05
Apr	8.77	1.05	1.02-1.07	8.79	1.05	1.002-1.09	8.78	1.05	1.01-1.08
May	8.83	1.01	0.99-1.03	9.41	1.08	1.04-1.13	8.64	0.99	0.95-1.03
Jun	8.44	1.01	0.99-1.04	8.70	1.04	1.01-1.09	8.47	1.02	0.98-1.06
Jul	8.49	0.99	0.97-1.01	8.51	0.99	0.95-1.04	8.37	0.98	0.94-1.01
Aug	8.16	0.97	0.95-0.99	8.20	0.98	0.94-1.02	8.14	0.97	0.93-1.00
Sep	8.12	1.00	0.97-1.02	7.94	0.96	0.92-1.01	8.10	1.00	0.96-1.03
Oct	8.05	0.95	0.92-0.97	8.08	0.96	0.92-1.00	8.20	0.96	0.93-0.99
Nov	7.61	0.98	0.96-1.00	7.43	0.96	0.91-1.00	7.65	0.99	0.95-1.02
Dec	8.11	1.01	0.99-1.03	8.01	1.00	0.96-1.04	8.30	1.04	1.00-1.07
Month	**UC**			**SLE**			**CD**		
	Birth %	**OR**	**95% CI**	**Birth %**	**OR**	**95% CI**	**Birth %**	**OR**	**95% CI**
Jan	8.63	1.02	0.97-1.06	9.57	1.14	1.03-1.27	8.99	1.06	1.01-1.11
Feb	8.07	1.04	0.99-1.09	7.86	1.00	0.90-1.13	7.84	1.01	0.96-1.06
Mar	8.90	1.00	0.96-1.05	8.75	0.98	0.88-1.09	8.94	1.01	0.96-1.06
Apr	8.92	1.06	1.02-1.11	8.85	1.05	0.95-1.18	8.54	1.02	0.97-1.07
May	8.73	1.00	0.96-1.05	9.71	1.12	1.01-1.24	8.40	0.96	0.91-1.01
Jun	8.25	0.99	0.94-1.04	7.61	0.90	0.81-1.02	8.44	1.02	0.97-1.07
Jul	8.44	0.99	0.94-1.03	8.58	1.00	0.90-1.12	8.75	1.03	0.98-1.08
Aug	8.15	0.97	0.93-1.02	8.40	1.00	0.90-1.12	8.13	0.97	0.92-1.02
Sep	8.18	1.00	0.96-1.05	7.51	0.91	0.81-1.02	8.45	1.04	0.99-1.10
Oct	7.91	0.93	0.88-0.97	7.93	0.94	0.84-1.05	7.89	0.92	0.87-0.97
Nov	7.69	0.99	0.95-1.04	7.09	0.91	0.81-1.03	7.78	1.00	0.95-1.06
Dec	8.13	1.01	0.97-1.06	8.13	1.02	0.91-1.14	7.84	0.97	0.93-1.03

I next investigated whether the monthly risk of ID inversely correlated with predicted gestational UVB exposure and vitamin D status during different trimesters of pregnancy. Based on the Nimbus 7 satellite, in the United Kingdom UVB radiation reaches the minimum and maximum levels during winter (December-January) and summer (June-July) respectively. The highest and lowest 25-OH-D levels were collected during September and February respectively (165). Figure 7 shows the direct

relation between UVB radiation and vitamin D status and the amount of time required for a change in UVB to impact on vitamin D metabolism. The peak and the trough of 25-OH-D levels are shifted approximately 2-3 months later than UVB radiation (two months lag: Spearman's rho=0.91, p <2.2e-16; three months lag: Spearman's rho=0.88, p=0.002). This is consistent with previous reports (168).

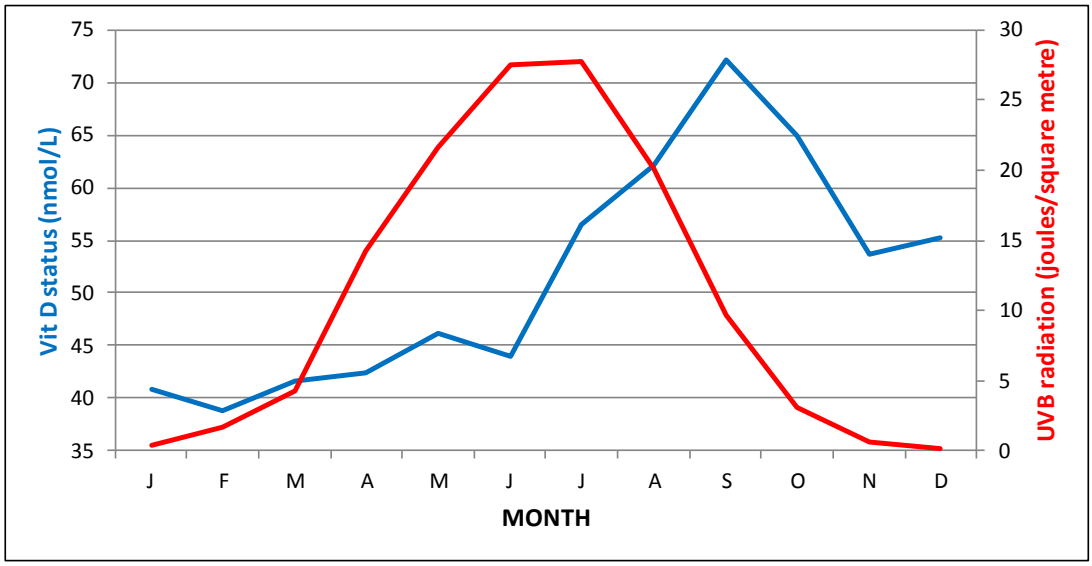

Figure 7: Correlation between monthly UVB radiation and 25-OH-D levels. The seasonal distribution of 25-OH-D levels is shifted approximately 2-3 months later than that of UVB radiation.

I found that the monthly risk of ID inversely correlated with predicted UVB exposure during the second trimester of pregnancy (Spearman's rho=-0.51, p=0.00003). Similarly, predicted maternal 25-OH-D levels were also inversely associated with risk of ID but the negative correlation was shifted to the third trimester (Spearman's rho=-0.45, p=0.0002) (figure 8).

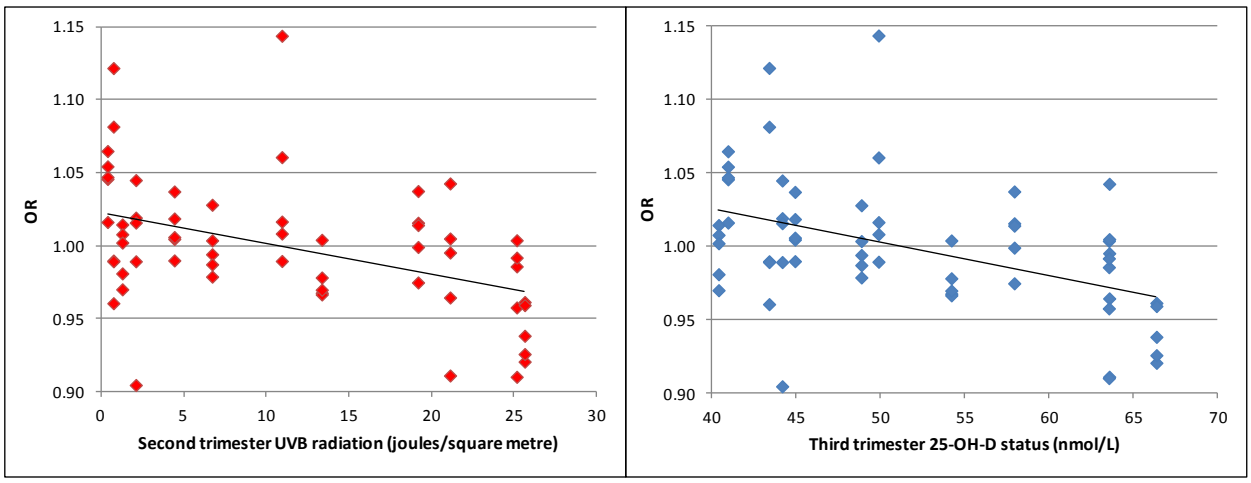

Figure 8: Inverse correlation between risk of ID and predicted second trimester UVB exposure (left panel) and third trimester vitamin D status (right panel).

Month of birth influences thymic output

I observed a significantly greater number of sjTRECs in CD4+ and CD8+ cells from individuals born in May as compared to November (p=5.5x10^{-5} and 1.2x10^{-6} respectively, table 6). Individuals born in May had significantly less circulating 25-OH-D than those born in November (P=0.016), and there was also an inverse correlation between 25-OH-D and CD4+ and CD8+ sjTRECs (Spearman's rho=-0.37, p=0.009 and Spearman's rho=-0.4, p=0.004).

Table 6: Mean CD4+ and CD8+ sjTRECS and 25-hydroxyvitamin D in May and November born individuals

	Mean sjTREC/100,000 cells (95% Confidence Interval)	**Monthly Mean 25-hydroxyvitamin D nmol/L**
CD4+ Nov	11,089 (10,205-11,973)	50.9
CD8+ Nov	10,414 (9,509-11,319)	
CD4+ May	19,547 (17,609-21,485)	38.4
CD8+ May	25,520 (22,488-28,552)	

DISCUSSION

I report here the largest study performed on ID and seasonality of birth. When patients suffering from different conditions were grouped together, I observed a clear seasonal birth distribution with peak in April and trough exactly six months later in October. The effect size of being born at the "wrong time" appears very low, with the highest ORs being under 1.1. However, considering the increased risk of all ID in rest of the year vs October born individuals and the proportion of population born in months other than October, the population proportional attributable risk percent (PPAR) is 5.05%. This suggests that approximately 5% of ID cases could be prevented by ameliorating the risk factor responsible for the seasonal distribution of ID births. The season of birth effect was particularly clear in Scotland as compared to England but no significant differences between the two sites could be observed.

That the risk of MS varies by month of birth had already been shown in a number of regions including Canada, Denmark, Sweden, Sardinia, Finland, England, Scotland and Australia (51-58). I further confirmed these findings by increasing the sample size of a previously analyzed cohort of UK MS patients (51). Based on the Cosinor test also RA, UC and SLE births followed a clear seasonal distribution. Notably, all the predicted sinusoids peaked around the same period with phases ranging from 2.69 to 4.12 (late winter-spring). In contrast to other ID the distribution of CD births was not seasonal.

A recent Australian study reported an inverse association between the risk of MS and UVB exposure during the first trimester of gestation (58). However, the sample size was relatively small (n=1,524) and thus analysis had to be performed using bi-monthly periods. Furthermore, the seasonal variation of 25-OH-D levels was not investigated and no other studies have tried to answer the same question in ID other than MS. I found that the risk of ID was inversely associated with predicted second trimester UVB exposure and third trimester vitamin D status.

By looking at TRECs, I observed that month of birth has a significant effect on thymic output and that this is inversely associated with the vitamin D status at the end of the pregnancy. One possible hypothesis is that birth month influences T cell production and may impair T cell central tolerance and/or T regulatory / T effector cell balance, predisposing to MS and other immune conditions. Notably, genetic regions associated with immune mediated diseases including MS are significantly enriched for vitamin D receptor binding sites and vitamin D is able to regulate the expression of hundreds of genes, many of which play roles in T and B cell development (94). Animal studies have shown that in utero vitamin D deficiency has a significant effect on the developing immune system (169, 170). These findings represent the first direct evidence suggesting that the same happens in the developing human immune system.

This study has limitations. Information on sex and ethnicity was not available and this may have confounded the results. The data I gathered from the Scottish NHS and the English HES could not be restricted to UK born but only to UK resident individuals. However, the large sample size (115,172 ID cases), the relatively homogeneous Scottish population and the strong a priori evidence for a month of birth effect in MS make the risk of a spurious association less likely. Furthermore, it is striking that the ID analyzed (apart from CD) show a similar seasonal risk distribution which is also the one reported in T1D patients (147, 148).

I was limited to using average UVB radiation and general population vitamin D measures which may differ from the individual maternal exposures. It is important to note that the UVB and vitamin D correlation analysis does not prove causation and that although the vitamin D hypothesis is supported by both epidemiological and functional observations, seasonality dominates many features of the global environment and other seasonal factors may play a role in determining the risk of ID. Climate, temperature, infectious disease and maternal nutrition are all characterized by seasonality and thus represent excellent candidate factors.

In summary, the susceptibility to different ID in addition to MS in the United Kingdom appears to be influenced by the season of birth. This is particularly clear in patients suffering from MS, RA, UC and SLE and suggests that at least some proportion of ID risk is preventable. In addition, I provide evidence for a role of vitamin D underlying the season of birth association with MS risk through effects on thymic development and T cell production. Long term prospective studies should be initiated to investigate the effect of gestational vitamin D supplementation on thymic output, and ultimately, the subsequent risk of MS and other ID.

CHAPTER 3: EBV, LATITUDE, VITAMIN D AND MS

*-**Disanto G**, Pakpoor J, Morahan JM, Meier UC, Giovannoni G, Ramagopalan SV. Epstein barr virus, latitude and multiple sclerosis. Multiple sclerosis journal (2013)19(3):362-5*
*-**Disanto G**, Handel AE, Damoiseaux J, Hupperts R, Giovannoni G, Smolders J, Ramagopalan SV (2013) Vitamin D supplementation and antibodies against the Epstein-Barr virus in multiple sclerosis patients. Multiple Sclerosis journal (in press)*

INTRODUCTION

In addition to vitamin D deficiency, month of birth and smoking history, epidemiological evidence strongly implicates the Epstein-Barr virus (EBV) in MS aetiology (35). The risk of MS is extremely low among EBV negative individuals and augmented in those with either high anti-EBV antibody titres or a history of symptomatic EBV infection (infectious mononucleosis) (98, 101, 105). Although the vast majority of adult MS patients are seropositive for EBV antibodies, frequency of seropositivity varies from study to study ranging from 80 to 100%. Since the prevalence of MS increases with latitude, I reviewed all studies assessing EBV seroprevalence in MS patients and controls and investigated whether this also varies according to latitude. The *a-priori* hypothesis was that EBV seroprevalence would increase with distance from the equator. Latitude is a major determinant of vitamin D status and it is well known that vitamin D is a potent modulator of the immune system which plays an important role against bacterial and viral infections (171, 172). I therefore also investigated whether vitamin D supplementation has any effect on the humoral response against EBV.

METHODS

PubMed searches were undertaken to identify studies investigating the association between MS and EBV using the terms ("herpesvirus 4, human"[MeSH Terms] OR "human herpesvirus 4"[All Fields] OR ("epstein"[All Fields] AND "barr"[All Fields] AND "virus"[All Fields]) OR "epstein barr virus"[All Fields]) AND ("multiple sclerosis"[MeSH Terms] OR ("multiple"[All Fields] AND "sclerosis"[All Fields]) OR

"multiple sclerosis"[All Fields]). EMBASE was searched for abstracts using the terms “multiple sclerosis” [All fields] AND “epstein barr virus” [All fields]. Twenty two studies assessing EBV infection in adult MS patients and unrelated controls were found (supplementary table 1). After excluding those studies in which latitude could not be precisely determined (n=2) (99, 173) and studies performed in the southern hemisphere (n=1) (174), 19 studies were used for analysis (104, 175-192). The relation between latitude and EBV seroprevalence was investigated using logistic regression and assessing differences between proportions. Serum samples were collected from 15 relapsing remitting MS patients that were supplemented with 20,000 IU/day vitamin D3 for 12 weeks (193) and their circulating antibody levels against EBNA1 were measured using the Liaison® quantitative chemiluminescent assay (194). Statistical analyses were performed using R.

RESULTS

All studies were performed between 29^{o} and 65^{o} latitude. There was no evidence of either study heterogeneity (I^2=1.5% (95%CI=0%-44.5%)) or publication bias (Egger p=0.06). I first investigated whether MS status and latitude were significant predictors of EBV status using logistic regression. As expected the strongest predictor of EBV seroprevalence was MS status (OR=4.41, 95%CI=2.4-8.08, p<0.001). However, latitude was also positively associated with EBV status independently of MS status (OR=1.06, 95%CI=1.02-1.09, p=0.002). When mean age and sex ratio were also included in the regression model, the effect of latitude was attenuated but still present (OR=1.05, 95%CI= 1.0003=1.119, p=0.038). Notably, among MS patients 7 of the 9 studies performed at the highest latitudes reported an EBV seroprevalence of 100% as compared to 2 of the remaining 10 studies performed at the lowest latitudes (figure 9a).

Given the association between vitamin D deficiency and MS risk (80), I grouped studies into those conducted below and above 42^{o} latitude, as above 42^{o} latitude most UVB radiation is absorbed by

the atmosphere and vitamin D production is more difficult (195, 196). EBV seroprevalence was substantially higher in the ≥42^{o} group in both MS patients and controls (MS *p*=0.001, controls *p*=7.17E-11) (figure 9c). Different methods for antibody detection have different sensitivity and specificity for EBV (197). The most frequently used method was enzyme-linked immunosorbent assay (ELISA) tests. Performing a subgroup analysis of those studies using ELISA (n=9) did not change the results (MS *p*=4.91E-09, controls *p*=4.57E-06) (figure 9d). The proportion of EBV positive cases and controls with 95% CI according to latitude groups and their differences are presented in table 7.

I then aimed to investigate whether vitamin D may be the factor driving the association between latitude and EBV serology by measuring antibody levels against EBNA1 pre- and post-vitamin D supplementation. I found that the initial antibody levels (mean=77.98 arbritrary units/ml (AU/ml), 95%CI=57.40-98.56) were significantly reduced after vitamin D supplementation (mean=62.27, 95%CI=46.73-77.80) (mean difference=15.72; 95%CI=3.37-28.07; paired T test *p*=0.016) (figure 9) (supplementary table 2). Notably, the effect of vitamin D was particularly strong in those patients with higher antibody levels and a positive correlation was present between pre-supplementation antibody levels and their decrease following supplementation (Pearson's rho=0.66, p=0.007) (figure 10).

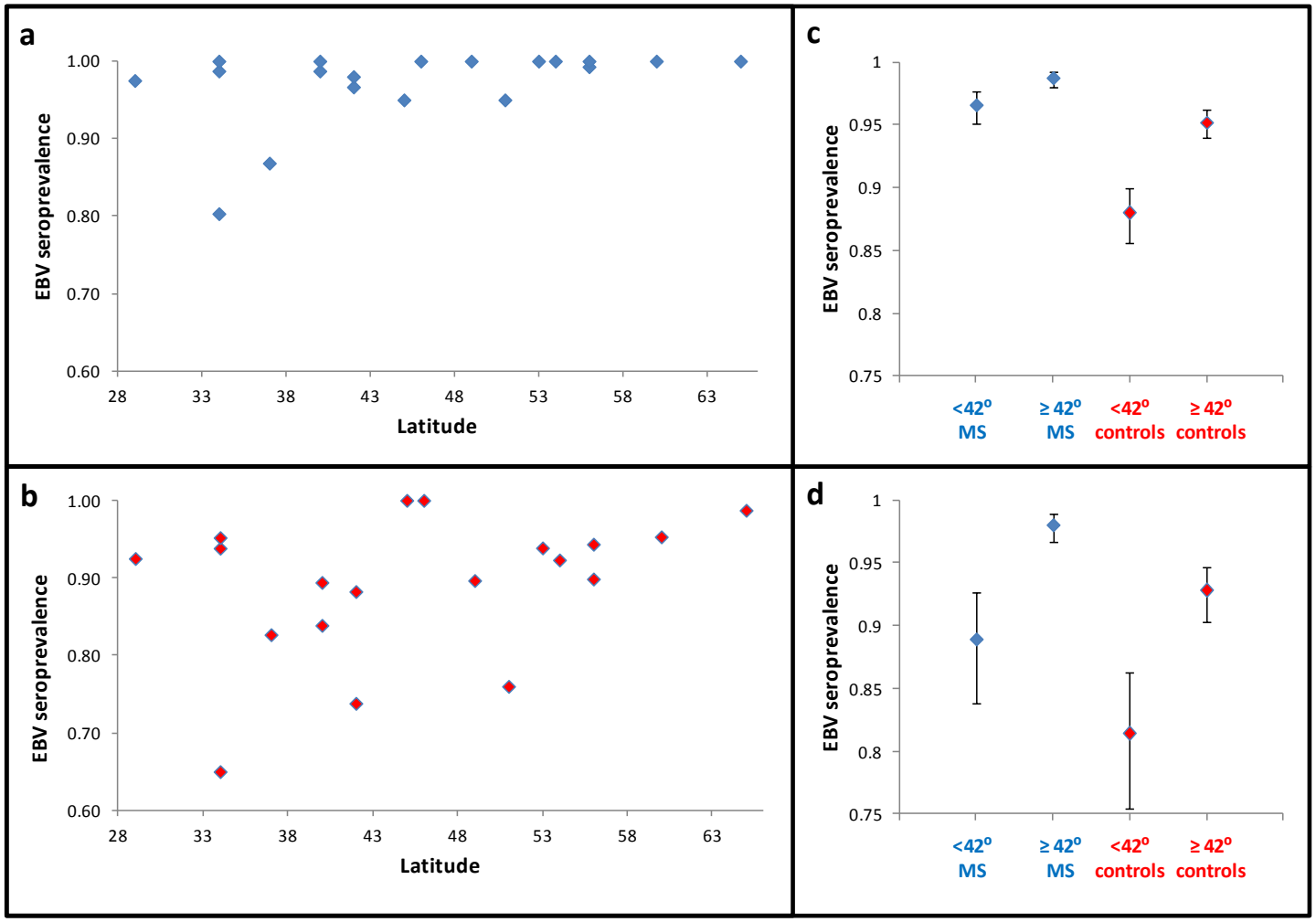

Figure 9: a) Distribution of EBV seroprevalence in MS patients according to latitude; b) Distribution of EBV seroprevalence in controls according to latitude; c) EBV seroprevalence with 95% CI in ≥42° vs <42° latitude groups among MS patients (blue) and controls (red); d) Subgroup analysis of studies using ELISA.

Table 7: Difference between proportions of EBV positive MS patients and controls by latitude group (<42° vs ≥42°).

	MS				Controls			
	<42° (n=884)	**≥42° (n=1,291)**	**Difference**	***p*-value**	**<42° (n=899)**	**≥42° (n=1,584)**	**Difference**	***p*-value**
Proportion EBV +	0.966	0.988	0.022	0.001	0.880	0.952	0.072	7.17E-11
95% CI lower limit	0.951	0.979	0.007		0.856	0.940	0.048	
95% CI upper limit	0.977	0.993	0.036		0.900	0.962	0.097	
	MS (ELISA)				**Controls (ELISA)**			
	<42° (n=217)	**≥42° (n=763)**	**Difference**	***p*-value**	**<42° (n=215)**	**≥42° (n=595)**	**Difference**	***p*-value**
Proportion EBV +	0.889	0.980	0.091	4.91E-09	0.814	0.928	0.114	4.57E-06
95% CI lower limit	0.838	0.967	0.045		0.754	0.903	0.055	
95% CI upper limit	0.926	0.989	0.137		0.862	0.947	0.173	

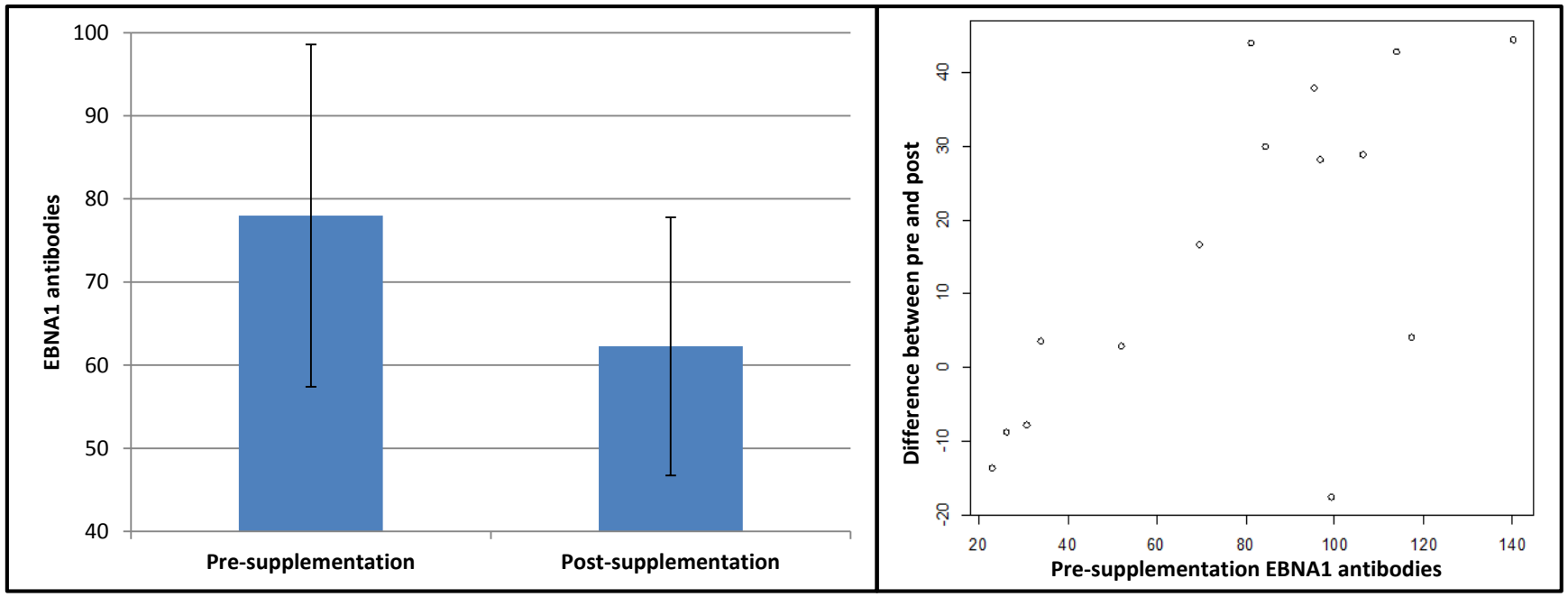

Figure 10: Pre- and post- vitamin D supplementation mean anti EBNA1 antibody levels with 95%CI (left panel); Positive correlation between pre-supplementation antibody levels and decrease following supplementation (right panel).

DISCUSSION

I carried out a systematic review of EBV studies in MS and showed that EBV seroprevalence in both MS patients and controls is positively associated with latitude. This effect was still present when the analysis was corrected for MS status, mean age and sex ratio with an OR of 1.05 per degree increase. Interestingly, the effect appeared particularly strong when studies were grouped into those conducted above and below a latitude threshold which has a functional effect on vitamin D production. Notably, the latitude effect was present in both MS cases and controls and this suggests that the relation between latitude and EBV is not specific to MS but likely represents a general phenomenon acting at the population level.

The main limitation of this study is the variation in methodologies used and antibodies tested to detect past EBV infection. This variation represents a potential confounding factor and future studies should investigate EBV status across different latitudes using the same method and testing the same antibodies. However, the increase by latitude is still present when a more homogeneous group of studies (those using ELISA based tests) was considered. Nevertheless, as compared to gold standard methods (indirect immunofluorescence and anti complement immunofluorescence) ELISA tests are approximately 90% sensitive and specific (197).

These findings suggest a role for latitude related factors influencing EBV infection. Genetic, social and climatic differences are all likely to contribute to this observation. If socioeconomic factors were primarily involved, one would expect to see an inverse relation between latitude and EBV status since worldwide hygiene generally increases with latitude. Among other candidate factors is vitamin D deficiency and this hypothesis is strongly supported by the observation that vitamin D supplementation considerably decreases the antibody levels against EBV and that this effect is particularly strong in those individuals with the highest antibody levels. It is important to note that this mechanism appears to be specific for anti-EBNA1 and/or for other specific antibodies, since no

change in total Ig-load and B cell differentiation profile after vitamin D supplementation was found in these patients (198).

Vitamin D deficiency could impair the immune response to EBV and increase either susceptibility to EBV infection or its control after infection has occurred, leading to viral reactivation and augmented antibody production. Therefore, if EBV was a causative factor in MS, vitamin D deficiency could enhance its pathogenic role. In line with this hypothesis is the observation that a statistical interaction between infectious mononucleosis prevalence and UVB radiation could explain 72% of the variance of MS prevalence across England (81). These considerations are further supported by recent studies showing that anti-EBNA1 antibody levels and vitamin D status are inversely correlated in young individuals (199) and that people with high serum vitamin D levels are less likely to have EBV shedding in saliva (200). In line with these findings is the observation that EBV antibody levels are significantly higher in black-African American (at high risk of vitamin D deficiency) than in white American youth (although a role for genetic differences cannot be excluded) (201, 202). Future studies are needed to understand how vitamin D influences the production of antibodies against EBV, the immune response against this virus and whether this mechanism is specific to EBV or shared with other viruses. These studies will elucidate how these two environmental factors are linked in the causal cascade leading to MS.

CHAPTER 4: GENOMIC REGIONS ASSOCIATED TO MS ARE ACTIVE IN B CELL LINES

-***Disanto G**, Morahan JM, Barnett MH, Giovannoni G, Ramagopalan SV: The evidence for a role of B cells in multiple sclerosis. Neurology 2012, 78:823-832.*
-***Disanto G**, Sandve GK, Berlanga-Taylor AJ, Morahan JM, Dobson R, Giovannoni G et al. Genomic regions associated to multiple sclerosis are active in B cells. PLoS One 2012;7(3):e32281*

INTRODUCTION

A recent genome wide association study (GWAS) involving more than nine thousand MS patients found evidence for association of MS with 57 genomic regions (31). However, there remains limited understanding as to how these variants are involved in MS development. Although T cells have traditionally been thought to mediate MS pathophysiology, attention to the role of B cells is increasing (203, 204). The presence of oligoclonal bands in the cerebrospinal fluid is the most consistent immunological finding in MS, and this indicates abnormal B cell activation within the CNS of MS patients (24). Furthermore B cell abnormalities influence both conversion to clinically definite MS, MRI activity, onset of relapses and disease progression (14, 25-29). Possibly the strongest evidence for B cells in MS comes from clinical trials showing that MRI activity and onset of relapses are significantly decreased after depletion of CD20+ B cells (205, 206). Further details can be found in figure 11.

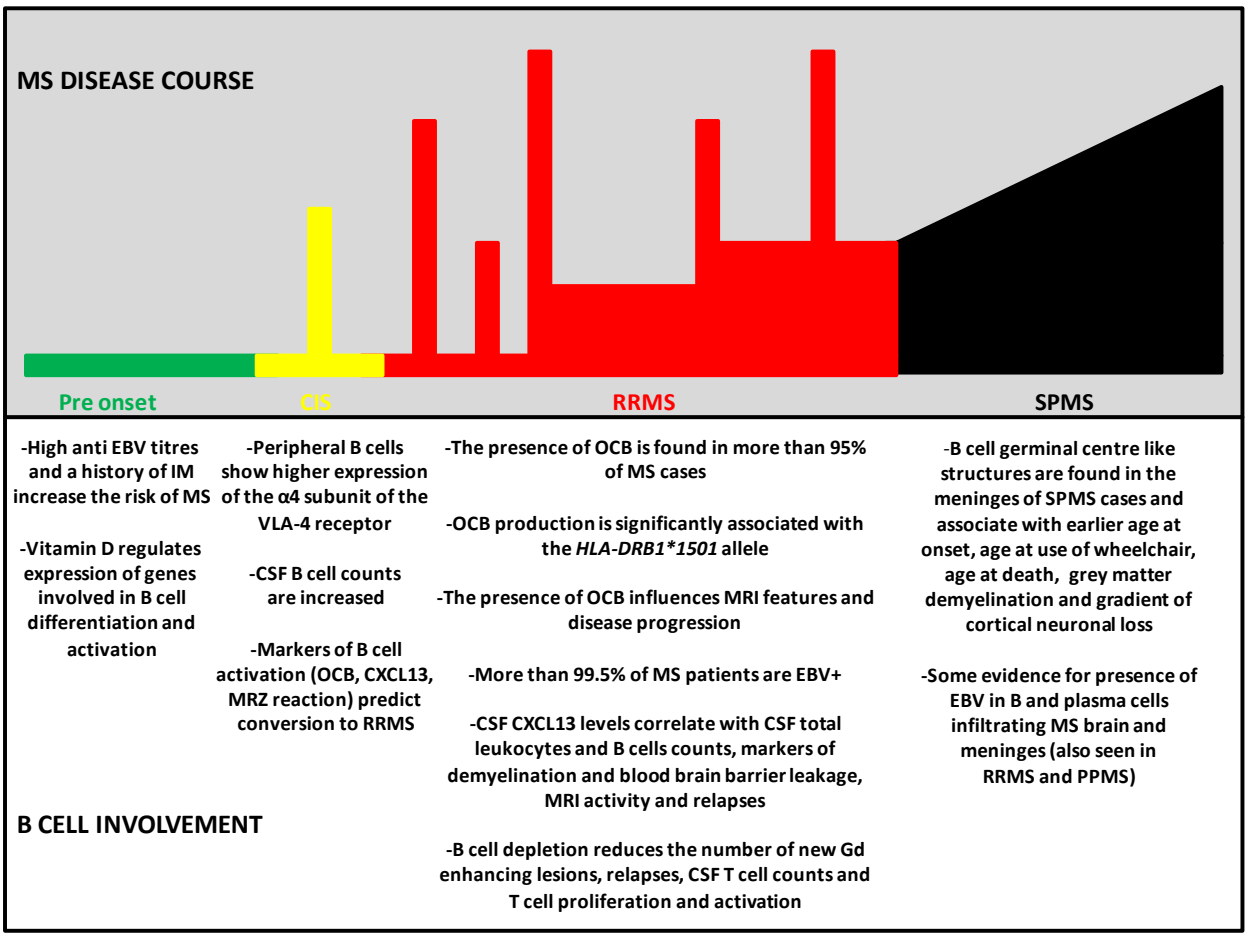

Figure 11: B cell involvement in different phases of MS. Environmental risk factors influence B cells prior to disease onset; B cell abnormalities within the CNS are already present at time of CIS diagnosis and predict conversion to clinically definite MS; during RRMS markers of B cell activation correlate with disease course and MRI activity while B cell depletion significantly reduces MRI lesions and onset of relapses; organized B cell structures are found in the meningeal space of SPMS patients and their presence is associated with a worse outcome.

The regulation of genes can be just as important as the proteins they encode. Regulatory elements in the genome are much harder to identify than protein-coding genes because they lack distinguishing sequence signatures. Moreover, many regulatory elements function only in certain cell types and conditions (207). Chromatin profiling is a powerful means of genome annotation and detection of regulatory activity. The chromatin landscape of a cell is distinctive for a specific cell type and among other roles determines which regions of the genome are accessible to the binding of transcription

factors and whether transcription occurs or is repressed. A recent study mapped a number of chromatin marks across nine cell types to systematically characterize regulatory elements and their cell-type specificities. These included enhancer elements (DNA sequences able to modulate gene expression through the binding of transcription factors to them), promoter regions (DNA regions located near the transcription start site of a gene which facilitate the binding of RNA polymerase and the initiation of transcription), polycomb repressed (DNA regions in which gene expression is actively repressed by the binding of polycomb group proteins), heterochromatin (large portions of DNA which are densely packed and therefore less accessible to transcription factors), insulator sites (DNA elements bound by the zinc finger protein CTCF which functions as an enhancer-blocking element), transcribed regions and finally repetitive / copy number variations (CNV, DNA regions characterized by a variable number of copies between individuals). Among the cell types profiled were cell lines derived from B cells, hepatocytes, fibroblasts and keratinocytes (208).

A large proportion of SNPs associated with MS do not lie in the coding regions of genes and therefore are likely to influence disease risk through a gene regulatory role. It is plausible that genetic variants associated with a certain disease act through influencing the particular cell type(s) that trigger disease onset. Therefore, one would expect to observe an overlap between genomic regions associated with disease risk and those which are active in the causative cell type(s). The aim of this study was to assess whether genomic regions that have been associated with MS significantly overlap with active regulatory regions in B cell lines and whether this overlap is higher than that observed in non-immunological cell types. This potentially provides us with relevant information regarding the importance of the immune system and B cells as mediators of disease in MS.

METHODS

Data acquisition

Genetic variants with confirmed association to MS risk were obtained from the recent GWAS performed by the International Multiple Sclerosis Genetics Consortium (IMSGC) and the Wellcome Trust Case Control Consortium 2 (WTCCC2) (n=57) (31). MS regions were defined as genomic intervals of 0.25cM centred on the lead associated SNP. The chromatin profiles of B-lymphoblastoid cells (GM12878), hepatocellular carcinoma cells (HepG2), normal lung fibroblasts (NHLF) and normal epidermal keratinocytes (NHEK) were obtained from the ENCODE project (208). Briefly, chromatin immunoprecipitation followed by massively parallel DNA sequencing (ChIP-seq) and expression data were used to identify different classes of chromatin states: active promoter (AP), weak promoter (WP), poised promoter (PP), strong enhancer (SE), weak enhancer (WE), polycomb repressed (PR), heterochromatic (H), insulator (I), strongly transcribed (ST), weakly transcribed (WT) and repetitive/CNV (Rep/CNV) (208).

Overlap analysis

All analyses were performed using the Genomic Hyperbrowser (http://hyperbrowser.uio.no/hb/) (209). The enrichment of MS regions in a certain chromatin state (e.g. AP) was calculated as the ratio of the proportion of AP intervals covered by MS regions, to the proportion of non-AP intervals covered by MS regions. In order to assess whether the overlap between MS regions and a certain chromatin state was higher than expected by chance, a permutation based analysis was performed. I defined a null model for which the location of individual chromatin intervals varied randomly, while preserving the empirical segment and inter-segment length distribution of chromatin intervals. MS regions were fixed. The number of overlapping base pairs between the two tracks was calculated for the real data, as well as for 20,000 Monte Carlo samples from the null model. The p-value was

calculated in the usual way, i.e. as the proportion of Monte Carlo samples being equal to or more extreme than the observed overlap. These analyses were performed on both a global (whole genome) and local (chromosome arms) scale and for each cell type. For local analyses p-values were adjusted to a FDR of 10%.

When comparing B cell lines to non-immunological cell types, case-control tracks were created for each chromatin state by removing all parts of chromatin intervals that overlapped between B and control cells and marking the remaining intervals as case (B cell specific intervals) and control (other cell type specific intervals). P-values were computed by a Monte Carlo procedure, in which the case-control labels of chromatin intervals were randomly permuted. The observed base pair overlap between case intervals and MS regions were compared against the corresponding distribution for 20,000 Monte Carlo samples in the usual way. The fold enrichment difference in overlap between B and control cells was calculated as the ratio between the proportions of case and control intervals that overlapped with MS regions.

Finally I tested whether MS associated SNPs (primary SNPs) and SNPs in perfect linkage disequilibrium (LD) with primary SNPs (r^2=1) were located within certain chromatin states more than expected by chance as described above for MS regions.

RESULTS

Active chromatin states in B cell lines overlap with MS regions

The first aim was to assess whether and where in the genome a particular chromatin state in B cell lines significantly overlapped with MS regions. The enrichment of MS regions in different chromatin states and the significance of the overlap are presented in table 8. On a global scale (whole genome) enrichment values varied considerably between different chromatin states ranging from 0.34 in H to 3.07 in SE regions. When testing statistical significance, MS regions overlapped with promoter (AP

and WP $p<0.00005$; PP $p=0.0005$), enhancer ($p<0.00005$) and transcribed ($p<0.00005$) regions more than expected by chance.

In order to assess whether the significant global overlap was homogeneously distributed across the genome or resulted from particularly highly enriched regions, the same analysis was performed on a local scale by dividing the whole genome into chromosome arms. This resulted in 43 different 'bins' of which 17 had to be excluded due to the absence of MS associated regions in those chromosome arms leaving a total of 26 bins. Out of these 26 bins, statistically significant overlap was found in 18 for AP, 15 for WP, 7 for PP, 23 for SE, 24 for WE, 3 for PR, 0 for H, 2 for I, 9 for ST, 9 for WT and 1 for Rep/CNV chromatin states (table 8). As expected, the chromatin states with significant overlap on a global scale were those with the highest number of significant bins. SE and WE regions showed the most homogeneously distributed overlap, being significant in all but 3 and 2 bins respectively. The overlap of promoter and transcribed regions appeared more dependent on particular bins.

Active chromatin states in B cell lines overlap with MS regions more than in non immune cell types

On its own, the presence of an overlap between MS regions and active chromatin states in B cell lines is not sufficient to indicate that these are relevant to MS pathogenesis. Both MS regions and active chromatin states could just be more likely to be near commonly transcribed genes, giving rise to co-localization in the absence of any direct relationship between the MS-associated regions and chromatin states. To rule out this hypothesis I tested 3 additional cell types (hepatocytes, fibroblasts and keratinocytes) that, based on current knowledge, are not implicated in MS pathogenesis. Enrichment values, global significance and number of significant bins are presented in table 9.

Similarly to findings observed in B cell lines, promoter, enhancer and transcribed regions overlapped with MS regions more than expected by chance in all control cell types. However, the number of significant bins as well as the enrichment values tended to be higher in B than in other cell types. I

explored this further by directly comparing the overlap in B cell lines with that of the control cells (table 10). Strikingly the overlap between MS regions and AP, SE, WE and ST regions was significantly higher in B cell lines than in any of the other cell types. The highest fold enrichment differences were observed for higher activity states (AP, SE and ST).

MS associated SNPs preferentially land in active promoter and enhancer regions

The presence of significant overlap between MS regions and certain active chromatin states in B cells supports an important role for the immune system in MS but does not provide any insight into how MS risk variants may be acting. I attempted to answer this question by looking at where in the genome MS associated SNPs, and SNPs in perfect LD (r^2=1), preferentially land. A total of 452 SNPs were tested and enrichment of chromatin states for MS SNPs and significance of overlap are presented in table 11. MS SNPs were located within AP, SE and WE intervals more than expected by chance. Weak evidence for overlap was also observed for WP, ST and WT regions. The full list of SNPs and relative chromatin states is presented in supplementary table 3.

When examined in the light of the chromatin data, several regions seemed particularly interesting. For example MS associated SNPs in the regions of *CLECL1*, *CD86*, *TYK2* and *CD58* land in promoter and enhancer regions which are present in B but not in other cell types (figure 12). I also looked at the position of MS SNPs in regions in which no candidate genes have been identified. Interestingly, rs12466022 on chromosome 2 and respective SNPs in LD landed in WE and SE intervals, while rs13192841 on chromosome 6 and respective SNPs in LD were located in WE and WT regions.

Table 8: Enrichment of MS regions in chromatin states on both a global and local (chromosome arms) scale in B cell lines. Red cells indicate significant overlap

BINS	AP	WP	PP	SE	WE	PR	H	I	ST	WT	Rep/ CNV	MS SNPs and putative associated genes
Whole genome	2.72	2.13	2.35	3.07	2.22	1.35	0.34	1.2	2.57	1.92	1.013	
chr1p	1.28	1.20	0.41	2.12	1.48	0.45	0.69	0.96	1.10	1.47	4.39	rs4648356 (*MMEL1/TNFRSF14*); rs11810217 (*EVI5*); rs11581062 (*VCAM1*); rs1335532 (*CD58*)
chr1q	2.04	0.46	0.00	4.65	3.12	2.78	0.33	1.72	0.90	1.57	0.00	rs1323292 (*RGS1*); rs7522462 (*C1orf106/KIF21B*)
chr2p	3.22	2.73	4.12	12.78	3.35	1.47	0.31	1.72	0.44	1.12	0.00	rs12466022 (no gene); rs7595037 (*PLEK*)
chr2q	3.18	3.26	1.89	1.45	1.89	2.28	0.32	1.23	3.16	1.69	0.50	rs17174870 (*MERTK*); rs882300 (*CXCR4*); rs10201872 (*SP140*)
chr3p	0.00	0.00	20.01	0.00	0.28	3.18	2.33	0.15	0.00	0.12	0.54	rs11129295 (*EOMES*); rs669607 (no gene)
chr3q	3.11	1.99	1.05	7.35	2.27	0.22	0.14	0.87	5.88	2.24	0.65	rs2028597 (*CBLB*); rs2293370 (*TMEM39A/CD80*); rs9282641 (*CD86*); rs2243123 (*IL12A*)
chr4q	6.20	4.16	0.00	4.16	1.40	0.02	0.15	0.70	9.65	3.66	0.94	rs228614 (*NFKB1*)
chr5p	0.61	0.91	4.10	1.71	3.55	3.41	0.59	2.38	0.23	1.79	0.00	rs6897932 (*IL7R*); rs4613763 (*PTGER4*)
chr5q	5.85	4.21	1.17	1.42	1.51	0.82	0.48	1.22	2.74	1.06	0.35	rs2546890 (*IL12B*)
chr6q	0.51	0.82	2.71	3.21	2.28	4.03	0.59	1.49	0.04	1.12	0.55	rs12212193 (*BACH2*); rs802734 (*THEMIS*); rs11154801 (*MYB/AHI1*); rs17066096 (*IL22RA2*); rs13192841 (no gene); rs1738074 (*TAGAP*)
chr7q	2.87	1.67	0.00	1.65	0.91	2.52	0.22	0.95	6.34	1.71	0.00	rs354033 (*ZNF746*)
chr8q	0.87	1.04	0.00	0.32	1.60	0.42	0.66	0.61	2.53	1.47	0.13	rs1520333 (*IL7*); rs4410871 (*MYC*); rs2019960 (*PVT1*)
chr10p	4.40	7.65	4.42	4.98	7.49	2.93	0.13	1.35	1.98	1.85	0.51	rs3118470 (*IL2RA*)
chr10q	3.03	2.46	7.89	3.02	2.34	1.90	0.11	0.89	8.78	1.64	0.66	rs1250550 (*ZMIZ1*); rs7923837 (*HHEX*)
chr11q	2.55	2.98	8.53	1.93	3.01	1.34	0.17	0.93	3.62	2.63	0.21	rs650258 (*CD6*); rs630923 (*CXCR5*)
chr12p	4.55	1.94	0.00	3.97	2.31	0.66	0.34	1.34	2.03	1.79	0.67	rs1800693 (*TNFRSF1A*); rs10466829 (*CLECL1*)
chr12q	3.66	2.52	3.68	1.66	1.73	0.73	0.26	1.09	3.20	2.35	0.64	rs12368653 (*CYP27B1*); rs949143 (*ARL6IP4*)
chr14q	2.56	1.37	0.00	3.42	2.24	0.23	0.60	1.25	0.20	1.93	1.08	rs4902647 (*ZFP36L1*); rs2300603 (*BATF*); rs2119704 (*GALC/GPR65*)
chr16p	1.33	0.35	0.00	6.65	3.54	0.00	0.14	0.66	1.63	2.64	0.34	rs2744148 (*SOX8*); rs7200786 (*CLEC16A/CIITA*)
chr16q	0.78	1.27	0.00	16.42	8.29	0.00	0.10	1.65	0.00	3.07	3.47	rs13333054 (*IRF8*)
chr17q	2.48	2.07	0.48	2.38	1.77	0.37	0.21	0.74	2.66	1.81	1.02	rs9891119 (*STAT3*); rs180515 (*RPS6KB1*)
chr18q	7.70	5.81	4.78	3.77	3.58	2.16	0.08	1.06	11.92	2.36	0.00	rs7238078 (*MALT1*);
chr19p	1.66	1.87	4.79	2.54	2.33	2.13	0.19	1.90	1.54	1.66	0.29	rs1077667 (*TNFSF14*); rs8112449 (*TYK2/ICAM3*); rs874628 (*MPV17L2/IL12RB1*)
chr19q	0.25	0.62	2.82	2.17	1.51	2.26	1.78	1.62	0.00	0.43	0.14	rs2303759 (*DKKL1/CD37*)
chr20q	3.12	2.53	5.08	1.37	1.92	2.52	0.31	1.72	1.30	2.00	0.18	rs2425752 (*CD40*); rs2248359 (*CYP24A1*); rs6062314 (*TNFRSF6B*)
chr22q	3.15	2.79	0.31	1.50	1.26	0.33	0.32	1.01	2.48	1.66	1.09	rs2283792 (*MAPK1*); rs140522 (*SCO2*)

Table 9: Enrichment, global significance and number of significant bins in B and control cells

CHROMATIN STATE	MEASURE	B CELLS	HEPATOCYTES	FIBROBLASTS	KERATINOCYTES
Active promoter	**Global significance**	YES (p<0.00005)	YES (p<0.00005)	YES (p<0.00005)	YES (p<0.00015)
	Significant bins	18	9	11	6
	Global enrichment	2.721	2.114	2.207	1.95
Weak promoter	**Global significance**	YES (p<0.00005)	YES (p<0.00005)	YES (p=0.0001)	YES (p<0.00005)
	Significant bins	15	11	7	10
	Global enrichment	2.134	2.121	1.803	1.949
Poised promoter	**Global significance**	YES (p<0.0005)	Maybe (p=0.01975)	YES (p=0.0022)	YES (p=0.0013)
	Significant bins	7	5	4	4
	Global enrichment	2.352	1.935	2.057	2.011
Strong enhancer	**Global significance**	YES (p<0.00005)	YES (p<0.0016)	YES (p<0.00005)	YES (p=0.00065)
	Significant bins	23	1	6	7
	Global enrichment	3.074	1.908	1.886	1.577
Weak enhancer	**Global significance**	YES (p<0.00005)	YES (p<0.00005)	YES (p=0.00035)	YES (p<0.00005)
	Significant bins	24	12	4	10
	Global enrichment	2.222	1.747	1.436	1.528
Polycomb repressed	**Global significance**	Maybe (p=0.07035)	NO (p=0.6002)	YES (p=0.0056)	NO (p=0.1098)
	Significant bins	3	1	8	5
	Global enrichment	1.355	1.161	1.574	1.314
Heterochromatic	**Global significance**	NO (p=1)	NO (p=1)	NO (p=1)	NO (p=1)
	Significant bins	0	0	0	0
	Global enrichment	0.3363	0.4589	0.4273	0.4829
Insulator	**Global significance**	Maybe (p=0.08100)	Maybe (p=0.02090)	YES (p=0.0015)	YES (p=0.00025)
	Significant bins	2	5	4	7
	Global enrichment	1.205	1.267	1.348	1.376
Strong transcribed	**Global significance**	YES (p<0.00005)	YES (p=0.0004)	YES (p=0.0007)	YES (p=0.005)
	Significant bins	9	7	8	7
	Global enrichment	2.575	2.171	2.133	1.868
Weak transcribed	**Global significance**	YES (p<0.00005)	YES (p=0.0008)	YES (p<0.00005)	YES (p<0.00005)
	Significant bins	9	3	8	7
	Global enrichment	1.919	1.668	1.836	1.811
Repetitive / CNV	**Global significance**	NO (p=0.3547)	Maybe (p=0.02565)	NO (p=0.5321)	NO (p=0.1582)
	Significant bins	1	1	1	1
	Global enrichment	1.013	1.965	0.8118	1.351

Table 10: Comparison of overlaps between B cell lines and other cell types

Chromatin state	Is overlap in B cells > Hepatocytes?		Is overlap in B cells > Fibroblasts?		Is overlap in B cells > Keratinocytes?	
	Significance	Fold difference	Significance	Fold difference	Significance	Fold difference
Active promoter	YES (p<0.0001)	1.958	YES (p=0.0006)	1.888	YES (p<0.00005)	2.747
Weak promoter	NO (p=0.5972)	1.022	YES (p=0.00075)	1.239	YES (p=0.0025)	1.068
Poised promoter	YES (p=0.00945)	1.259	Maybe (p=0.07815)	1.208	NO (p=0.2274)	1.255
Strong enhancer	YES (p<0.00005)	1.709	YES (p<0.00005)	1.772	YES (p<0.00005)	2.196
Weak enhancer	YES (p<0.00005)	1.306	YES (p<0.00005)	1.626	YES (p<0.00005)	1.513
Polycomb repressed	NO (p=0.1456)	1.213	NO (p=1)	0.8860	NO (p=0.9866)	1.119
Heterochromatic	NO (p=1)	0.5286	NO (p=1)	0.5938	NO (p=1)	0.4705
Insulator	NO (p=0.7215)	0.8227	NO (p=0.3529)	0.9084	NO (p=0.9425)	0.8002
Strong transcribed	YES (p=0.0015)	1.486	YES (p<0.00005)	1.417	YES (p<0.00005)	2.027
Weak transcribed	Maybe (p=0.07585)	1.230	NO (p=0.1152)	1.096	NO (p=0.1358)	1.132
Repetitive / CNV	NO (p=1)	0.1713	NO (p=0.2893)	2.492	NO (p=0.9999)	0.4362

Table 11: Enrichment and significance of the overlap between MS SNPs and chromatin states in B cell lines

Chromatin state	MS SNPs		Significance	Enrichment
	Number	Percentage		
Active promoter	30	6.64	YES (p<0.00005)	9.202
Weak promoter	9	1.99	Maybe (p=0.0413)	2.907
Poised promoter	2	0.44	NO (p=0.181)	2.72
Strong enhancer	35	7.74	YES (p=0.0006)	4.878
Weak enhancer	57	12.61	YES (p<0.00005)	4.580
Polycomb repressed	9	1.99	NO (p=0.7278)	0.63
Heterochromatic	162	35.84	NO (p=1)	0.216
Insulator	4	0.88	NO (p=0.2574)	1.634
Strong transcribed	59	13.05	Maybe (p=0.0529)	2.271
Weak transcribed	85	18.81	Maybe (p=0.0357)	1.973
Repetitive / CNV	0	0.00	NO (p=1)	0

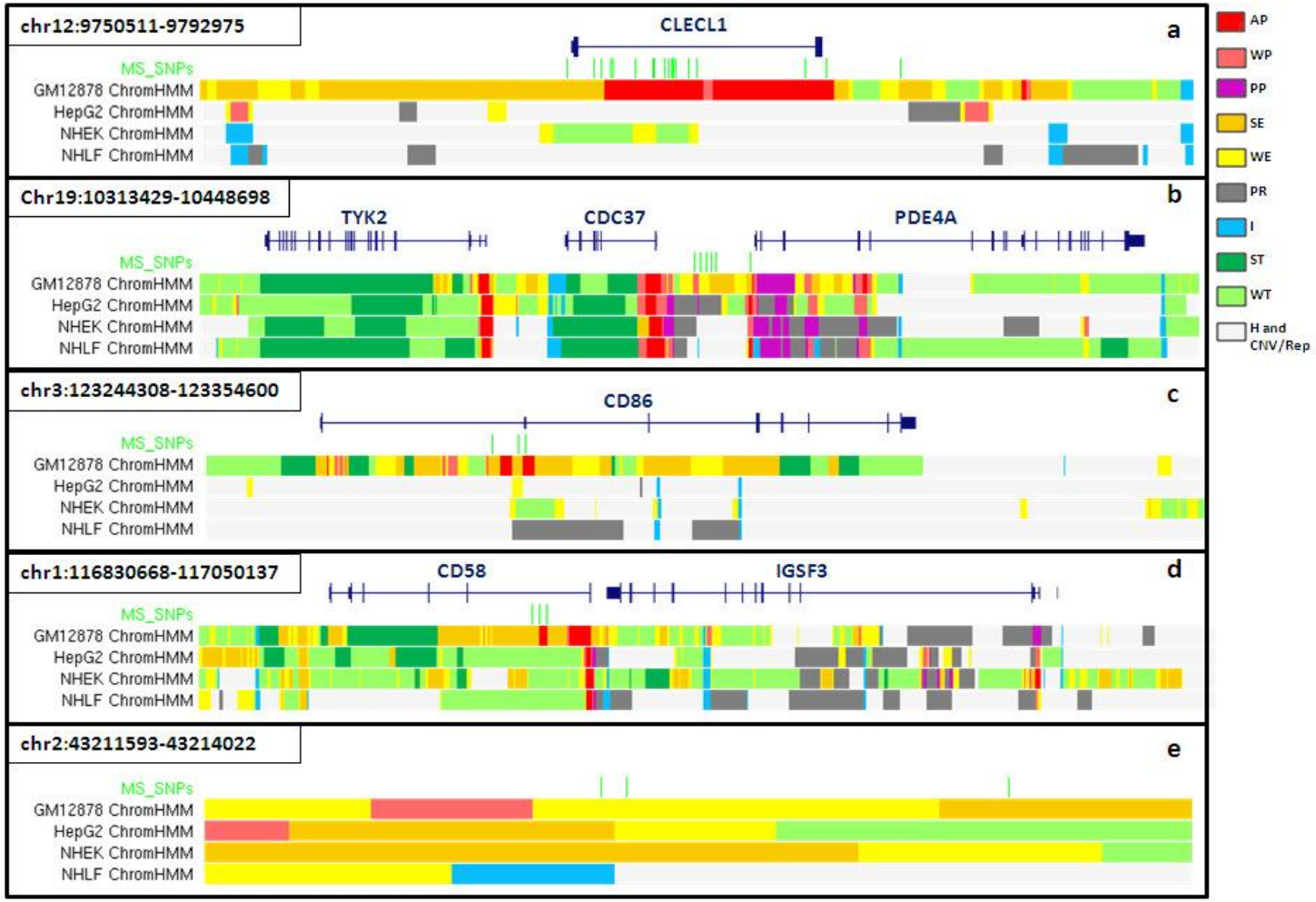

Figure 12: MS SNPs land in B cell specific promoter and enhancer elements in the regions of CLECL1, TYK2, CD86, and CD58 (a-d); MS SNPs in a region with no candidate gene land in enhancer intervals (e). Chromatin states: AP=active promoter; WP=weak promoter; PP=poised promoter; SE=strong enhancer; WE=weak enhancer; PR=polycomb repressed; H=heterochromatic; I=insulator; ST=strong transcribed; WT=weak transcribed; CNV/Rep= CNV/repetitive. Cell lines: GM12878=B cells; HepG2=hepatocytes; NHEK=keratinocytes; NHLF=fibroblasts.

DISCUSSION

MS is a complex disorder of unknown aetiology. Here I show that genomic regions associated with MS overlap with AP, SE, WE and ST regions in B cell lines and that this occurs more than would be expected by chance, and more than was observed in 3 other cell types unrelated to MS pathogenesis. Notably, the overlap was particularly striking in SE and WE regions for which significance was reached in 23 and 24 out of the 26 analyzed bins respectively. This is in accordance with the previous observations that tissue-specific genes appear more dependent on enhancer than promoter elements (208). Furthermore, I provide evidence that the associated SNPs preferentially land in promoter, enhancer and to a lesser extent transcribed regions. These findings have several important implications.

Firstly, this work further supports the immunological aetiology of MS (19). These findings are in agreement with those of a gene-ontology analysis of the genes located within MS associated regions, which showed a substantial overrepresentation of immune-related processes (31). As compared to this type of analysis, the chromatin approach has the relative advantage of being independent of what is currently known on genes and cell types.

Secondly, these observations provide further support for an important role for B cells in the pathogenesis of MS (210). The presence of oligoclonal bands in the cerebrospinal fluid represents the most consistent immunological finding in MS patients and B cell abnormalities influence both conversion to clinically definite MS, MRI activity, onset of relapses and disease progression (14, 24-29). Furthermore clinical trials have shown that antibody mediated depletion of CD20+ B cells is highly effective in diminishing MRI activity and onset of clinical relapses (205, 206).

However, it must be considered that certain chromatin features may be shared between B and other immune cell types, in particular T cells. Unfortunately, a similarly detailed chromatin profile of T cells is not yet available and therefore a direct comparison between B and T cells could not be performed.

Even if similar chromatin profiles exist between T and B cells, the attempts to understand the effects of genetic risk variants on T cell function (211) may be missing important effects in B cells. Given the increasing evidence for B-T cell interactions in MS (212-214), this analysis has the potential to greatly help the dissection of the roles played by these two cell types.

When MS SNPs rather than MS regions were analyzed, I found that MS SNPs were significantly more likely to land in AP, SE and WE regions than expected by chance perhaps suggesting that many of the associated SNPs may influence the risk of MS by modifying the binding of transcription factors and transcription in general. This is in agreement with previous observations (208). For MS associated SNPs landing in non-genic regions, for the first time I was able to show a likely functional role in gene regulation. However these findings should be interpreted with caution for two reasons. First, the observed overlap between MS SNPs and active chromatin states may be consequent to the fact that MS SNPs land in MS regions, themselves enriched for active chromatin states. Secondly, a conclusive answer to this question can only come from functional studies which should investigate if and how MS variants affect the chromatin landscape and gene expression.

To conclude, genomic regions associated to MS susceptibility are active in B cell lines and causative SNPs may act by changing the chromatin landscape. Further similar analyses in other immunological cell types relevant to MS and functional studies are required to fully understand in which cells, at which stage and how MS genetic variants are acting.

CHAPTER 5: VITAMIN D RECEPTOR BINDING, CHROMATIN STATES AND ASSOCIATION WITH MS

***-Disanto G**, Sandve GK, Berlanga-Taylor AJ, Ragnedda G, Morahan JM, Giovannoni G, Ebers GC, Ramagopalan SV. Vitamin d receptor binding, chromatin states and association with multiple sclerosis. Hum mol genetics (2012) 21: 3575-3586*

INTRODUCTION

I have shown that MS associated genomic regions overlap with strong enhancer, active promoter and strongly transcribed regions in B cell lines, more than expected by chance and more so than in non-immune cell types (215). Like all complex traits, genes are not the only determinant of MS susceptibility. Several lines of evidence support a role for vitamin D deficiency in influencing the risk of MS (see above). As mentioned in the introduction of this thesis, vitamin D acts in the genome via its cognate nuclear receptor, the vitamin D receptor (VDR). The presence of 2,776 different genomic locations bound by the VDR in B cell lines has been recently shown. Notably, genetic loci associated with MS were substantially enriched for VDR binding sites (94).

However, the VDR-MS enrichment analysis was performed using old data on MS associated genomic regions and needs to be updated in light of the latest GWAS findings (31). Furthermore, the relation between VDR binding and chromatin states has never been explored. In the present study the aim was to investigate how genomic regions with VDR binding in B cell lines (VDR regions) (94), chromatin states in B cell lines (208) and MS associated genomic regions (MS regions) (31) are related to each other.

METHODS

Data acquisition

Genetic variants and chromatin profiles were obtained as described in chapter 4. Since VDR binding could regulate gene expression at several kb of distance I decided to define MS regions as genomic

intervals of 0.25 cM centred on the lead associated SNP ± 100 kb each side. Genomic regions with VDR binding in B cell lines after stimulation for 36 h with 0.1 mM calcitriol were identified as previously described (94).

Assessing significance and enrichment of base pair overlap

To assess statistical significance of overlap and enrichment I used the same strategies as described in chapter 4. For the following analyses I used 50,000 Monte Carlo samples.

Taking into account genes as a confounding track

To control for a possibly confounding effect of gene proximity to the VDR-MS relation, I assessed VDR-MS overlap conditioning on their common relation to genes, using the approach to handle confounder tracks previously described (209). Briefly, I estimated the occurrence frequency of observed VDR binding in exponentially increasing ranges of distance to nearest gene, and under the null model sampled VDR binding with intensity given by the estimated occurrence frequency at a given distance (range) from the nearest gene. MS regions were kept at their observed locations.

RESULTS

Global overlap between VDR regions and chromatin states in B cell lines

The first aim was to assess in which chromatin states in B cell lines VDR binding was more likely to occur. Chromatin states analyzed were active promoter (AP), weak promoter (WP), poised promoter (PP), strong enhancer (SE), weak enhancer (WE), polycomb repressed (PR), heterochromatic (H), insulator (I), strongly transcribed (ST), weakly transcribed (WT) and repetitive/CNV (Rep/CNV) (208). The position of VDR regions was fixed while chromatin intervals were randomized as described in the methods. On the global scale, the overlap between VDR regions and chromatin states was higher

than expected by chance in 5 out of the 11 chromatin states analyzed (SE, WE, AP, WP and I). However, the overlap appeared particularly strong in chromatin states of high regulatory activity (SE and AP). Approximately 44% and 33% of VDR regions were in fact covered by SE and AP respectively. The enrichment values of VDR regions within these chromatin states were 45.26 for SE (p<2.0e-05) and 63.41 for AP (p<2.0e-05). Furthermore, SE and AP were the only chromatin states in which overlap was significantly higher than expected in all chromosome arms (41 out of 41) (table 12).

Overlap between VDR regions and chromatin states of B cell lines vs control cell lines

On its own, the presence of significant overlap between VDR regions and certain chromatin states in B cell lines does not necessarily imply biological significance. Both VDR regions and chromatin intervals could just be more likely to be near commonly transcribed genes and a similar degree of overlap could be present between VDR regions and chromatin states of other cell types. To exclude this hypothesis I tested whether the previously identified B cell chromatin states with significant overlap (SE, WE, AP, WP and I) co-localized with VDR regions more than the same chromatin states from 3 additional cell types (hepatocytes, fibroblasts and keratinocytes). I found that B cell specific SE and AP elements overlapped with VDR regions more than SE and AP elements from control cell types (p<2.0e-05 for all comparisons, table 13).

Overlap at the cytoband level: shared preference of bins vs detailed base pair overlap

By simply looking at the proportion of the genome which is covered individually by two sets of genomic intervals (tracks) one can calculate the expected base pair overlap (e.g. if two tracks cover 10% of the genome each, the expected overlap would be 1%). However this assumption is valid only if the two tracks are independent from each other, which is not always the case. For example, tracks could be preferentially located within the same regions of the genome and this could considerably

increase their expected base pair overlap. This appears to be the case for both VDR-SE and VDR-AP relationships as suggested by the strong correlations observed between the proportional coverage of SE or AP versus VDR regions within each cytoband (figure 13).

In order to understand whether the strong overlap observed between VDR regions and SE and AP is not only due to shared preference for the same genomic regions but also due to a tendency towards detailed base pair overlap I calculated and compared the following values: 1) The expected VDR-SE and VDR-AP base pair overlaps if all these elements were independent (independency of tracks); 2) The proportional coverage of VDR, SE and AP within each cytoband; 3) The expected VDR-SE and VDR-AP base pair overlaps after fixing the proportional coverage of VDR, SE and AP elements per cytoband (so that each of them is required to occur with the same base pair coverage within each cytoband as observed); 4) The actual VDR-SE and VDR-AP observed base pair overlaps. All values are shown in table 14.

As compared to independency of tracks, fixing cytoband preferences increased the expected overlap by approximately two fold (from 0.0011% to 0.0023% for VDR-SE and from 0.0005% to 0.0011% for VDR-AP). This further indicates the strong bias of VDR, SE and AP elements towards occurring within the same regions of the genome. However, the actual observed VDR-SE and VDR-AP overlaps were a further 13 and 21 times greater than the expected overlap after fixing cytoband preferences (p<0.001, for both VDR-SE and VDR-AP). This shows that both the VDR-SE and VDR-AP overlaps are not only due to shared preference for the same cytobands but also due to a more detailed overlap at the base pair level.

Global overlap between VDR regions and MS regions

I next investigated whether VDR regions were more likely than chance to occur within MS associated regions. This has been previously observed this using three year old data on MS associated regions

(94). In the updated analysis, out of 58 MS associated regions, 35 (60.3%) had VDR binding inside (table 15). The positions of MS regions were kept fixed, while VDR intervals were randomized as described in the methods. The observed base pair overlap was considerably higher than expected by chance (p<2.0e-05), with a fold enrichment of 3.67.

However both VDR binding sites and MS regions could have a tendency to be located near genes. This shared relation could itself lead to a higher overlap between VDR and MS regions than expected by independency of these tracks. As expected, when position relative to genes was taken into account in the null model the expected number of VDR regions falling inside MS regions increased from 30 to 45. However the observed number of VDR regions falling inside MS regions was 2.2 fold higher than expected (99 vs 45) (p<0.001).

VDR-AP and VDR-SE overlaps within MS regions

Given the significant co-occurrence of VDR-SE and VDR-AP at the genome wide level and the observed overlap between VDR and MS regions, I investigated the extent of the overlap between VDR and SE and between VDR and AP within MS regions. Similar to how I analyzed the genome wide VDR-SE and VDR-AP overlaps, I calculated and compared the following values: 1) The expected VDR-SE and VDR-AP base pair overlaps within MS regions if all these elements were independent; 2) The proportional coverage of VDR, SE and AP within each MS region; 3) The expected VDR-SE and VDR-AP base pair overlaps after fixing the proportional coverage of VDR, SE and AP elements per MS region; 4) The actual VDR-SE and VDR-AP observed base pair overlaps within MS regions. All values are shown in table 16.

As compared to independency of tracks, fixing proportional coverage of MS regions slightly increased the expected overlap (from 0.0133% to 0.0204% for VDR-SE and from 0.0055% to 0.0079% for VDR-AP). The actual observed VDR-SE and VDR-AP overlaps were a further 4.75 and 14.93 times higher

than the expected overlap after fixing for MS regions (p<0.001, for both VDR-SE and VDR-AP). Overlap between VDR and either SE or AP elements was present in several MS associated regions and near a number of putative candidate genes including *SP140, CD86, IL22RA2, CXCR5, TNFRSF1A, ZFP36L1, BATF, IRF8, CD40, TNFRSF6B* in addition to a region on chromosome 6 with no candidate genes (figure 14 and 15).

Table 12: Enrichment of VDR regions in B cell lines within chromatin states in B cell lines (SE=strong enhancer; WE=weak enhancer; AP=active promoter; WP=weak promoter; PP=poised promoter; PR=polycomb repressed; H=heterochromatic; I=insulator; ST=strong transcribed; WT=weak transcribed)

Variable	SE	WE	AP	WP	PP	R	H	I	ST	WT	CNV
Global significance	*p*<2.0e-05	*p*<2.0e-05	*p*< 2.0e-05	*p*<2.0e-05	*p*=0.65	*p*=1	*p*=1	*p*=0.047	*p*=0.85	*p*=1	*p*=0.3
Proportion of VDR regions covered	0.438	0.068	0.329	0.039	0.002	0.004	0.022	0.008	0.065	0.022	0.003
Enrichment	45.26	2.33	63.41	5.88	0.963	0.125	0.009	1.47	1.06	0.19	1.22
Significant chromosome arms	41/41	20/41	41/41	33/41	0/41	0/41	0/41	0/41	0/41	0/41	0/41

Table 13: Comparison of overlap between VDR regions in B cell lines and chromatin states in B cell lines vs control cell types

Chromatin states	Is overlap in B cell > Hepatocytes?			Is overlap in B cell > Fibroblasts?			Is overlap in B cell > Keratinocytes?		
	Global p-value	Significant bins	Fold difference	Global p-value	Significant bins	Fold difference	Global p-value	Significant bins	Fold difference
SE	*p*<2.0e-05	41/41	8.069	*p*<2.0e-05	41/41	8.34	*p*<2.0e-05	41/41	7.84
WE	*p*=0.998	0/41	0.729	*p*=1	0/41	0.702	*p*=0.989	0/41	0.741
AP	*p*<2.0e-05	19/41	4.955	*p*<2.0e-05	15/41	4.047	*p*<2.0e-05	15/41	5.047
WP	*p*=1	0/41	0.688	*p*=0.961	0/41	0.768	*p*=0.665	0/41	0.768
I	*p*=0.989	0/41	0.499	*p*=0.996	0/41	0.423	*p*=0.987	0/41	0.542

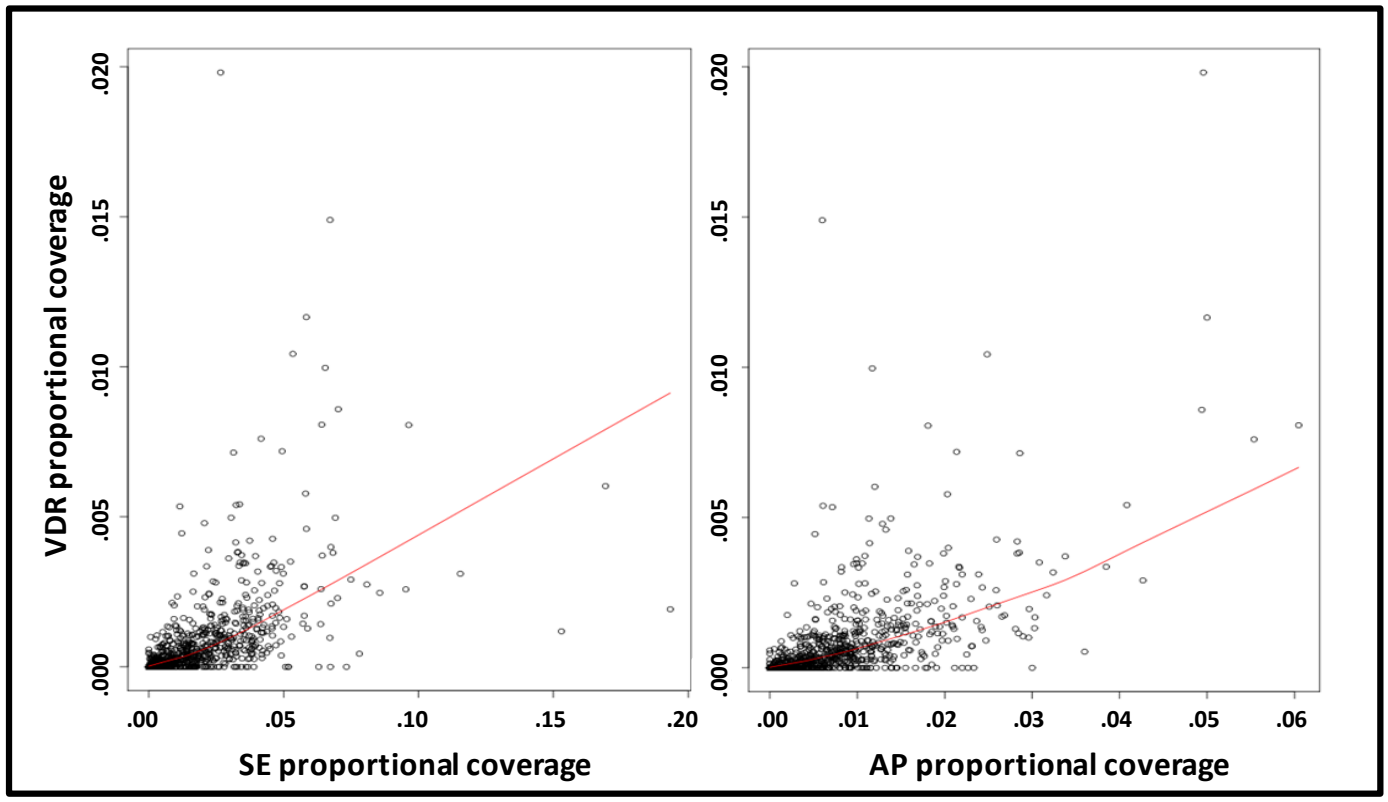

Figure 13: Correlation between VDR and SE/AP proportional coverage per cytoband (SE-VDR: Pearson correlation coefficient=0.53, p<1e-10; AP-VDR: Pearson correlation coefficient=0.61, p<1e-10).

Table 14: Overlap between VDR and SE and between VDR and AP regions at the cytoband level

	SE **VDR regions**	**AP** **VDR regions**
Expected overlap given coverage of the whole genome	0.0011 %	0.0005 %
Expected overlap after fixing individual coverage of cytobands	0.0023 %	0.0011 %
Observed overlap	0.0307 %	0.0231 %
Observed / expected given coverage of whole genome	27.9	46.2
Observed / expected after fixing coverage of cytobands	13.35	21

Table 15: MS associated regions, candidate genes and presence of VDR binding inside

MS regions	Candidate gene	VDR binding?	MS regions	Candidate gene	VDR binding?
chr1:2296586-2927369	*MMEL1*	***YES***	chr8:128722713-129112679	*MYC*	
chr1:91914277-93406499	*EVI5*	***YES***	chr8:129088889-129466891	*PVT1*	***YES***
chr1:100883315-101555310	*VCAM1*	***YES***	chr10:5970200-6335251	*IL2RA*	
chr1:116732232-117009542	*CD58*		chr10:80530397-80912694	*ZMIZ1*	
chr1:190633439-190914781	*RGS1*	***YES***	chr10:94086576-94591896	*HHEX*	
chr1:199028354-199436605	*C1orf106*		chr11:60381885-60707448	*CD6*	
chr2:43065922-43316721	*No gene*		chr11:117719848-118503512	*CXCR5*	***YES***
chr2:68287193-68634474	*PLEK*		chr12:6185549-6440964	*TNFRSF1A*	***YES***
chr2:112261731-113061046	*MERTK*		chr12:9220354-10011731	*CLECL1*	***YES***
chr2:136516329-136792725	*CXCR4*	***YES***	chr12:55815445-56993768	*CYP27B1*	***YES***
chr2:230686908-231043610	*SP140*	***YES***	chr12:121854616-122622299	*ARL6IP4*	***YES***
chr3:27596539-27915543	*EOMES*		chr14:68106919-68485659	*ZFP36L1*	***YES***
chr3:27917551-28262135	*No gene*		chr14:74914385-75200333	*BATF*	***YES***
chr3:106755817-107254331	*CBLB*		chr14:87181291-87857780	*GALC*	
chr3:120484165-120881518	*TMEM39A*	***YES***	chr16:885320-1148646	*SOX8*	
chr3:122997536-123554717	*CD86*	***YES***	chr16:10825550-11369956	*CLEC16A*	***YES***
chr3:160964425-161340687	*IL12A*		chr16:84420930-84680066	*IRF8*	***YES***
chr4:103508587-104694517	*NFKB1*	***YES***	chr17:37471803-38274639	*STAT3*	***YES***
chr5:35574799-36288926	*IL7R*		chr17:54683128-55516497	*RPS6KB1*	***YES***
chr5:40169348-40818595	*PTGER4*	***YES***	chr18:54376982-54781323	*MALT1*	
chr5:158333068-158858750	*IL12B*		chr19:6463807-6736403	*TNFSF14*	***YES***
chr6:90766998-91248913	*BACH2*	***YES***	chr19:10173627-10589218	*TYK2*	***YES***
chr6:127760104-128484638	*THEMIS*		chr19:17904486-18371361	*MPV17L2*	***YES***
chr6:135530175-136350729	*MYB*	***YES***	chr19:54306574-54720371	*DKKL1*	***YES***
chr6:137248718-137655201	*IL22RA2*	***YES***	chr20:43865041-44348988	*CD40*	***YES***
chr6:137815794-138295369	*No gene*	***YES***	chr20:52101726-52353768	*CYP24A1*	
chr6:159142319-159561276	*TAGAP*		chr20:61562867-62060032	*TNFRSF6B*	***YES***
chr7:148654372-149158702	*ZNF746*	***YES***	chr22:20129932-20787848	*MAPK1*	***YES***
chr8:78728530-79923754	*IL7*		chr22:49179169-49523510	*SCO2*	***YES***

Table 16: Overlap between VDR and SE and between VDR and AP regions within MS regions

	SE **VDR regions**	**AP** **VDR regions**
Expected overlap given coverage of all MS regions	0.0133 %	0.0055 %
Expected overlap after fixing individual coverage of MS regions	0.0204 %	0.0079 %
Observed overlap	0.0970 %	0.1180 %
Observed / expected given coverage of all MS regions	7.29	21.45
Observed / expected after fixing coverage of MS regions	4.75	14.93

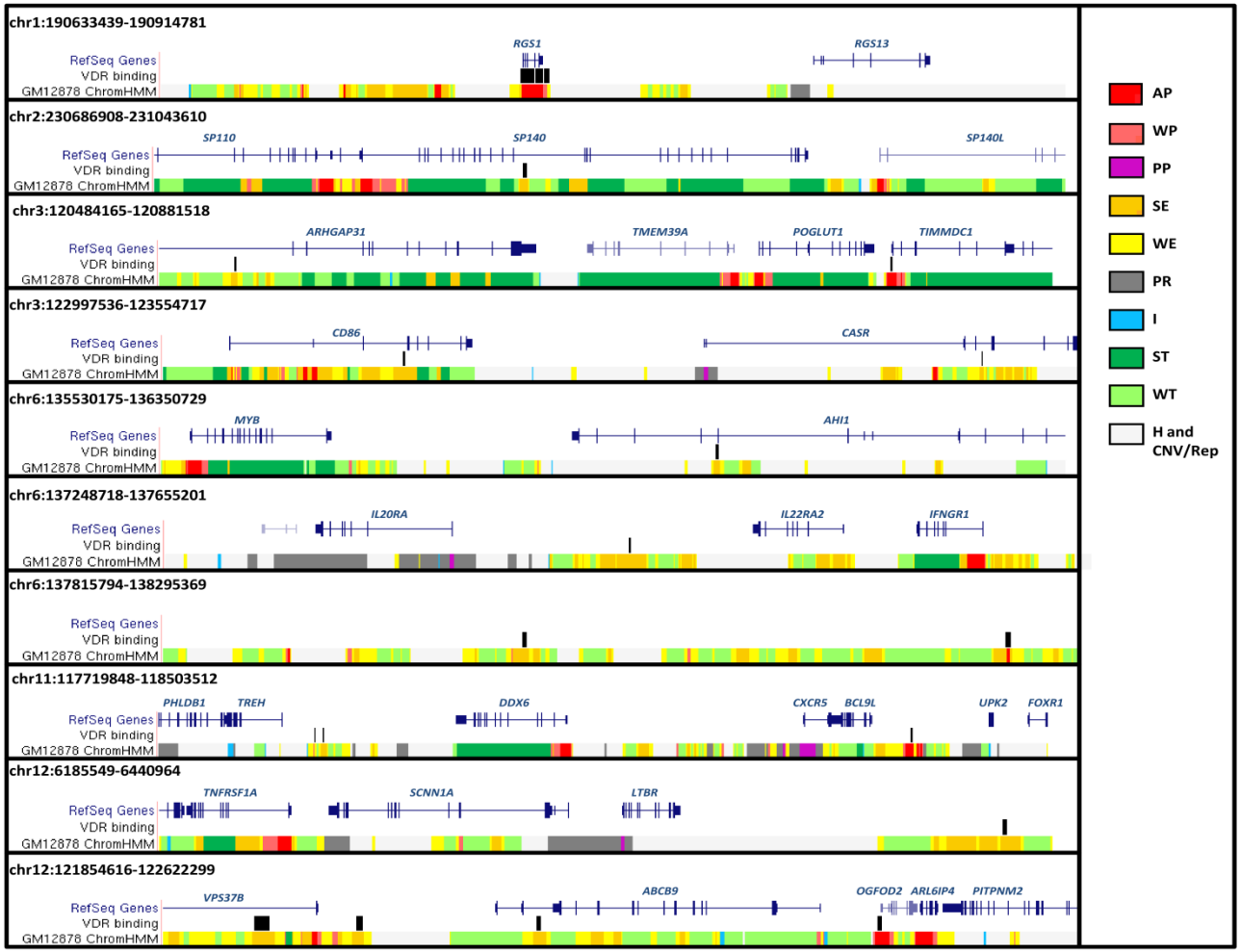

Figure 14: Overlap between VDR regions and SE/AP elements within MS associated genomic regions (I)

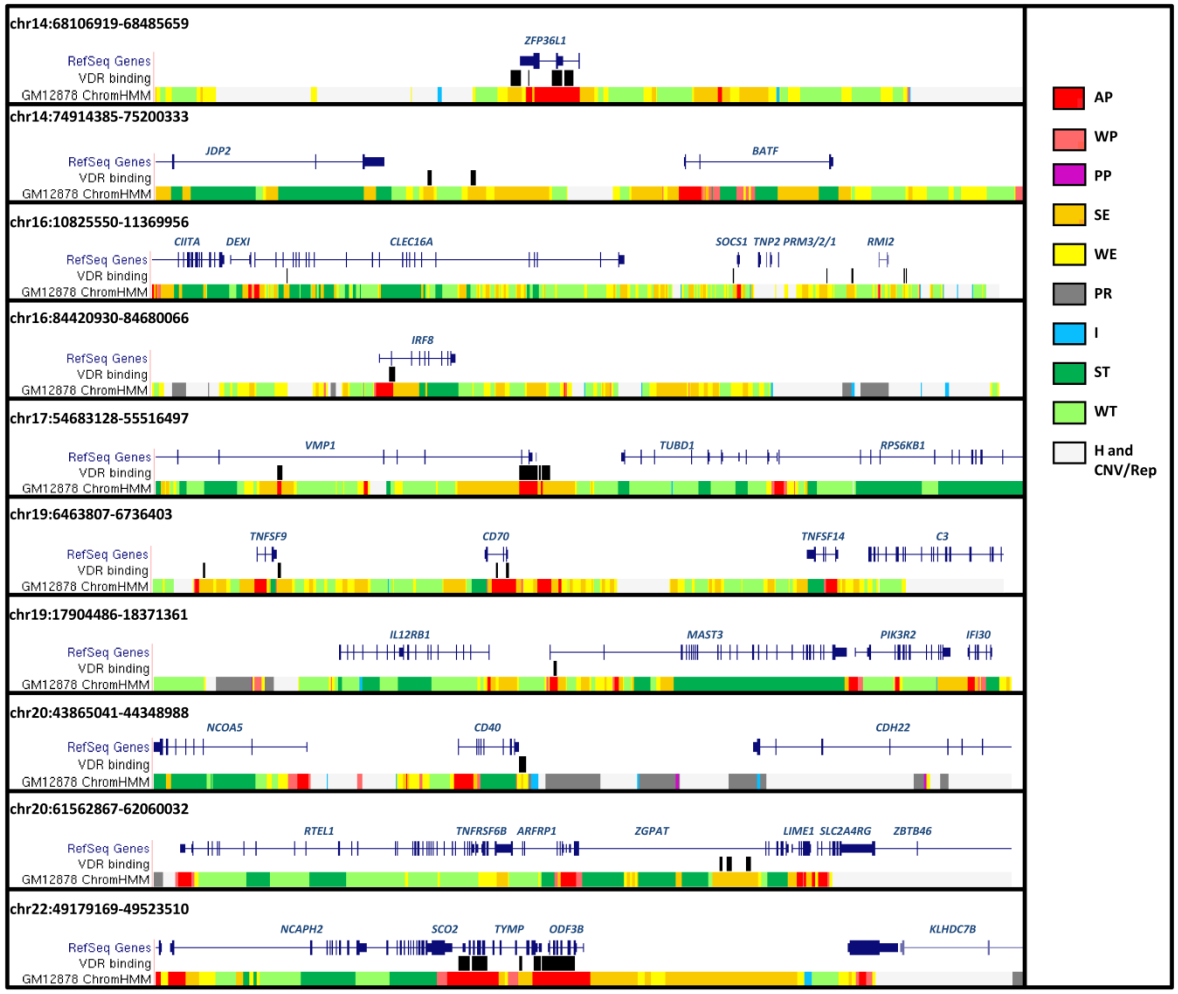

Figure 15: Overlap between VDR regions and AP/SE elements within MS associated genomic regions (II)

DISCUSSION

Assessing whether two or three sets of genomic intervals overlap may appear like a simple question to answer. However, performing the analysis without bias is more complicated. For example one needs to distinguish between whether the overlap is due to shared preference for the same genomic regions or due to detailed overlap at the base pair level. Furthermore, two different tracks could share a similar tendency to be located in proximity to a third set of genomic intervals (for example genes), and when this happens the effect of these confounding tracks needs to be taken into account. In chapter 6 I showed how MS regions overlap with SE and AP elements of B cell lines more than expected by chance and more than in non-immune cell types (215). I followed up these findings and investigated how VDR binding is involved in the relation between MS regions and chromatin states. I found that:

1) Regions with VDR binding in B cell lines overlap with AP and SE intervals in B cell lines more than expected by chance and more than in non-immune cell types. The extent of this overlap is astonishing with more than 40% and 30% of VDR regions covered by SE and AP respectively.

2) The AP-VDR and SE-VDR overlap is present both at the global (genome wide) and local (cytoband) scales. Notably, I was able to show that these elements are not only preferentially located within the same genomic regions but also tend to overlap at base pair level. This suggests that the VDR directly interacts with enhancer and promoter elements and likely modifies their action.

3) VDR binding is also more likely to occur within MS regions as compared to the rest of the genome and more than 60% of MS regions are bound by the VDR. These include the genomic regions containing the candidate genes *EVI5, VCAM1, CXCR4, SP140, TMEM39A, CD86, NFKB1, CXCR5, TNFRSF1A, CLECL1, CYP27B1, CLEC16A, IRF8, STAT3, TNFSF14, TYK2, CD40* and *TNFRSF6B*. This overlap is also significantly higher than expected when genes are considered as a confounding track.

4) The SE-VDR and AP-VDR overlaps are also present with the same characteristics within MS regions suggesting that these relations are relevant for disease aetiology. Candidate genes with either SE-VDR or AP-VDR overlap nearby included *CD86, CXCR5, TNFRSF1A, RGS1, SP140, TMEM39A, IL22RA2, BATF, CLEC16A, TNFRSF6B, IRF8* and *CD40*. Interestingly, one of the MS associated regions located on chromosome 6, which was void of candidate genes, was characterized by two sites of overlap of SE-VDR and AP-VDR, perhaps suggesting a long-range or *trans* regulatory role.

Taken together these findings have a number of important implications. First, they suggest that VDR binding influences or interacts with the chomatin profile of a cell. The regulatory role played by vitamin D at the genomic level likely underlies its modulatory actions in the immune system. Both the VDR and the CYP27B1 enzyme are expressed in an exceptionally wide range of immune cells, including B cells, but also Th1 and Th17 subsets and FOXP3+ regulatory T cells (91). In B cells, vitamin D is able to modulate cell proliferation, induce apoptosis and inhibit plasma cell differentiation and immunoglobulin secretion (216).

Second, the fact that VDR-SE and VDR-AP overlaps are seen not only at the genome wide level, but also inside genomic regions associated with MS risk, provides a potential biological explanation for the epidemiological observations linking vitamin D deficiency to both risk and clinical course of MS (80, 87, 217-219).

Finally, these data provide insights into the cell types that are primarily driving the onset of MS. It would be logical to hypothesize that genetic and environmental risk factors would have a direct influence on the cell type triggering the abnormal immune response in MS patients. Therefore, the observation that both genomic regions associated with MS and VDR binding intervals overlap with genomic regions which are active and play a regulatory role in B cell lines provide further support for an important role for this cell type in the pathogenesis of this devastating disease (see above).

One possible confounding factor is that both chromatin and VDR binding data come from B cells immortalized via Epstein Barr virus (EBV). While it is acknowledged that the EBV mediated immortalization process could influence both chromatin profile and VDR binding, this could itself be interesting for MS since EBV infection is another environmental risk factor which is strongly implicated in the aetiology of this complex disease (97, 98, 101). It would therefore be particularly interesting to investigate how EBV infection can change the chromatin features of infected B cells, and whether this modifies their relation with MS associated regions and VDR binding.

To conclude, genomic regions with VDR binding in B cell lines substantially overlap with promoter and enhancer elements that are active in B cell lines. This is also true in MS associated regions and is therefore relevant for MS aetiology. Further similar analyses in other immunological cell types and functional studies will be required to fully understand how vitamin D can influence the chromatin landscape and gene expression outside and inside genomic regions which play a role in the aetiology of MS.

CHAPTER 6: DNASE HYPERSENSITIVITY SITES AND ASSOCIATION WITH MS

INTRODUCTION

I have shown in chapter 4 how MS associated genetic regions overlap with sites of active and regulatory chromatin in immortalized B cell lines more than in non-immune cell types. However, it is plausible that the chromatin state is influenced by the immortalization process. Furthermore, it would be important to investigate the relation between MS associations and active chromatin in different components of the immune system. Deoxyribonuclease I (DNase I) hypersensitive sites (DHSs) represent precise and reliable indicators of open and active chromatin across the genome. Recent studies have shown that these marks are highly cell specific and can be used to identify pathogenic cell types in complex traits (208, 220). ENCODE researchers have recently investigated the overlap between MS associated genetic variants and DHSs across a number of cell types (220). They used data from a recently published meta-analysis of genome wide association studies (GWAS) in MS (73) and compared the proportion of SNPs meeting increasingly significant association *p*-values thresholds that were located inside DHSs of each cell type to the proportion of all SNPs inside DHSs of the same cell type. The authors concluded that MS associated variants were particularly active in CD3+ cord blood T cells and CD19+/CD20+ B cells (220). However, this strategy is biased by the considerable variability of LD across the genome. Indeed, associated SNPs may be in strong LD with either a very large number (even hundreds) or a small number (only a few) SNPs. The inevitable consequence of this disparity is that some associated SNPs will count more than others in the overall estimate. For example, the association between cord blood T and B cell DHSs and MS SNPs was particularly evident for SNPs with *p*-values lower than 10^{-10}. In MS, such small *p*-values are only reached by a very large number of major histocompatibility complex (MHC) SNPs and an equally large number of SNPs across this region reach *p* values suggestive of association (e.g. between 10^{-4} and 10^{-}

[8)]) (73). Furthermore, many of these SNPs are near genes not involved in disease but are only associated through LD with the MS causative *HLA-DRB1* gene (67, 221).

Although most genetic susceptibility to MS resides within the MHC, more than 50 genetic loci located outside the MHC have now been associated with MS risk and many others show suggestive association (31, 73). I therefore aimed to provide a more reasonable estimate of the overlap between MS associated genetic variants and cell specific DHSs. In order to avoid inflating our analysis with MHC variants, I decided to consider MS associated regions (MS regions) rather than SNPs. In addition, I investigated the relationship between cell specific genetic variants and heritability and assessed which cell type/s explain most of the genetic susceptibility to MS. I finally looked at additional diseases and assessed to what extent different immune mediated traits share common pathogenic cell types.

METHODS

Data collection

I collected genetic association data from the "Catalog of Published Genome-Wide Association Studies" (http://www.genome.gov/gwastudies/) for the following diseases: multiple sclerosis (MS), type 1 diabetes (T1D), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), systemic sclerosis (SS), psoriasis, Crohn's disease (CD), ulcerative colitis (UC), celiac disease, leprosy, schizophrenia and colorectal cancer (CC). Only associations reported from GWAS with the largest sample size were included in the analyses (supplementary table 4). I decided to further analyse MS associated SNPs by dividing them in those with confirmed and suggestive genetic association using data from the two largest and most recent GWAS performed to date (31, 73). In Sawcer et al., confirmed associated SNPs were defined as those that either were a replication of previous GWAS findings or had a discovery stage $p < 1x10^{-4.5}$, a replication $p < 0.05$ and a p-combined $< 5x10^{-7}$. SNPs

suggestive of association were those having a discovery stage $p < 1x10^{-4.5}$ and in which the same allele was overrepresented in cases compared to controls in both discovery and replication stages (31). In Patsopoulos et al., confirmed associated SNPs were defined as either those with *p*-values $< 5x10^{-8}$ or previously identified associated SNPs that replicated. Suggestive SNPs were those with *p* values ranging between 5×10^{-8} and 5×10^{-6} (73). Coordinates of cell specific DHSs for 112 different cell types were provided by the ENCODE project (supplementary table 5) (220).

Cell specific DHSs – MS regions overlap analysis

All analyses were performed using the Genomic Hyperbrowser (209). For each SNP we defined an association region as the genomic interval comprising 50 kb each side of the associated SNPs. In order to assess whether cell specific DHSs fell more in the proximity of SNPs associated to a given disease than the general tendency of DHSs across all cell types to fall in the same proximity, we applied a three-track approach. A general DHSs track was created by aggregating all DHS sites across all cell types/tissues. Enrichment for a particular cell type was calculated as the ratio of cell specific DHSs coverage to aggregated DHSs coverage within disease associated regions, divided by the corresponding ratio for the full genome. This is equivalent to testing whether cell specific DHSs overlap with disease associated regions more than the general tendency of DHSs across all cell types to do this. The aggregated DHSs track contains overlapping regions (sites from different cell types that have overlapping genome coordinates), and in such cases the coverage for a given base pair were calculated as the count of regions covering the base pair. Instead of counting uniformly within associate regions, a weight inversely proportional to the distance of the position from the associated SNP was assigned (figure 16). Total coverage within disease associated regions (as referred to above) was then defined as the sum of the weighted DHSs densities for all SNPs. Statistical hypothesis testing was performed using a permutation based approach to compute p-values. A null model was defined

for which the location of individual disease-associated SNPs were uniformly randomized within the chromosome arms they were located in, whereas DHSs were kept fixed. The distance-based enrichment, as described above, was used as a test statistic in the hypothesis testing.

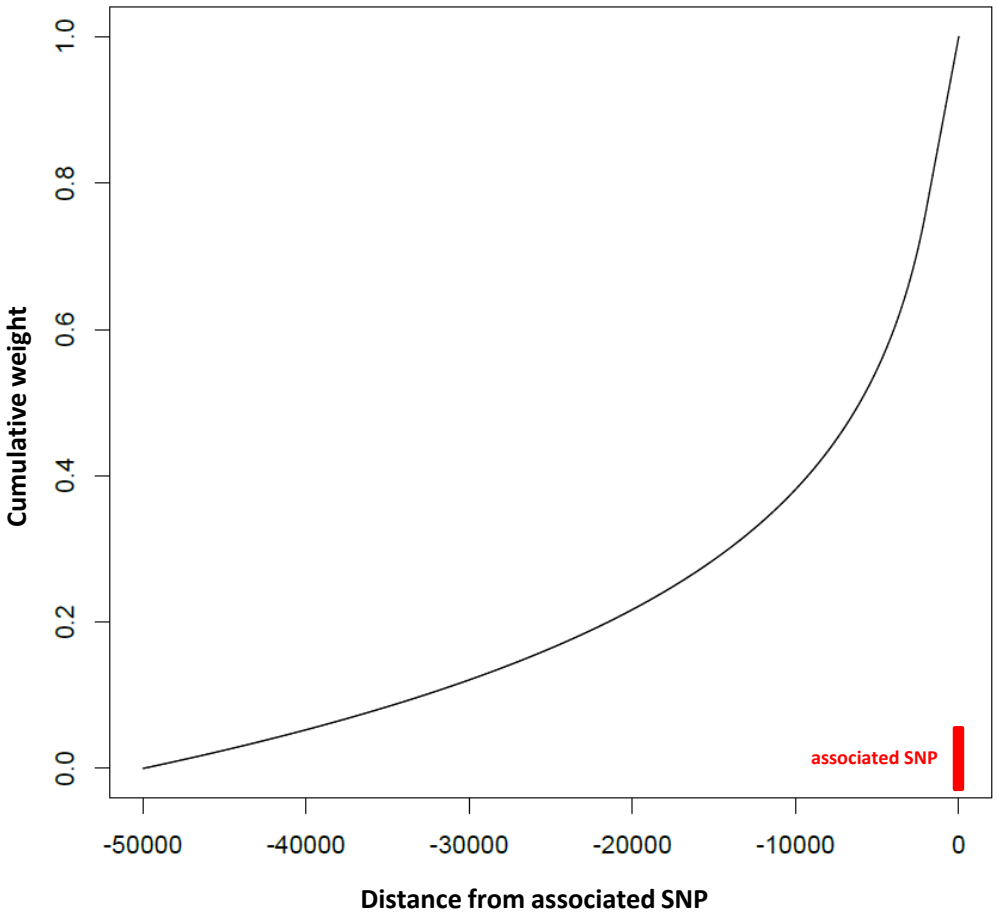

Figure 16: Plot showing how the total weight assigned to nucleotides is distributed as a function of the distance from the GWAS associated SNP. The intervals between 50kb and 30kb away from the associated SNPs make up a total of around 10% of the total weight assigned to positions when evaluating DHSs coverage. Correspondingly, the broader regions from 50kb to 10kb away from the SNPs (at both sides) make up a total of around 40% of the total weight, while the remaining 60% of the weight is then necessarily assigned to positions closer than 10kb from the SNP.

Estimation of phenotypic variance explained by DHSs specific SNPs

As a consequence of the very large number of tests performed in GWAS, the criteria used to define genetic associations need to be very stringent and a Bonferroni correction is strongly suggested. As a consequence, several truly associated variants fail to reach genome wide significance and are not recognized. This represents one of the reasons why associated SNPs usually explain a relatively small proportion of the phenotypic variance between cases and controls. However, this fraction considerably increases if all genotyped SNPs rather than only those that reach genome wide significance are considered. I used raw GWAS data from a total of 1,854 MS patients and 5,164 unrelated controls from the United Kingdom generated by the Wellcome Trust Case Control Consortium 2 (WTCCC2) (31). I estimated the proportion of phenotypic variance between MS cases and controls explained by all genotyped SNPs located within cell specific active chromatin using the Genome-wide Complex Trait Analysis (GCTA) tool (222, 223). I first estimated the genetic relationship between individuals from GWAS SNP data and this pair-wise information was stored in a genetic relationship matrix (GRM). The total phenotypic variance between cases and controls is a variable depending on the GRM, the polygenic additive genetic variance explained by the SNPs (σ_u) and error variance (σ_e). The narrow sense heritability of a trait (proportion of phenotypic variance explained by additive genetic effects or h^2) can be defined as $h^2 = \sigma_u / (\sigma_u + \sigma_e)$ and was calculated via residual maximum likelihood (REML) analysis. These variances are on the observed 0–1 scale. However, heritability is most interpretable on a continuous scale and the estimated variance was therefore transformed on the liability scale assuming a disease prevalence of 0.001 (222, 223).

RESULTS

Overlap between cell specific DHSs and MS associated genomic regions

My first aim was to assess in which cell types and tissues DHSs overlapped with genomic regions associated with MS more than expected from the general tendency of DHSs across all cell types. The overlap between MS regions and DHSs was higher than expected only for a number of immune cell types. The highest enrichment values were observed for various T cell types (in particular Th1, Th17 and CD8+ cytotoxic T cells). CD19+ B cells, CD56+ natural killer (NK) and fetal thymus also showed high enrichment values (figure 17 and supplementary table 6). I then calculated enrichment considering confirmed and suggestive SNPs as defined above. There was a clear overrepresentation of immune cell types in both groups and there was a considerable correlation between confirmed and suggestive MS regions (Pearson's rho=0.86; p<2.2e-16) (figure 18) (supplementary table 7).

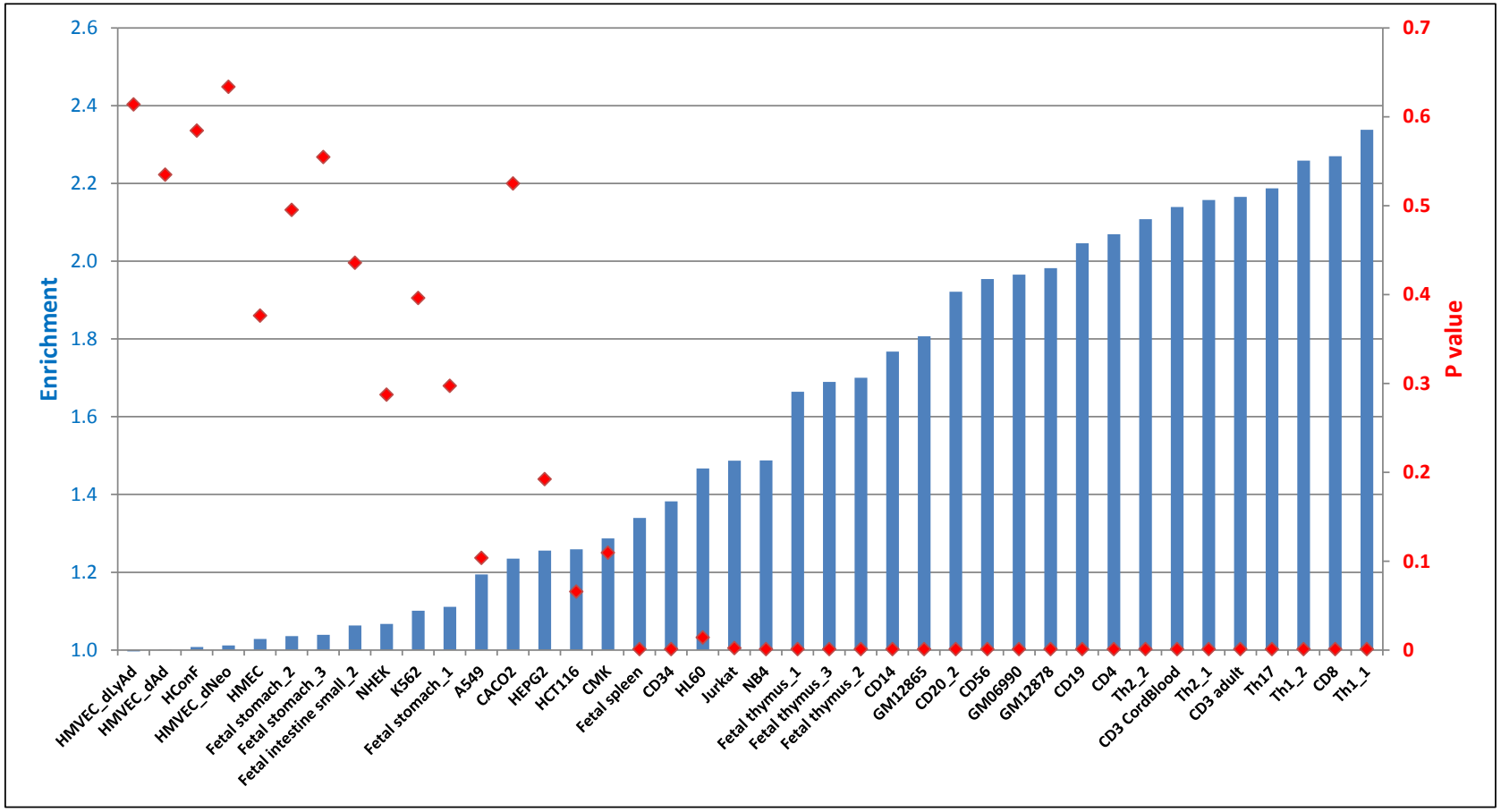

Figure 17: Enrichment and significance of overlap between MS associated genomic regions and cell specific DHSs. Only the top 40 cell types are shown in the figure. For cell types with more than 1 replicate available, the number after the underscore indicates the number of replicate (e.g. Th1_2= Th1 second replicate)

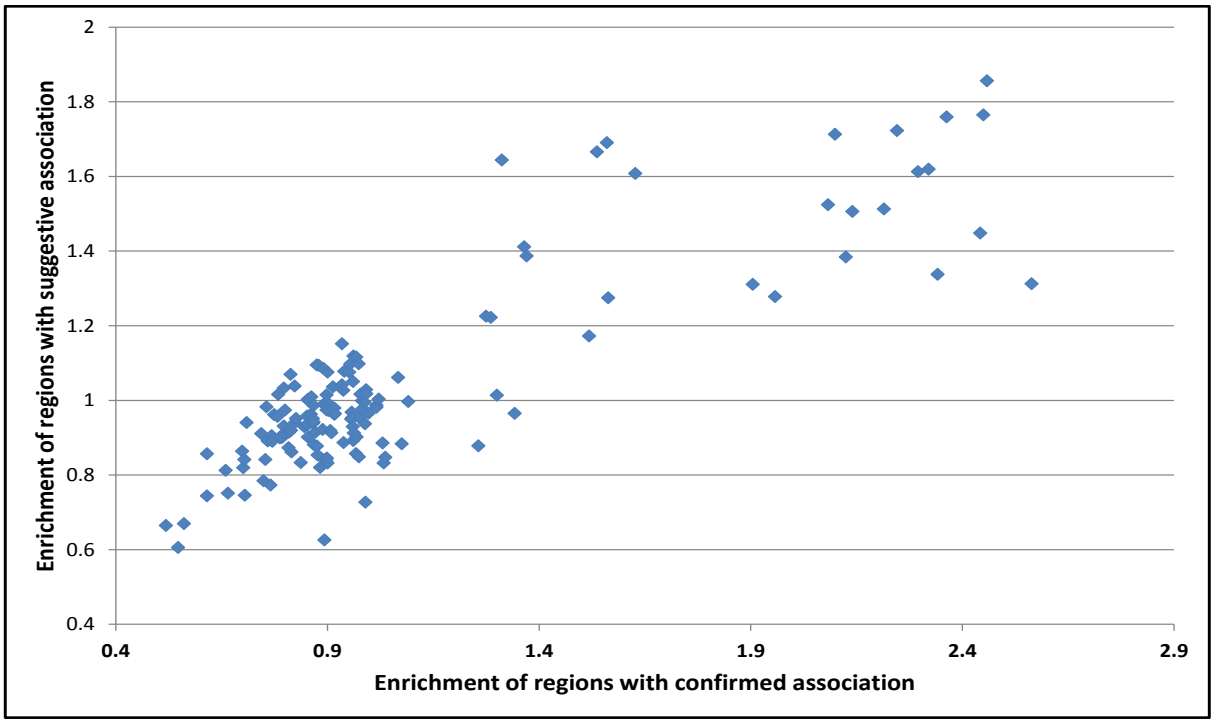

Figure 18: Correlation between enrichment of regions with confirmed vs suggestive association with MS

I further confirmed these results and the immunological nature of suggestive associations by using the software GRAIL (Gene Relationships Across Implicated Loci) (224). This tool is able to examine relationships between genes in disease associated loci based on the Novartis Gene Expression Atlas, a database containing expression measurements for more than 30 thousand probes across multiple tissues and cell types (224, 225). I observed that both confirmed and suggestive MS candidate genes were expressed in the several immune cell types that were also suggested by the DHSs analysis (CD4+ and CD8+ T, B and NK cells). In addition, I found a large number of connections between genes with confirmed and suggestive association with MS and 8 genes located within suggestive regions were connected within the network more than expected by chance only (*CARD11, FCRL2, CHST12, SYK, TCF7, SOCS1, NFKBIZ* and *NPAS1*) (figure 19).

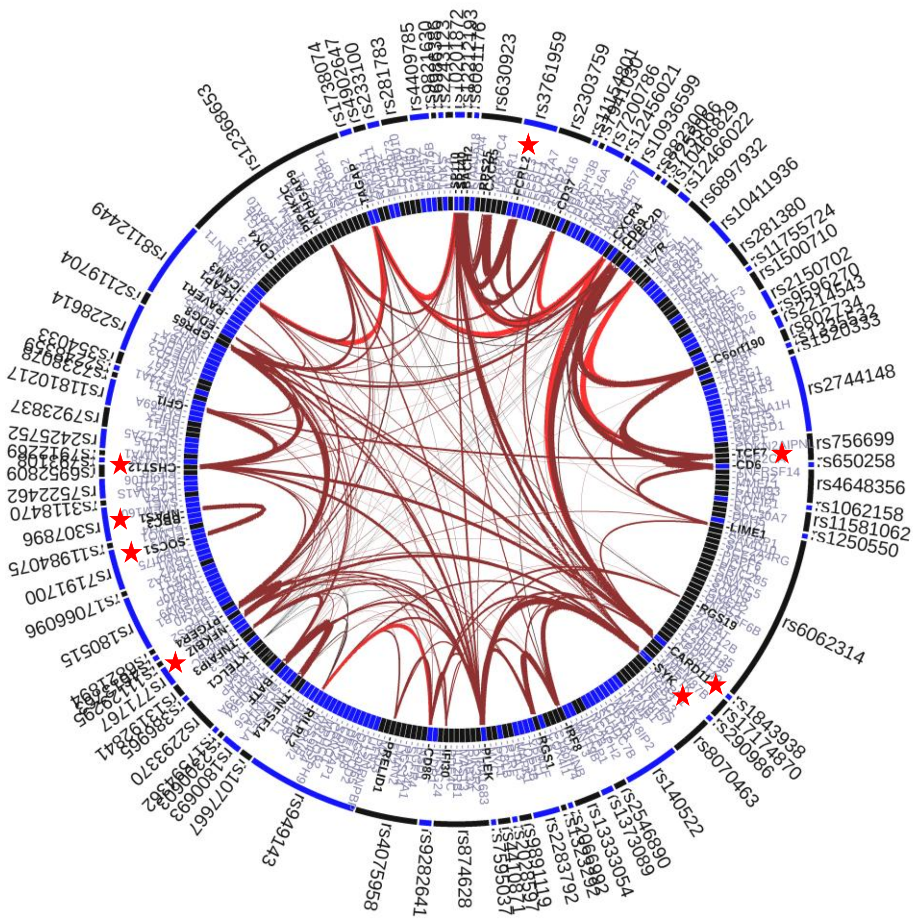

Figure 19: Gene-gene connections found by GRAIL based on the Novartis Gene Expression Atlas. Those genes that participate in the network are indicated in bold while red stars indicate genes located within regions of suggestive association that are connected within the network more than expected by chance.

Estimation of phenotypic variance explained by DHSs specific SNPs

I next investigated cell type specific genetic variants and MS heritability. All SNPs genotyped by the WTCCC2 were grouped based on their location relative to cell specific DHSs. Several studies have suggested that T and B cells play an important role in the etiology of MS. I therefore decided to focus on CD4+, Th1, Th17, CD8+ T and CD19+/CD20+ B cell DHSs (immune group). The overall proportion of variance on the liability scale that was explained by all autosomic SNPs in the immune group (n=35,336) was 18.5% (standard error (SE) = 1.2%). I then pooled DHSs from CD4+ T, Th1 and Th17

into a single T helper group and DHSs from CD19+ and CD20+ cells into a single B cell group. The proportion of variance explained by SNPs in B DHSs, CD8+ T DHSs and CD4+ helper T DHSs was 16.5% (SE=1.1%), 9.8% (SE=0.8%) and 13.9% (SE=0.9%) respectively. I reasoned that many of these SNPs may well be shared across more than one cell type. I therefore grouped SNPs into those that were located within: 1) only B DHSs; 2) only T helper DHSs; 3) only CD8+ T DHSs; 4) T helper and B but not CD8+ T DHSs; 5) CD8+ T and B but not T helper DHSs; 6) T helper and CD8+ T but not B DHSs; 7) both T helper, CD8+ T and B DHSs. The proportion of phenotypic variance explained by SNPs located in only B DHSs (4.63%, SE=1.10%) was higher than that explained by SNPs in only CD8+ T DHSs (0.02%, SE=0.22%) and only T helper DHSs (2.48%, SE=0.69%). However, the highest proportion of variance was explained by SNPs located in DHSs that were shared across all cell types (5.96%, SE=0.73%) (figure 20).

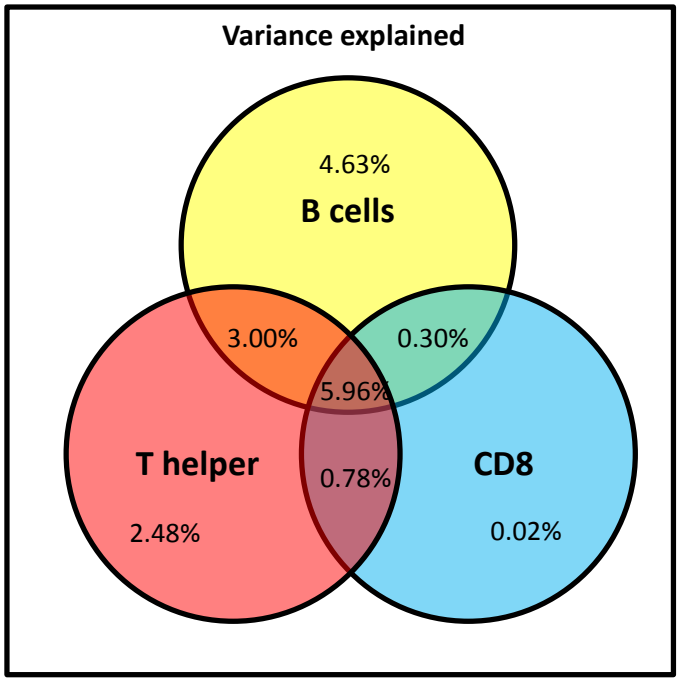

Figure 20: Proportion of phenotypic variance explained by SNPs within 1) only B DHSs; 2) only T helper DHSs; 3) only CD8+ T DHSs; 4) T helper and B but not CD8+ T DHSs; 5) CD8+ T and B but not T helper DHSs; 6) T helper and CD8+ T but not B DHSs; 7) both T helper, CD8+ T and B DHSs.

DHSs in other complex diseases

I finally investigated the DHSs profile of 11 additional complex traits. Among immune-mediated diseases, the top enriched cell types were CD8+ T cells in psoriasis, Th17 in T1D and RA, Th1 in celiac disease and CD, cord blood CD3+ T and CD19+ B cells in SS, B cells in SLE, CD14+ monocytes and colorectal carcinoma cells in UC. Among non-immune mediated diseases, the top cell types were hepatocellular carcinoma cells and fetal intestine in CC, CD14+ monocytes in leprosy and fetal brain in schizophrenia (figure 21, table 17 and supplementary table 6).

Table 17: Table 1: Top enriched cell types and tissues in all 17 conditions tested (GM12878=B lymphoblastoid cell line; HEPG2= hepatocellular carcinoma cells).

Disease	Top Cell type / tissue	Enrichment	P value
MS	Th1	2.338	0.001
RA	Th17	3.26	0.001
T1D	Th17	2.678	0.001
Celiac disease	Th1	2.774	0.001
CD	Th1	2.441	0.001
UC	CD14+ monocytes	2.153	0.001
SLE	GM12878	2.582	0.001
Psoriasis	CD8+ T	2.867	0.001
SS	CD3+ cord blood T	3.52	0.001
Leprosy	CD14+ monocytes	2.721	0.0015
CC	HEPG2	1.837	0.021
Schizophrenia	Fetal brain	1.486	0.0419

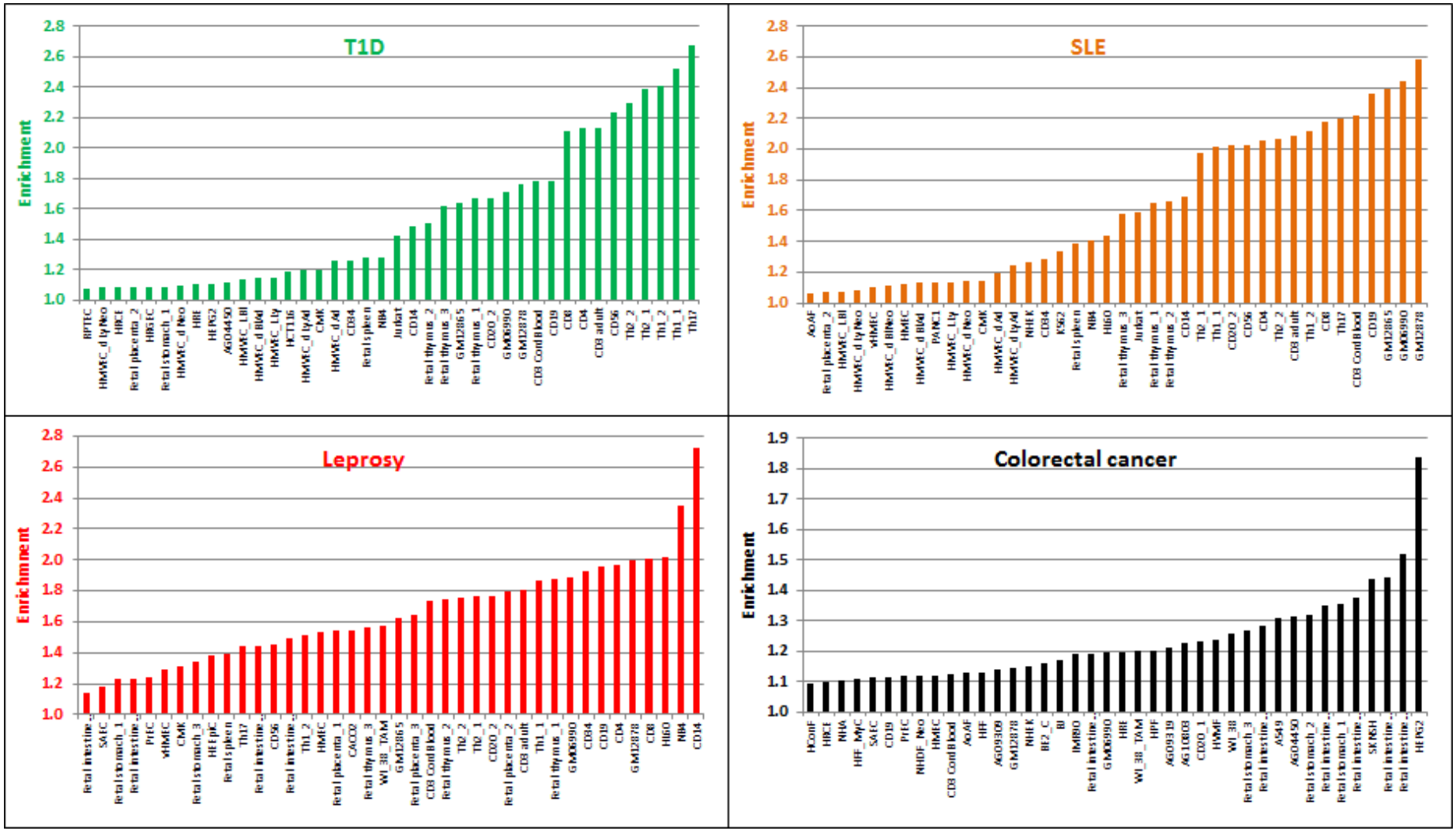

Figure 21: Enrichment of overlap between genomic regions associated with T1D, SLE, leprosy and colorectal cancer and cell specific DHSs. Only the top 40 cell types are shown in the figure. For cell types with more than 1 replicate available, the number after the underscore indicates the number of replicate (e.g. Th1_2= Th1 second replicate)

DISCUSSION

I observed that genomic regions with association with MS overlap with immune specific DHSs more than expected by chance only. These findings add further support to the hypothesis that the pathogenesis of MS is primarily immune-mediated (19, 31). By considering all genomic regions associated to MS (rather than only MHC variants), I was able to show that DHSs from a variety of T cell subtypes (in particular Th1, Th17 and CD8+ cytotoxic T cells) but also CD19+ B cells, CD56+ NK cells and fetal thymic tissue have a strong tendency to overlap with MS regions. This is substantially different from the previous observation that pointed at cord blood CD3+ T cells as the major determinant of MS pathogenesis (220). These results are different and I believe more reasonable because I considered and assigned equal weight to each MS associated variant.

Notably, the same overrepresentation of DHSs from immune cell types was found when associated SNPs were divided into those with confirmed and suggestive evidence of association. This finding strongly suggests that many genetic variants that influence the risk of MS actually fail to reach genome-wide significance and that additional common variants associated to MS remain to be uncovered. I also investigated the proportion of phenotypic variance in MS and controls that was explained by SNPs in immune specific DHSs. I found that the 35,336 SNPs located inside either CD19+/CD20+ B, CD4+ helper T or CD8+ T DHSs explained more than 18% of the variance. Notably, when all autosomal genotyped SNPs are considered (n=475,806), the proportion of explained variance is approximately 30% (226). Since DHSs are likely to vary according to activation states and across subsets of B and helper and cytotoxic T cells, it was remarkable that the SNPs I identified (7.4% of the total genotyped SNPs) actually explain more than 50% of the variance explained by all SNPs. B cell specific SNPs appeared to play a prominent role, however the highest proportion of variance was explained by SNPs located within DHSs shared across T helper, CD8+ T and B cells.

I performed similar overlap analyses between cell specific DHSs and genomic regions associated with 11 additional complex conditions. The method used was sensitive enough to discriminate between diseases driven by different components of the immune system. For example, genomic regions associated with SLE and psoriasis were particularly active in B and CD8+ T cells respectively, while MS, RA, CD and celiac disease showed a marked T helper profile. The observation of a high enrichment of CD14+ and colorectal DHSs in UC associated regions adds further evidence for host-microbe interactions at the colonic level in shaping the risk of UC (227). Genetic variants active in CD14+ monocytes also influence the risk of leprosy and this is expected given the role of the innate immune response against this infectious condition (228). I also show that schizophrenia associated variants tend to be active in the developing brain further confirming the neurodevelopmental origin of this disease (229). The future identification of new associated loci and chromatin profiling in additional cell types will provide further insights in the pathogenesis of these conditions.

That an important role in MS is played by T cells is perhaps unremarkable, although I here show the importance of both CD4+ helper T and CD8+ T subsets. MS associated genomic regions are enriched for genes involved in T cell proliferation and activation (31). Furthermore, the main MS genetic risk factor resides in the MHC class II region, which plays a central role in the development of CD4+ helper T cell central tolerance (18, 230). The importance of CD8+ T cells is increasingly being recognised, especially given the fact that clonally expanded CD8+ T cells typically outnumber CD4+ T cells in active MS lesions and that HLA class I genes (necessary for CD8+ T central tolerance development) also influence MS risk (30, 31, 231, 232). The role of B cells in the pathogenesis of MS has been more debated. I am now able to confirm that my previous findings (see chapter 4) were not biased by the immortalization process and that MS associated variants considerably overlap with genomic regions that are active in B cells. In addition, I found that SNPs located within B cell DHSs explain a high

proportion of the phenotypic variance between MS patients and controls and this adds further evidence for the important role played by this cell type in MS.

To conclude, I show how the integration of GWAS and functional genomics information can help our understanding of complex diseases with relevant clinical implications. Genomic regions with confirmed and suggestive association with MS are active in immunological cell types and in particular T helper, T cytotoxic and B cells. The same strategy applied to other complex traits highlights differences in the pathogenesis of diseases driven by different components of the immune system as well as of non-immune mediated conditions. Future functional studies should aim to identify how MS associated variants can modify the chromatin landscape and gene expression. Since these effects are likely to be cell specific, our data should be considered when planning these experiments.

CONCLUSIONS:

In the course of my DPhil, I have been involved in a variety of projects aimed to elucidate the role played by genetic and environmental factors in the aetiology of MS. In particular, my aim was to investigate not only how each MS associated factor independently influences disease susceptibility but also how these factors interact in the causal cascade leading to disease.

In chapter one I was able to confirm the central role played by the *HLA-DRB1*1501* allele by showing that this genetic factor increases the risk of conversion to MS in an unselected population of children experiencing a first demyelinating event. In addition, I provided evidence that EBV negative paediatric cases of MS represent a separate entity characterized by lower age at disease onset, a considerably lower female to male ratio and a trend towards a lower frequency of the *HLA-DRB1*1501* allele. Although the presence of a co-association between EBV and MS cannot be excluded, I believe these findings indicate the peculiarity of EBV negative MS patients and further strengthen the hypothesis of a necessary role for EBV in the aetiology of MS.

In chapter two I investigated the potential role of another environmental factor, gestational vitamin D deficiency, underlying the seasonality of birth that characterizes MS patients. I was able to show that not only MS but also RA, UC and SLE are influenced by the season of birth suggesting a common mechanism behind this intriguing epidemiological phenomenon. Furthermore, I provided evidence for the presence of differences in the developing immune system of individuals born in different times of the year. By measuring TRECs, I could show that those individuals who were born in the high risk month for MS (May) had a considerably higher number of recently produced T cells compared to those born in the low risk month (November). Furthermore, the inverse correlation found between cord blood vitamin D and TRECs levels suggests that the seasonal nature of vitamin D status during gestation is the agent driving these observations.

In chapter three I wondered whether vitamin D could also be involved in the association observed between MS and EBV infection. I observed that EBV seroprevalence in both MS patients and controls is positively associated with latitude and I also demonstrated that vitamin D supplementation can effectively influence the immune response against EBV and results in a decreased production of antibodies against this virus. These results add further support to the increasing evidence suggesting the presence of a cascade linking these two environmental agents in the causal model of MS, whereby vitamin D deficiency leads to an altered immune response against EBV that in turn influences the development of MS.

The aim of chapter four was to investigate how non-MHC genetic variants associated with MS may be acting. I found that the genomic regions associated to MS susceptibility are located in regions of the genome which are active and have a regulatory function (enhancers and promoters) in immune cells and in particular immortalized B cell lines. I therefore concluded that the causative associated SNPs influence the risk of MS by influencing the chromatin landscape and regulating the expression of genes and that this happens specifically in the immune system.

In chapter five I aimed to provide a biological link between vitamin D and the genetics of MS. I found that a very large proportion of VDR binding in B cell lines localizes to regulatory active genomic regions (again enhancer and promoters) and that this overlap is particularly evident within MS associated regions. These findings indicate how the roles played by vitamin D and by MS associated genetic variants converge into the regulation of the expression of the same genes that likely drive the abnormal immune response in MS.

Finally, in chapter six I used publicly available chromatin data across more than 100 different cell types to define the pathogenic cell types driving the onset of MS. I observed that MS associated genetic variants are active in both T helper and CD8+ cytotoxic T cells, B cells and to a lesser extent NK cells and fetal thymic tissue. I found a similar overrepresentation of immune cell types when I

looked at genetic variants with only suggestive association with MS and this strongly suggests that a large number of SNPs influencing MS risk remain to be uncovered. Lastly, I showed how similar cell types are involved in the pathogenesis of immunological diseases other than MS, a finding that further confirms the presence of shared mechanisms across multiple immune mediated conditions.

Multiple pathways appear to link genetics and environment in MS. Based on what is currently known in this disease and based on the novel observations resulting from this thesis, we can conclude that MS has a strong genetic component and that the main genetic factor in MS plays an equally important role in adult and paediatric patients. MS associated genetic variants act by influencing the function of regulatory regions that are active in immune cell types and in particular T helper, T cytotoxic and B cells. Susceptibility to MS is further modified by environmental factors. One of them is vitamin D which is likely to influence the risk of MS through the action of its cognate nuclear receptor VDR that binds to the genome in proximity of a very large proportion of MS associated genes. The importance of vitamin D in the very early stages of the developing immune system is supported by the fact that both the risk of MS and the activity of the immune system are influenced by the month of birth and correlated with gestational vitamin D exposure. Vitamin D also appears to play an important role in modifying the immune response against EBV which is also associated with MS. The observation that vitamin D treatment has a considerable effect on the production of antibodies against this virus suggests the presence of a causal cascade linking vitamin D deficiency, EBV and the onset of MS.

Further work is needed to fully elucidate how these causative factors and their interactions lead to MS. For example, new EBV- PMS samples should be collected and genetic and demographic features of this particular subgroup of patients investigated with a larger sample size and therefore greater statistical power to detect differences (in particular genetic differences). Furthermore, I have shown a correlation between vitamin D, month of birth, immune system and risk of MS but intervention

studies are needed to prove causation. Therefore, I believe future studies should assess the effect of vitamin D supplementation in pregnant women on the developing immune system and the risk of immune-mediated diseases in their children. In addition, the effect of vitamin D supplementation on the immune response against EBV should also be studied. A recent study has shown that the CD8+ T-cell response to EBV correlates with disease activity in RRMS patients (233). Therefore, I propose that the effect of vitamin D on both the humoral and CD8+ T cell response against EBV as well as other viruses should be investigated in randomized placebo controlled trials of vitamin D supplementation in both MS patients and healthy controls. Finally, I have used epigenetic marks to define pathogenic cells in MS and I believe that the same marks in addition to DNA methylation could be used to assess if and how the epigenome of MS patients in disease relevant cell types is different from that of healthy controls and whether these differences are corrected by currently available treatments. These studies will be important because it is only by improving our knowledge of the mechanisms driving the onset of this often devastating condition that we will be able to develop effective strategies of disease treatment and prevention.

REFERENCES

1. Compston A, Coles A. Multiple sclerosis. Lancet. 2008;372(9648):1502-17. Epub 2008/10/31.
2. Koch-Henriksen N, Sorensen PS. The changing demographic pattern of multiple sclerosis epidemiology. Lancet neurology. 2010;9(5):520-32. Epub 2010/04/20.
3. Miller DH, Chard DT, Ciccarelli O. Clinically isolated syndromes. Lancet neurology. 2012;11(2):157-69. Epub 2012/01/24.
4. Fisniku LK, Brex PA, Altmann DR, Miszkiel KA, Benton CE, Lanyon R, et al. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. Brain : a journal of neurology. 2008;131(Pt 3):808-17. Epub 2008/02/01.
5. Compston A, Coles A. Multiple sclerosis. Lancet. 2002;359(9313):1221-31. Epub 2002/04/17.
6. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Annals of neurology. 2011;69(2):292-302. Epub 2011/03/10.
7. Barnett MH, Parratt JD, Pollard JD, Prineas JW. MS: is it one disease? International MS journal / MS Forum. 2009;16(2):57-65. Epub 2009/08/13.
8. Lassmann H, Bruck W, Lucchinetti CF. The immunopathology of multiple sclerosis: an overview. Brain Pathol. 2007;17(2):210-8. Epub 2007/03/29.
9. Gay FW, Drye TJ, Dick GW, Esiri MM. The application of multifactorial cluster analysis in the staging of plaques in early multiple sclerosis. Identification and characterization of the primary demyelinating lesion. Brain : a journal of neurology. 1997;120 (Pt 8):1461-83. Epub 1997/08/01.
10. Tzartos JS, Friese MA, Craner MJ, Palace J, Newcombe J, Esiri MM, et al. Interleukin-17 production in central nervous system-infiltrating T cells and glial cells is associated with active disease in multiple sclerosis. The American journal of pathology. 2008;172(1):146-55. Epub 2007/12/25.
11. Calabrese M, Filippi M, Gallo P. Cortical lesions in multiple sclerosis. Nature reviews Neurology. 2010;6(8):438-44. Epub 2010/07/14.
12. Calabrese M, Romualdi C, Poretto V, Favaretto A, Morra A, Rinaldi F, et al. The changing clinical course of multiple sclerosis: A matter of grey matter. Annals of neurology. 2013. Epub 2013/03/16.
13. Dalton CM, Chard DT, Davies GR, Miszkiel KA, Altmann DR, Fernando K, et al. Early development of multiple sclerosis is associated with progressive grey matter atrophy in patients presenting with clinically isolated syndromes. Brain : a journal of neurology. 2004;127(Pt 5):1101-7. Epub 2004/03/05.
14. Howell OW, Reeves CA, Nicholas R, Carassiti D, Radotra B, Gentleman SM, et al. Meningeal inflammation is widespread and linked to cortical pathology in multiple sclerosis. Brain : a journal of neurology. 2011;134(Pt 9):2755-71. Epub 2011/08/16.
15. 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. Epub 2004/06/15.
16. Barnett MH, Prineas JW. Relapsing and remitting multiple sclerosis: pathology of the newly forming lesion. Annals of neurology. 2004;55(4):458-68. Epub 2004/03/30.
17. Henderson AP, Barnett MH, Parratt JD, Prineas JW. Multiple sclerosis: distribution of inflammatory cells in newly forming lesions. Annals of neurology. 2009;66(6):739-53. Epub 2009/12/26.
18. Ramagopalan SV, Ebers GC. Multiple sclerosis: major histocompatibility complexity and antigen presentation. Genome medicine. 2009;1(11):105. Epub 2009/11/10.

19. Kasper LH, Shoemaker J. Multiple sclerosis immunology: The healthy immune system vs the MS immune system. Neurology. 2010;74 Suppl 1:S2-8. Epub 2010/02/06.
20. Segal BM, Constantinescu CS, Raychaudhuri A, Kim L, Fidelus-Gort R, Kasper LH. Repeated subcutaneous injections of IL12/23 p40 neutralising antibody, ustekinumab, in patients with relapsing-remitting multiple sclerosis: a phase II, double-blind, placebo-controlled, randomised, dose-ranging study. Lancet neurology. 2008;7(9):796-804. Epub 2008/08/16.
21. van Oosten BW, Lai M, Hodgkinson S, Barkhof F, Miller DH, Moseley IF, et al. Treatment of multiple sclerosis with the monoclonal anti-CD4 antibody cM-T412: results of a randomized, double-blind, placebo-controlled, MR-monitored phase II trial. Neurology. 1997;49(2):351-7. Epub 1997/08/01.
22. Brok HP, van Meurs M, Blezer E, Schantz A, Peritt D, Treacy G, et al. Prevention of experimental autoimmune encephalomyelitis in common marmosets using an anti-IL-12p40 monoclonal antibody. J Immunol. 2002;169(11):6554-63. Epub 2002/11/22.
23. Waldor MK, Sriram S, Hardy R, Herzenberg LA, Lanier L, Lim M, et al. Reversal of experimental allergic encephalomyelitis with monoclonal antibody to a T-cell subset marker. Science. 1985;227(4685):415-7. Epub 1985/01/25.
24. Freedman MS, Thompson EJ, Deisenhammer F, Giovannoni G, Grimsley G, Keir G, et al. Recommended standard of cerebrospinal fluid analysis in the diagnosis of multiple sclerosis: a consensus statement. Archives of neurology. 2005;62(6):865-70. Epub 2005/06/16.
25. Brettschneider J, Czerwoniak A, Senel M, Fang L, Kassubek J, Pinkhardt E, et al. The chemokine CXCL13 is a prognostic marker in clinically isolated syndrome (CIS). PloS one. 2010;5(8):e11986. Epub 2010/08/12.
26. Brettschneider J, Tumani H, Kiechle U, Muche R, Richards G, Lehmensiek V, et al. IgG antibodies against measles, rubella, and varicella zoster virus predict conversion to multiple sclerosis in clinically isolated syndrome. PloS one. 2009;4(11):e7638. Epub 2009/11/06.
27. Khademi M, Kockum I, Andersson ML, Iacobaeus E, Brundin L, Sellebjerg F, et al. Cerebrospinal fluid CXCL13 in multiple sclerosis: a suggestive prognostic marker for the disease course. Mult Scler. 2011;17(3):335-43. Epub 2010/12/08.
28. Sellebjerg F, Bornsen L, Khademi M, Krakauer M, Olsson T, Frederiksen JL, et al. Increased cerebrospinal fluid concentrations of the chemokine CXCL13 in active MS. Neurology. 2009;73(23):2003-10. Epub 2009/12/10.
29. Joseph FG, Hirst CL, Pickersgill TP, Ben-Shlomo Y, Robertson NP, Scolding NJ. CSF oligoclonal band status informs prognosis in multiple sclerosis: a case control study of 100 patients. Journal of neurology, neurosurgery, and psychiatry. 2009;80(3):292-6. Epub 2008/10/03.
30. Friese MA, Fugger L. Pathogenic CD8(+) T cells in multiple sclerosis. Annals of neurology. 2009;66(2):132-41. Epub 2009/09/11.
31. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9. Epub 2011/08/13.
32. Chanvillard C, Jacolik RF, Infante-Duarte C, Nayak RC. The role of natural killer cells in multiple sclerosis and their therapeutic implications. Frontiers in immunology. 2013;4:63. Epub 2013/03/16.
33. Kaur G, Trowsdale J, Fugger L. Natural killer cells and their receptors in multiple sclerosis. Brain : a journal of neurology. 2012. Epub 2012/06/27.
34. Zozulya AL, Wiendl H. The role of regulatory T cells in multiple sclerosis. Nature clinical practice Neurology. 2008;4(7):384-98. Epub 2008/06/26.
35. Ramagopalan SV, Dobson R, Meier UC, Giovannoni G. Multiple sclerosis: risk factors, prodromes, and potential causal pathways. Lancet neurology. 2010;9(7):727-39. Epub 2010/07/09.

36. Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC. Twin concordance and sibling recurrence rates in multiple sclerosis. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(22):12877-82. Epub 2003/10/22.
37. Ebers GC, Sadovnick AD, Risch NJ. A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. Nature. 1995;377(6545):150-1. Epub 1995/09/14.
38. Ebers GC, Yee IM, Sadovnick AD, Duquette P. Conjugal multiple sclerosis: population-based prevalence and recurrence risks in offspring. Canadian Collaborative Study Group. Annals of neurology. 2000;48(6):927-31. Epub 2000/12/16.
39. Hawkes CH, Macgregor AJ. Twin studies and the heritability of MS: a conclusion. Mult Scler. 2009;15(6):661-7. Epub 2009/06/02.
40. Mumford CJ, Wood NW, Kellar-Wood H, Thorpe JW, Miller DH, Compston DA. The British Isles survey of multiple sclerosis in twins. Neurology. 1994;44(1):11-5. Epub 1994/01/01.
41. Hansen T, Skytthe A, Stenager E, Petersen HC, Bronnum-Hansen H, Kyvik KO. Concordance for multiple sclerosis in Danish twins: an update of a nationwide study. Mult Scler. 2005;11(5):504-10. Epub 2005/10/01.
42. Hansen T, Skytthe A, Stenager E, Petersen HC, Kyvik KO, Bronnum-Hansen H. Risk for multiple sclerosis in dizygotic and monozygotic twins. Mult Scler. 2005;11(5):500-3. Epub 2005/10/01.
43. Ristori G, Cannoni S, Stazi MA, Vanacore N, Cotichini R, Alfo M, et al. Multiple sclerosis in twins from continental Italy and Sardinia: a nationwide study. Annals of neurology. 2006;59(1):27-34. Epub 2005/10/22.
44. Islam T, Gauderman WJ, Cozen W, Hamilton AS, Burnett ME, Mack TM. Differential twin concordance for multiple sclerosis by latitude of birthplace. Annals of neurology. 2006;60(1):56-64. Epub 2006/05/11.
45. Kurtzke JF. Geographic distribution of multiple sclerosis: An update with special reference to Europe and the Mediterranean region. Acta neurologica Scandinavica. 1980;62(2):65-80. Epub 1980/08/01.
46. Pugliatti M, Sotgiu S, Rosati G. The worldwide prevalence of multiple sclerosis. Clinical neurology and neurosurgery. 2002;104(3):182-91. Epub 2002/07/20.
47. Simpson S, Jr., Blizzard L, Otahal P, Van der Mei I, Taylor B. Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis. Journal of neurology, neurosurgery, and psychiatry. 2011;82(10):1132-41. Epub 2011/04/12.
48. Koch-Henriksen N, Sorensen PS. Why does the north-south gradient of incidence of multiple sclerosis seem to have disappeared on the northern hemisphere? Journal of the neurological sciences. 2011;311(1-2):58-63. Epub 2011/10/11.
49. Taylor BV, Lucas RM, Dear K, Kilpatrick TJ, Pender MP, van der Mei IA, et al. Latitudinal variation in incidence and type of first central nervous system demyelinating events. Mult Scler. 2010;16(4):398-405. Epub 2010/02/20.
50. Elian M, Nightingale S, Dean G. Multiple sclerosis among United Kingdom-born children of immigrants from the Indian subcontinent, Africa and the West Indies. Journal of neurology, neurosurgery, and psychiatry. 1990;53(10):906-11. Epub 1990/10/01.
51. Willer CJ, Dyment DA, Sadovnick AD, Rothwell PM, Murray TJ, Ebers GC. Timing of birth and risk of multiple sclerosis: population based study. BMJ. 2005;330(7483):120. Epub 2004/12/09.
52. Bayes HK, Weir CJ, O'Leary C. Timing of birth and risk of multiple sclerosis in the Scottish population. European neurology. 2010;63(1):36-40. Epub 2009/12/24.
53. Templer DI, Trent NH, Spencer DA, Trent A, Corgiat MD, Mortensen PB, et al. Season of birth in multiple sclerosis. Acta neurologica Scandinavica. 1992;85(2):107-9. Epub 1992/02/01.
54. Salzer J, Svenningsson A, Sundstrom P. Season of birth and multiple sclerosis in Sweden. Acta neurologica Scandinavica. 2010;121(1):20-3. Epub 2009/11/26.

55. Fernandes de Abreu DA, Babron MC, Rebeix I, Fontenille C, Yaouanq J, Brassat D, et al. Season of birth and not vitamin D receptor promoter polymorphisms is a risk factor for multiple sclerosis. Mult Scler. 2009;15(10):1146-52. Epub 2009/12/08.
56. Saastamoinen KP, Auvinen MK, Tienari PJ. Month of birth is associated with multiple sclerosis but not with HLA-DR15 in Finland. Mult Scler. 2012;18(5):563-8. Epub 2011/11/02.
57. Sotgiu S, Pugliatti M, Sotgiu MA, Fois ML, Arru G, Sanna A, et al. Seasonal fluctuation of multiple sclerosis births in Sardinia. Journal of neurology. 2006;253(1):38-44. Epub 2005/07/16.
58. Staples J, Ponsonby AL, Lim L. Low maternal exposure to ultraviolet radiation in pregnancy, month of birth, and risk of multiple sclerosis in offspring: longitudinal analysis. BMJ. 2010;340:c1640. Epub 2010/10/30.
59. Ebers GC, Sadovnick AD, Dyment DA, Yee IML, Willer CJ, Risch N. Parent-of-origin effect in multiple sclerosis: observations in half-siblings. Lancet. 2004;363(9423):1773-4.
60. Dean G, Kurtzke JF. On the risk of multiple sclerosis according to age at immigration to South Africa. British medical journal. 1971;3(5777):725-9. Epub 1971/09/25.
61. Dean G, Elian M. Age at immigration to England of Asian and Caribbean immigrants and the risk of developing multiple sclerosis. Journal of neurology, neurosurgery, and psychiatry. 1997;63(5):565-8. Epub 1998/01/04.
62. Orton SM, Herrera BM, Yee IM, Valdar W, Ramagopalan SV, Sadovnick AD, et al. Sex ratio of multiple sclerosis in Canada: a longitudinal study. Lancet neurology. 2006;5(11):932-6. Epub 2006/10/21.
63. Handel AE, Giovannoni G, Ebers GC, Ramagopalan SV. Environmental factors and their timing in adult-onset multiple sclerosis. Nature reviews Neurology. 2010;6(3):156-66. Epub 2010/02/17.
64. Jersild C, Fog T, Hansen GS, Thomsen M, Svejgaard A, Dupont B. Histocompatibility determinants in multiple sclerosis, with special reference to clinical course. Lancet. 1973;2(7840):1221-5. Epub 1973/12/01.
65. Dyment DA, Herrera BM, Cader MZ, Willer CJ, Lincoln MR, Sadovnick AD, et al. Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. Human molecular genetics. 2005;14(14):2019-26. Epub 2005/06/03.
66. Ramagopalan SV, Morris AP, Dyment DA, Herrera BM, DeLuca GC, Lincoln MR, et al. The inheritance of resistance alleles in multiple sclerosis. PLoS genetics. 2007;3(9):1607-13. Epub 2007/09/12.
67. Oksenberg JR, Barcellos LF, Cree BA, Baranzini SE, Bugawan TL, Khan O, et al. Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. American journal of human genetics. 2004;74(1):160-7. Epub 2003/12/12.
68. Matsuoka T, Matsushita T, Osoegawa M, Kawano Y, Minohara M, Mihara F, et al. Association of the HLA-DRB1 alleles with characteristic MRI features of Asian multiple sclerosis. Mult Scler. 2008;14(9):1181-90. Epub 2008/10/28.
69. Marrosu MG, Murru R, Murru MR, Costa G, Zavattari P, Whalen M, et al. Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. Human molecular genetics. 2001;10(25):2907-16. Epub 2001/12/14.
70. Lincoln MR, Ramagopalan SV, Chao MJ, Herrera BM, Deluca GC, Orton SM, et al. Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(18):7542-7. Epub 2009/04/22.
71. De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L, Aggarwal NT, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nature genetics. 2009;41(7):776-82. Epub 2009/06/16.

72. ANZgene. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nature genetics. 2009;41(7):824-8. Epub 2009/06/16.
73. Patsopoulos NA, Esposito F, Reischl J, Lehr S, Bauer D, Heubach J, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Annals of neurology. 2011;70(6):897-912. Epub 2011/12/23.
74. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. Nature reviews Genetics. 2009;10(4):241-51. Epub 2009/03/19.
75. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nature reviews Genetics. 2010;11(6):446-50. Epub 2010/05/19.
76. D'Netto MJ, Ward H, Morrison KM, Ramagopalan SV, Dyment DA, DeLuca GC, et al. Risk alleles for multiple sclerosis in multiplex families. Neurology. 2009;72(23):1984-8. Epub 2009/06/10.
77. Gourraud PA, McElroy JP, Caillier SJ, Johnson BA, Santaniello A, Hauser SL, et al. Aggregation of multiple sclerosis genetic risk variants in multiple and single case families. Annals of neurology. 2011;69(1):65-74. Epub 2011/02/01.
78. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nature genetics. 2008;40(6):695-701. Epub 2008/05/30.
79. Handel AE, Disanto G, Ramagopalan SV. Next-generation sequencing in understanding complex neurological disease. Expert review of neurotherapeutics. 2013;13(2):215-27. Epub 2013/02/02.
80. Ascherio A, Munger KL, Simon KC. Vitamin D and multiple sclerosis. Lancet neurology. 2010;9(6):599-612. Epub 2010/05/25.
81. Ramagopalan SV, Handel AE, Giovannoni G, Rutherford Siegel S, Ebers GC, Chaplin G. Relationship of UV exposure to prevalence of multiple sclerosis in England. Neurology. 2011;76(16):1410-4. Epub 2011/04/20.
82. Ebers GC. Environmental factors and multiple sclerosis. Lancet neurology. 2008;7(3):268-77. Epub 2008/02/16.
83. Orton SM, Wald L, Confavreux C, Vukusic S, Krohn JP, Ramagopalan SV, et al. Association of UV radiation with multiple sclerosis prevalence and sex ratio in France. Neurology. 2011;76(5):425-31. Epub 2011/02/02.
84. van der Mei IA, Ponsonby AL, Dwyer T, Blizzard L, Simmons R, Taylor BV, et al. Past exposure to sun, skin phenotype, and risk of multiple sclerosis: case-control study. BMJ. 2003;327(7410):316. Epub 2003/08/09.
85. Islam T, Gauderman WJ, Cozen W, Mack TM. Childhood sun exposure influences risk of multiple sclerosis in monozygotic twins. Neurology. 2007;69(4):381-8. Epub 2007/07/25.
86. Munger KL, Zhang SM, O'Reilly E, Hernan MA, Olek MJ, Willett WC, et al. Vitamin D intake and incidence of multiple sclerosis. Neurology. 2004;62(1):60-5. Epub 2004/01/14.
87. Munger KL, Levin LI, Hollis BW, Howard NS, Ascherio A. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. JAMA : the journal of the American Medical Association. 2006;296(23):2832-8. Epub 2006/12/21.
88. Lawlor DA, Davey Smith G, Bruckdorfer KR, Kundu D, Ebrahim S. Observational versus randomised trial evidence. Lancet. 2004;364(9436):755. Epub 2004/09/01.
89. Ramagopalan SV, Dyment DA, Cader MZ, Morrison KM, Disanto G, Morahan JM, et al. Rare variants in the CYP27B1 gene are associated with multiple sclerosis. Annals of neurology. 2011;70(6):881-6. Epub 2011/12/23.
90. Holick MF. Vitamin D deficiency. The New England journal of medicine. 2007;357(3):266-81. Epub 2007/07/20.

91. Kamen DL, Tangpricha V. Vitamin D and molecular actions on the immune system: modulation of innate and autoimmunity. J Mol Med (Berl). 2010;88(5):441-50. Epub 2010/02/02.
92. Panda DK, Miao D, Tremblay ML, Sirois J, Farookhi R, Hendy GN, et al. Targeted ablation of the 25-hydroxyvitamin D 1alpha -hydroxylase enzyme: evidence for skeletal, reproductive, and immune dysfunction. Proceedings of the National Academy of Sciences of the United States of America. 2001;98(13):7498-503. Epub 2001/06/21.
93. Cantorna MT, Hayes CE, DeLuca HF. 1,25-Dihydroxyvitamin D3 reversibly blocks the progression of relapsing encephalomyelitis, a model of multiple sclerosis. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(15):7861-4. Epub 1996/07/23.
94. Ramagopalan SV, Heger A, Berlanga AJ, Maugeri NJ, Lincoln MR, Burrell A, et al. A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. Genome research. 2010;20(10):1352-60. Epub 2010/08/26.
95. Klein G, Klein E, Kashuba E. Interaction of Epstein-Barr virus (EBV) with human B-lymphocytes. Biochemical and biophysical research communications. 2010;396(1):67-73. Epub 2010/05/25.
96. Luzuriaga K, Sullivan JL. Infectious mononucleosis. The New England journal of medicine. 2010;362(21):1993-2000. Epub 2010/05/28.
97. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. Annals of neurology. 2007;61(4):288-99. Epub 2007/04/21.
98. Pakpoor J, Disanto G, Gerber JE, Dobson R, Meier UC, Giovannoni G, et al. The risk of developing multiple sclerosis in individuals seronegative for Epstein-Barr virus: a meta-analysis. Mult Scler. 2012. Epub 2012/06/29.
99. Levin LI, Munger KL, O'Reilly EJ, Falk KI, Ascherio A. Primary infection with the Epstein-Barr virus and risk of multiple sclerosis. Annals of neurology. 2010;67(6):824-30. Epub 2010/06/03.
100. Levin LI, Munger KL, Rubertone MV, Peck CA, Lennette ET, Spiegelman D, et al. Temporal relationship between elevation of epstein-barr virus antibody titers and initial onset of neurological symptoms in multiple sclerosis. JAMA : the journal of the American Medical Association. 2005;293(20):2496-500. Epub 2005/05/26.
101. Munger KL, Levin LI, O'Reilly EJ, Falk KI, Ascherio A. Anti-Epstein-Barr virus antibodies as serological markers of multiple sclerosis: a prospective study among United States military personnel. Mult Scler. 2011;17(10):1185-93. Epub 2011/06/21.
102. Nielsen TR, Rostgaard K, Nielsen NM, Koch-Henriksen N, Haahr S, Sorensen PS, et al. Multiple sclerosis after infectious mononucleosis. Archives of neurology. 2007;64(1):72-5. Epub 2007/01/11.
103. Ahlgren C, Toren K, Oden A, Andersen O. A population-based case-control study on viral infections and vaccinations and subsequent multiple sclerosis risk. European journal of epidemiology. 2009;24(9):541-52. Epub 2009/07/28.
104. Sundstrom P, Juto P, Wadell G, Hallmans G, Svenningsson A, Nystrom L, et al. An altered immune response to Epstein-Barr virus in multiple sclerosis: a prospective study. Neurology. 2004;62(12):2277-82. Epub 2004/06/24.
105. Handel AE, Williamson AJ, Disanto G, Handunnetthi L, Giovannoni G, Ramagopalan SV. An updated meta-analysis of risk of multiple sclerosis following infectious mononucleosis. PloS one. 2010;5(9). Epub 2010/09/09.
106. Ramagopalan SV, Hoang U, Seagroatt V, Handel A, Ebers GC, Giovannoni G, et al. Geography of hospital admissions for multiple sclerosis in England and comparison with the geography of hospital admissions for infectious mononucleosis: a descriptive study. Journal of neurology, neurosurgery, and psychiatry. 2011;82(6):682-7. Epub 2011/01/08.

107. Santon A, Cristobal E, Aparicio M, Royuela A, Villar LM, Alvarez-Cermeno JC. High frequency of co-infection by Epstein-Barr virus types 1 and 2 in patients with multiple sclerosis. Mult Scler. 2011;17(11):1295-300. Epub 2011/07/16.
108. Banwell B, Bar-Or A, Arnold DL, Sadovnick D, Narayanan S, McGowan M, et al. Clinical, environmental, and genetic determinants of multiple sclerosis in children with acute demyelination: a prospective national cohort study. Lancet neurology. 2011;10(5):436-45. Epub 2011/04/05.
109. Ramagopalan SV, Meier UC, Conacher M, Ebers GC, Giovannoni G, Crawford DH, et al. Role of the HLA system in the association between multiple sclerosis and infectious mononucleosis. Archives of neurology. 2011;68(4):469-72. Epub 2011/04/13.
110. Antonovsky A, Leibowitz U, Smith HA, Medalie JM, Balogh M, Kats R, et al. Epidemiologic Study of Multiple Sclerosis in Israel. I. An Overall Review of Methods and Findings. Archives of neurology. 1965;13:183-93. Epub 1965/08/01.
111. Ghadirian P, Dadgostar B, Azani R, Maisonneuve P. A case-control study of the association between socio-demographic, lifestyle and medical history factors and multiple sclerosis. Canadian journal of public health Revue canadienne de sante publique. 2001;92(4):281-5. Epub 2002/04/23.
112. Hedstrom AK, Baarnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. Neurology. 2009;73(9):696-701. Epub 2009/09/02.
113. Pekmezovic T, Drulovic J, Milenkovic M, Jarebinski M, Stojsavljevic N, Mesaros S, et al. Lifestyle factors and multiple sclerosis: A case-control study in Belgrade. Neuroepidemiology. 2006;27(4):212-6. Epub 2006/11/11.
114. Zorzon M, Zivadinov R, Nasuelli D, Dolfini P, Bosco A, Bratina A, et al. Risk factors of multiple sclerosis: a case-control study. Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology. 2003;24(4):242-7. Epub 2003/12/06.
115. Riise T, Nortvedt MW, Ascherio A. Smoking is a risk factor for multiple sclerosis. Neurology. 2003;61(8):1122-4. Epub 2003/10/29.
116. Villard-Mackintosh L, Vessey MP. Oral contraceptives and reproductive factors in multiple sclerosis incidence. Contraception. 1993;47(2):161-8. Epub 1993/02/01.
117. Thorogood M, Hannaford PC. The influence of oral contraceptives on the risk of multiple sclerosis. British journal of obstetrics and gynaecology. 1998;105(12):1296-9. Epub 1999/01/12.
118. Hernan MA, Olek MJ, Ascherio A. Cigarette smoking and incidence of multiple sclerosis. American journal of epidemiology. 2001;154(1):69-74. Epub 2001/06/28.
119. Handel AE, Williamson AJ, Disanto G, Dobson R, Giovannoni G, Ramagopalan SV. Smoking and multiple sclerosis: an updated meta-analysis. PloS one. 2011;6(1):e16149. Epub 2011/01/21.
120. Mikaeloff Y, Caridade G, Tardieu M, Suissa S. Parental smoking at home and the risk of childhood-onset multiple sclerosis in children. Brain : a journal of neurology. 2007;130(Pt 10):2589-95. Epub 2007/09/11.
121. Palacios N, Alonso A, Bronnum-Hansen H, Ascherio A. Smoking and increased risk of multiple sclerosis: parallel trends in the sex ratio reinforce the evidence. Annals of epidemiology. 2011;21(7):536-42. Epub 2011/05/10.
122. Hersey P, Prendergast D, Edwards A. Effects of cigarette smoking on the immune system. Follow-up studies in normal subjects after cessation of smoking. The Medical journal of Australia. 1983;2(9):425-9. Epub 1983/10/29.
123. Moszczynski P, Zabinski Z, Moszczynski P, Jr., Rutowski J, Slowinski S, Tabarowski Z. Immunological findings in cigarette smokers. Toxicology letters. 2001;118(3):121-7. Epub 2001/01/04.
124. Smith KJ, Kapoor R, Felts PA. Demyelination: the role of reactive oxygen and nitrogen species. Brain Pathol. 1999;9(1):69-92. Epub 1999/02/16.

125. Moscarello MA, Wood DD, Ackerley C, Boulias C. Myelin in multiple sclerosis is developmentally immature. The Journal of clinical investigation. 1994;94(1):146-54. Epub 1994/07/01.
126. Mastronardi FG, Noor A, Wood DD, Paton T, Moscarello MA. Peptidyl argininedeiminase 2 CpG island in multiple sclerosis white matter is hypomethylated. Journal of neuroscience research. 2007;85(9):2006-16. Epub 2007/05/01.
127. Makrygiannakis D, Hermansson M, Ulfgren AK, Nicholas AP, Zendman AJ, Eklund A, et al. Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. Annals of the rheumatic diseases. 2008;67(10):1488-92. Epub 2008/04/17.
128. Banwell B, Ghezzi A, Bar-Or A, Mikaeloff Y, Tardieu M. Multiple sclerosis in children: clinical diagnosis, therapeutic strategies, and future directions. Lancet neurology. 2007;6(10):887-902. Epub 2007/09/22.
129. Mikaeloff Y, Suissa S, Vallee L, Lubetzki C, Ponsot G, Confavreux C, et al. First episode of acute CNS inflammatory demyelination in childhood: prognostic factors for multiple sclerosis and disability. The Journal of pediatrics. 2004;144(2):246-52. Epub 2004/02/05.
130. Boiko AN, Gusev EI, Sudomoina MA, Alekseenkov AD, Kulakova OG, Bikova OV, et al. Association and linkage of juvenile MS with HLA-DR2(15) in Russians. Neurology. 2002;58(4):658-60. Epub 2002/02/28.
131. Idrissova Zh R, Boldyreva MN, Dekonenko EP, Malishev NA, Leontyeva IY, Martinenko IN, et al. Acute disseminated encephalomyelitis in children: clinical features and HLA-DR linkage. European journal of neurology : the official journal of the European Federation of Neurological Societies. 2003;10(5):537-46. Epub 2003/08/28.
132. Banwell B, Krupp L, Kennedy J, Tellier R, Tenembaum S, Ness J, et al. Clinical features and viral serologies in children with multiple sclerosis: a multinational observational study. Lancet neurology. 2007;6(9):773-81. Epub 2007/08/11.
133. Waubant E, Mowry EM, Krupp L, Chitnis T, Yeh EA, Kuntz N, et al. Common viruses associated with lower pediatric multiple sclerosis risk. Neurology. 2011;76(23):1989-95. Epub 2011/06/08.
134. Pohl D, Krone B, Rostasy K, Kahler E, Brunner E, Lehnert M, et al. High seroprevalence of Epstein-Barr virus in children with multiple sclerosis. Neurology. 2006;67(11):2063-5. Epub 2006/12/13.
135. Alotaibi S, Kennedy J, Tellier R, Stephens D, Banwell B. Epstein-Barr virus in pediatric multiple sclerosis. JAMA : the journal of the American Medical Association. 2004;291(15):1875-9. Epub 2004/04/22.
136. Krupp LB, Banwell B, Tenembaum S. Consensus definitions proposed for pediatric multiple sclerosis and related disorders. Neurology. 2007;68(16 Suppl 2):S7-12. Epub 2007/04/18.
137. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Annals of neurology. 2005;58(6):840-6. Epub 2005/11/12.
138. Sadovnick AD, Risch NJ, Ebers GC. Canadian collaborative project on genetic susceptibility to MS, phase 2: rationale and method. Canadian Collaborative Study Group. The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques. 1998;25(3):216-21. Epub 1998/08/26.
139. Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. Tissue antigens. 1992;39(5):225-35. Epub 1992/05/01.
140. Yeh EA, Chitnis T, Krupp L, Ness J, Chabas D, Kuntz N, et al. Pediatric multiple sclerosis. Nature reviews Neurology. 2009;5(11):621-31. Epub 2009/10/15.

141. Shapira Y, Agmon-Levin N, Shoenfeld Y. Defining and analyzing geoepidemiology and human autoimmunity. Journal of autoimmunity. 2010;34(3):J168-77. Epub 2009/12/26.
142. Branca F, Pastore G, Demissie T, Ferro-Luzzi A. The nutritional impact of seasonality in children and adults of rural Ethiopia. European journal of clinical nutrition. 1993;47(12):840-50. Epub 1993/12/01.
143. Watson PE, McDonald BW. Seasonal variation of nutrient intake in pregnancy: effects on infant measures and possible influence on diseases related to season of birth. European journal of clinical nutrition. 2007;61(11):1271-80. Epub 2007/02/15.
144. Bates CJ, Prentice AM, Paul AA. Seasonal variations in vitamins A, C, riboflavin and folate intakes and status of pregnant and lactating women in a rural Gambian community: some possible implications. European journal of clinical nutrition. 1994;48(9):660-8. Epub 1994/09/01.
145. Barker DJ. The Wellcome Foundation Lecture, 1994. The fetal origins of adult disease. Proceedings Biological sciences / The Royal Society. 1995;262(1363):37-43. Epub 1995/10/23.
146. Langley-Evans SC, McMullen S. Developmental origins of adult disease. Medical principles and practice : international journal of the Kuwait University, Health Science Centre. 2010;19(2):87-98. Epub 2010/02/06.
147. Kahn HS, Morgan TM, Case LD, Dabelea D, Mayer-Davis EJ, Lawrence JM, et al. Association of type 1 diabetes with month of birth among U.S. youth: The SEARCH for Diabetes in Youth Study. Diabetes care. 2009;32(11):2010-5. Epub 2009/08/14.
148. Rothwell PM, Staines A, Smail P, Wadsworth E, McKinney P. Seasonality of birth of patients with childhood diabetes in Britain. BMJ. 1996;312(7044):1456-7. Epub 1996/06/08.
149. Ekbom A, Zack M, Adami HO, Helmick C. Is there clustering of inflammatory bowel disease at birth? American journal of epidemiology. 1991;134(8):876-86. Epub 1991/10/15.
150. Haslam N, Mayberry JF, Hawthorne AB, Newcombe RG, Holmes GK, Probert CS. Measles, month of birth, and Crohn's disease. Gut. 2000;47(6):801-3. Epub 2000/11/15.
151. Sorensen HT, Pedersen L, Norgard B, Fonager K, Rothman KJ. Does month of birth affect risk of Crohn's disease in childhood and adolescence? BMJ. 2001;323(7318):907. Epub 2001/10/23.
152. Card TR, Sawczenko A, Sandhu BK, Logan RF. No seasonality in month of birth of inflammatory bowel disease cases: a prospective population based study of British under 20 year olds. Gut. 2002;51(6):814-5. Epub 2002/11/13.
153. Chowers Y, Odes S, Bujanover Y, Eliakim R, Bar Meir S, Avidan B. The month of birth is linked to the risk of Crohn's disease in the Israeli population. The American journal of gastroenterology. 2004;99(10):1974-6. Epub 2004/09/28.
154. Bai A, Guo Y, Shen Y, Xie Y, Zhu X, Lu N. Seasonality in flares and months of births of patients with ulcerative colitis in a Chinese population. Digestive diseases and sciences. 2009;54(5):1094-8. Epub 2008/12/04.
155. Buchanan WW, Gregoire LG, Buchanan HM. Month of birth and rheumatoid arthritis. Lancet. 1987;2(8557):517. Epub 1987/08/29.
156. Angelucci E, Cocco A, Cesarini M, Crudeli A, Necozione S, Caprilli R, et al. Monthly and seasonal birth patterns and the occurrence of Crohn's disease. The American journal of gastroenterology. 2009;104(6):1608-9. Epub 2009/06/06.
157. Sonnenberg A. Date of birth in the occurrence of inflammatory bowel disease. Inflammatory bowel diseases. 2009;15(2):206-11. Epub 2008/10/04.
158. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. The New England journal of medicine. 2011;365(23):2205-19. Epub 2011/12/14.
159. Tsokos GC. Systemic lupus erythematosus. The New England journal of medicine. 2011;365(22):2110-21. Epub 2011/12/02.

160. Kaser A, Zeissig S, Blumberg RS. Inflammatory bowel disease. Annual review of immunology. 2010;28:573-621. Epub 2010/03/03.
161. Baranzini SE. The genetics of autoimmune diseases: a networked perspective. Current opinion in immunology. 2009;21(6):596-605. Epub 2009/11/10.
162. Shoenfeld N, Amital H, Shoenfeld Y. The effect of melanism and vitamin D synthesis on the incidence of autoimmune disease. Nature clinical practice Rheumatology. 2009;5(2):99-105. Epub 2009/02/03.
163. Beach RS, Gershwin ME, Hurley LS. Reversibility of development retardation following murine fetal zinc deprivation. The Journal of nutrition. 1982;112(6):1169-81. Epub 1982/06/01.
164. Barnett AG, Dobson AJ. Analysing Seasonal Health Data: Springer; 2010.
165. Hypponen E, Power C. Hypovitaminosis D in British adults at age 45 y: nationwide cohort study of dietary and lifestyle predictors. The American journal of clinical nutrition. 2007;85(3):860-8. Epub 2007/03/09.
166. Spits H. Development of alphabeta T cells in the human thymus. Nature reviews Immunology. 2002;2(10):760-72. Epub 2002/10/03.
167. Lynch HE, Sempowski GD. Molecular measurement of T cell receptor excision circles. Methods Mol Biol. 2013;979:147-59. Epub 2013/02/12.
168. Parfitt AM, Gallagher JC, Heaney RP, Johnston CC, Neer R, Whedon GD. Vitamin D and bone health in the elderly. The American journal of clinical nutrition. 1982;36(5 Suppl):1014-31. Epub 1982/11/01.
169. Harvey L, Burne TH, McGrath JJ, Eyles DW. Developmental vitamin D3 deficiency induces alterations in immune organ morphology and function in adult offspring. The Journal of steroid biochemistry and molecular biology. 2010;121(1-2):239-42. Epub 2010/03/23.
170. Yu S, Cantorna MT. Epigenetic reduction in invariant NKT cells following in utero vitamin D deficiency in mice. J Immunol. 2011;186(3):1384-90. Epub 2010/12/31.
171. Hewison M. Antibacterial effects of vitamin D. Nature reviews Endocrinology. 2011;7(6):337-45. Epub 2011/01/26.
172. Beard JA, Bearden A, Striker R. Vitamin D and the anti-viral state. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology. 2011;50(3):194-200. Epub 2011/01/19.
173. Ascherio A, Munger KL, Lennette ET, Spiegelman D, Hernan MA, Olek MJ, et al. Epstein-Barr virus antibodies and risk of multiple sclerosis: a prospective study. JAMA : the journal of the American Medical Association. 2001;286(24):3083-8. Epub 2002/01/05.
174. Ponsonby AL, van der Mei I, Dwyer T, Blizzard L, Taylor B, Kemp A, et al. Exposure to infant siblings during early life and risk of multiple sclerosis. JAMA : the journal of the American Medical Association. 2005;293(4):463-9. Epub 2005/01/27.
175. Bray PF, Bloomer LC, Salmon VC, Bagley MH, Larsen PD. Epstein-Barr virus infection and antibody synthesis in patients with multiple sclerosis. Archives of neurology. 1983;40(7):406-8. Epub 1983/07/01.
176. Haahr S, Plesner AM, Vestergaard BF, Hollsberg P. A role of late Epstein-Barr virus infection in multiple sclerosis. Acta neurologica Scandinavica. 2004;109(4):270-5. Epub 2004/03/16.
177. Larsen PD, Bloomer LC, Bray PF. Epstein-Barr nuclear antigen and viral capsid antigen antibody titers in multiple sclerosis. Neurology. 1985;35(3):435-8. Epub 1985/03/01.
178. Munch M, Hvas J, Christensen T, Moller-Larsen A, Haahr S. The implications of Epstein-Barr virus in multiple sclerosis--a review. Acta neurologica Scandinavica Supplementum. 1997;169:59-64. Epub 1997/01/01.
179. Myhr KM, Riise T, Barrett-Connor E, Myrmel H, Vedeler C, Gronning M, et al. Altered antibody pattern to Epstein-Barr virus but not to other herpesviruses in multiple sclerosis: a population based

case-control study from western Norway. Journal of neurology, neurosurgery, and psychiatry. 1998;64(4):539-42. Epub 1998/05/12.
180. Shirodaria PV, Haire M, Fleming E, Merrett JD, Hawkins SA, Roberts SD. Viral antibody titers. Comparison in patients with multiple sclerosis and rheumatoid arthritis. Archives of neurology. 1987;44(12):1237-41. Epub 1987/12/01.
181. Sumaya CV, Myers LW, Ellison GW. Epstein-Barr virus antibodies in multiple sclerosis. Archives of neurology. 1980;37(2):94-6. Epub 1980/02/01.
182. Sumaya CV, Myers LW, Ellison GW, Ench Y. Increased prevalence and titer of Epstein-Barr virus antibodies in patients with multiple sclerosis. Annals of neurology. 1985;17(4):371-7. Epub 1985/04/01.
183. Wagner HJ, Hennig H, Jabs WJ, Siekhaus A, Wessel K, Wandinger KP. Altered prevalence and reactivity of anti-Epstein-Barr virus antibodies in patients with multiple sclerosis. Viral immunology. 2000;13(4):497-502. Epub 2001/02/24.
184. Ferrante P, Castellani P, Barbi M, Bergamini F. The Italian Cooperative Multiple Sclerosis case-control study: preliminary results on viral antibodies. Italian journal of neurological sciences. 1987;Suppl 6:45-50. Epub 1987/06/01.
185. Jilek S, Schluep M, Meylan P, Vingerhoets F, Guignard L, Monney A, et al. Strong EBV-specific CD8+ T-cell response in patients with early multiple sclerosis. Brain : a journal of neurology. 2008;131(Pt 7):1712-21. Epub 2008/06/14.
186. Lindsey JW, Hatfield LM, Vu T. Epstein-Barr virus neutralizing and early antigen antibodies in multiple sclerosis. European journal of neurology : the official journal of the European Federation of Neurological Societies. 2010;17(10):1263-9. Epub 2010/04/21.
187. Villegas E, Santiago O, Carrillo JA, Sorlozano A, Guerrero M, Fernandez O, et al. Low intrathecal immune response of anti-EBNA-1 antibodies and EBV DNA from multiple sclerosis patients. Diagnostic microbiology and infectious disease. 2011;70(1):85-90. Epub 2011/03/11.
188. Zivadinov R, Nasuelli D, Tommasi MA, Serafin M, Bratina A, Ukmar M, et al. Positivity of cytomegalovirus antibodies predicts a better clinical and radiological outcome in multiple sclerosis patients. Neurological research. 2006;28(3):262-9. Epub 2006/05/12.
189. Riverol M, Sepulcre J, Fernandez-Diez B, Villoslada P, Fernandez-Alonso M, Rubio M, et al. Antibodies against Epstein-Barr virus and herpesvirus type 6 are associated with the early phases of multiple sclerosis. Journal of neuroimmunology. 2007;192(1-2):184-5. Epub 2007/09/18.
190. Gronen F, Ruprecht K, Weissbrich B, Klinker E, Kroner A, Hofstetter HH, et al. Frequency analysis of HLA-B7-restricted Epstein-Barr virus-specific cytotoxic T lymphocytes in patients with multiple sclerosis and healthy controls. Journal of neuroimmunology. 2006;180(1-2):185-92. Epub 2006/10/07.
191. Ingram G, Bugert JJ, Loveless S, Robertson NP. Anti-EBNA-1 IgG is not a reliable marker of multiple sclerosis clinical disease activity. European journal of neurology : the official journal of the European Federation of Neurological Societies. 2010;17(11):1386-9. Epub 2010/05/21.
192. Khaki M, Ghazavi A, Ghasami K, Rafiei M, Payani MA, Ghaznavi-Rad E, et al. Evaluation of viral antibodies in Iranian multiple sclerosis patients. Neurosciences (Riyadh). 2011;16(3):224-8. Epub 2011/06/17.
193. Smolders J, Peelen E, Thewissen M, Cohen Tervaert JW, Menheere P, Hupperts R, et al. Safety and T cell modulating effects of high dose vitamin D3 supplementation in multiple sclerosis. PloS one. 2010;5(12):e15235. Epub 2010/12/24.
194. Farrell RA, Antony D, Wall GR, Clark DA, Fisniku L, Swanton J, et al. Humoral immune response to EBV in multiple sclerosis is associated with disease activity on MRI. Neurology. 2009;73(1):32-8. Epub 2009/05/22.

195. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. Annals of neurology. 2007;61(6):504-13. Epub 2007/05/12.
196. Webb AR, Kline L, Holick MF. Influence of season and latitude on the cutaneous synthesis of vitamin D3: exposure to winter sunlight in Boston and Edmonton will not promote vitamin D3 synthesis in human skin. The Journal of clinical endocrinology and metabolism. 1988;67(2):373-8. Epub 1988/08/01.
197. de Ory F, Guisasola ME, Sanz JC, Garcia-Bermejo I. Evaluation of four commercial systems for the diagnosis of Epstein-Barr virus primary infections. Clinical and vaccine immunology : CVI. 2011;18(3):444-8. Epub 2010/12/31.
198. Knippenberg S, Smolders J, Thewissen M, Peelen E, Tervaert JW, Hupperts R, et al. Effect of vitamin D(3) supplementation on peripheral B cell differentiation and isotype switching in patients with multiple sclerosis. Mult Scler. 2011;17(12):1418-23. Epub 2011/06/22.
199. Salzer J, Nystrom M, Hallmans G, Stenlund H, Wadell G, Sundstrom P. Epstein-Barr virus antibodies and vitamin D in prospective multiple sclerosis biobank samples. Mult Scler. 2013. Epub 2013/04/04.
200. Zwart SR, Mehta SK, Ploutz-Snyder R, Bourbeau Y, Locke JP, Pierson DL, et al. Response to vitamin D supplementation during Antarctic winter is related to BMI, and supplementation can mitigate Epstein-Barr Virus Reactivation. The Journal of nutrition. 2011;141(4):692-7. Epub 2011/05/04.
201. Ford JL, Stowe RP. Racial-ethnic differences in Epstein-Barr virus antibody titers among U.S. children and adolescents. Annals of epidemiology. 2013;23(5):275-80. Epub 2013/04/30.
202. Looker AC, Johnson CL, Lacher DA, Pfeiffer CM, Schleicher RL, Sempos CT. Vitamin D status: United States, 2001-2006. NCHS data brief. 2011(59):1-8. Epub 2011/05/20.
203. von Budingen HC, Bar-Or A, Zamvil SS. B cells in multiple sclerosis: connecting the dots. Current opinion in immunology. 2011;23(6):713-20. Epub 2011/10/11.
204. Franciotta D, Salvetti M, Lolli F, Serafini B, Aloisi F. B cells and multiple sclerosis. Lancet neurology. 2008;7(9):852-8. Epub 2008/08/16.
205. Kappos L, Li D, Calabresi PA, O'Connor P, Bar-Or A, Barkhof F, et al. Ocrelizumab in relapsing-remitting multiple sclerosis: a phase 2, randomised, placebo-controlled, multicentre trial. Lancet. 2011;378(9805):1779-87. Epub 2011/11/04.
206. Hauser SL, Waubant E, Arnold DL, Vollmer T, Antel J, Fox RJ, et al. B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. The New England journal of medicine. 2008;358(7):676-88. Epub 2008/02/15.
207. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447(7146):799-816. Epub 2007/06/16.
208. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473(7345):43-9. Epub 2011/03/29.
209. Sandve GK, Gundersen S, Rydbeck H, Glad IK, Holden L, Holden M, et al. The Genomic HyperBrowser: inferential genomics at the sequence level. Genome biology. 2010;11(12):R121. Epub 2010/12/25.
210. Disanto G, Morahan JM, Barnett MH, Giovannoni G, Ramagopalan SV. The evidence for a role of B cells in multiple sclerosis. Neurology. 2012;78(11):823-32. Epub 2012/03/14.
211. De Jager PL, Baecher-Allan C, Maier LM, Arthur AT, Ottoboni L, Barcellos L, et al. The role of the CD58 locus in multiple sclerosis. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(13):5264-9. Epub 2009/02/25.

212. Bar-Or A, Fawaz L, Fan B, Darlington PJ, Rieger A, Ghorayeb C, et al. Abnormal B-cell cytokine responses a trigger of T-cell-mediated disease in MS? Annals of neurology. 2010;67(4):452-61. Epub 2010/05/04.
213. Piccio L, Naismith RT, Trinkaus K, Klein RS, Parks BJ, Lyons JA, et al. Changes in B- and T-lymphocyte and chemokine levels with rituximab treatment in multiple sclerosis. Archives of neurology. 2010;67(6):707-14. Epub 2010/06/19.
214. Schneider R, Mohebiany AN, Ifergan I, Beauseigle D, Duquette P, Prat A, et al. B cell-derived IL-15 enhances CD8 T cell cytotoxicity and is increased in multiple sclerosis patients. J Immunol. 2011;187(8):4119-28. Epub 2011/09/14.
215. Disanto G, Sandve GK, Berlanga-Taylor AJ, Morahan JM, Dobson R, Giovannoni G, et al. Genomic regions associated with multiple sclerosis are active in B cells. PloS one. 2012;7(3):e32281. Epub 2012/03/08.
216. Chen S, Sims GP, Chen XX, Gu YY, Chen S, Lipsky PE. Modulatory effects of 1,25 - dihydroxyvitamin D3 on human B cell differentiation. Journal of Immunology. 2007 ; 179 (3) : 1634-47.
217. Salzer J, Hallmans G, Nystrom M, Stenlund H, Wadell G, Sundstrom P. Vitamin D as a protective factor in multiple sclerosis. Neurology. 2012;79(21):2140-5. Epub 2012/11/22.
218. Simpson S, Jr., Taylor B, Blizzard L, Ponsonby AL, Pittas F, Tremlett H, et al. Higher 25-hydroxyvitamin D is associated with lower relapse risk in multiple sclerosis. Annals of neurology. 2010;68(2):193-203. Epub 2010/08/10.
219. Mowry EM, Waubant E, McCulloch CE, Okuda DT, Evangelista AA, Lincoln RR, et al. Vitamin D status predicts new brain magnetic resonance imaging activity in multiple sclerosis. Annals of neurology. 2012;72(2):234-40. Epub 2012/08/29.
220. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190-5. Epub 2012/09/08.
221. Lincoln MR, Montpetit A, Cader MZ, Saarela J, Dyment DA, Tiislar M, et al. A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis. Nature genetics. 2005;37(10):1108-12. Epub 2005/09/28.
222. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. American journal of human genetics. 2011;88(1):76-82. Epub 2010/12/21.
223. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. American journal of human genetics. 2011;88(3):294-305. Epub 2011/03/08.
224. Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, Purcell SM, Sklar P, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS genetics. 2009;5(6):e1000534. Epub 2009/06/27.
225. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(16):6062-7. Epub 2004/04/13.
226. Watson CT, Disanto G, Breden F, Giovannoni G, Ramagopalan SV. Estimating the proportion of variation in susceptibility to multiple sclerosis captured by common SNPs. Scientific reports. 2012;2:770. Epub 2012/10/30.
227. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119-24. Epub 2012/11/07.
228. Modlin RL. The innate immune response in leprosy. Current opinion in immunology. 2010;22(1):48-54. Epub 2010/01/12.

229. Piper M, Beneyto M, Burne TH, Eyles DW, Lewis DA, McGrath JJ. The neurodevelopmental hypothesis of schizophrenia: convergent clues from epidemiology and neuropathology. The Psychiatric clinics of North America. 2012;35(3):571-84. Epub 2012/08/30.
230. Klein L, Hinterberger M, Wirnsberger G, Kyewski B. Antigen presentation in the thymus for positive selection and central tolerance induction. Nature reviews Immunology. 2009;9(12):833-44. Epub 2009/11/26.
231. Babbe H, Roers A, Waisman A, Lassmann H, Goebels N, Hohlfeld R, et al. Clonal expansions of CD8(+) T cells dominate the T cell infiltrate in active multiple sclerosis lesions as shown by micromanipulation and single cell polymerase chain reaction. The Journal of experimental medicine. 2000;192(3):393-404. Epub 2000/08/10.
232. Skulina C, Schmidt S, Dornmair K, Babbe H, Roers A, Rajewsky K, et al. Multiple sclerosis: brain-infiltrating CD8+ T cells persist as clonal expansions in the cerebrospinal fluid and blood. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(8):2428-33. Epub 2004/02/26.
233. Angelini DF, Serafini B, Piras E, Severa M, Coccia EM, Rosicarelli B, et al. Increased CD8+ T Cell Response to Epstein-Barr Virus Lytic Antigens in the Active Phase of Multiple Sclerosis. PLoS pathogens. 2013;9(4):e1003220. Epub 2013/04/18.

Supplementary table 1: List of studies assessing EBV seropositivity in MS cases and controls

STUDY		LOCATION		CASES		CONTROLS		EBV DETECTION	
Author	**Year**	**City and Country**	**Latitude**	**Total**	**EBV+**	**Total**	**EBV+**	**Method**	**Antibodies tested**
Sumaya	1980	Los Angeles (US)	34 N	157	0.987	81	0.938	IIF	VCA
Bray	1983	Salt lake city (US)	40 N	313	0.987	406	0.894	IIF	VCA IgG
Larsen	1985	Salt lake city (US)	40 N	93	1.000	93	0.839	IIF + ACIF	VCA, EBNA
Sumaya	1985	Los Angeles (US)	34 N	104	1.000	104	0.952	IIF + ACIF	VCA, EA, EBNA
Ferrante	1987	Italy	42 N	30	0.967	42	0.738	-	-
Shirodaria	1987	Belfast (Northern Ireland)	54 N	26	1.000	26	0.923	IIF + ACIF	VCA IgG, EBNA
Munch	1997	Aarhus (Denmark)	56 N	138	0.993	138	0.899	ELISA	EBNA-1, EA IgG and IgM
Myhr	1998	Hordaland (Norway)	60 N	144	1.000	170	0.953	IIF + ELISA	VCA IgM and IgG, EBNA IgG and EA IgG
Wagner	2000	Lubeck (Germany)	53 N	107	1.000	163	0.939	ELISA	EBNA IgG, EA IgG and IgM
Ascherio	2001	USA	-	144	0.993	287	0.937	IIF + ACIF	VCA IgG, EA IgG, EBNA-complex, EBNA1 and EBNA2 IgG
Haahr	2004	Aarhus (Denmark)	56 N	153	1.000	53	0.943	ELISA	EBNA-1 IgG and EA IgM
Sundstrom	2004	Västerbotten (Sweden)	65 N	234	1.000	702	0.987	IIF + ELISA	EBNA-1 IgG and VCA IgG
Ponsonby	2005	Tasmania (Australia)	42 S	136	1.000	261	0.966	ELISA	EBNA IgG and VCA IgG
Gronen	2006	Würzburg (Germany)	49 N	24	1.000	29	0.897	IIF + ELISA	VCA IgG, EBNA IgG and EA IgG
Riverol	2007	Pamplona (Spain)	42 N	150	0.980	85	0.882	ELISA	EBNA-1 IgG, EA IgG
Jilek	2008	Lausanne (Switzerland)	46 N	70	1.000	20	1.000	ELISPOT + ELISA	Cellular response + EBNA IgG, VCA IgG
Zivadinov	2009	Trieste (Italy)	45 N	140	0.950	131	1.000	ELISA	VCA IgG
Ingram	2010	Cardiff (Wales)	51 N	75	0.950	25	0.760	ELISA	EBNA-1 IgG
Lindsey	2010	Houston (US)	29 N	80	0.975	80	0.925	ELISA	EBNA-1 IgG
Levin	2010	USA	-	305	1.000	610	0.963	IIF + ACIF	VCA IgG, EBNA-complex, EBNA1 and EBNA2 IgG
Khaki	2011	Arak (Iran)	34 N	61	0.803	60	0.650	ELISA	Anti-EBV IgG and IgM
Villegas	2011	Granada (Spain)	37 N	76	0.868	75	0.827	ELISA	EBNA-1 IgG

IIF= indirect immuno fluorescence

ELISA= enzyme-linked immuno sorbent assay

ACIF= anti complement immunofluorescence

Supplementary table 2: Pre and post vitamin D supplementation anti EBNA levels

	EBNA1 antibody		
Individual	**Pre-supplementation**	**Post-supplementation**	**Difference**
1	95.438	57.466	37.972
2	140.102	95.618	44.484
3	113.984	71.16	42.824
4	99.186	116.754	-17.568
5	22.824	36.496	-13.672
6	33.91	30.328	3.582
7	81.162	37.114	44.048
8	117.418	113.302	4.116
9	69.642	52.982	16.66
10	84.282	54.396	29.886
11	106.304	77.386	28.918
12	51.816	48.924	2.892
13	96.84	68.662	28.178
14	26.078	34.804	-8.726
15	30.752	38.602	-7.85

Supplementary table 3: List of MS associated SNPs stratified by chromatin state

Chromosome	Position	SNPs (primary SNP in yellow)	Candidate Gene	Chromatin state
1	2527324	rs4074789	MMEL1(TNFRSF14)	H
1	2527698	rs41300092	MMEL1(TNFRSF14)	H
1	2528069	rs10910105	MMEL1(TNFRSF14)	H
1	2528209	rs10797436	MMEL1(TNFRSF14)	H
1	2528746	rs10910106	MMEL1(TNFRSF14)	H
1	2532513	rs10910111	MMEL1(TNFRSF14)	H
1	2534500	rs10752748	MMEL1(TNFRSF14)	H
1	2536089	rs6684865	MMEL1(TNFRSF14)	H
1	2541520	rs10797441	MMEL1(TNFRSF14)	H
1	2543484	rs3890745	MMEL1(TNFRSF14)	H
1	2543618	rs4474198	MMEL1(TNFRSF14)	H
1	2545500	rs28568531	MMEL1(TNFRSF14)	H
1	2545985	rs7368072	MMEL1(TNFRSF14)	H
1	2546408	rs7368105	MMEL1(TNFRSF14)	H
1	2546569	rs7364820	MMEL1(TNFRSF14)	H
1	2549449	rs12730932	MMEL1(TNFRSF14)	H
1	2551146	rs28532547	MMEL1(TNFRSF14)	H
1	2553110	rs12749591	MMEL1(TNFRSF14)	H
1	2554043	rs6698817	MMEL1(TNFRSF14)	H
1	2554325	rs12752515	MMEL1(TNFRSF14)	H
1	2557360	rs6666788	MMEL1(TNFRSF14)	H
1	2559883	rs4648565	MMEL1(TNFRSF14)	H
1	2561683	rs28690427	MMEL1(TNFRSF14)	H
1	2563938	rs9970196	MMEL1(TNFRSF14)	AP
1	2688810	rs6666430	MMEL1(TNFRSF14)	H
1	2692967	rs6657596	MMEL1(TNFRSF14)	H
1	2699024	rs4648356	MMEL1(TNFRSF14)	H
1	2706484	rs897628	MMEL1(TNFRSF14)	H
1	92913252	rs11809055	EVI5	H
1	92920965	rs11810217	EVI5	H
1	92925852	rs111807611	EVI5	H
1	92926683	rs35612297	EVI5	H
1	92928783	rs2051841	EVI5	H
1	92930692	rs72724574	EVI5	WT
1	92932826	rs11577426	EVI5	WT
1	92944512	rs12749786	EVI5	H
1	92953113	rs12749480	EVI5	H
1	92965326	rs12743005	EVI5	H
1	101104124	rs12048904	2nd hit in VCAM1	H
1	101165997	rs58029001	VCAM1	ST
1	101169721	rs56402739	VCAM1	WT
1	101180107	rs11581062	VCAM1	ST
1	101181521	rs3850453	VCAM1	ST
1	101182560	rs6577216	VCAM1	WT
1	101189459	rs11582211	VCAM1	ST
1	101191157	rs6675403	VCAM1	ST
1	101194997	rs17525451	VCAM1	WT
1	101197149	rs6693456	VCAM1	WT
1	116902480	rs1335532	CD58	SE
1	116904172	rs1414273	CD58	AP
1	116905738	rs2300747	CD58	AP
1	190786488	rs1323298	RGS1	SE
1	190790892	rs2816305	RGS1	SE
1	190791899	rs2760521	RGS1	WE
1	190802176	rs1359063	RGS1	H
1	190803008	rs2816317	RGS1	H
1	190803381	rs2760527	RGS1	H
1	190803408	rs2760528	RGS1	H
1	190803436	rs2816316	RGS1	H

1	190803509	rs2816315	RGS1	H
1	190804023	rs1323297	RGS1	H
1	190804131	rs1323296	RGS1	H
1	190807644	rs1323292	RGS1	H
1	190810460	rs1547624	RGS1	WE
1	190811418	rs2984920	RGS1	AP
1	190811722	rs7535818	RGS1	AP
1	199140852	rs35730213	C1orf106(KIF21B)	WT
1	199141351	rs55838263	C1orf106(KIF21B)	WE
1	199141718	rs12132298	C1orf106(KIF21B)	WE
1	199141865	rs12132349	C1orf106(KIF21B)	WE
1	199144185	rs7554511	C1orf106(KIF21B)	WE
1	199144473	rs41299637	C1orf106(KIF21B)	WT
1	199145350	rs12131796	C1orf106(KIF21B)	WT
1	199148015	rs10800746	C1orf106(KIF21B)	H
1	199148218	rs7522462	C1orf106(KIF21B)	H
2	43212565	rs12466022	-	WE
2	43212631	rs12471190	-	WE
2	43213569	rs28498283	-	SE
2	68500040	rs12622670	PLEK	WE
2	68500287	rs7592330	PLEK	WE
2	68500505	rs7592560	PLEK	WE
2	68500599	rs7595037	PLEK	WE
2	112373248	rs6729037	MERTK	PP
2	112375250	rs1516629	MERTK	I
2	112376009	rs2001444	MERTK	H
2	112377577	rs17835605	MERTK	H
2	112377918	rs55785997	MERTK	H
2	112381551	rs79559916	MERTK	H
2	112381672	rs17174870	MERTK	H
2	112382567	rs55728465	MERTK	H
2	112383393	rs55708724	MERTK	H
2	112384675	rs75069292	MERTK	H
2	112385115	rs12477716	MERTK	H
2	112386421	rs7568632	MERTK	H
2	112386517	rs1400321	MERTK	H
2	112387203	rs58641606	MERTK	H
2	112390105	rs7575552	MERTK	H
2	112391276	rs80088736	MERTK	H
2	136692725	rs882300	CXCR4	PR
2	230796328	rs13384787	SP140	WE
2	230813890	rs10198539	SP140	ST
2	230814968	rs10201872	SP140	ST
2	230815179	rs10202244	SP140	ST
2	230818826	rs28445040	SP140	ST
2	230824034	rs9989746	SP140	WT
3	27763784	rs11129295	EOMES	H
3	27766618	rs12330493	EOMES	H
3	28046448	rs669607	-	H
3	107041527	rs2028597	CBLB	ST
3	107043568	rs78795990	CBLB	ST
3	107050729	rs76314485	CBLB	ST
3	107058747	rs75471618	CBLB	ST
3	120702624	rs2293370	TMEM39A(CD80)	WP
3	120705146	rs1131265	TMEM39A(CD80)	ST
3	120711198	rs9843355	TMEM39A(CD80)	ST
3	120720488	rs62264485	TMEM39A(CD80)	ST
3	120727283	rs57271503	TMEM39A(CD80)	ST
3	120727734	rs13092998	TMEM39A(CD80)	ST
3	120734898	rs34844599	TMEM39A(CD80)	ST
3	123143354	rs4285028	2nd hit in CD86	ST
3	123275877	rs4308217	3rd hit in CD86	SE

3	123278779	rs11922347	CD86	SE
3	123279458	rs9282641	CD86	AP
3	161189880	rs7615589	IL12A	SE
3	161192345	rs2243123	IL12A	WP
3	161197241	rs2243140	IL12A	WT
3	161198105	rs2243148	IL12A	WT
4	103791368	rs228623	NFKB1	ST
4	103791799	rs228625	NFKB1	WT
4	103792130	rs228627	NFKB1	WT
4	103796513	rs228613	NFKB1	WT
4	103797685	rs228614	NFKB1	ST
4	103798508	rs228615	NFKB1	ST
4	103798739	rs228616	NFKB1	ST
4	103806025	rs227361	NFKB1	ST
5	35887588	rs1494564	IL7R	WE
5	35888068	rs6890853	IL7R	SE
5	35893461	rs11567694	IL7R	SE
5	35893607	rs10213865	IL7R	WE
5	35896909	rs11567705	IL7R	WT
5	35898598	rs7717955	IL7R	WE
5	35910332	rs6897932	IL7R	WT
5	35914852	rs6881270	IL7R	WT
5	35914913	rs6881706	IL7R	WT
5	35917133	rs11742240	IL7R	WT
5	35917200	rs11742270	IL7R	WT
5	40355634	rs1445002	PTGER4	H
5	40356444	rs73099712	PTGER4	H
5	40357238	rs73099715	PTGER4	H
5	40362541	rs12374574	PTGER4	H
5	40366841	rs348603	PTGER4	H
5	40367121	rs348604	PTGER4	H
5	40368755	rs73099728	PTGER4	H
5	40373021	rs114616683	PTGER4	H
5	40380421	rs13183025	PTGER4	H
5	40382448	rs73099741	PTGER4	H
5	40383226	rs6889364	PTGER4	H
5	40389945	rs73082840	PTGER4	H
5	40396215	rs11743463	PTGER4	H
5	40411916	rs56244034	PTGER4	WE
5	40413623	rs17227583	PTGER4	WE
5	40419818	rs895123	PTGER4	WP
5	40425044	rs1124233	PTGER4	H
5	40425609	rs12187530	PTGER4	H
5	40427983	rs2371685	PTGER4	H
5	40428485	rs4613763	PTGER4	H
5	40429958	rs10512735	PTGER4	H
5	40430428	rs6897633	PTGER4	H
5	40432182	rs11749040	PTGER4	H
5	40433947	rs113957080	PTGER4	H
5	40435777	rs11738106	PTGER4	H
5	40437266	rs17234657	PTGER4	WE
5	40437990	rs6879283	PTGER4	WE
5	40438434	rs6883975	PTGER4	WT
5	40438950	rs1373694	PTGER4	WT
5	40442742	rs6896243	PTGER4	SE
5	40457379	rs115948982	PTGER4	H
5	158692478	rs2546890	IL12B	WE
6	91044593	rs2174281	BACH2	PR
6	91053070	rs4142967	BACH2	WE
6	91053490	rs12212193	BACH2	I
6	128320491	rs802734	THEMIS	H
6	135781048	rs11154801	MYB(AHI1)	H

6	137494601	rs17066096	IL22RA2	H
6	138002706	rs17066662	-	WT
6	138004746	rs12524433	-	I
6	138007111	rs13207033	-	WT
6	138007580	rs4896289	-	WT
6	138008553	rs34810205	-	WT
6	138008907	rs13192841	-	WT
6	138008945	rs12527282	-	WT
6	138010292	rs2056224	-	WE
6	138010330	rs2056223	-	WE
6	138013551	rs13199053	-	H
6	138016529	rs66469519	-	H
6	138022567	rs67178054	-	WT
6	138039025	rs4896295	-	H
6	138040853	rs34996815	-	WE
6	138044330	rs10499194	-	WE
6	138047121	rs13205649	-	WT
6	138050372	rs66499821	-	WT
6	138051593	rs34654849	-	H
6	138052844	rs12525643	-	H
6	159385965	rs1738074	TAGAP	AP
6	159388553	rs212411	TAGAP	WT
6	159389562	rs182429	TAGAP	H
7	148855291	rs58026905	ZNF746	H
7	148877613	rs3735320	ZNF746	ST
7	148896661	rs56707870	ZNF746	ST
7	148910553	rs11766910	ZNF746	ST
7	148916854	rs354031	ZNF746	ST
7	148920397	rs354033	ZNF746	ST
7	148921430	rs354034	ZNF746	ST
7	148928037	rs354042	ZNF746	SE
7	148928265	rs354043	ZNF746	SE
7	148931541	rs354045	ZNF746	ST
8	79544710	rs10808825	IL7	H
8	79546415	rs11985754	IL7	H
8	79555183	rs6990902	IL7	H
8	79563593	rs1520333	IL7	H
8	79571297	rs6982372	IL7	H
8	79571904	rs10086475	IL7	H
8	79885663	rs2587156	Previously reported IL7	WT
8	128878739	rs6470578	MYC	AP
8	128884211	rs4410871	MYC	ST
8	129236322	rs13249280	PVT1	H
8	129239778	rs4733839	PVT1	H
8	129239958	rs4733840	PVT1	H
8	129240264	rs13255672	PVT1	H
8	129241743	rs36111777	PVT1	H
8	129242943	rs55886082	PVT1	H
8	129243466	rs4733842	PVT1	H
8	129244227	rs13276381	PVT1	H
8	129244642	rs4733843	PVT1	H
8	129245193	rs7826401	PVT1	H
8	129245841	rs7017155	PVT1	H
8	129246952	rs11989574	PVT1	H
8	129248272	rs13280095	PVT1	H
8	129248278	rs13280365	PVT1	H
8	129248970	rs7830477	PVT1	H
8	129249491	rs36099148	PVT1	WE
8	129249512	rs34746976	PVT1	WE
8	129249889	rs1995218	PVT1	WE
8	129250829	rs4733598	PVT1	WT
8	129252026	rs7016362	PVT1	WT

8	129252266	rs16902685	PVT1	WT
8	129254581	rs6989829	PVT1	WE
8	129255133	rs6470605	PVT1	WE
8	129255178	rs6470606	PVT1	WE
8	129255251	rs6470607	PVT1	WE
8	129255607	rs7825733	PVT1	WE
8	129255730	rs7826019	PVT1	WE
8	129255818	rs7826413	PVT1	WE
8	129256930	rs6987197	PVT1	WE
8	129257152	rs6987519	PVT1	WE
8	129257607	rs4733844	PVT1	WT
8	129257639	rs4733845	PVT1	WT
8	129259726	rs12156002	PVT1	WT
8	129261453	rs2019960	PVT1	WT
8	129262797	rs11997042	PVT1	WT
8	129267472	rs59602790	PVT1	WP
8	129268067	rs16902700	PVT1	WE
8	129271758	rs34409623	PVT1	WT
8	129273373	rs6996790	PVT1	WT
8	129279508	rs62527078	PVT1	WT
10	6140731	rs3134883	IL2RA	WP
10	6141719	rs3118470	IL2RA	WE
10	6150835	rs7090512	2nd hit in IL2RA	WE
10	80730323	rs1250550	ZMIZ1	ST
10	94459960	rs11187144	HHEX	H
10	94461575	rs4933736	HHEX	WE
10	94463609	rs7087591	HHEX	WT
10	94464077	rs10786053	HHEX	WT
10	94467199	rs10748582	HHEX	PR
10	94469087	rs1112718	HHEX	PR
10	94471897	rs7923837	HHEX	PR
10	94472056	rs7923866	HHEX	PR
11	60588858	rs650258	CD6	WT
11	118259563	rs630923	CXCR5	PP
12	6310270	rs1800693	TNFRSF1A	SE
12	6317038	rs1860545	TNFRSF1A	SE
12	9766166	rs10844597	CLECL1	SE
12	9767358	rs10466829	CLECL1	SE
12	9767617	rs11052710	CLECL1	SE
12	9768035	rs10743819	CLECL1	AP
12	9768144	rs10772085	CLECL1	AP
12	9769090	rs7307111	CLECL1	AP
12	9769850	rs2401391	CLECL1	AP
12	9769922	rs7976584	CLECL1	AP
12	9770351	rs2401393	CLECL1	AP
12	9770502	rs2192437	CLECL1	AP
12	9770635	rs10844617	CLECL1	AP
12	9770693	rs2401394	CLECL1	AP
12	9770807	rs2401395	CLECL1	AP
12	9771381	rs7970116	CLECL1	AP
12	9771708	rs7973638	CLECL1	AP
12	9776349	rs10844622	CLECL1	AP
12	9777266	rs10492166	CLECL1	AP
12	9780424	rs7956831	CLECL1	SE
12	56419523	rs12368653	CYP27B1	WE
12	56421687	rs12371356	CYP27B1	AP
12	122013881	rs4275659	ARL6IP4	WT
12	122016971	rs61955196	ARL6IP4	WP
12	122026915	rs884956	ARL6IP4	WP
12	122030232	rs3759115	ARL6IP4	AP
12	122032064	rs55742290	ARL6IP4	AP
12	122067925	rs12425850	ARL6IP4	H

12	122086024	rs11613937	ARL6IP4	H
12	122095009	rs11608811	ARL6IP4	PR
12	122107559	rs10773921	ARL6IP4	I
12	122110831	rs7957096	ARL6IP4	WT
12	122134572	rs4148862	ARL6IP4	SE
12	122135328	rs4148863	ARL6IP4	WE
12	122138448	rs10848428	ARL6IP4	WT
12	122141695	rs1727307	ARL6IP4	WE
12	122152641	rs1790094	ARL6IP4	H
12	122157549	rs1790106	ARL6IP4	WE
12	122161116	rs949143	ARL6IP4	H
12	122170445	rs1790121	ARL6IP4	WE
12	122172692	rs1790122	ARL6IP4	H
14	68320644	rs12435329	ZFP36L1	WT
14	68323096	rs11158763	ZFP36L1	SE
14	68323117	rs12434551	ZFP36L1	SE
14	68323944	rs4902647	ZFP36L1	SE
14	75071810	rs2359737	BATF	ST
14	75075310	rs2300603	BATF	SE
14	87464613	rs7141097	GALC(GPR65)	WT
14	87476282	rs112634971	GALC(GPR65)	WE
14	87483638	rs17123925	GALC(GPR65)	WT
14	87487851	rs8019944	GALC(GPR65)	WE
14	87489111	rs2301123	GALC(GPR65)	WE
14	87495397	rs116150779	GALC(GPR65)	WT
14	87496271	rs8015102	GALC(GPR65)	WT
14	87497777	rs80094596	GALC(GPR65)	WT
14	87502745	rs7141758	GALC(GPR65)	WT
14	87516953	rs17203370	GALC(GPR65)	WT
14	87529548	rs73312836	GALC(GPR65)	AP
14	87530522	rs8007714	GALC(GPR65)	SE
14	87532614	rs73312845	GALC(GPR65)	WE
14	87532759	rs73312850	GALC(GPR65)	WE
14	87533141	rs8003806	GALC(GPR65)	WT
14	87533845	rs57564449	GALC(GPR65)	WT
14	87536592	rs58824819	GALC(GPR65)	WT
14	87536980	rs117262875	GALC(GPR65)	WT
14	87538233	rs4418995	GALC(GPR65)	WE
14	87542931	rs1372363	GALC(GPR65)	AP
14	87544206	rs68177277	GALC(GPR65)	WP
14	87550770	rs7156702	GALC(GPR65)	WT
14	87551047	rs7156254	GALC(GPR65)	WT
14	87551745	rs58961470	GALC(GPR65)	WT
14	87552685	rs74929665	GALC(GPR65)	WT
14	87553945	rs7152721	GALC(GPR65)	WT
14	87554297	rs8016427	GALC(GPR65)	WT
14	87557442	rs2119704	GALC(GPR65)	WT
14	87557842	rs56972319	GALC(GPR65)	WT
14	87559576	rs59761017	GALC(GPR65)	WT
14	87559964	rs74074328	GALC(GPR65)	WE
14	87567480	rs67412305	GALC(GPR65)	H
14	87569798	rs17123978	GALC(GPR65)	H
14	87570398	rs60623686	GALC(GPR65)	H
14	87570899	rs66469373	GALC(GPR65)	H
14	87570953	rs56168158	GALC(GPR65)	H
14	87571482	rs10220369	GALC(GPR65)	H
14	87571605	rs10220371	GALC(GPR65)	H
14	87572918	rs66475691	GALC(GPR65)	H
14	87575490	rs59246400	GALC(GPR65)	H
14	87575495	rs58839614	GALC(GPR65)	H
14	87576031	rs7157875	GALC(GPR65)	H
14	87577583	rs28424269	GALC(GPR65)	H
14	87577695	rs28645517	GALC(GPR65)	H

14	87582448	rs17123987	GALC(GPR65)	H
14	87583770	rs7151916	GALC(GPR65)	H
14	87593775	rs57103834	GALC(GPR65)	H
14	87599632	rs74850043	GALC(GPR65)	H
14	87601968	rs60410555	GALC(GPR65)	H
14	87602495	rs59578734	GALC(GPR65)	H
14	87603654	rs73318427	GALC(GPR65)	H
14	87604985	rs1441810	GALC(GPR65)	H
14	87605194	rs8013837	GALC(GPR65)	H
14	87613716	rs17124024	GALC(GPR65)	H
14	87614954	rs57668430	GALC(GPR65)	H
14	87626072	rs76306664	GALC(GPR65)	H
16	1013373	rs2573148	SOX8	H
16	1013532	rs63194866	SOX8	H
16	1013553	rs2744148	SOX8	H
16	1014444	rs449295	SOX8	H
16	1014459	rs421245	SOX8	H
16	1014527	rs400354	SOX8	H
16	1014932	rs416055	SOX8	H
16	11085302	rs7200786	CLEC16A(CIITA)	WT
16	84567261	rs12232384	IRF8	H
16	84568534	rs13333054	IRF8	H
17	37761506	rs9891119	STAT3	SE
17	37773416	rs957970	STAT3	WT
17	37773451	rs957971	STAT3	WT
17	37783361	rs1026916	STAT3	SE
17	37786227	rs8068748	STAT3	ST
17	37787601	rs7211777	STAT3	WE
17	55379057	rs180515	RPS6KB1	ST
18	54504665	rs4940738	MALT1	SE
18	54511871	rs7239068	MALT1	ST
18	54522363	rs7230925	MALT1	ST
18	54522999	rs9946837	MALT1	ST
18	54526781	rs8088079	MALT1	ST
18	54526838	rs4940743	MALT1	ST
18	54534303	rs9959997	MALT1	SE
18	54535031	rs7234035	MALT1	ST
18	54535172	rs7238078	MALT1	ST
18	54535685	rs7238879	MALT1	ST
18	54540412	rs11873030	MALT1	ST
18	54546377	rs9966968	MALT1	ST
18	54550898	rs9963281	MALT1	ST
18	54551806	rs28740963	MALT1	ST
18	54553775	rs7236784	MALT1	ST
18	54557937	rs55653859	MALT1	ST
18	54573848	rs4545915	MALT1	WT
19	6619972	rs1077667	TNFSF14	AP
19	10380410	rs17697965	TYK2(ICAM3)	WP
19	10381064	rs8112449	TYK2(ICAM3)	WE
19	10381891	rs12978984	TYK2(ICAM3)	WE
19	10382548	rs35951601	TYK2(ICAM3)	SE
19	10383176	rs7253917	TYK2(ICAM3)	SE
19	10387831	rs35916055	TYK2(ICAM3)	SE
19	18158608	rs62120363	MPV17L2(IL12RB1)	WT
19	18158656	rs62120364	MPV17L2(IL12RB1)	WT
19	18163239	rs55990904	MPV17L2(IL12RB1)	SE
19	18163261	rs56128769	MPV17L2(IL12RB1)	SE
19	18165700	rs874628	MPV17L2(IL12RB1)	AP
19	54560863	rs2303759	DKKL1(CD37)	H
20	44114260	rs12624433	CD40	H
20	44126005	rs13037326	CD40	SE
20	44135527	rs2425752	CD40	ST

20	52223150	rs2248137	CYP24A1	PR
20	52224925	rs2248359	CYP24A1	PR
20	61880157	rs6062314	TNFRSF6B	H
22	20436618	rs5749806	MAPK1	H
22	20445004	rs6928	MAPK1	ST
22	20445498	rs3810610	MAPK1	ST
22	20446202	rs13943	MAPK1	ST
22	20458678	rs2006893	MAPK1	ST
22	20461125	rs2283792	MAPK1	WT
22	20475101	rs5999515	MAPK1	WT
22	20476399	rs5999521	MAPK1	ST
22	20486789	rs2266967	MAPK1	ST
22	20487661	rs5749986	MAPK1	ST
22	20493425	rs9607287	MAPK1	WT
22	20501567	rs3788332	MAPK1	SE
22	20515347	rs6518986	MAPK1	WT
22	20516479	rs1892846	MAPK1	WT
22	20516562	rs1892848	MAPK1	WT
22	20518009	rs5999750	MAPK1	WT
22	49318132	rs140522	SCO2	AP

Supplementary table 4: List of GWAS SNPs from the GWAS catolog

PUBMEDID	First Author	Journal	SNPs	Disease
21833088	Sawcer S	Nature	rs7200786	Multiple sclerosis
21833088	Sawcer S	Nature	rs10466829	Multiple sclerosis
21833088	Sawcer S	Nature	rs630923	Multiple sclerosis
21833088	Sawcer S	Nature	rs233100	Multiple sclerosis
21833088	Sawcer S	Nature	rs2303759	Multiple sclerosis
21833088	Sawcer S	Nature	rs11129295	Multiple sclerosis
21833088	Sawcer S	Nature	rs11810217	Multiple sclerosis
21833088	Sawcer S	Nature	rs12048904	Multiple sclerosis
21833088	Sawcer S	Nature	rs2119704	Multiple sclerosis
21833088	Sawcer S	Nature	rs7923837	Multiple sclerosis
21833088	Sawcer S	Nature	rs2546890	Multiple sclerosis
21833088	Sawcer S	Nature	rs17066096	Multiple sclerosis
21833088	Sawcer S	Nature	rs6897932	Multiple sclerosis
21833088	Sawcer S	Nature	rs13333054	Multiple sclerosis
21833088	Sawcer S	Nature	rs17594362	Multiple sclerosis
21833088	Sawcer S	Nature	rs7238078	Multiple sclerosis
21833088	Sawcer S	Nature	rs281380	Multiple sclerosis
21833088	Sawcer S	Nature	rs228614	Multiple sclerosis
21833088	Sawcer S	Nature	rs2283792	Multiple sclerosis
21833088	Sawcer S	Nature	rs17174870	Multiple sclerosis
21833088	Sawcer S	Nature	rs4648356	Multiple sclerosis
21833088	Sawcer S	Nature	rs2248359	Multiple sclerosis
21833088	Sawcer S	Nature	rs806321	Multiple sclerosis
21833088	Sawcer S	Nature	rs3761959	Multiple sclerosis
21833088	Sawcer S	Nature	rs2243123	Multiple sclerosis
21833088	Sawcer S	Nature	rs7090512	Multiple sclerosis
21833088	Sawcer S	Nature	rs386965	Multiple sclerosis
21833088	Sawcer S	Nature	rs4792814	Multiple sclerosis
21833088	Sawcer S	Nature	rs874628	Multiple sclerosis
21833088	Sawcer S	Nature	rs6062314	Multiple sclerosis
21833088	Sawcer S	Nature	rs793108	Multiple sclerosis
21833088	Sawcer S	Nature	rs9321490	Multiple sclerosis
21833088	Sawcer S	Nature	rs4410871	Multiple sclerosis
21833088	Sawcer S	Nature	rs12466022	Multiple sclerosis
21833088	Sawcer S	Nature	rs669607	Multiple sclerosis
21833088	Sawcer S	Nature	rs13192841	Multiple sclerosis
21833088	Sawcer S	Nature	rs2425752	Multiple sclerosis
21833088	Sawcer S	Nature	rs1062158	Multiple sclerosis
21833088	Sawcer S	Nature	rs140522	Multiple sclerosis
21833088	Sawcer S	Nature	rs1520333	Multiple sclerosis
21833088	Sawcer S	Nature	rs9821630	Multiple sclerosis
21833088	Sawcer S	Nature	rs7595037	Multiple sclerosis
21833088	Sawcer S	Nature	rs4613763	Multiple sclerosis
21833088	Sawcer S	Nature	rs7255066	Multiple sclerosis
21833088	Sawcer S	Nature	rs2019960	Multiple sclerosis
21833088	Sawcer S	Nature	rs4075958	Multiple sclerosis
21833088	Sawcer S	Nature	rs180515	Multiple sclerosis
21833088	Sawcer S	Nature	rs11755724	Multiple sclerosis
21833088	Sawcer S	Nature	rs11581062	Multiple sclerosis
21833088	Sawcer S	Nature	rs2744148	Multiple sclerosis
21833088	Sawcer S	Nature	rs9891119	Multiple sclerosis
21833088	Sawcer S	Nature	rs290986	Multiple sclerosis
21833088	Sawcer S	Nature	rs1738074	Multiple sclerosis
21833088	Sawcer S	Nature	rs1800693	Multiple sclerosis
21833088	Sawcer S	Nature	rs1077667	Multiple sclerosis
21833088	Sawcer S	Nature	rs4902647	Multiple sclerosis
21833088	Sawcer S	Nature	rs1250550	Multiple sclerosis
21833088	Sawcer S	Nature	rs354033	Multiple sclerosis
21833088	Sawcer S	Nature	rs10936599	Multiple sclerosis
21833088	Sawcer S	Nature	rs4409785	Multiple sclerosis

21833088	Sawcer S	Nature	rs771767	Multiple sclerosis
21833088	Sawcer S	Nature	rs802734	Multiple sclerosis
21833088	Sawcer S	Nature	rs1323292	Multiple sclerosis
21833088	Sawcer S	Nature	rs307896	Multiple sclerosis
21833088	Sawcer S	Nature	rs10201872	Multiple sclerosis
21833088	Sawcer S	Nature	rs756699	Multiple sclerosis
21833088	Sawcer S	Nature	rs12048904	Multiple sclerosis
21833088	Sawcer S	Nature	rs4285028	Multiple sclerosis
21833088	Sawcer S	Nature	rs4308217	Multiple sclerosis
21833088	Sawcer S	Nature	rs3118470	Multiple sclerosis
21833088	Sawcer S	Nature	DRB*15:01	Multiple sclerosis
21833088	Sawcer S	Nature	DRB*13:03	Multiple sclerosis
21833088	Sawcer S	Nature	DRB*03:01	Multiple sclerosis
21833088	Sawcer S	Nature	A*02:01	Multiple sclerosis
21833088	Sawcer S	Nature	DRB*08:01	Multiple sclerosis
21833088	Sawcer S	Nature	rs12368653	Multiple sclerosis
21833088	Sawcer S	Nature	rs11154801	Multiple sclerosis
21833088	Sawcer S	Nature	rs12456021	Multiple sclerosis
21833088	Sawcer S	Nature	rs12212193	Multiple sclerosis
21833088	Sawcer S	Nature	rs2300603	Multiple sclerosis
21833088	Sawcer S	Nature	rs7522462	Multiple sclerosis
21833088	Sawcer S	Nature	rs2293370	Multiple sclerosis
21833088	Sawcer S	Nature	rs650258	Multiple sclerosis
21833088	Sawcer S	Nature	rs1335532	Multiple sclerosis
21833088	Sawcer S	Nature	rs9282641	Multiple sclerosis
21833088	Sawcer S	Nature	rs8112449	Multiple sclerosis
21833088	Sawcer S	Nature	rs6952809	Multiple sclerosis
23273568	Yang W	Am J Hum Genet	rs2785197	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs9270984	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs742108	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs10276619	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs729302	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs7574865	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs6705628	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs4852324	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs6804441	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs4948496	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs12822507	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs10845606	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs34330	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs4622329	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs10911390	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs11717455	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs12529935	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs2252996	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs12917712	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs2303745	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs4522865	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs2254546	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs877819	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs6590330	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs1385374	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs7329174	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs13385731	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs10036748	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs5754217	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs3734266	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs12599402	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs9937837	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs2230926	Systemic lupus erythematosus
23273568	Yang W	Am J Hum Genet	rs2205960	Systemic lupus erythematosus
20453842	Stahl EA	Nat Genet	rs840016	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs7155603	Rheumatoid arthritis

20453842	Stahl EA	Nat Genet	rs934734	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs13315591	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs3093023	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs10488631	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs11676922	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs706778	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs13119723	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs3184504	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs2872507	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs11203203	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs6910071	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs13031237	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs3087243	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs4750316	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs4810485	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs17374222	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs2476601	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs10865035	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs6859219	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs874040	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs12131057	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs6920220	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs951005	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs3890745	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs7574865	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs3761847	Rheumatoid arthritis
20453842	Stahl EA	Nat Genet	rs26232	Rheumatoid arthritis
21779181	Gorlova O	PLoS Genet	rs11642873	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs12540874	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs11047102	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs987870	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs3129763	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs10488631	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs10488631	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs11171747	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs10488631	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs3821236	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs2056626	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs9275390	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs443198	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs11047102	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs3129882	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs9296015	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs10488631	Systemic sclerosis
21779181	Gorlova O	PLoS Genet	rs9275390	Systemic sclerosis
19430480	Barrett JC	Nat Genet	rs11755527	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs2290400	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs7111341	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs12708716	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs229541	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs10517086	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs7020673	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs1465788	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs4788084	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs7221109	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs2664170	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs12444268	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs12251307	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs4505848	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs425105	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs9268645	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs7804356	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs16956936	Type 1 diabetes

19430480	Barrett JC	Nat Genet	rs2476601	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs1990760	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs3087243	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs11258747	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs2292239	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs3184504	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs3825932	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs1893217	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs11203203	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs3024505	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs9388489	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs4948088	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs10509540	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs4763879	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs4900384	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs7202877	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs2281808	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs5753037	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs2269241	Type 1 diabetes
19430480	Barrett JC	Nat Genet	rs1534422	Type 1 diabetes
21102463	Franke A	Nat Genet	rs3180018	Crohn's disease
21102463	Franke A	Nat Genet	rs4809330	Crohn's disease
21102463	Franke A	Nat Genet	rs2476601	Crohn's disease
21102463	Franke A	Nat Genet	rs1893217	Crohn's disease
21102463	Franke A	Nat Genet	rs11742570	Crohn's disease
21102463	Franke A	Nat Genet	rs694739	Crohn's disease
21102463	Franke A	Nat Genet	rs6568421	Crohn's disease
21102463	Franke A	Nat Genet	rs2076756	Crohn's disease
21102463	Franke A	Nat Genet	rs4409764	Crohn's disease
21102463	Franke A	Nat Genet	rs6738825	Crohn's disease
21102463	Franke A	Nat Genet	rs11167764	Crohn's disease
21102463	Franke A	Nat Genet	rs6908425	Crohn's disease
21102463	Franke A	Nat Genet	rs3091315	Crohn's disease
21102463	Franke A	Nat Genet	rs3792109	Crohn's disease
21102463	Franke A	Nat Genet	rs6651252	Crohn's disease
21102463	Franke A	Nat Genet	rs4871611	Crohn's disease
21102463	Franke A	Nat Genet	rs359457	Crohn's disease
21102463	Franke A	Nat Genet	rs4656940	Crohn's disease
21102463	Franke A	Nat Genet	rs415890	Crohn's disease
21102463	Franke A	Nat Genet	rs4077515	Crohn's disease
21102463	Franke A	Nat Genet	rs10181042	Crohn's disease
21102463	Franke A	Nat Genet	rs7554511	Crohn's disease
21102463	Franke A	Nat Genet	rs7927997	Crohn's disease
21102463	Franke A	Nat Genet	rs1847472	Crohn's disease
21102463	Franke A	Nat Genet	rs13073817	Crohn's disease
21102463	Franke A	Nat Genet	rs7702331	Crohn's disease
21102463	Franke A	Nat Genet	rs17309827	Crohn's disease
21102463	Franke A	Nat Genet	rs736289	Crohn's disease
21102463	Franke A	Nat Genet	rs1736020	Crohn's disease
21102463	Franke A	Nat Genet	rs3764147	Crohn's disease
21102463	Franke A	Nat Genet	rs11564258	Crohn's disease
21102463	Franke A	Nat Genet	rs713875	Crohn's disease
21102463	Franke A	Nat Genet	rs3197999	Crohn's disease
21102463	Franke A	Nat Genet	rs11871801	Crohn's disease
21102463	Franke A	Nat Genet	rs2413583	Crohn's disease
21102463	Franke A	Nat Genet	rs1799964	Crohn's disease
21102463	Franke A	Nat Genet	rs10758669	Crohn's disease
21102463	Franke A	Nat Genet	rs7714584	Crohn's disease
21102463	Franke A	Nat Genet	rs12722489	Crohn's disease
21102463	Franke A	Nat Genet	rs151181	Crohn's disease
21102463	Franke A	Nat Genet	rs11209026	Crohn's disease
21102463	Franke A	Nat Genet	rs2058660	Crohn's disease
21102463	Franke A	Nat Genet	rs6556412	Crohn's disease

21102463	Franke A	Nat Genet	rs2797685	Crohn's disease
21102463	Franke A	Nat Genet	rs3024505	Crohn's disease
21102463	Franke A	Nat Genet	rs1456896	Crohn's disease
21102463	Franke A	Nat Genet	rs2838519	Crohn's disease
21102463	Franke A	Nat Genet	rs2872507	Crohn's disease
21102463	Franke A	Nat Genet	rs740495	Crohn's disease
21102463	Franke A	Nat Genet	rs780093	Crohn's disease
21102463	Franke A	Nat Genet	rs8005161	Crohn's disease
21102463	Franke A	Nat Genet	rs281379	Crohn's disease
21102463	Franke A	Nat Genet	rs102275	Crohn's disease
21102463	Franke A	Nat Genet	rs2549794	Crohn's disease
21102463	Franke A	Nat Genet	rs13428812	Crohn's disease
21102463	Franke A	Nat Genet	rs1998598	Crohn's disease
21102463	Franke A	Nat Genet	rs12242110	Crohn's disease
21102463	Franke A	Nat Genet	rs10761659	Crohn's disease
21102463	Franke A	Nat Genet	rs1250550	Crohn's disease
21102463	Franke A	Nat Genet	rs4902642	Crohn's disease
21102463	Franke A	Nat Genet	rs181359	Crohn's disease
21102463	Franke A	Nat Genet	rs1819658	Crohn's disease
21102463	Franke A	Nat Genet	rs12720356	Crohn's disease
21102463	Franke A	Nat Genet	rs7517810	Crohn's disease
21102463	Franke A	Nat Genet	rs3810936	Crohn's disease
21102463	Franke A	Nat Genet	rs2062305	Crohn's disease
21102463	Franke A	Nat Genet	rs10495903	Crohn's disease
21102463	Franke A	Nat Genet	rs212388	Crohn's disease
21102463	Franke A	Nat Genet	rs7423615	Crohn's disease
21102463	Franke A	Nat Genet	rs17293632	Crohn's disease
21102463	Franke A	Nat Genet	rs12521868	Crohn's disease
21297633	Anderson CA	Nat Genet	rs6451493	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs6871626	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs6920220	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs12261843	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs907611	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs3024505	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs1297265	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs6499188	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs2872507	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs6017342	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs2836878	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs5771069	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs35675666	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs7524102	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs2310173	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs254560	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs4246905	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs798502	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs4728142	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs10758669	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs10781499	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs2155219	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs941823	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs11209026	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs17085007	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs2297441	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs2838519	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs7554511	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs6426833	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs1801274	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs7608910	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs9268853	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs4676406	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs9822268	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs17388568	Ulcerative colitis

21297633	Anderson CA	Nat Genet	rs11739663	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs4510766	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs6584283	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs7134599	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs734999	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs11676348	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs267939	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs3194051	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs943072	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs6911490	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs678170	Ulcerative colitis
21297633	Anderson CA	Nat Genet	rs16940202	Ulcerative colitis
20953190	Strange A	Nat Genet	rs27524	Psoriasis
20953190	Strange A	Nat Genet	rs240993	Psoriasis
20953190	Strange A	Nat Genet	rs12720356	Psoriasis
20953190	Strange A	Nat Genet	rs3213094	Psoriasis
20953190	Strange A	Nat Genet	rs610604	Psoriasis
20953190	Strange A	Nat Genet	rs8016947	Psoriasis
20953190	Strange A	Nat Genet	rs458017,rs465	Psoriasis
20953190	Strange A	Nat Genet	rs11209026	Psoriasis
20953190	Strange A	Nat Genet	rs2066808	Psoriasis
20953190	Strange A	Nat Genet	rs2235617,rs49	Psoriasis
20953190	Strange A	Nat Genet	rs6809854	Psoriasis
20953190	Strange A	Nat Genet	rs702873	Psoriasis
20953190	Strange A	Nat Genet	rs4649203	Psoriasis
20953190	Strange A	Nat Genet	rs17716942	Psoriasis
20953190	Strange A	Nat Genet	rs4112788	Psoriasis
20953190	Strange A	Nat Genet	rs280519	Psoriasis
20953190	Strange A	Nat Genet	rs10484554	Psoriasis
20190752	Dubois PC	Nat Genet	rs13151961	Celiac disease
20190752	Dubois PC	Nat Genet	rs859637	Celiac disease
20190752	Dubois PC	Nat Genet	rs653178	Celiac disease
20190752	Dubois PC	Nat Genet	rs3748816	Celiac disease
20190752	Dubois PC	Nat Genet	rs10903122	Celiac disease
20190752	Dubois PC	Nat Genet	rs6974491	Celiac disease
20190752	Dubois PC	Nat Genet	rs4899260	Celiac disease
20190752	Dubois PC	Nat Genet	rs296547	Celiac disease
20190752	Dubois PC	Nat Genet	rs13003464	Celiac disease
20190752	Dubois PC	Nat Genet	rs2327832	Celiac disease
20190752	Dubois PC	Nat Genet	rs17035378	Celiac disease
20190752	Dubois PC	Nat Genet	rs13314993	Celiac disease
20190752	Dubois PC	Nat Genet	rs11712165	Celiac disease
20190752	Dubois PC	Nat Genet	rs10806425	Celiac disease
20190752	Dubois PC	Nat Genet	rs802734	Celiac disease
20190752	Dubois PC	Nat Genet	rs1250552	Celiac disease
20190752	Dubois PC	Nat Genet	rs11221332	Celiac disease
20190752	Dubois PC	Nat Genet	rs4819388	Celiac disease
20190752	Dubois PC	Nat Genet	rs1033180	Celiac disease
20190752	Dubois PC	Nat Genet	rs2298428	Celiac disease
20190752	Dubois PC	Nat Genet	rs917997	Celiac disease
20190752	Dubois PC	Nat Genet	rs4675374	Celiac disease
20190752	Dubois PC	Nat Genet	rs13098911	Celiac disease
20190752	Dubois PC	Nat Genet	rs17810546	Celiac disease
20190752	Dubois PC	Nat Genet	rs2074404	Celiac disease
20190752	Dubois PC	Nat Genet	rs6806528	Celiac disease
20190752	Dubois PC	Nat Genet	rs12727642	Celiac disease
20190752	Dubois PC	Nat Genet	rs10936599	Celiac disease
20190752	Dubois PC	Nat Genet	rs2762051	Celiac disease
20190752	Dubois PC	Nat Genet	rs1893217	Celiac disease
20190752	Dubois PC	Nat Genet	rs9792269	Celiac disease
20190752	Dubois PC	Nat Genet	rs12928822	Celiac disease
20190752	Dubois PC	Nat Genet	rs2816316	Celiac disease
20190752	Dubois PC	Nat Genet	rs864537	Celiac disease

20190752	Dubois PC	Nat Genet	rs1738074	Celiac disease
20190752	Dubois PC	Nat Genet	rs5979785	Celiac disease
20190752	Dubois PC	Nat Genet	rs13010713	Celiac disease
20190752	Dubois PC	Nat Genet	rs1464510	Celiac disease
20190752	Dubois PC	Nat Genet	rs6691768	Celiac disease
20190752	Dubois PC	Nat Genet	rs2187668	Celiac disease
22634755	Dunlop MG	Nat Genet	rs3824999	Colorectal cancer
22634755	Dunlop MG	Nat Genet	rs5934683	Colorectal cancer
22634755	Dunlop MG	Nat Genet	rs1321311	Colorectal cancer
21761138	Peters U	Hum Genet	rs7315438	Colorectal cancer
21761138	Peters U	Hum Genet	rs16892766	Colorectal cancer
21761138	Peters U	Hum Genet	rs4779584	Colorectal cancer
21761138	Peters U	Hum Genet	rs3802842	Colorectal cancer
21761138	Peters U	Hum Genet	rs4939827	Colorectal cancer
20972440	Houlston RS	Nat Genet	rs6691170	Colorectal cancer
20972440	Houlston RS	Nat Genet	rs4925386	Colorectal cancer
20972440	Houlston RS	Nat Genet	rs11169552	Colorectal cancer
20972440	Houlston RS	Nat Genet	rs10936599	Colorectal cancer
21926974	Ripke S	Nat Genet	rs10894294	Schizophrenia
21926974	Ripke S	Nat Genet	rs4624519	Schizophrenia
21926974	Ripke S	Nat Genet	rs7914558	Schizophrenia
21926974	Ripke S	Nat Genet	rs11191580	Schizophrenia
21926974	Ripke S	Nat Genet	rs17512836	Schizophrenia
21926974	Ripke S	Nat Genet	rs16915157	Schizophrenia
21926974	Ripke S	Nat Genet	rs2239547	Schizophrenia
21926974	Ripke S	Nat Genet	rs13025591	Schizophrenia
21926974	Ripke S	Nat Genet	rs6703335	Schizophrenia
21926974	Ripke S	Nat Genet	rs1869901	Schizophrenia
21926974	Ripke S	Nat Genet	rs2021722	Schizophrenia
21926974	Ripke S	Nat Genet	rs7004633	Schizophrenia
21926974	Ripke S	Nat Genet	rs10503253	Schizophrenia
21926974	Ripke S	Nat Genet	rs12699131	Schizophrenia
21926974	Ripke S	Nat Genet	rs1879248	Schizophrenia
21926974	Ripke S	Nat Genet	rs17662626	Schizophrenia
21926974	Ripke S	Nat Genet	rs1625579	Schizophrenia
21926974	Ripke S	Nat Genet	rs2252865	Schizophrenia
21926974	Ripke S	Nat Genet	rs12966547	Schizophrenia
21926974	Ripke S	Nat Genet	rs4775413	Schizophrenia
21926974	Ripke S	Nat Genet	rs4765905	Schizophrenia
21926974	Ripke S	Nat Genet	rs548181	Schizophrenia
21926974	Ripke S	Nat Genet	rs11130874	Schizophrenia
21926974	Ripke S	Nat Genet	rs4356203	Schizophrenia
21926974	Ripke S	Nat Genet	rs1009080	Schizophrenia
22019778	Zhang F	Nat Genet	rs3762318	Leprosy
22019778	Zhang F	Nat Genet	rs16948876	Leprosy
22019778	Zhang F	Nat Genet	rs2275606	Leprosy
22019778	Zhang F	Nat Genet	rs538147	Leprosy
20018961	Zhang FR	N Engl J Med	rs6478108	Leprosy
20018961	Zhang FR	N Engl J Med	rs3764147	Leprosy
20018961	Zhang FR	N Engl J Med	rs42490	Leprosy
20018961	Zhang FR	N Engl J Med	rs602875	Leprosy
20018961	Zhang FR	N Engl J Med	rs40457	Leprosy
20018961	Zhang FR	N Engl J Med	rs9302752	Leprosy

Supplementary table 5: List of cell types profiled for DHSs and used in the analysis

Name	Description	Number of replicates
A549	Epithelial cell line derived from a lung carcinoma tissue	1
AG04449	Fetal buttock/thigh fibroblast	1
AG04450	Fetal lung fibroblast	1
AG09309	Adult human toe fibroblast	1
AG09319	Adult human gum tissue fibroblast	1
AG10803	Adult human abdominal skin fibroblast	1
AoAF	Normal Human Aortic Adventitial Fibroblast Cells	1
BE2_C	Human Neuroblastoma cell line	1
BJ	Skin fibroblast	1
Caco-2	Colorectal adenocarcinoma	1
CD14+	Monocytes	1
CD19+	B lymphocytes	1
CD20+	B lymphocytes	2
CD3+	T lymphocytes adult	1
CD34+	Mobilized hematopoietic progenitor cells	1
CD3+_Cord blood	T lymphocytes cord blood	1
CD4+	T helper cells	1
CD56+	CD56+ lymphocytes	1
CD8+	Cytotoxic T cells	1
CMK	Acute megakaryocytic leukaemia cells	1
LCLs	Lymphoblastoid cell lines (GM12865, GM06990, GM12878)	3
H1_P18	H1-derived embryonic stem cells	1
H7-hESC	Undifferentiated embryonic stem cells	1
H9_P42	H9-derived embryonic stem cells	1
HAEpiC	Human Amniotic Epithelial Cells	1
HAc	Human astrocytes cerebellar	1
HAh	Human astrocytes hippocampal	1
HAsp	Human astrocytes spinal cord	1
HBMEC	Human Brain Microvascular Endothelial Cells	1
HCF	Human Cardiac Fibroblasts	1
HCFaa	Human Cardiac Fibroblasts adult atrial	1
HCM	Human cardiomyocytes	1
HCPEpiC	Human Choroid Plexus Epithelial Cells	1
HCT-116	Colon adenocarcinoma	1
HConF	Human Conjunctival Fibroblasts	1
HEEpiC	Human Esophageal Epithelial Cells	1
HepG2	Hepatocellular carcinoma	1
HESC	Human H1 Embryonic Stem Cell line	1
HFF	Human foreskin fibroblasts	1

HFF_Myc	Human foreskin fibroblast_Myc transgene	1
HGF	Human gingival fibroblasts	1
HIPEpiC	Human Iris Pigment Epithelial Cells	1
HL-60	Human promyelocytic leukaemia cells	1
HMEC	Human mammary epithelial cells	1
HMF	Human Mammary Fibroblasts	1
HMVEC-LBl	Human Lung Blood Microvascular Endothelial Cells	1
HMVEC-LLY	Human Lung Lymphatic Microvascular Endothelial Cells	1
HMVEC-dAd	Adult Human Dermal Microvascular Endothelial Cells	1
HMVEC-dBl-Ad	Adult Human Dermal Blood Microvascular Endothelial Cells	1
HMVEC-dBl-Neo	Neonatal Human Dermal Blood Microvascular Endothelial Cells	1
HMVEC-dLy-Ad	Adult Human Dermal Lymphatic Microvascular Endothelial Cells	1
HMVEC-dLy-Neo	Neonatal Human Dermal Lymphatic Microvascular Endothelial Cells	1
HMVEC-dNeo	Neonatal Human Dermal Microvascular Endothelial Cells	1
HNPCEpiC	Human Non-Pigment Ciliary Epithelial Cells	1
HPAEC	Human Pulmonary Artery Endothelial Cells	1
HPAF	Human Pulmonary Artery Fibroblasts	1
HPF	Human Pulmonary Fibroblasts	1
HPdLF	Human Periodontal Ligament Fibroblasts	1
HRCEpiC	Human renal cortical epithelial cells (normal)	1
HRE	Human renal epithelial cells (normal)	1
HRGEC	Human Renal Glomerular Endothelial Cells	1
HRPEpiC	Human Retinal Pigment Epithelial Cells	1
HSMM	Human Skeletal Muscle Myoblasts	1
HSMM_D	Human Skeletal Muscle Myoblasts_differentiated	1
HUVEC	Human umbilical vein endothelial cells	1
HVMF	Human Villous Mesenchymal Fibroblasts	1
HeLa-S3	Cervical carcinoma	1
IMR90	Fibroblast	1
Jurkat	T lymphoblastoid cell line derived from acute T cell leukaemia	1
K562	Chronic myeloid leukaemia	1
LNCaP	Prostate adenocarcinoma cell line	1
MCF-7	Mammary gland adenocarcinoma	1
Mesendoderm	H1 derived mesendoderm cells	1
NB4	Acute Promyelocytic Leukaemia cell line	1
NH-A	Normal Human Astrocytes	1
NHDF-Ad	Adult Human Dermal Fibroblasts	1
NHDF-neo	Neonatal Human Dermal Fibroblasts	1
NHEK	Normal Human Epidermal Keratinocytes	1
NHLF	Normal Human Lung Fibroblasts	1
NPC	H1 derived neuroprogenitor cells	1

NT2-D1	Human malignant pluripotent embryonal cancer cell line Induced by RA to neuronal cells	1
PANC-1	Pancreatic carcinoma cell line	1
PrEC	Human prostate epithelial cell line	1
RPTEC	Human renal proximal tubule cells	1
SAEC	Small airway epithelial cells	1
SK-N-SH_RA	Neuroblastoma cell lines differentiated with retinoic acid	1
SK_N_MC	Neuroepithelioma cell line derived from a metastatic supra-orbital human brain tumor	1
SKMC	Human skeletal muscle cells	1
WERI-Rb1	Retinoblastoma cell line	1
WI-38	Embryonic lung fibroblasts immortalized hTERT	1
WI-38_TAM	Embryonic lung fibroblasts immortalized hTERT_Tamoxifen treated	1
fAdrenal	Fetal adrenal gland	3
fBrain	Fetal brain	3
fHeart	Fetal heart	3
fIntestine_Lg	Fetal intestine large	3
fIntestine_Sm	Fetal intestine small	3
fKidney cortex	Fetal kidney cortex	3
fKidney pelvis	Fetal kidney pelvis	3
fLung	Fetal lung	3
fMuscle	Fetal muscle	3
fPlacenta	Fetal placenta	3
fSkin	Fetal skin	3
fSpinal_cord	Fetal spinal cord	3
fSpleen	Fetal spleen	1
fStomach	Fetal stomach	3
fTestes	Fetal testicle tissue	1
fThymus	Fetal thymus	3
Th1	Human primary T helper 1	2
Th17	Human primary T helper 17	1
Th2	Human primary T helper 2	2
iPS_19_11	Induced pluripotent stem cells	3
vHMEC	Human mammary epithelial cells	1

Supplementary table 6: Enrichment and significance of overlap between GWAS associations and cell specific DHSs

Multiple sclerosis			Type 1 diabetes			Systemic lupus erythematosus		
Cell type	**Enrichment**	**P value**	**Cell type**	**Enrichment**	**P value**	**Cell type**	**Enrichment**	**P value**
Th1_1	2.3375	<0.0001	Th17	2.6779	<0.001	GM12878	2.5818	<0.001
CD8	2.2695	<0.0001	Th1_1	2.5221	<0.001	GM06990	2.4424	<0.001
Th1_2	2.2584	<0.0001	Th1_2	2.4087	<0.001	GM12865	2.3946	<0.001
Th17	2.1875	<0.0001	Th2_1	2.3905	<0.001	CD19	2.3566	<0.001
CD3 adult	2.1654	<0.0001	Th2_2	2.2984	<0.001	CD3 CordBlood	2.2227	<0.001
Th2_1	2.1574	<0.0001	CD56	2.2347	<0.001	Th17	2.2015	<0.001
CD3 CordBlood	2.1392	<0.0001	CD3 adult	2.1335	<0.001	CD8	2.1765	<0.001
Th2_2	2.1079	<0.0001	CD4	2.1275	<0.001	Th1_2	2.1168	<0.001
CD4	2.0694	<0.0001	CD8	2.1135	<0.001	CD3 adult	2.0914	<0.001
CD19	2.0459	<0.0001	CD19	1.7812	0.0023	Th2_2	2.0708	<0.001
GM12878	1.9819	<0.0001	CD3 CordBlood	1.7805	0.0040	CD4	2.0550	<0.001
GM06990	1.9654	<0.0001	GM12878	1.7581	0.0056	CD56	2.0262	<0.001
CD56	1.9539	<0.0001	GM06990	1.7098	0.0072	CD20_2	2.0258	<0.001
CD20_2	1.9217	<0.0001	CD20_2	1.6739	0.0034	Th1_1	2.0182	<0.001
GM12865	1.8071	<0.0001	Fetal thymus_1	1.6719	0.0031	Th2_1	1.9721	0.0011
CD14	1.7677	<0.001	GM12865	1.6401	0.0064	CD14	1.6925	0.0051
Fetal thymus_2	1.7000	<0.001	Fetal thymus_3	1.6228	0.0059	Fetal thymus_2	1.6619	0.0033
Fetal thymus_3	1.6896	<0.001	Fetal thymus_2	1.5043	0.0247	Fetal thymus_1	1.6480	0.0035
Fetal thymus_1	1.6640	<0.001	CD14	1.4818	0.0247	Jurkat	1.5878	0.0159
NB4	1.4878	<0.001	Jurkat	1.4271	0.0306	Fetal thymus_3	1.5749	0.0123
Jurkat	1.4870	0.0020	NB4	1.2789	0.1095	HL60	1.4368	0.0764
HL60	1.4667	0.0140	Fetal spleen	1.2785	0.0108	NB4	1.4091	0.0627
CD34	1.3821	0.0006	CD34	1.2639	0.0956	Fetal spleen	1.3817	0.0051
Fetal spleen	1.3398	<0.001	HMVEC_dAd	1.2590	0.0764	K562	1.3369	0.1144
CMK	1.2867	0.1095	CMK	1.2039	0.2376	CD34	1.2889	0.0916
HCT116	1.2588	0.0655	HMVEC_dLyAd	1.1986	0.1589	NHEK	1.2638	0.1523
HEPG2	1.2557	0.1921	HCT116	1.1911	0.1523	HMVEC_dLyAd	1.2416	0.0897
CACO2	1.2351	0.5248	HMVEC_LLy	1.1529	0.2178	HMVEC_dAd	1.1904	0.1523
A549	1.1946	0.1036	HMVEC_dBlAd	1.1502	0.1457	CMK	1.1384	0.3762
Fetal stomach_1	1.1109	0.2970	HMVEC_LBl	1.1327	0.1589	HMVEC_dNeo	1.1372	0.2119
K562	1.1011	0.3960	AG04450	1.1123	0.2079	HMVEC_LLy	1.1289	0.2475
NHEK	1.0674	0.2871	HEPG2	1.1114	0.5743	PANC1	1.1286	0.2871
Fetal intestine small_2	1.0630	0.4356	HRE	1.1051	0.1244	HMVEC_dBlAd	1.1276	0.1656
Fetal stomach_3	1.0389	0.5545	HMVEC_dNeo	1.1002	0.3564	HMEC	1.1185	0.2079
Fetal stomach_2	1.0355	0.4950	Fetal stomach_1	1.0899	0.3069	HMVEC_dBlNeo	1.1078	0.1921
HMEC	1.0284	0.3762	HRGEC	1.0888	0.1523	vHMEC	1.1042	0.2277
HMVEC_dNeo	1.0117	0.6337	Fetal placenta_2	1.0886	0.4356	HMVEC_dLyNeo	1.0773	0.3168
HConF	1.0080	0.5842	HRCE	1.0878	0.1457	HMVEC_LBl	1.0703	0.3861

HMVEC_dAd	1.0001	0.5347	HMVEC_dLyNeo	1.0839	0.2871	Fetal placenta_2	1.0675	0.3960
HMVEC_dLyAd	0.9964	0.6139	RPTEC	1.0746	0.2185	AoAF	1.0579	0.3663
Fetal intestine large_2	0.9944	0.5941	HPF	1.0594	0.2772	Fetal intestine small_2	1.0553	0.4554
NPC	0.9942	0.7327	Fetal muscle_3	1.0571	0.3366	RPTEC	1.0499	0.3267
HPF	0.9940	0.4554	Fetal placenta_1	1.0537	0.4158	HUVEC	1.0466	0.2475
HMVEC_dLyNeo	0.9890	0.6535	HMVEC_dBlNeo	1.0499	0.3168	HRGEC	1.0449	0.2871
HMVEC_LLy	0.9838	0.7030	Fetal stomach_3	1.0440	0.5149	Fetal stomach_3	1.0335	0.5347
AG09319	0.9834	0.4356	Fetal placenta_3	1.0429	0.4257	HVMF	1.0311	0.3564
PANC1	0.9815	0.5545	Fetal stomach_2	1.0391	0.4554	HPAF	1.0274	0.3861
AG04450	0.9719	0.5248	HCF	1.0306	0.3069	Fetal stomach_1	1.0274	0.4554
Fetal intestine large_1	0.9717	0.7426	WI_38	1.0282	0.3267	HCFaa	1.0186	0.3267
HFF	0.9714	0.6634	K562	1.0235	0.4059	Fetal intestine large_2	1.0183	0.5248
HRE	0.9710	0.4950	HPAEC	1.0213	0.5446	HRCE	1.0137	0.4059
HMF	0.9688	0.5545	HMF	1.0199	0.3564	HMF	1.0137	0.4455
Fetal kidney pelvis_3	0.9644	0.8119	PANC1	1.0198	0.3960	HPAEC	1.0118	0.3663
HPAEC	0.9616	0.6832	IMR90	1.0180	0.2970	SAEC	1.0036	0.3267
SKNSH	0.9610	0.7822	HFF_MyC	1.0134	0.4356	HEPG2	0.9948	0.5644
HRCE	0.9586	0.4356	A549	1.0126	0.6535	Fetal intestine large_1	0.9880	0.5941
HMVEC_LBl	0.9573	0.6040	Fetal intestine large_1	1.0070	0.6238	Fetal placenta_3	0.9868	0.5149
Fetal intestine large_3	0.9569	0.6832	WI_38_TAM	0.9867	0.4455	HFF	0.9867	0.4653
HPAF	0.9569	0.5545	HL60	0.9856	0.5743	Fetal intestine large_3	0.9864	0.6040
HFF_MyC	0.9555	0.7426	HPAF	0.9853	0.4752	HCPEpiC	0.9771	0.5347
Fetal placenta_2	0.9513	0.8317	HUVEC	0.9818	0.3861	HFF_MyC	0.9741	0.5149
Fetal intestine small_1	0.9459	0.7030	Fetal intestine small_1	0.9812	0.5446	Fetal intestine small_1	0.9677	0.6337
HGF	0.9417	0.7030	HVMF	0.9793	0.4356	HAEpiC	0.9563	0.5347
PrEC	0.9405	0.6634	Fetal intestine large_3	0.9760	0.5842	HCT116	0.9549	0.6337
HCF	0.9401	0.7426	HConF	0.9709	0.6238	WI_38	0.9536	0.6040
WI_38	0.9394	0.6139	AG09319	0.9661	0.6040	IMR90	0.9516	0.6634
CD20_1	0.9392	0.8119	Fetal kidney pelvis_3	0.9599	0.7030	HRE	0.9513	0.6040
HIPEpiC	0.9381	0.6634	Fetal lung_1	0.9578	0.6337	HCM	0.9450	0.5347
MCF7	0.9365	0.8218	Fetal muscle_2	0.9566	0.7129	BJ	0.9434	0.6238
Fetal placenta_1	0.9346	0.8614	Fetal kidney cortex_3	0.9554	0.7228	CD20_1	0.9414	0.5941
Fetal kidney cortex_3	0.9334	0.8614	HFF	0.9541	0.6733	HPF	0.9395	0.6733
Fetal placenta_3	0.9330	0.8713	Fetal muscle_1	0.9539	0.4752	Fetal intestine small_3	0.9347	0.6436
SAEC	0.9327	0.6238	Fetal intestine large_2	0.9503	0.7030	AG09309	0.9306	0.6931
Fetal adrenal_3	0.9313	0.8515	Fetal intestine small_2	0.9468	0.7921	Fetal stomach_2	0.9282	0.7921
AG10803	0.9288	0.7327	HAEpiC	0.9465	0.5644	HEEpiC	0.9279	0.6337
IMR90	0.9269	0.7525	Fetal intestine small_3	0.9453	0.7426	AG09319	0.9246	0.5941
Fetal muscle_3	0.9231	0.8119	HCPEpiC	0.9403	0.5248	NHDF_Ad	0.9223	0.6337
Fetal intestine small_3	0.9228	0.8218	vHMEC	0.9351	0.6436	WI_38_TAM	0.9201	0.6634
HEEpiC	0.9226	0.7426	CD20_1	0.9240	0.7030	HCF	0.9170	0.7228
WI_38_TAM	0.9221	0.7129	HCM	0.9200	0.6931	AG10803	0.9155	0.6238

vHMEC	0.9201	0.7624	NHLF	0.9171	0.6040	HSMM_D	0.9138	0.6733
HCPEpiC	0.9197	0.8416	NHEK	0.9158	0.6634	Fetal placenta_1	0.9121	0.8218
HVMF	0.9179	0.5842	HAc	0.9151	0.6040	HPdLF	0.9079	0.7129
HAEpiC	0.9151	0.8218	NHA	0.9144	0.5842	HBMEC	0.9074	0.6931
H1_P18	0.9149	0.9109	HBMEC	0.9135	0.6337	HConF	0.8999	0.7525
HCM	0.9100	0.8515	HIPEpiC	0.9104	0.6733	HIPEpiC	0.8987	0.7624
HMVEC_dBlAd	0.9094	0.9010	AoAF	0.9101	0.7030	HGF	0.8979	0.7129
HMVEC_dBlNeo	0.9065	0.8020	Fetal kidney pelvis_1	0.9090	0.8416	PrEC	0.8965	0.6733
HNPCEpiC	0.9041	0.7426	Fetal kidney pelvis_2	0.9061	0.8515	AG04449	0.8962	0.7426
HAc	0.8965	0.8218	AG10803	0.9061	0.7921	Fetal heart_1	0.8947	0.6832
RPTEC	0.8932	0.7327	Fetal lung_3	0.9056	0.8218	HAh	0.8876	0.7228
Fetal lung_3	0.8913	0.9604	HMEC	0.9054	0.8119	NHLF	0.8864	0.7525
NHA	0.8903	0.7624	Fetal adrenal_3	0.9028	0.8119	HSMM	0.8828	0.7723
AoAF	0.8849	0.8911	SkMC	0.9021	0.6535	Fetal kidney pelvis_3	0.8811	0.8614
NHLF	0.8801	0.8317	Fetal lung_2	0.8897	0.7822	NHA	0.8792	0.7129
HPdLF	0.8797	0.8218	HPdLF	0.8896	0.7228	Fetal adrenal_3	0.8733	0.8119
iPS_6_9	0.8778	0.9604	HCFaa	0.8877	0.5248	Fetal testes	0.8695	0.8614
Fetal muscle_2	0.8762	0.9505	HNPCEpiC	0.8844	0.6931	A549	0.8692	0.7228
AG09309	0.8748	0.8416	Fetal kidney cortex_1	0.8690	0.9307	AG04450	0.8678	0.8020
H9_P42	0.8745	0.9406	AG04449	0.8652	0.7921	Fetal heart_3	0.8635	0.7228
HESC	0.8744	0.9901	Fetal adrenal_2	0.8647	0.9010	Fetal skin_3	0.8587	0.8020
Fetal skin_1	0.8707	0.8614	MCF7	0.8615	0.8317	HAsp	0.8585	0.7426
Fetal lung_1	0.8675	0.9307	BJ	0.8614	0.7129	HNPCEpiC	0.8559	0.8119
Fetal kidney pelvis_1	0.8673	0.9802	NHDF_Neo	0.8608	0.8317	Hela	0.8513	0.7426
Fetal kidney pelvis_2	0.8654	0.9802	PrEC	0.8561	0.8515	Fetal kidney cortex_3	0.8474	0.8911
HUVEC	0.8646	0.8515	HGF	0.8537	0.8515	Fetal heart_2	0.8460	0.8515
AG04449	0.8622	0.8911	CACO2	0.8534	0.9010	Fetal kidney pelvis_2	0.8458	0.8713
Fetal testes	0.8574	0.9208	Fetal skin_1	0.8466	0.8416	Fetal lung_1	0.8408	0.9208
HBMEC	0.8568	0.9505	fetal kidney cortex_2	0.8405	0.9010	HAc	0.8311	0.8317
HCFaa	0.8556	0.8119	Fetal adrenal_1	0.8362	0.9406	HRPEpiC	0.8303	0.9010
BJ	0.8505	0.9307	Fetal skin_2	0.8350	0.8317	CACO2	0.8254	0.8119
HSMM_D	0.8397	0.9901	HSMM	0.8338	0.8020	NHDF_Neo	0.8186	0.8317
Fetal skin_2	0.8379	0.9208	Fetal skin_3	0.8268	0.8911	Fetal muscle_3	0.8135	0.9208
NHDF_Neo	0.8326	0.9208	AG09309	0.8264	0.8911	Fetal lung_2	0.8132	0.9010
Fetal kidney cortex_1	0.8305	1.0000	HSMM_D	0.8244	0.9406	Fetal skin_1	0.8111	0.8911
HRGEC	0.8291	0.9010	NHDF_Ad	0.8118	0.8911	Fetal kidney pelvis_1	0.8033	1.0000
NHDF_Ad	0.8284	0.9703	HAsp	0.8099	0.7426	Fetal skin_2	0.7879	0.9307
iPS_19_7	0.8216	0.9901	SKNSH	0.8082	0.8614	Fetal lung_3	0.7868	0.9505
Fetal lung_2	0.8199	1.0000	Fetal testes	0.8078	0.8911	Fetal kidney cortex_1	0.7701	0.9901
Fetal adrenal_2	0.8187	1.0000	HAh	0.8054	0.9307	fetal kidney cortex_2	0.7575	0.9901
HAh	0.8152	1.0000	HEEpiC	0.8016	0.9604	Fetal adrenal_2	0.7553	0.9703
Fetal adrenal_1	0.8142	1.0000	HESC	0.7984	0.9901	Fetal muscle_1	0.7503	1.0000

Fetal muscle_1	0.8119	0.9703	NT2_D1	0.7934	0.9505	BE2_C	0.7445	0.9109
NT2_D1	0.8118	0.9901	SAEC	0.7854	0.9010	SkMC	0.7434	0.9406
Fetal skin_3	0.8095	0.9802	Hela	0.7838	0.7822	NPC	0.7388	0.9406
Mesendoderm	0.8033	0.9901	HRPEpiC	0.7786	0.9604	Fetal adrenal_1	0.7341	0.9802
HRPEpiC	0.8014	1.0000	Mesendoderm	0.7754	0.9901	SK_N_MC	0.7336	0.8812
iPS_4_7	0.7964	1.0000	iPS_19_7	0.7728	0.9901	Fetal muscle_2	0.7325	1.0000
SkMC	0.7922	0.9703	NPC	0.7716	0.9802	Mesendoderm	0.7061	1.0000
Hela	0.7753	0.9505	H9_P42	0.7693	0.9604	SKNSH	0.7021	0.9406
HSMM	0.7728	0.9802	Fetal heart_1	0.7400	0.9802	HESC	0.6836	1.0000
fetal kidney cortex_2	0.7659	1.0000	iPS_4_7	0.7364	0.9901	H9_P42	0.6610	1.0000
HAsp	0.7621	0.9703	H1_P18	0.7237	0.9901	MCF7	0.6411	0.9802
Fetal spinal cord_2	0.7591	1.0000	Fetal heart_3	0.7117	0.9505	LNCap	0.6394	0.9604
hESCT0	0.7529	1.0000	LNCap	0.7000	0.9703	Fetal spinal cord_1	0.6393	1.0000
Fetal heart_1	0.7527	0.9901	Fetal spinal cord_2	0.6965	1.0000	NT2_D1	0.6324	1.0000
SK_N_MC	0.7296	0.9901	iPS_6_9	0.6962	1.0000	iPS_6_9	0.6292	1.0000
Fetal heart_3	0.7183	0.9901	Fetal spinal cord_3	0.6841	1.0000	Fetal spinal cord_2	0.6278	1.0000
Fetal heart_2	0.7075	1.0000	hESCT0	0.6771	1.0000	Fetal spinal cord_3	0.6214	1.0000
BE2_C	0.7073	1.0000	BE2_C	0.6628	0.9802	iPS_19_7	0.6174	1.0000
Fetal spinal cord_3	0.6942	1.0000	Fetal heart_2	0.6620	1.0000	H1_P18	0.6094	1.0000
LNCap	0.6879	1.0000	Fetal spinal cord_1	0.6144	1.0000	WERI_Rb1	0.6009	0.9901
WERI_Rb1	0.6777	1.0000	SK_N_MC	0.5980	0.9703	iPS_4_7	0.5862	1.0000
Fetal spinal cord_1	0.6716	1.0000	WERI_Rb1	0.5580	1.0000	Fetal brain_3	0.5552	1.0000
Fetal brain_1	0.5843	1.0000	Fetal brain_2	0.5219	1.0000	hESCT0	0.5545	1.0000
Fetal brain_2	0.5555	1.0000	Fetal brain_3	0.5068	1.0000	Fetal brain_2	0.4804	1.0000
Fetal brain_3	0.5500	1.0000	Fetal brain_1	0.4954	1.0000	Fetal brain_1	0.4766	1.0000

Rheumatoid arthritis			Psoriasis			Celiac disease		
Cell type	**Enrichment**	**P value**	**Cell type**	**Enrichment**	**P value**	**Cell type**	**Enrichment**	**P value**
Th17	3.2603	<0.001	CD8	2.8673	<0.001	Th1_1	2.7736	<0.001
GM06990	2.9596	<0.001	CD4	2.7236	<0.001	Th1_2	2.7023	<0.001
CD3 adult	2.6052	<0.001	CD3 adult	2.4539	0.0015	Th17	2.6724	<0.001
Th1_1	2.5906	<0.001	CD14	2.4309	<0.001	CD8	2.4476	<0.001
CD8	2.5755	<0.001	CD56	2.2805	0.0010	CD3 adult	2.4149	<0.001
Th2_2	2.5723	<0.001	GM12878	2.1950	0.0067	Th2_1	2.3230	<0.001
GM12865	2.5664	<0.001	CD3 CordBlood	2.1258	0.0042	Th2_2	2.3107	<0.001
CD4	2.5540	<0.001	GM12865	2.1088	0.0050	CD3 CordBlood	2.2958	<0.001
Th2_1	2.5257	<0.001	Th1_1	2.0975	0.0108	CD56	2.2351	<0.001
Th1_2	2.5232	<0.001	Th2_1	2.0895	0.0064	CD4	2.2052	<0.001
GM12878	2.4857	<0.001	CD19	2.0698	0.0108	CD19	2.1265	<0.001
CD3 CordBlood	2.3545	<0.001	GM06990	1.9768	0.0142	CD20_2	2.0244	<0.001
CD19	2.2969	<0.001	CD20_2	1.9750	0.0074	GM12878	2.0052	0.0010
CD20_2	2.2473	<0.001	NHEK	1.9439	0.0086	GM06990	1.9735	<0.001

CD56	2.2153	<0.001	Th2_2	1.8660	0.0135	GM12865	1.9225	<0.001
Fetal thymus_2	1.8285	0.0010	Th1_2	1.8536	0.0323	Fetal thymus_3	1.9195	<0.001
Fetal thymus_1	1.8207	0.0027	HL60	1.8447	0.0262	Fetal thymus_1	1.9124	<0.001
Fetal thymus_3	1.8159	0.0029	NB4	1.7726	0.0123	CD14	1.8091	0.0010
CD14	1.7093	0.0042	Th17	1.7011	0.0598	Fetal thymus_2	1.7645	0.0010
Fetal spleen	1.4948	0.0010	HMEC	1.6837	0.0055	Jurkat	1.6627	0.0060
CD34	1.4142	0.0100	Fetal thymus_2	1.6528	0.0182	NB4	1.4946	0.0079
Jurkat	1.3444	0.1194	Fetal thymus_1	1.6456	0.0357	HL60	1.4015	0.0655
HMVEC_dAd	1.3382	0.0218	Fetal thymus_3	1.6324	0.0381	CD34	1.3189	0.0102
NB4	1.3352	0.0532	vHMEC	1.5813	0.0127	Fetal spleen	1.2904	0.0145
CMK	1.3179	0.0627	CD34	1.5168	0.0381	CMK	1.2557	0.1393
HMVEC_dLyAd	1.3125	0.0599	HEEpiC	1.4981	0.0145	Fetal placenta_3	1.1399	0.1987
HL60	1.2345	0.1144	Fetal intestine small_2	1.3530	0.0698	Fetal placenta_2	1.1319	0.1987
HMVEC_LLy	1.1738	0.1854	Jurkat	1.3322	0.2178	PrEC	1.1302	0.1788
HMVEC_dBlAd	1.1543	0.1921	Fetal spleen	1.3180	0.0627	NHEK	1.1164	0.3267
HMVEC_dNeo	1.1398	0.1854	K562	1.2873	0.2574	HMEC	1.1162	0.2277
HRCE	1.1236	0.2277	HEPG2	1.2709	0.2574	HCT116	1.0982	0.2970
HMVEC_dLyNeo	1.1075	0.2178	Fetal intestine large_2	1.2519	0.1391	SAEC	1.0921	0.2376
HMVEC_LBl	1.1063	0.2871	Fetal intestine small_1	1.2262	0.1493	K562	1.0921	0.3564
HPAEC	1.0802	0.2475	SAEC	1.2219	0.1194	HMVEC_dLyAd	1.0833	0.2079
HMVEC_dBlNeo	1.0674	0.3762	HPAEC	1.1598	0.2178	Fetal placenta_1	1.0810	0.1656
HRE	1.0664	0.3465	Fetal intestine small_3	1.1585	0.2178	HMVEC_LBl	1.0763	0.2673
Fetal intestine large_2	1.0432	0.3069	CMK	1.1363	0.3366	vHMEC	1.0667	0.2574
Fetal intestine large_3	1.0254	0.4059	CD20_1	1.1294	0.3366	HRGEC	1.0643	0.4257
HRGEC	1.0220	0.4653	Fetal intestine large_3	1.1148	0.2970	HMVEC_dLyNeo	1.0532	0.3168
Fetal intestine small_1	1.0168	0.3960	Fetal placenta_3	1.1053	0.3465	HEEpiC	1.0527	0.3267
RPTEC	0.9979	0.3663	PrEC	1.0998	0.2574	PANC1	1.0490	0.3465
Fetal intestine small_3	0.9868	0.5248	Fetal placenta_2	1.0737	0.4356	HRE	1.0427	0.3861
PANC1	0.9736	0.4653	CACO2	1.0674	0.3861	A549	1.0403	0.2673
Fetal intestine small_2	0.9610	0.5149	HRGEC	1.0445	0.3663	HRCE	1.0395	0.3960
WI_38_TAM	0.9571	0.6535	Fetal intestine large_1	1.0406	0.4752	HMVEC_dAd	1.0336	0.3168
HCF	0.9470	0.6733	HMVEC_LBl	1.0374	0.2970	Fetal stomach_1	1.0202	0.3663
Fetal muscle_3	0.9459	0.6436	Fetal placenta_1	1.0320	0.4851	HMVEC_dBlAd	1.0036	0.5644
HESC	0.9438	0.5644	HPAF	1.0304	0.3564	HMVEC_dNeo	0.9998	0.4455
HGF	0.9433	0.7723	A549	1.0295	0.3663	HEPG2	0.9980	0.5050
HEPG2	0.9317	0.5842	HAEpiC	1.0166	0.4653	HIPEpiC	0.9973	0.6139
HPAF	0.9301	0.7822	HRCE	1.0140	0.3861	HMVEC_LLy	0.9958	0.4356
CD20_1	0.9298	0.6139	WI_38_TAM	1.0137	0.3762	Fetal stomach_3	0.9856	0.5347
HUVEC	0.9284	0.7723	HMVEC_dAd	1.0106	0.4554	HPAEC	0.9763	0.4455
K562	0.9267	0.7525	Fetal stomach_1	1.0062	0.4950	AG04450	0.9653	0.6238
NT2_D1	0.9229	0.6238	HCPEpiC	1.0058	0.4554	HMVEC_dBlNeo	0.9650	0.6436
Fetal muscle_2	0.9196	0.7426	HMVEC_dLyAd	1.0033	0.4257	HVMF	0.9633	0.7030

HCM	0.9122	0.7327	HCFaa	0.9903	0.3663	WI_38	0.9565	0.6337
SAEC	0.9068	0.7426	HMVEC_dBlAd	0.9899	0.5446	WI_38_TAM	0.9535	0.7030
HFF	0.9061	0.7525	HUVEC	0.9724	0.5149	AG04449	0.9500	0.7822
Fetal intestine large_1	0.8985	0.6634	AoAF	0.9724	0.5248	HUVEC	0.9439	0.6040
Hela	0.8960	0.6040	Hela	0.9721	0.4851	HMF	0.9426	0.7228
HMEC	0.8880	0.7228	IMR90	0.9711	0.5149	Fetal intestine small_2	0.9406	0.6634
PrEC	0.8880	0.8020	MCF7	0.9693	0.6436	IMR90	0.9394	0.6535
Fetal placenta_1	0.8868	0.8614	RPTEC	0.9601	0.4950	CACO2	0.9391	0.2970
HConF	0.8843	0.7426	HMVEC_dLyNeo	0.9363	0.6337	Fetal intestine large_2	0.9359	0.6832
CACO2	0.8766	0.5842	HCF	0.9315	0.6238	HConF	0.9356	0.7921
HIPEpiC	0.8712	0.8416	HMF	0.9314	0.5842	HPF	0.9326	0.7030
HFF_MyC	0.8698	0.8020	HRE	0.9309	0.5842	HAEpiC	0.9264	0.7723
HSMM_D	0.8653	0.8812	Fetal stomach_3	0.9295	0.6931	HCPEpiC	0.9252	0.7426
Fetal kidney pelvis_2	0.8639	0.8713	PANC1	0.9187	0.5743	HBMEC	0.9234	0.8119
A549	0.8620	0.6733	AG09309	0.9172	0.6733	RPTEC	0.9202	0.7327
BE2_C	0.8591	0.6436	HMVEC_dBlNeo	0.9145	0.6238	Fetal stomach_2	0.9176	0.8416
Fetal kidney pelvis_3	0.8587	0.8515	Fetal adrenal_3	0.9111	0.6040	MCF7	0.9171	0.6040
iPS_19_7	0.8573	0.7822	HCT116	0.9102	0.6931	Fetal skin_3	0.9151	0.8119
Fetal stomach_3	0.8573	0.8614	Fetal stomach_2	0.9094	0.7327	HNPCEpiC	0.9149	0.7921
HCPEpiC	0.8554	0.9010	HPdLF	0.9040	0.6436	Fetal intestine large_1	0.9130	0.6733
fetal kidney cortex_2	0.8506	0.9208	Fetal adrenal_2	0.8955	0.7228	Fetal intestine small_3	0.9130	0.7327
vHMEC	0.8478	0.7921	HMVEC_dNeo	0.8930	0.6535	HCFaa	0.9079	0.7822
AG09319	0.8466	0.7327	HGF	0.8830	0.7030	SkMC	0.9007	0.8515
Fetal kidney cortex_1	0.8457	0.8812	HFF	0.8823	0.7624	HFF_MyC	0.8956	0.8515
HAh	0.8442	0.9109	WI_38	0.8815	0.6238	Fetal intestine small_1	0.8940	0.8317
HEEpiC	0.8434	0.9109	HAc	0.8732	0.7327	CD20_1	0.8920	0.7822
HAc	0.8388	0.9307	HFF_MyC	0.8588	0.7822	HGF	0.8903	0.8218
Fetal muscle_1	0.8381	0.9307	Fetal adrenal_1	0.8571	0.7228	HPdLF	0.8891	0.8218
Fetal stomach_2	0.8376	0.8911	HIPEpiC	0.8538	0.7822	Fetal intestine large_3	0.8877	0.8416
H1_P18	0.8374	0.7822	Fetal kidney pelvis_3	0.8536	0.7822	HPAF	0.8873	0.8416
AoAF	0.8370	0.8119	Fetal kidney cortex_3	0.8520	0.8218	BJ	0.8843	0.8416
MCF7	0.8356	0.8020	AG04450	0.8469	0.8119	AG09319	0.8798	0.8119
IMR90	0.8336	0.9406	HESC	0.8467	0.8119	HFF	0.8790	0.9406
LNCap	0.8304	0.8515	HCM	0.8460	0.8020	Fetal muscle_3	0.8695	0.8614
HMF	0.8280	0.8515	HSMM	0.8456	0.7426	HCF	0.8568	0.8416
HBMEC	0.8272	0.9406	HNPCEpiC	0.8384	0.8119	Fetal muscle_2	0.8546	0.9604
HCT116	0.8239	0.7822	HSMM_D	0.8352	0.7921	NPC	0.8544	0.6139
Fetal adrenal_3	0.8215	0.8515	NT2_D1	0.8274	0.8416	AG10803	0.8543	0.9010
NHDF_Neo	0.8212	0.8713	NHDF_Ad	0.8273	0.8515	NHLF	0.8528	0.9406
Fetal kidney pelvis_1	0.8154	0.9406	Mesendoderm	0.8169	0.8020	Fetal skin_1	0.8448	0.9307
Fetal kidney cortex_3	0.8144	0.8911	HBMEC	0.8144	0.8020	NHDF_Neo	0.8422	0.9109
HNPCEpiC	0.8142	0.9109	H1_P18	0.8064	0.7426	SKNSH	0.8400	0.7228

HCFaa	0.8134	0.8812	HConF	0.8026	0.8515	HSMM_D	0.8393	0.9604
Fetal placenta_2	0.8100	0.8911	AG04449	0.8015	0.8614	NHA	0.8386	0.8515
Fetal stomach_1	0.8077	0.8317	HPF	0.7984	0.8713	HCM	0.8382	0.9703
Fetal spinal cord_2	0.8064	0.9010	Fetal kidney pelvis_2	0.7879	0.8713	AG09309	0.8349	0.9010
Fetal lung_1	0.8061	0.9604	BE2_C	0.7866	0.7525	Fetal kidney cortex_3	0.8292	0.9505
H9_P42	0.8038	0.8218	iPS_19_7	0.7845	0.8713	hESCT0	0.8274	0.9505
HPF	0.8009	0.9505	HAh	0.7837	0.8911	HSMM	0.8223	0.9703
Fetal adrenal_1	0.7963	0.9208	AG10803	0.7816	0.8614	Fetal skin_2	0.8217	0.9802
iPS_6_9	0.7925	0.9109	NHA	0.7808	0.8614	HAsp	0.8216	0.9307
SKNSH	0.7881	0.7723	hESCT0	0.7784	0.9505	Fetal kidney pelvis_3	0.8203	0.9010
HVMF	0.7861	0.9505	HMVEC_LLy	0.7763	0.8317	HESC	0.8199	0.8713
NHA	0.7851	0.9505	Fetal skin_3	0.7712	0.8812	Fetal testes	0.8162	0.9406
WERI_Rb1	0.7830	0.8416	H9_P42	0.7704	0.8218	Fetal lung_2	0.8131	0.9901
Fetal skin_1	0.7828	0.9604	SkMC	0.7696	0.9010	Hela	0.8046	0.8911
HSMM	0.7827	0.9406	NHLF	0.7562	0.9208	HAh	0.8045	0.9901
NHDF_Ad	0.7815	0.9604	HRPEpiC	0.7551	0.9604	Fetal lung_1	0.8016	0.9802
HPdLF	0.7727	0.9109	Fetal skin_1	0.7511	0.9109	H9_P42	0.8012	0.9208
HAsp	0.7698	0.9604	Fetal kidney pelvis_1	0.7458	0.9208	AoAF	0.7978	0.9802
Fetal lung_3	0.7683	0.9703	NPC	0.7403	0.8713	HAc	0.7939	0.9802
Fetal skin_3	0.7657	0.9406	Fetal muscle_3	0.7400	0.9604	Fetal adrenal_3	0.7938	0.9109
HAEpiC	0.7632	0.9406	iPS_6_9	0.7369	0.9505	NT2_D1	0.7923	1.0000
WI_38	0.7611	0.9406	AG09319	0.7333	0.9010	Fetal muscle_1	0.7833	0.9802
Fetal lung_2	0.7610	0.9604	Fetal muscle_2	0.7263	0.9703	iPS_6_9	0.7765	0.9505
Fetal testes	0.7579	0.9406	HVMF	0.7259	0.9010	NHDF_Ad	0.7749	0.9901
Fetal adrenal_2	0.7467	0.9505	SKNSH	0.7239	0.8218	Fetal heart_1	0.7668	0.9901
NHEK	0.7454	0.8713	Fetal kidney cortex_1	0.7186	0.9604	Fetal kidney cortex_1	0.7596	0.9703
BJ	0.7443	0.9703	fetal kidney cortex_2	0.7095	0.9307	HRPEpiC	0.7590	0.9901
hESCT0	0.7440	1.0000	LNCap	0.7081	0.9208	H1_P18	0.7572	0.9406
NHLF	0.7432	0.9703	Fetal skin_2	0.7034	0.8911	iPS_4_7	0.7567	0.9604
AG10803	0.7408	0.9406	Fetal testes	0.6998	0.9406	Fetal lung_3	0.7561	0.9703
AG09309	0.7363	0.9802	BJ	0.6990	0.9505	iPS_19_7	0.7559	0.9406
Fetal placenta_3	0.7300	0.9010	Fetal heart_3	0.6940	0.8614	Fetal kidney pelvis_1	0.7519	0.9703
SkMC	0.7275	0.9901	Fetal heart_1	0.6892	0.9109	Mesendoderm	0.7499	0.9802
Fetal spinal cord_3	0.7252	0.9802	iPS_4_7	0.6822	0.9901	Fetal kidney pelvis_2	0.7498	0.9802
AG04450	0.7238	0.9703	SK_N_MC	0.6797	0.8713	LNCap	0.7343	0.9703
Mesendoderm	0.7188	1.0000	Fetal heart_2	0.6725	0.9307	Fetal adrenal_1	0.7343	0.9505
Fetal brain_3	0.7124	0.9505	HAsp	0.6668	0.9703	fetal kidney cortex_2	0.7207	0.9901
Fetal spinal cord_1	0.7059	0.9802	WERI_Rb1	0.6484	0.9406	Fetal adrenal_2	0.7097	0.9802
NPC	0.7042	0.8911	NHDF_Neo	0.6238	0.9901	Fetal heart_3	0.6892	0.9802
AG04449	0.6967	0.9406	Fetal lung_3	0.6204	0.9802	Fetal heart_2	0.6866	1.0000
Fetal skin_2	0.6879	0.9901	Fetal lung_1	0.5871	0.9901	WERI_Rb1	0.6594	1.0000
iPS_4_7	0.6707	1.0000	Fetal lung_2	0.5563	1.0000	Fetal spinal cord_2	0.6301	1.0000

Fetal heart_2	0.6597	1.0000	Fetal brain_1	0.5493	1.0000	SK_N_MC	0.6227	1.0000
HRPEpiC	0.6536	1.0000	Fetal muscle_1	0.5386	1.0000	BE2_C	0.6219	1.0000
Fetal brain_2	0.6407	1.0000	Fetal spinal cord_2	0.5367	1.0000	Fetal spinal cord_1	0.5955	1.0000
Fetal brain_1	0.6393	0.9901	Fetal brain_2	0.5156	1.0000	Fetal spinal cord_3	0.5916	1.0000
Fetal heart_3	0.6333	0.9901	Fetal spinal cord_3	0.4750	1.0000	Fetal brain_1	0.5445	1.0000
Fetal heart_1	0.6316	0.9802	Fetal brain_3	0.4385	1.0000	Fetal brain_2	0.5438	1.0000
SK_N_MC	0.5982	0.9703	Fetal spinal cord_1	0.4344	1.0000	Fetal brain_3	0.5164	1.0000

Crohn's disease			Ulcerative colitis			Systemic sclerosis		
Cell type	**Enrichment**	**P value**	**Cell type**	**Enrichment**	**P value**	**Cell type**	**Enrichment**	**P value**
Th1_1	2.4405	<0.001	CD14	2.1528	<0.001	CD3 CordBlood	3.5197	<0.001
Th2_1	2.4395	<0.001	CACO2	2.0740	0.0024	CD19	3.3328	<0.001
Th17	2.4369	<0.001	Th2_1	1.9384	<0.001	CD20_2	3.0701	<0.001
Th2_2	2.1288	<0.001	Th1_1	1.8337	<0.001	GM12878	3.0324	0.0019
Th1_2	2.0480	<0.001	CD20_2	1.8223	<0.001	Th17	2.8720	0.0026
CD3 adult	2.0236	<0.001	GM12878	1.7895	<0.001	Th1_2	2.8445	0.0042
CD8	2.0083	<0.001	GM06990	1.7607	<0.001	CD8	2.7702	0.0028
CD56	1.9506	<0.001	CD19	1.7588	<0.001	Th2_1	2.6776	0.0035
CD14	1.9413	<0.001	NB4	1.7438	<0.001	GM06990	2.6472	0.0038
CD4	1.9377	<0.001	CD3 CordBlood	1.7314	<0.001	CD3 adult	2.5763	0.0042
HL60	1.7478	<0.001	HL60	1.7162	0.0013	CD56	2.5500	0.0019
GM06990	1.6754	<0.001	CD8	1.6073	0.0058	Th1_1	2.3003	0.0131
Fetal thymus_1	1.6566	<0.001	GM12865	1.6038	0.0026	Th2_2	2.2484	0.0100
NB4	1.6563	<0.001	CD34	1.5989	<0.001	GM12865	2.2260	0.0107
CD19	1.6356	<0.001	CD4	1.5593	0.0033	CD14	2.1714	0.0046
GM12878	1.5958	0.0041	Th17	1.5494	0.0093	CD4	2.1634	0.0074
CD20_2	1.5930	<0.001	Th1_2	1.5371	0.0150	Fetal spleen	1.6500	0.0100
CD3 CordBlood	1.5534	0.0033	CD3 adult	1.5064	0.0163	Fetal thymus_3	1.6397	0.0731
Fetal thymus_3	1.5471	<0.001	Th2_2	1.5035	0.0155	Fetal thymus_1	1.5921	0.1144
GM12865	1.5136	0.0022	Fetal spleen	1.5033	<0.001	Fetal thymus_2	1.5720	0.0598
CD34	1.4846	<0.001	Fetal intestine large_2	1.4819	0.0016	CD34	1.4297	0.0837
Fetal thymus_2	1.4807	0.0028	CMK	1.4713	0.0349	NB4	1.3738	0.2277
CMK	1.4627	0.0155	Fetal intestine large_1	1.4445	<0.001	Jurkat	1.3008	0.1722
Fetal spleen	1.3995	<0.001	CD56	1.4435	0.0182	Fetal intestine small_2	1.2718	0.1542
Jurkat	1.2521	0.0549	Fetal intestine small_1	1.4268	0.0022	Fetal stomach_1	1.2524	0.1294
CACO2	1.2518	0.2178	Fetal intestine small_2	1.4239	0.0054	NPC	1.2084	0.2475
K562	1.2440	0.1244	Fetal intestine large_3	1.4194	0.0010	HL60	1.1999	0.3366
HEPG2	1.2408	0.1393	Fetal thymus_2	1.3797	0.0182	H9_P42	1.1632	0.1987
Fetal intestine small_2	1.1725	0.1036	HEPG2	1.3554	0.0510	Fetal intestine large_2	1.1495	0.3663
Fetal placenta_2	1.1608	0.1095	Fetal intestine small_3	1.3518	0.0052	Fetal intestine large_1	1.1430	0.2574
Fetal intestine large_2	1.0906	0.2079	Fetal stomach_1	1.3155	0.0087	Fetal stomach_2	1.1394	0.1589
A549	1.0860	0.3069	CD20_1	1.2935	0.0145	HCT116	1.1240	0.4257

HCT116	1.0841	0.3564	Fetal thymus_1	1.2603	0.0996	CMK	1.1083	0.3861
Fetal stomach_1	1.0812	0.3663	Fetal stomach_3	1.2348	0.0323	CACO2	1.0997	0.2277
Fetal placenta_1	1.0744	0.2970	HVMF	1.2275	0.0532	Fetal stomach_3	1.0892	0.3168
NHEK	1.0739	0.3267	Fetal stomach_2	1.2135	0.0247	H1_P18	1.0812	0.3267
Fetal placenta_3	1.0703	0.2772	HAEpiC	1.2064	0.0366	iPS_19_7	1.0796	0.3762
Fetal intestine large_3	1.0644	0.2574	Fetal placenta_2	1.1795	0.0956	MCF7	1.0731	0.2871
Fetal stomach_3	1.0609	0.3564	Fetal placenta_1	1.1748	0.0731	HPAEC	1.0683	0.4653
HAEpiC	1.0584	0.2079	Fetal kidney pelvis_3	1.1600	0.1095	Fetal intestine small_3	1.0524	0.3267
Fetal intestine large_1	1.0578	0.2772	Fetal kidney cortex_3	1.1109	0.1788	NT2_D1	1.0487	0.5248
HMVEC_dNeo	1.0547	0.3762	Fetal muscle_2	1.1104	0.2079	CD20_1	1.0458	0.4950
HMVEC_dAd	1.0392	0.5347	Fetal thymus_3	1.1086	0.2574	SKNSH	1.0446	0.2871
Fetal intestine small_1	1.0366	0.4554	Fetal muscle_3	1.0883	0.2376	PANC1	1.0384	0.6535
HMEC	1.0311	0.3069	Fetal kidney cortex_1	1.0882	0.2475	Fetal intestine large_3	1.0295	0.5644
HMVEC_dLyNeo	1.0274	0.3564	Fetal kidney pelvis_2	1.0603	0.3267	Fetal intestine small_1	1.0289	0.4752
Fetal intestine small_3	1.0168	0.3762	SKNSH	1.0486	0.4158	iPS_6_9	1.0152	0.5644
WI_38_TAM	1.0130	0.4059	K562	1.0445	0.5248	K562	1.0101	0.5743
HMVEC_dBlAd	1.0104	0.4257	Fetal kidney pelvis_1	1.0382	0.3663	HFF	1.0080	0.3960
HEEpiC	1.0092	0.2277	HESC	1.0254	0.4158	HConF	0.9953	0.4059
HVMF	1.0048	0.4950	HRCE	1.0241	0.4257	HESC	0.9867	0.5050
Fetal stomach_2	1.0042	0.6040	Fetal placenta_3	1.0225	0.3861	iPS_4_7	0.9784	0.6832
AG04450	1.0011	0.4851	HMEC	1.0188	0.4158	Fetal adrenal_3	0.9714	0.3861
vHMEC	0.9992	0.4554	HMVEC_dLyAd	1.0169	0.3663	HMVEC_dBlAd	0.9583	0.6139
CD20_1	0.9992	0.5842	NPC	1.0080	0.4158	HMVEC_dLyNeo	0.9470	0.5743
HMVEC_LBl	0.9911	0.4752	HMVEC_LLy	1.0043	0.4257	HUVEC	0.9424	0.6139
HRE	0.9886	0.5347	Fetal lung_3	0.9849	0.5842	Fetal kidney pelvis_3	0.9349	0.5842
HMVEC_dLyAd	0.9866	0.5743	Hela	0.9831	0.6436	Fetal kidney cortex_3	0.9266	0.6238
SAEC	0.9823	0.4554	PrEC	0.9713	0.5446	HMVEC_dLyAd	0.9089	0.6436
HPAEC	0.9814	0.5941	NHEK	0.9700	0.5347	Fetal placenta_2	0.9081	0.6139
HMVEC_LLy	0.9786	0.6436	HMVEC_dAd	0.9516	0.5941	Fetal adrenal_1	0.9061	0.6040
HRCE	0.9749	0.4455	WI_38_TAM	0.9495	0.6436	HMVEC_dBlNeo	0.9051	0.6040
HMF	0.9748	0.4653	HMVEC_LBl	0.9469	0.6832	HEPG2	0.9039	0.5644
HESC	0.9723	0.6832	H1_P18	0.9458	0.6733	HMVEC_dAd	0.9037	0.6040
PrEC	0.9671	0.5743	HRE	0.9426	0.7327	HGF	0.9023	0.6238
HMVEC_dBlNeo	0.9660	0.5347	HMVEC_dNeo	0.9406	0.5248	HMVEC_dNeo	0.8960	0.5644
Fetal muscle_3	0.9630	0.8515	HSMM_D	0.9389	0.7822	RPTEC	0.8912	0.6535
HBMEC	0.9577	0.5050	HPF	0.9377	0.7129	HMVEC_LBl	0.8871	0.5842
HPAF	0.9563	0.6436	Jurkat	0.9369	0.6634	AG09319	0.8858	0.6436
IMR90	0.9549	0.6238	HSMM	0.9355	0.7030	Fetal testes	0.8828	0.6436
HIPEpiC	0.9547	0.5050	HCM	0.9338	0.7525	Fetal adrenal_2	0.8733	0.5842
Hela	0.9536	0.4950	HCF	0.9325	0.7525	HCM	0.8732	0.6931
HPF	0.9495	0.7525	Fetal adrenal_2	0.9181	0.7723	HRGEC	0.8730	0.7228
WI_38	0.9476	0.6634	vHMEC	0.9140	0.7030	HMVEC_LLy	0.8714	0.7129

NPC	0.9455	0.6931	Fetal adrenal_3	0.9114	0.7624	Fetal kidney cortex_1	0.8713	0.7129
Fetal adrenal_3	0.9452	0.8317	HPAF	0.9080	0.8218	A549	0.8631	0.6832
Fetal muscle_2	0.9398	0.7822	HMF	0.9069	0.8119	Fetal lung_1	0.8559	0.7129
HCF	0.9377	0.7624	HMVEC_dBlAd	0.9052	0.7228	Hela	0.8510	0.6436
PANC1	0.9321	0.6733	SkMC	0.8988	0.7822	Fetal muscle_3	0.8487	0.6733
HRGEC	0.9299	0.6337	HFF_MyC	0.8970	0.8713	BE2_C	0.8478	0.6337
HCPEpiC	0.9284	0.7030	HCPEpiC	0.8933	0.8317	Fetal heart_1	0.8444	0.6634
AoAF	0.9280	0.6931	HIPEpiC	0.8902	0.8218	Fetal placenta_3	0.8403	0.7327
iPS_6_9	0.9246	0.7624	fetal kidney cortex_2	0.8899	0.8812	HMEC	0.8390	0.7228
HConF	0.9190	0.8713	HPAEC	0.8865	0.7921	Fetal placenta_1	0.8354	0.7822
HFF_MyC	0.9121	0.8515	iPS_6_9	0.8838	0.7525	HRCE	0.8343	0.8218
AG04449	0.9085	0.7822	HUVEC	0.8829	0.7030	HCF	0.8335	0.6931
HFF	0.9065	0.8812	RPTEC	0.8804	0.7822	Fetal kidney pelvis_1	0.8335	0.8020
MCF7	0.9058	0.7426	HMVEC_dLyNeo	0.8769	0.7921	HEEpiC	0.8259	0.8020
HCM	0.9024	0.8812	HRGEC	0.8715	0.7624	fetal kidney cortex_2	0.8248	0.8515
H1_P18	0.8997	0.9505	Fetal lung_1	0.8680	0.9010	WERI_Rb1	0.8243	0.8515
AG09309	0.8931	0.8713	H9_P42	0.8679	0.8020	AG10803	0.8205	0.7426
iPS_19_7	0.8912	0.9406	HFF	0.8652	0.9010	Mesendoderm	0.8190	0.7624
HCFaa	0.8900	0.7723	SAEC	0.8605	0.8614	Fetal lung_2	0.8179	0.8119
Fetal kidney pelvis_3	0.8896	0.9505	Fetal adrenal_1	0.8579	0.8911	Fetal muscle_1	0.8164	0.6733
SkMC	0.8877	0.7624	HEEpiC	0.8548	0.9208	vHMEC	0.8112	0.8020
HNPCEpiC	0.8824	0.8416	HCT116	0.8530	0.7030	Fetal heart_3	0.8064	0.7327
HSMM_D	0.8770	0.9208	Fetal lung_2	0.8467	0.9604	SAEC	0.8063	0.8911
H9_P42	0.8763	0.9109	NT2_D1	0.8387	0.9901	HFF_MyC	0.8057	0.8713
HAc	0.8753	0.9307	iPS_19_7	0.8380	0.9109	Fetal kidney pelvis_2	0.8047	0.8614
AG10803	0.8751	0.8614	Fetal muscle_1	0.8316	0.9406	hESCT0	0.7950	0.8020
HUVEC	0.8728	0.6733	HBMEC	0.8315	0.9307	Fetal lung_3	0.7926	0.8317
BJ	0.8700	0.7624	HMVEC_dBlNeo	0.8291	0.9109	Fetal muscle_2	0.7895	0.7723
HGF	0.8696	0.9307	WI_38	0.8184	0.9703	Fetal spinal cord_3	0.7867	0.6832
SKNSH	0.8696	0.8713	Fetal testes	0.8163	0.9703	NHEK	0.7817	0.8416
Fetal adrenal_2	0.8681	0.9505	BJ	0.8139	0.9010	LNCap	0.7789	0.8416
NHLF	0.8667	0.8515	AoAF	0.8127	0.9505	HPAF	0.7777	0.8317
NHA	0.8606	0.9208	A549	0.8118	0.7723	HPdLF	0.7709	0.8317
AG09319	0.8593	0.9307	NHDF_Neo	0.8092	0.9406	HMF	0.7695	0.8119
Fetal skin_3	0.8573	0.9109	HConF	0.8044	0.9604	Fetal spinal cord_1	0.7675	0.7723
Fetal skin_1	0.8553	0.9109	AG04450	0.7930	0.9802	HVMF	0.7594	0.8020
HSMM	0.8544	0.9208	IMR90	0.7914	1.0000	HAEpiC	0.7590	0.8416
Fetal kidney cortex_3	0.8538	0.9901	hESCT0	0.7895	0.9901	WI_38	0.7435	0.9208
HPdLF	0.8531	0.8119	MCF7	0.7868	0.9703	SK_N_MC	0.7406	0.8416
Mesendoderm	0.8522	0.9604	LNCap	0.7848	0.9505	Fetal heart_2	0.7349	0.8416
NT2_D1	0.8435	0.9307	HCFaa	0.7839	0.9703	AG04450	0.7317	0.9109
Fetal kidney pelvis_2	0.8413	0.9703	HGF	0.7818	0.9505	AG04449	0.7262	0.8812

RPTEC	0.8397	0.9208	Fetal heart_1	0.7785	0.9208	HIPEpiC	0.7247	0.9307
Fetal kidney cortex_1	0.8321	1.0000	NHDF_Ad	0.7755	1.0000	Fetal spinal cord_2	0.7228	0.8515
NHDF_Ad	0.8288	0.9307	NHLF	0.7640	1.0000	HRE	0.7171	0.9406
Fetal adrenal_1	0.8271	0.9901	Fetal skin_3	0.7543	0.9802	IMR90	0.7170	0.9307
Fetal lung_3	0.8242	0.9802	Mesendoderm	0.7541	1.0000	BJ	0.7099	0.9010
Fetal skin_2	0.8202	0.9802	AG09309	0.7529	0.9901	NHDF_Ad	0.7029	0.9406
HAh	0.8167	1.0000	iPS_4_7	0.7502	1.0000	Fetal skin_1	0.7027	0.9703
iPS_4_7	0.8162	0.9604	Fetal heart_3	0.7497	0.9505	HNPCEpiC	0.6996	0.9604
HAsp	0.8139	0.8317	Fetal skin_1	0.7491	0.9802	AG09309	0.6986	0.9505
Fetal lung_1	0.8106	1.0000	NHA	0.7457	0.9802	HCPEpiC	0.6931	0.9703
Fetal muscle_1	0.8068	0.9505	WERI_Rb1	0.7452	1.0000	PrEC	0.6930	0.9406
Fetal kidney pelvis_1	0.8053	0.9802	HNPCEpiC	0.7407	1.0000	HPF	0.6915	0.9505
NHDF_Neo	0.7818	0.9703	AG04449	0.7382	0.9802	NHDF_Neo	0.6914	0.9505
Fetal heart_1	0.7656	0.9901	Fetal spinal cord_2	0.7379	1.0000	Fetal skin_3	0.6856	0.9406
hESCT0	0.7575	1.0000	AG09319	0.7217	1.0000	WI_38_TAM	0.6845	0.9010
Fetal lung_2	0.7571	1.0000	HAh	0.7206	1.0000	HAh	0.6758	0.9802
HRPEpiC	0.7562	1.0000	HAc	0.7175	1.0000	Fetal brain_3	0.6678	0.9109
fetal kidney cortex_2	0.7491	0.9901	BE2_C	0.7157	0.9604	HCFaa	0.6659	0.9208
Fetal heart_3	0.7426	1.0000	Fetal skin_2	0.7152	0.9901	HSMM_D	0.6645	0.9505
Fetal testes	0.7413	1.0000	HRPEpiC	0.7123	1.0000	AoAF	0.6499	0.9505
LNCap	0.7078	1.0000	Fetal heart_2	0.6964	1.0000	HBMEC	0.6406	0.9802
Fetal heart_2	0.6966	1.0000	HPdLF	0.6926	1.0000	NHLF	0.6303	0.9703
WERI_Rb1	0.6853	1.0000	AG10803	0.6891	1.0000	SkMC	0.6134	0.9802
BE2_C	0.6644	1.0000	HAsp	0.6707	1.0000	HAc	0.5984	0.9802
Fetal spinal cord_2	0.6628	1.0000	PANC1	0.6249	0.9901	HAsp	0.5936	0.9604
Fetal spinal cord_3	0.6004	1.0000	Fetal spinal cord_1	0.6149	1.0000	HSMM	0.5858	0.9505
Fetal spinal cord_1	0.5955	1.0000	Fetal spinal cord_3	0.5793	1.0000	Fetal skin_2	0.5840	0.9802
SK_N_MC	0.5866	1.0000	SK_N_MC	0.5757	0.9901	NHA	0.5674	0.9802
Fetal brain_3	0.4652	1.0000	Fetal brain_1	0.4948	1.0000	Fetal brain_2	0.5652	0.9505
Fetal brain_1	0.4547	1.0000	Fetal brain_2	0.4841	1.0000	Fetal brain_1	0.5573	0.9406
Fetal brain_2	0.4323	1.0000	Fetal brain_3	0.4675	1.0000	HRPEpiC	0.5181	1.0000

Leprosy			Schizophrenia			Colorectal cancer		
Cell type	**Enrichment**	**P value**	**Cell type**	**Enrichment**	**P value**	**Cell type**	**Enrichment**	**P value**
CD14	2.7205	0.0015	Fetal brain_3	1.4855	0.0419	HEPG2	1.8372	0.0210
NB4	2.3519	0.0113	Fetal heart_1	1.4579	0.0419	Fetal intestine small_2	1.5185	0.0323
HL60	2.0174	0.0323	Fetal heart_3	1.3968	0.0574	Fetal intestine large_2	1.4420	0.0698
CD8	2.0073	0.0419	Fetal brain_2	1.3966	0.0798	SKNSH	1.4388	0.0916
GM12878	1.9994	0.0314	Fetal brain_1	1.3762	0.0599	Fetal intestine large_3	1.3735	0.0731
CD4	1.9706	0.0219	Fetal heart_2	1.3371	0.0627	Fetal stomach_1	1.3535	0.0466
CD19	1.9555	0.0381	HCPEpiC	1.2459	0.0576	Fetal intestine small_3	1.3491	0.0764
CD34	1.9311	0.0084	HSMM_D	1.2183	0.1244	Fetal stomach_2	1.3191	0.0599

GM06990	1.8893	0.0300	Fetal adrenal_2	1.2116	0.1294	AG04450	1.3152	0.0764
Fetal thymus_1	1.8749	0.0262	HAsp	1.2036	0.0837	A549	1.3112	0.2673
Th1_1	1.8668	0.0488	Fetal adrenal_1	1.1921	0.1194	Fetal intestine large_1	1.2839	0.1194
CD3 adult	1.8074	0.0524	IMR90	1.1858	0.0876	Fetal stomach_3	1.2663	0.0876
Fetal placenta_2	1.7993	0.0200	HAc	1.1819	0.1095	WI_38	1.2588	0.1144
CD20_2	1.7679	0.0439	HIPEpiC	1.1704	0.1294	HVMF	1.2377	0.0837
Th2_1	1.7632	0.0524	AG04449	1.1585	0.1788	CD20_1	1.2331	0.1854
Th2_2	1.7522	0.0599	Fetal thymus_2	1.1576	0.2475	AG10803	1.2249	0.1391
Fetal thymus_2	1.7496	0.0419	NHA	1.1542	0.1656	AG09319	1.2140	0.1921
CD3 CordBlood	1.7388	0.0524	HCF	1.1541	0.1523	HPF	1.2023	0.1343
Fetal placenta_3	1.6404	0.0698	Fetal spinal cord_2	1.1372	0.2673	WI_38_TAM	1.2014	0.2475
GM12865	1.6246	0.0764	WI_38_TAM	1.1251	0.2871	HRE	1.1981	0.1391
WI_38_TAM	1.5752	0.0598	Fetal thymus_3	1.1228	0.2970	GM06990	1.1978	0.2970
Fetal thymus_3	1.5676	0.0956	hESCT0	1.1127	0.3366	Fetal intestine small_1	1.1903	0.2119
CACO2	1.5458	0.3069	Fetal lung_1	1.1091	0.2178	IMR90	1.1894	0.2079
Fetal placenta_1	1.5421	0.0655	CACO2	1.1077	0.2178	BJ	1.1712	0.1854
HMEC	1.5339	0.0466	HFF_MyC	1.1050	0.2871	BE2_C	1.1582	0.2970
Th1_2	1.5140	0.1045	Fetal thymus_1	1.1027	0.2772	NHEK	1.1516	0.2970
Fetal intestine large_1	1.4945	0.0574	BE2_C	1.1027	0.2277	GM12878	1.1450	0.3267
CD56	1.4529	0.0876	WI_38	1.1021	0.2178	AG09309	1.1380	0.2277
Fetal intestine small_2	1.4394	0.0876	AG04450	1.0999	0.2574	HFF	1.1303	0.1722
Th17	1.4389	0.1589	Fetal adrenal_3	1.0979	0.2772	AoAF	1.1293	0.2673
Fetal spleen	1.3920	0.0574	HFF	1.0969	0.2185	CD3 CordBlood	1.1267	0.3663
HEEpiC	1.3858	0.1391	NT2_D1	1.0932	0.4752	HMEC	1.1206	0.3069
Fetal stomach_3	1.3406	0.0698	HCM	1.0911	0.1921	NHDF_Neo	1.1202	0.3168
CMK	1.3088	0.2970	HMVEC_dLyAd	1.0891	0.2574	PrEC	1.1181	0.3366
vHMEC	1.2942	0.1457	NHLF	1.0888	0.3069	CD19	1.1157	0.4059
PrEC	1.2409	0.2772	Fetal lung_3	1.0881	0.2178	SAEC	1.1138	0.3267
Fetal intestine large_2	1.2347	0.3168	HPdLF	1.0869	0.2475	HFF_MyC	1.1080	0.2772
Fetal stomach_1	1.2337	0.1921	Mesendoderm	1.0807	0.2871	NHA	1.1037	0.3564
SAEC	1.1818	0.2673	AoAF	1.0759	0.2871	HRCE	1.0971	0.2871
Fetal intestine small_3	1.1380	0.3168	HPAEC	1.0758	0.2871	HConF	1.0915	0.2772
HRE	1.1084	0.2475	H1_P18	1.0719	0.2772	HCT116	1.0898	0.3168
Fetal stomach_2	1.0736	0.3267	HMVEC_LLy	1.0717	0.2475	CD20_2	1.0892	0.3762
NHEK	1.0602	0.3465	HMVEC_dAd	1.0711	0.3267	Fetal skin_2	1.0880	0.2673
H1_P18	1.0591	0.4158	Fetal muscle_2	1.0693	0.2673	HAEpiC	1.0851	0.2673
Fetal intestine large_3	1.0566	0.3465	HMVEC_LBl	1.0652	0.2673	CD8	1.0826	0.3564
Fetal adrenal_3	1.0519	0.5050	HAh	1.0613	0.3465	NHDF_Ad	1.0820	0.3069
CD20_1	1.0423	0.4257	iPS_4_7	1.0606	0.3267	Fetal skin_1	1.0798	0.3564
LNCap	1.0116	0.3267	HCFaa	1.0595	0.3960	Th17	1.0787	0.2970
Fetal kidney pelvis_3	1.0028	0.5149	HMVEC_dBlNeo	1.0592	0.3366	NHLF	1.0724	0.3465
HMVEC_dNeo	0.9871	0.4950	NPC	1.0544	0.4257	CD3 adult	1.0700	0.3465

HEPG2	0.9847	0.4752	H9_P42	1.0497	0.4455	Hela	1.0687	0.4554
Fetal lung_3	0.9839	0.5347	BJ	1.0462	0.3861	HMF	1.0675	0.3564
HESC	0.9815	0.4455	HUVEC	1.0427	0.4356	CACO2	1.0593	0.4851
HRCE	0.9791	0.5446	HPAF	1.0387	0.2475	HPdLF	1.0588	0.3168
Fetal intestine small_1	0.9731	0.5050	HNPCEpiC	1.0345	0.4059	GM12865	1.0549	0.4257
HConF	0.9660	0.5149	Fetal spleen	1.0344	0.3762	Fetal adrenal_3	1.0546	0.3465
iPS_6_9	0.9587	0.5446	iPS_6_9	1.0341	0.2772	Fetal thymus_2	1.0528	0.3267
HCT116	0.9562	0.5743	Fetal skin_3	1.0293	0.4653	PANC1	1.0467	0.5050
Fetal skin_3	0.9445	0.5644	Fetal testes	1.0264	0.4356	Fetal skin_3	1.0415	0.4455
AG04450	0.9426	0.5545	AG10803	1.0253	0.4950	HEEpiC	1.0394	0.4059
Fetal muscle_2	0.9390	0.6337	Fetal spinal cord_1	1.0224	0.4950	HNPCEpiC	1.0388	0.4554
Fetal lung_1	0.9336	0.5545	HSMM	1.0190	0.4554	Mesendoderm	1.0377	0.4257
HMVEC_LBl	0.9329	0.5248	Hela	1.0116	0.4950	HMVEC_dLyNeo	1.0344	0.3366
HMVEC_LLy	0.9314	0.4950	HBMEC	1.0086	0.4455	Fetal kidney pelvis_2	1.0344	0.3564
HPdLF	0.9259	0.6139	Fetal lung_2	1.0078	0.4653	AG04449	1.0291	0.3366
HMVEC_dAd	0.9189	0.5743	NHDF_Ad	1.0078	0.4851	HGF	1.0284	0.3960
Fetal kidney pelvis_1	0.9146	0.6337	AG09309	1.0053	0.5050	Fetal kidney pelvis_3	1.0144	0.3663
PANC1	0.9096	0.3168	Fetal spinal cord_3	1.0045	0.5842	HPAF	1.0042	0.4653
Fetal muscle_3	0.9087	0.6238	HGF	1.0018	0.4257	HMVEC_dBlAd	1.0025	0.4653
Fetal skin_2	0.8951	0.7030	HESC	0.9986	0.5347	SkMC	1.0010	0.4059
SKNSH	0.8918	0.6040	HRGEC	0.9986	0.4356	CD4	0.9992	0.4752
SkMC	0.8815	0.5842	WERI_Rb1	0.9978	0.5446	NPC	0.9979	0.3960
HMVEC_dLyAd	0.8787	0.5743	HVMF	0.9973	0.4158	HBMEC	0.9901	0.3564
WI_38	0.8768	0.5545	HMVEC_dLyNeo	0.9933	0.5248	HMVEC_LBl	0.9872	0.6040
Fetal adrenal_2	0.8766	0.7228	Fetal stomach_3	0.9925	0.5248	H1_P18	0.9840	0.5149
iPS_19_7	0.8750	0.6436	CMK	0.9896	0.4257	Fetal placenta_1	0.9820	0.5347
H9_P42	0.8750	0.5347	HEPG2	0.9881	0.3366	MCF7	0.9788	0.6535
Fetal kidney cortex_1	0.8716	0.7624	HPF	0.9864	0.4455	Fetal kidney cortex_3	0.9767	0.4851
HBMEC	0.8669	0.6931	Fetal kidney cortex_3	0.9862	0.5050	CMK	0.9750	0.4950
HPAEC	0.8616	0.6337	HMVEC_dBlAd	0.9857	0.4554	LNCap	0.9701	0.5149
Fetal testes	0.8591	0.6733	Fetal kidney cortex_1	0.9826	0.5545	Fetal testes	0.9701	0.5842
HIPEpiC	0.8588	0.6634	A549	0.9814	0.2970	iPS_19_7	0.9687	0.4257
A549	0.8506	0.4257	HConF	0.9768	0.5149	Fetal lung_3	0.9664	0.4752
Fetal kidney pelvis_2	0.8483	0.7723	iPS_19_7	0.9766	0.5446	Fetal adrenal_2	0.9656	0.5545
HAEpiC	0.8480	0.7030	HMVEC_dNeo	0.9713	0.4653	HCFaa	0.9629	0.5050
K562	0.8425	0.5941	SkMC	0.9707	0.5644	Fetal placenta_2	0.9606	0.4851
Fetal spinal cord_3	0.8409	0.7327	HAEpiC	0.9694	0.4851	Fetal muscle_3	0.9580	0.5149
Fetal kidney cortex_3	0.8379	0.6931	Fetal skin_1	0.9684	0.5743	RPTEC	0.9521	0.5050
Mesendoderm	0.8375	0.7030	HRPEpiC	0.9675	0.5545	Fetal kidney cortex_1	0.9515	0.6238
Fetal skin_1	0.8297	0.7228	Fetal kidney pelvis_2	0.9665	0.6634	iPS_6_9	0.9506	0.5347
IMR90	0.8267	0.7426	Fetal stomach_2	0.9658	0.5347	vHMEC	0.9466	0.6238
NHDF_Ad	0.8255	0.7921	Fetal kidney pelvis_3	0.9504	0.6436	Fetal adrenal_1	0.9463	0.5842

AG09309	0.8238	0.7228	Fetal kidney pelvis_1	0.9469	0.7030	HMVEC_dAd	0.9455	0.5149
Fetal adrenal_1	0.8222	0.7723	HMF	0.9449	0.6436	HCPEpiC	0.9414	0.5842
HGF	0.8123	0.7624	AG09319	0.9432	0.5545	Fetal muscle_2	0.9399	0.5248
HNPCEpiC	0.8094	0.7723	Th2_2	0.9414	0.5842	HCF	0.9378	0.5347
HFF	0.8034	0.8020	Fetal placenta_1	0.9317	0.6931	HESC	0.9361	0.5842
fetal kidney cortex_2	0.8009	0.7822	Fetal muscle_3	0.9302	0.5545	HMVEC_dBlNeo	0.9326	0.5941
NT2_D1	0.8002	0.8317	LNCap	0.9277	0.7228	NB4	0.9295	0.5743
Fetal lung_2	0.7941	0.8119	HRE	0.9239	0.7822	HL60	0.9213	0.6040
HMVEC_dLyNeo	0.7922	0.7129	fetal kidney cortex_2	0.9164	0.7426	HSMM_D	0.9206	0.5347
NHLF	0.7883	0.7129	Fetal intestine large_1	0.9155	0.6832	H9_P42	0.9190	0.5743
Fetal spinal cord_2	0.7879	0.8614	CD19	0.9088	0.6139	NT2_D1	0.9183	0.6337
iPS_4_7	0.7850	0.8416	SKNSH	0.9078	0.5149	Fetal placenta_3	0.9119	0.5545
MCF7	0.7823	0.6139	CD56	0.9027	0.6139	Fetal thymus_3	0.9082	0.4950
HRPEpiC	0.7812	0.7426	Fetal muscle_1	0.8961	0.6535	CD14	0.9061	0.5644
HSMM_D	0.7810	0.8416	CD14	0.8894	0.6337	Fetal kidney pelvis_1	0.9058	0.6733
HCFaa	0.7802	0.6634	Fetal skin_2	0.8841	0.8317	HCM	0.9031	0.6238
HVMF	0.7801	0.7525	Th17	0.8787	0.5050	hESCT0	0.8988	0.7030
AG04449	0.7785	0.8218	K562	0.8785	0.6535	iPS_4_7	0.8965	0.6238
AG10803	0.7753	0.7822	CD34	0.8764	0.6634	CD34	0.8915	0.6040
HSMM	0.7598	0.8119	Th1_1	0.8648	0.6238	HMVEC_dNeo	0.8895	0.5446
Hela	0.7549	0.7129	PANC1	0.8611	0.7129	HSMM	0.8873	0.6733
Fetal heart_2	0.7506	0.7525	RPTEC	0.8538	0.7921	Fetal heart_1	0.8869	0.5842
Fetal heart_3	0.7460	0.6931	HRCE	0.8536	0.8911	fetal kidney cortex_2	0.8759	0.6832
HCM	0.7424	0.8515	NHEK	0.8496	0.7228	HAsp	0.8756	0.6337
HCPEpiC	0.7418	0.8713	NHDF_Neo	0.8468	0.8317	Fetal lung_1	0.8753	0.6238
Fetal heart_1	0.7293	0.7327	HL60	0.8442	0.7327	Th2_1	0.8741	0.5743
RPTEC	0.7223	0.8218	CD20_2	0.8434	0.7426	K562	0.8727	0.6436
Fetal spinal cord_1	0.7199	0.8317	GM12878	0.8419	0.6337	HAc	0.8697	0.7030
BJ	0.7167	0.8317	Fetal placenta_2	0.8397	0.7822	Fetal lung_2	0.8673	0.7426
HMVEC_dBlAd	0.7096	0.6832	NB4	0.8307	0.7426	Fetal muscle_1	0.8671	0.6139
HPAF	0.7085	0.9010	Fetal intestine large_2	0.8303	0.8515	HRPEpiC	0.8613	0.6634
AG09319	0.7043	0.8713	CD3 CordBlood	0.8300	0.5743	HRGEC	0.8570	0.6436
HMF	0.7022	0.8515	HCT116	0.8208	0.8218	Th1_2	0.8525	0.4950
NPC	0.7005	0.7921	Jurkat	0.8118	0.7921	HAh	0.8521	0.8317
NHA	0.6953	0.8614	Fetal stomach_1	0.8101	0.8317	HUVEC	0.8418	0.6337
hESCT0	0.6926	0.9307	Fetal intestine small_2	0.8094	0.8812	WERI_Rb1	0.8349	0.6931
Fetal brain_1	0.6657	0.8218	Fetal intestine large_3	0.7998	0.8614	HIPEpiC	0.8346	0.6535
HPF	0.6596	0.9604	SK_N_MC	0.7985	0.7723	Th1_1	0.8294	0.6535
AoAF	0.6584	0.9010	Fetal intestine small_1	0.7884	0.9604	Fetal spinal cord_3	0.8281	0.7129
HMVEC_dBlNeo	0.6574	0.8911	Th2_1	0.7784	0.7426	Fetal spleen	0.8214	0.8020
HAh	0.6547	0.9802	GM06990	0.7737	0.7426	Fetal thymus_1	0.8138	0.6931
HAsp	0.6314	0.9505	CD8	0.7611	0.7525	HMVEC_LLy	0.8098	0.7228

HFF_MyC	0.6296	0.9703	vHMEC	0.7575	0.9406	HPAEC	0.8015	0.7822
HUVEC	0.6293	0.8119	SAEC	0.7466	0.9703	Jurkat	0.7885	0.7426
HCF	0.6255	0.9208	CD20_1	0.7449	0.9406	SK_N_MC	0.7754	0.8614
HAc	0.6148	0.9802	Fetal placenta_3	0.7418	0.8515	Th2_2	0.7564	0.7624
Fetal muscle_1	0.6088	0.9010	Fetal intestine small_3	0.7304	0.9703	Fetal heart_3	0.7529	0.7525
Fetal brain_3	0.6083	0.9406	Th1_2	0.7233	0.7624	Fetal heart_2	0.7500	0.8515
SK_N_MC	0.5716	0.9010	MCF7	0.7066	0.8317	HMVEC_dLyAd	0.7484	0.7327
NHDF_Neo	0.5300	0.9703	HEEpiC	0.7058	0.9901	Fetal spinal cord_2	0.7428	0.8812
HRGEC	0.5008	0.9109	GM12865	0.7043	0.8614	CD56	0.7344	0.7723
Jurkat	0.4972	0.9109	CD3 adult	0.7037	0.7624	Fetal spinal cord_1	0.7299	0.8317
Fetal brain_2	0.4942	0.9703	PrEC	0.7035	0.9703	Fetal brain_1	0.6491	0.9109
WERI_Rb1	0.4568	0.9901	HMEC	0.6689	0.9802	Fetal brain_2	0.6339	0.9604
BE2_C	0.3817	1.0000	CD4	0.6488	0.8515	Fetal brain_3	0.5725	0.9703

Supplementary table 7: Enrichment of MS confirmed and suggestive regions

Cell type	MS_confirmed_enrichment	MS_suggestive_enrichment
Th1_2	2.563770349	1.312385143
Th1_1	2.458228853	1.856194078
CD8	2.449870975	1.765101069
Th17	2.442527865	1.448288792
CD3 CordBlood	2.363011834	1.759645197
Th2_2	2.341522975	1.33773552
CD3 adult	2.320557001	1.619070676
Th2_1	2.295378958	1.612653123
CD19	2.245732468	1.722692295
CD4	2.214773384	1.512712951
GM12878	2.140208538	1.506025597
CD56	2.124595462	1.384192426
CD20_2	2.099424435	1.71307989
GM06990	2.08269932	1.524337814
GM12865	1.957816446	1.278199086
CD14	1.905027947	1.310916659
Fetal thymus_2	1.627580014	1.607850362
NB4	1.563661328	1.274785525
Fetal thymus_3	1.560137578	1.690691231
Fetal thymus_1	1.536998273	1.665789104
HL60	1.51809435	1.172108314
CD34	1.37053091	1.38707281
Fetal spleen	1.364782489	1.411526201
HCT116	1.342486946	0.964933819
Jurkat	1.312035688	1.644052044
CACO2	1.300311547	1.013858906
HEPG2	1.286222814	1.222388165
CMK	1.275032308	1.225186421
A549	1.256887885	0.87817096
K562	1.090834757	0.996974958
NHEK	1.075430326	0.883249594
Fetal stomach_1	1.067358983	1.061533123
HMEC	1.036977999	0.847154504
Fetal adrenal_3	1.033043038	0.832212581
PANC1	1.03036255	0.885421728
HConF	1.021076387	1.003145661
HPF	1.015898689	0.981280319
Fetal stomach_2	1.015848616	0.98676399
HPAF	0.997397722	0.96546027
Fetal stomach_3	0.9920174	1.017352032
HMVEC_LBl	0.991497372	1.028724203
MCF7	0.990057855	0.726763694
HMF	0.988591371	0.937012116
Fetal intestine small_2	0.987850075	0.994728945
HRCE	0.982774148	0.977788605
HRE	0.98213258	0.998903107
NPC	0.978388062	1.016604927
HIPEpiC	0.97621067	0.952401673
HCPEpiC	0.974692781	0.848385682
AG09319	0.973988451	1.097960572
HVMF	0.969192581	0.902145886
HMVEC_dNeo	0.968031808	1.11578838
HAEpiC	0.967407542	0.857164953
WI_38_TAM	0.962591558	0.912400809
HMVEC_dLyNeo	0.961673658	1.1185034
Fetal placenta_3	0.961083108	0.892026307
HFF	0.960463351	1.050437524
Fetal placenta_2	0.960184012	0.929624442
HCF	0.958614076	0.952703411
HFF_MyC	0.957187301	0.96811871

Fetal placenta_1	0.956540709	0.949892965
HMVEC_dAd	0.954094768	1.097515017
HMVEC_LLy	0.951669425	1.075446882
HMVEC_dLyAd	0.939852816	1.077501363
HCM	0.937956262	0.886639273
WI_38	0.937506317	1.026742189
HPAEC	0.934493119	1.151807154
AG04450	0.934350378	1.041790021
HGF	0.917216004	0.964395835
Fetal intestine large_2	0.916463084	0.97896145
IMR90	0.916383891	0.962002935
HMVEC_dBlAd	0.912982685	1.035973549
Fetal intestine large_3	0.909372771	0.912860629
RPTEC	0.906704566	0.919009483
HNPCEpiC	0.902664148	0.972632704
AG10803	0.900388168	1.075320137
HEEpiC	0.900293217	0.832245208
H1_P18	0.899741837	0.983248876
Fetal kidney cortex_3	0.89960963	0.993824449
vHMEC	0.89871821	0.845417809
Fetal intestine large_1	0.898317382	1.014930085
AoAF	0.898201662	0.974385458
SAEC	0.897777187	0.841276048
PrEC	0.8935349	0.84146259
Hela	0.893184572	0.625794914
Fetal muscle_3	0.891574134	0.989551138
Fetal kidney pelvis_3	0.890694642	1.085359475
NHA	0.888496964	0.921998096
Fetal adrenal_1	0.8829552	0.820084264
SKNSH	0.879415916	1.093433384
Fetal adrenal_2	0.877223885	0.853453009
CD20_1	0.875616617	0.877708594
HMVEC_dBlNeo	0.875016359	1.094891241
Fetal intestine small_1	0.870887483	0.913860933
HCFaa	0.867882518	0.881536954
AG09309	0.867175442	0.941904651
H9_P42	0.867125253	0.986577757
Fetal lung_3	0.866233924	0.940737818
HPdLF	0.865932756	0.938872422
Fetal muscle_2	0.86565054	0.950753217
iPS_6_9	0.861849205	1.009186521
HESC	0.860833671	0.963059385
HAc	0.855058694	0.999712938
Fetal intestine small_3	0.853940837	0.901684546
HBMEC	0.853811926	1.003013385
Fetal kidney pelvis_1	0.85318857	0.934225816
NHLF	0.851821166	0.957824527
NT2_D1	0.851803488	0.934245871
Fetal skin_1	0.8455331	0.929369832
HSMM_D	0.836883884	0.832918055
Fetal kidney pelvis_2	0.826084517	0.951173484
Fetal lung_1	0.8230249	0.940380934
HUVEC	0.822720349	1.038273799
HRPEpiC	0.815086813	0.861611092
iPS_4_7	0.814378027	0.919448711
HRGEC	0.813116473	1.069437457
Fetal kidney cortex_1	0.809100803	0.914741403
BJ	0.808125607	0.873063414
Fetal testes	0.799770981	0.973843713
AG04449	0.7977001	0.911192179
iPS_19_7	0.79686683	1.032765784
Fetal skin_2	0.796819053	0.930974553
Fetal lung_2	0.792836401	0.900264273

NHDF_Ad	0.787335709	0.899218174
Mesendoderm	0.783711896	1.015716774
HAh	0.782339093	0.956859814
NHDF_Neo	0.774247888	0.961629001
hESCT0	0.770161928	0.889971057
SkMC	0.768125085	0.905419162
Fetal spinal cord_2	0.765637497	0.77280259
Fetal skin_3	0.75938229	0.89132125
Fetal muscle_1	0.756117531	0.982325367
HAsp	0.753359737	0.841455102
HSMM	0.74940984	0.78427509
Fetal heart_1	0.743865283	0.911092372
fetal kidney cortex_2	0.70908137	0.940361074
Fetal spinal cord_3	0.704800439	0.745613441
Fetal heart_2	0.703808346	0.841213634
SK_N_MC	0.701365246	0.819325293
Fetal heart_3	0.698442677	0.863866392
Fetal spinal cord_1	0.665041058	0.751342652
BE2_C	0.659722015	0.81235976
LNCap	0.615726477	0.744077188
WERI_Rb1	0.615319812	0.856721329
Fetal brain_1	0.561309827	0.669570286
Fetal brain_3	0.54742086	0.605467092
Fetal brain_2	0.518316132	0.66471717

ACKNOWLEDGMENTS

It would not have been possible to perform this work and write this doctoral thesis without the help and support of several people around me that I wish to give particular mention here.

Thanks to my supervisor Dr Sreeram V Ramagopalan. Sreeram, it is so difficult to put into words the gratitude I feel towards you. I have learnt so much by just being around you and talking to you about science and I know I am not the only person who has experienced this and feels the same. Your passion for research has been an extraordinary inspiration for me and I have always been amazed by how your intelligence, humility and generosity are able to change the life of the people who have been lucky enough to meet you and (as in my case) work with you. Without your guidance, persistent help and friendship this DPhil would not have been possible.

Thanks to Prof George C Ebers for opening the door of his lab and giving me the chance to come to Oxford without knowing anything about me. It has been a real honour to work in your group and I will carry this memory for the rest of my life. I sincerely hope I have been able to pay back at least partly the unconditional trust you had in me.

Thanks to my colleagues and friends of the Ebers group: Julia Morahan, Katie Morrison, Antonio Berlanga, Lahiru Handunnetthi, Julia Pakpoor and Adam Handel. Thanks for the time spent discussing science. You have been amazing colleagues, but more important you have been and are among my best and closest friends.

Thanks to Dr Geir Kjetil Sandve from University of Oslo for his invaluable help in many of the analyses that are part of this thesis. It has been amazing to work with you.

Thanks to Merton College and to the Italian Multiple Sclerosis Society for their extraordinary financial support that has been indispensable for me to live and do research in Oxford.

Thanks to Young Chae and to Chiara and Roberto. I wonder whether I should have acknowledged you together with my family members because you really have been a second family for me in the United Kingdom.

Thanks to my family. Thanks to my parents Andrea Disanto and Donatella Spina, to my brother Filippo, my sister Maria Giulia and my grandma Anna. I am perfectly aware of the privations and efforts you had to go through to support me and give me the chance to study abroad. Thanks for teaching me that nothing comes without hard work, honesty and passion in what one does.

Giulio