Meta-analysis of genome-wide association studies reveals genetic overlap between Hodgkin lymphoma and multiple sclerosis

Pouya Khankhanian,1,2 Wendy Cozen,3 Daniel S Himmelstein,2 Lohith Madireddy,2 Lennox Din,2 Anke van den Berg,4 Takuya Matsushita,2 Sally L Glaser,5 Jayaji M Moré,2 Karin E Smedby,6 Sergio E Baranzini,2 Thomas M Mack,3 Antoine Lizée,2 Silvia de Sanjose,7 Pierre-Antoine Gourraud,2 Alexandra Nieters,8 Stephen L Hauser,2 Pierluigi Cocco,9 Marc Maynadie,10 Lenka Foretová,11 Anthony Staines,12 Manon Delahaye-Sourdeix,13 Dalin Li,3 Smita Bhatia,14 Mads Melbye,15 Kenan Onel,16 Ruth Jarrett,17 James D McKay,13 Jorge R Oksenberg2 and Henrik Hjalgrim15*

1Center for Neuroengineering and Thereapeutics at the University of Pennsylvania, Philadelphia, PA, USA, 2University of California, San Francisco, CA, USA, 3University of Southern California, Los Angeles, CA, USA, 4University Medical Center, Groningen, The Netherlands, 5Cancer Prevention Institute of California, Fremont, CA, USA, 6Karolinska Institute, Stockholm, Sweden, 7Catalan Institute of Oncology, L’Hospitalet de Llobregat, Catalonia, Spain, 8University Medical Center Freiburg, Freiburg, Germany, 9University of Cagliari, Cagliari, Sardinia, Italy, 10Centre Hospitalier Universitaire de Dijon, Dijon, France, 11Masaryk Memorial Cancer Institute, Brno, Czech Republic, 12University College Dublin, Dublin, Ireland, 13International Agency for Research on Cancer, Lyon, France, 14City of Hope National Medical Center, Duarte, CA, USA, 15Statens Serum Institut, Copenhagen S, Denmark, 16University of Chicago, Chicago, IL, USA, 17MRC, University of Glasgow Centre for Virus Research, Glasgow, UK

*Corresponding author. Statens Serum Institut, 5 Artillerivej, DK-2300 Copenhagen S, Denmark. E-mail: HHJ@ssi.dk

Accepted 18 December 2015

Abstract

Background: Based on epidemiological commonalities, multiple sclerosis (MS) and Hodgkin lymphoma (HL), two clinically distinct conditions, have long been suspected to be aetiologically related. MS and HL occur in roughly the same age groups, both are associated with Epstein-Barr virus infection and ultraviolet (UV) light exposure, and they cluster mutually in families (though not in individuals). We speculated if in addition to sharing environmental risk factors, MS and HL were also genetically related. Using data from genome-wide association studies (GWAS) of 1816 HL patients, 9772 MS patients and 25 255 controls, we therefore investigated the genetic overlap between the two diseases.

Methods: From among a common denominator of 404 K single nucleotide polymorphisms (SNPs) studied, we identified SNPs and human leukocyte antigen (HLA) alleles...
independently associated with both diseases. Next, we assessed the cumulative genome-wide effect of MS-associated SNPs on HL and of HL-associated SNPs on MS. To provide an interpretational frame of reference, we used data from published GWAS to create a genetic network of diseases within which we analysed proximity of HL and MS to autoimmune diseases and haematological and non-haematological malignancies.

**Results:** SNP analyses revealed genome-wide overlap between HL and MS, most prominently in the HLA region. Polygenic HL risk scores explained 4.44% of HL risk (Nagelkerke R²), but also 2.36% of MS risk. Conversely, polygenic MS risk scores explained 8.08% of MS risk and 1.94% of HL risk. In the genetic disease network, HL was closer to autoimmune diseases than to solid cancers.

**Conclusions:** HL displays considerable genetic overlap with MS and other autoimmune diseases.

**Introduction**

Hodgkin lymphoma (HL) is an immunologically active malignant neoplasm of B cells in a heterogeneous reactive cellular infiltrate. HL is roughly equally common in men and women, and in socioeconomically affluent populations HL occurrence displays a bimodal age distribution, with separate peaks in younger adults (15-34 years old) and older adults (over 50 years old), In socioeconomically deprived settings, in contrast, there is no young adult HL incidence peak, but rather one among children.

Multiple sclerosis (MS) is a debilitating disease of the central nervous system (CNS) characterized by chronic polyclonal inflammation (including T cells, monocytes and B cells), myelin loss, gliosis, axonal and oligodendrocyte pathology and accumulation of progressive neurological disability. Onset is typically between the ages of 20 to 40 years, with a female to male ratio of three to one.

Based on conspicuous epidemiological similarities between the two conditions, e.g. regarding age patterns and geographical distributions, Newell in 1970 proposed that HL and MS were somehow aetiologically related. Subsequent epidemiological studies have tested this hypothesis by assessing clustering of HL with MS in individuals and in families. Whereas previous studies generally suggest that patients suffering from either condition are not at increased risk of the other, two partially overlapping investigations have reported mutual clustering of the two diseases among first-degree relatives.

Familial clustering of HL and MS may reflect shared environmental and genetic risk factors. Evidence implicates infection with the Epstein-Barr virus (EBV) and levels of ultraviolet light exposure in the pathogenesis of the roughly one-third of HL cases that harbour EBV in the malignant cells (EBV-positive HL), as well as in the pathogenesis of MS. Moreover, common genetic risk factors have emerged in HL and MS. For example, HLA-A*02 appears to be associated with a decreased risk of both MS and EBV-positive HL, and DNA variants in the Relish oncogene (REL, a member of the NF-kappaB transcription factor family) have been associated with both MS and the EBV-negative subset of HL. This suggests that the relationship between HL and MS is not limited to either EBV-positive or EBV-negative HL.

Unveiling of aetiological commonalities for HL and MS could contribute to the understanding of their pathogenesis and might even have clinical implications. We therefore combined data from previous genome-wide association studies (GWAS) of MS and HL and evaluated the genetic overlap between the two diseases.

**Methods**

**Overview**

Analysis was performed on a total of 1816 HL patients, 9772 MS patients and 25255 controls, using 404 K single
nucleotide polymorphisms (Figure 1). We first sought to identify single nucleotide polymorphisms (SNPs) and HLA alleles that associated independently with both diseases. Next, we calculated polygenic risk scores to assess the cumulative genome-wide effect of MS-associated SNPs on HL, and of HL-associated SNPs on MS. To describe the overlap between HL and MS, we performed a protein interaction network-based pathway analysis (PINBPA) on associated genes from each disease, and then investigated the intersection of networks for biological relevance. To place the genetic similarity between HL and MS in context, we used data from previously reported GWAS to create a diseasome: a network of diseases. Within the diseasome, we analysed proximity of HL and MS to autoimmune diseases, haematological malignancies and non-haematological malignancies (‘solid’ cancers).

**HL and MS dataset characteristics**

The HL26,27,33,34 and MS28 cohorts have previously been described in detail. For MS, data from the Wellcome Trust Case Control Consortium 2 meta-analysis project totalled 9772 cases and 17 376 controls. Individuals in this dataset were of European descent and originated from 15 geographical regions, including the USA, Australia, New Zealand and numerous European countries. Included in this dataset were summary-level association results for a total of 464 434 SNPs. For HL, data from a recent meta-analysis of three GWAS studies consisted of 1816 cases and 7879 controls.34 Individuals in this dataset were of European descent and originated from locations in the USA and Europe. Summary-level meta-association results were included for a total of 1 036 304 SNPs. The cases were subdivided into nodular lymphocytic predominant and classical HL, and classical HL further divided into subgroups by EBV tumour status (EBV-positive and EBV-negative) as determined by immunohistochemistry or *in situ* hybridization as previously described27 and histology [mixed cellularity (MC), nodular sclerosis (NS) and other or unspecified], when such data were available. Summary characteristics are shown in Table S1 (available as Supplementary data at IJE online).

The HL and MS datasets were merged (by rsID), giving a final dataset containing summary-level results for 404 069 overlapping SNPs.

**Overlap between diseases: SNP-level**

We followed a procedure similar to that used in other meta-analyses of complex genetic diseases.35 To assess genetic
overlap between HL and MS, we sought SNPs that associated independently with both diseases. We identified top (P-value < 5 x 10^{-8}) independent (r^2 < 0.1 in CEU) MS-susceptibility SNPs, and determined how many of these SNPs are associated with HL after correction by Benjamini-Hochberg method (corrected P-value < 0.05). The process was repeated for increasingly liberal values of the P-value threshold for defining MS susceptibility SNPs (ranging from P-value < 5 x 10^{-8} to P-value < 5 x 10^{-3}). Corroponding analysis was repeated after switching roles for HL and MS. The HLA region was analysed in further detail using imputed classical HLA alleles (see Supplementary methods for details, available as Supplementary data at IJE online).

Supplementary analysis considered MS SNPs on subsets of the HL dataset as defined by tumour EBV status (EBV-positive HL and EBV-negative HL), tumour histology [nodular sclerosis (NS), mixed cellularity (MC)], or tumour histology combined with age (NS among 15-35 year olds), in order to explore possible heterogeneity among the HL samples.

Overlap between diseases: polygenic risk
Polygenic risk scores were calculated to test the cumulative effect of SNPs associated with HL on MS and vice versa, as described in detail in other complex genetic diseases. For each trait (HL and MS), sets of top independent SNPs were chosen as described above. Multiple sclerosis genetic burden (MSGB) and Hodgkin lymphoma genetic burden (HLGB) were calculated for each individual: the weighted sum of the number of risk alleles at each SNP in the set, weighted by the log-odds ratio of association for each SNP. We assessed the ability of MSGB to distinguish HL cases from controls and the ability of HLGB to distinguish MS cases from controls using the Nagelkerke’s R^2 (note that the P-values of the linear regression models will largely be driven by large sample sizes, and the biological significance lies in the R^2 value rather than the P-value). This analysis was repeated for subgroups of HL (EBV-positive, EBV-negative, MC and NS).

Protein interaction network-based pathway analysis (PINBPA)
To visualize the sets of interacting genes found to associate with both HL and MS, a protein interaction network-based pathway analysis (PINBPA) was performed using methods described previously. Sub-networks of aggregate score of three or greater were chosen as associated. Network discovery was performed independently in HL and MS, networks of score three or greater being chosen as associated. The intersection of the HL and MS networks was visualized. Gene ontology analysis was performed on genes in this intersection.

Diseasome analysis
To further assess genetic similarity between HL and MS, a representation of the human diseasome (network of diseases) was constructed in which diseases (nodes) were connected by the extent of their shared genetic aetiology (edges) as reported by the GWAS catalogue. This network is termed the diseasome. Diseases were manually classified as haematological malignancies, solid cancers or autoimmune diseases. Pairwise proximity measures between diseases were calculated as described in Supplementary methods. Relative distances between haematological malignancies, solid cancers, and autoimmune diseases were tested by t-test.

Results
SNP and HLA allele overlap between HL and MS
We identified SNPs associated with MS across multiple P-value thresholds ranging from P-value < 5 x 10^{-8} to P-value < 5 x 10^{-2}. Among these SNPs, we then identified those that were also associated with HL (FDR < 0.05; false discovery rate with Benjamini-Hochberg adjustment of P-values for the total number of SNPs tested at each threshold) (Table 1).

At a threshold of P-value < 5 x 10^{-8}, 429 SNPs were associated with MS; 36 of these 429 were independent (r^2 < 0.1), and three of these 36 were associated with HL at an FDR < 0.05 (Benjamini-Hochberg correction for 36 tests), summarized in row 1 of Table 1. Panel 1 of Table 1 shows results for other P-value thresholds and panel 2 shows results when top HL SNPs were tested for association in MS. SNPs found to be overlapping (final column of Table 1) are described in Table 2 (after HLA is removed). While the actual number and proportion of overlapping SNPs varied by the P-value threshold, the majority of overlapping SNPs belonged to genes in the HLA region of chromosome 6; however, several mutually associated non-HLA SNPs were also detected (Table 2; Figure S1, Table S2, available as Supplementary data at IJE online). It should be noted that the direction of association was not taken into account in this analysis, which reveals only shared genetic risk loci (see genetic burden analysis below, which accounts for direction of association).

The SNP-level overlap between diseases was repeated for each HL subgroup: MS versus EBV-positive HL, MS versus EBV-negative HL, MS versus NS-HL, MS versus NS-HL in...
15 to 35-year-olds and MS versus MC-HL. These analyses revealed no major differences between HL subgroups.

Given the strong genetic overlap at HLA, HLA alleles were imputed from SNP information via the HIBAG algorithm (see Supplementary methods, available as Supplementary data at IJE online). Figure 2 demonstrates overlap between risk alleles for EBV-negative HL and risk alleles for MS, as well as overlap between protective alleles for EBV-negative HL and protective alleles for MS. In contrast, there is no overlap between risk alleles for EBV-positive HL and risk alleles for MS, whereas there is overlap between protective alleles for EBV-positive HL and protective alleles for MS. Table S3 (available as Supplementary data at IJE online) provides details of HLA allelic P-values and odds ratios in each disease. Analysis of NS-HL revealed a pattern similar to EBV-negative HL.

### Polygenic risk overlap between diseases

To assess the extent of genetic risk overlap between HL and MS at the genome-wide level (including HLA), polygenic risk scores, termed MS genetic burden (MSGB) and HL genetic burden (HLGB) were calculated in all cases of MS, all cases of HL and all controls used in the MS study. Figure 3A shows the MSGB (y-axis) in each population. As expected, the MSGB was higher in MS cases than in controls ($P$-value $< 1.0 \times 10^{-200}$) and explained 8.08% of the risk for MS (Nagelkerke's $R^2$). However, the MSGB was also higher in HL cases than in controls ($P$-value $< 2.8 \times 10^{-35}$) and explained 1.94% of the risk for HL. Figure 3B shows the HLGB (y-axis) in each population. As expected, the HLGB was higher in HL cases than in controls ($P$-value $< 2.8 \times 10^{-81}$) and explained 4.44% of the risk for HL. However, HLGB was also higher in MS cases compared with controls ($P$-value $< 2.0 \times 10^{-121}$) and explained 2.36% of the MS risk. Results shown here use a threshold of $P$-value $< 5 \times 10^{-6}$ for including SNPs in the polygenic risk score, which results in 76 independent SNPs used for MSGB and 17 independent SNPs used for HLGB. Similar results held true at other $P$-value thresholds. We observed no major differences among HL subgroups.

### Pathway analysis

To generate hypotheses about potential functional pathways that are common to HL and MS, we carried out PINBPA in each independent dataset based on the
We found 100 associated modules (with threshold of score $> 3$) in MS comprising 1404 genes and 4050 edges, and 100 associated modules in HL comprising 1049 genes and 3652 edges. The intersection of the HL and MS networks yielded a network of 430 genes and 1066 edges. A gene ontology (GO) analysis of this intersection network using the software binGO revealed enrichment in JUN kinase activity, antigen processing and presentation, peptidyl-tyrosine phosphorylation and lymphocyte-mediated immunity (Figure 4; Table S4, available as Supplementary data at IJE online). When the analysis was repeated after a seven mega-base region surrounding the HLA was masked, the antigen processing and presentation pathway was no longer associated but JUN kinase activity, peptidyl-tyrosine phosphorylation and lymphocyte-mediated immunity remained.

**Diseasome analysis**

To assess the relative position of HL and MS among other autoimmune diseases and cancers (in terms of shared genetic risk), pairwise proximities were calculated among 37 complex autoimmune diseases, solid cancers and haematological malignancies (Table 3), where proximity is a network-based relatedness measure derived from shared GWAS loci between diseases. Autoimmune diseases showed more genetic

### Table 2. Non-HLA SNPs associated with both HL and MS at decreasing thresholds

<table>
<thead>
<tr>
<th>MS P-value threshold</th>
<th>5 x 10^{-8}</th>
<th>5 x 10^{-7}</th>
<th>5 x 10^{-6}</th>
<th>5 x 10^{-5}</th>
<th>5 x 10^{-4}</th>
<th>0.005</th>
<th>0.05</th>
<th>CHR</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12746893</td>
<td>HL</td>
<td>HL</td>
<td>HL</td>
<td>HL</td>
<td></td>
<td>1</td>
<td></td>
<td>92384464</td>
<td></td>
</tr>
<tr>
<td>rs13062287</td>
<td></td>
<td></td>
<td>HL</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>54795186</td>
<td></td>
</tr>
<tr>
<td>rs13192841</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td>138008907</td>
<td></td>
</tr>
<tr>
<td>rs2019960</td>
<td></td>
<td></td>
<td></td>
<td>HL</td>
<td>HL</td>
<td>8</td>
<td></td>
<td>129261453</td>
<td></td>
</tr>
<tr>
<td>rs7111520</td>
<td></td>
<td></td>
<td></td>
<td>HL</td>
<td>HL</td>
<td>11</td>
<td></td>
<td>110754821</td>
<td></td>
</tr>
<tr>
<td>rs2425752</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HL</td>
<td>20</td>
<td></td>
<td>44135527</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HL P-value threshold</th>
<th>5 x 10^{-4}</th>
<th>5 x 10^{-5}</th>
<th>5 x 10^{-6}</th>
<th>0.005</th>
<th>0.05</th>
<th>CHR</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11161570</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td>1</td>
<td>85494499</td>
</tr>
<tr>
<td>rs6603986</td>
<td>MS</td>
<td>MS</td>
<td></td>
<td>1</td>
<td></td>
<td>92789809</td>
<td></td>
</tr>
<tr>
<td>rs2941584</td>
<td>MS</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td>54735125</td>
<td></td>
</tr>
<tr>
<td>rs10496164</td>
<td></td>
<td></td>
<td>MS</td>
<td>2</td>
<td></td>
<td>68463624</td>
<td></td>
</tr>
<tr>
<td>rs3806624</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>27739627</td>
<td></td>
</tr>
<tr>
<td>rs433317</td>
<td>MS</td>
<td>MS</td>
<td>MS</td>
<td>3</td>
<td></td>
<td>28035460</td>
<td></td>
</tr>
<tr>
<td>rs13062287</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>54795186</td>
<td></td>
</tr>
<tr>
<td>rs4285028</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>12314354</td>
<td></td>
</tr>
<tr>
<td>rs10941508</td>
<td>MS</td>
<td>MS</td>
<td>MS</td>
<td>5</td>
<td></td>
<td>40494824</td>
<td></td>
</tr>
<tr>
<td>rs10065637</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td>55474608</td>
<td></td>
</tr>
<tr>
<td>rs13188158</td>
<td>MS</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td>119902205</td>
<td></td>
</tr>
<tr>
<td>rs2064430</td>
<td></td>
<td></td>
<td>MS</td>
<td>6</td>
<td></td>
<td>13568449</td>
<td></td>
</tr>
<tr>
<td>rs9576268</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td>13757444</td>
<td></td>
</tr>
<tr>
<td>rs1002658</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td>13802377</td>
<td></td>
</tr>
<tr>
<td>rs2214543</td>
<td>MS</td>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td>10763417</td>
<td></td>
</tr>
<tr>
<td>rs2019960</td>
<td>MS</td>
<td>MS</td>
<td>MS</td>
<td>8</td>
<td></td>
<td>129261453</td>
<td></td>
</tr>
<tr>
<td>rs7111520</td>
<td>MS</td>
<td>MS</td>
<td>MS</td>
<td>11</td>
<td></td>
<td>110754821</td>
<td></td>
</tr>
<tr>
<td>rs715412</td>
<td>MS</td>
<td></td>
<td></td>
<td>11</td>
<td></td>
<td>118189820</td>
<td></td>
</tr>
<tr>
<td>rs11172482</td>
<td></td>
<td>MS</td>
<td></td>
<td>12</td>
<td></td>
<td>56846693</td>
<td></td>
</tr>
<tr>
<td>rs17119756</td>
<td></td>
<td></td>
<td>MS</td>
<td>14</td>
<td></td>
<td>83704185</td>
<td></td>
</tr>
<tr>
<td>rs2725544</td>
<td>MS</td>
<td></td>
<td></td>
<td>15</td>
<td></td>
<td>46780189</td>
<td></td>
</tr>
<tr>
<td>rs13330041</td>
<td>MS</td>
<td></td>
<td></td>
<td>16</td>
<td></td>
<td>10996309</td>
<td></td>
</tr>
<tr>
<td>rs12927773</td>
<td>MS</td>
<td></td>
<td></td>
<td>16</td>
<td></td>
<td>11311464</td>
<td></td>
</tr>
<tr>
<td>rs12456021</td>
<td>MS</td>
<td></td>
<td></td>
<td>18</td>
<td></td>
<td>54364370</td>
<td></td>
</tr>
<tr>
<td>rs2425752</td>
<td></td>
<td>MS</td>
<td></td>
<td>20</td>
<td></td>
<td>44135527</td>
<td></td>
</tr>
<tr>
<td>rs2248359</td>
<td></td>
<td></td>
<td>MS</td>
<td>20</td>
<td></td>
<td>52224925</td>
<td></td>
</tr>
</tbody>
</table>

Top: a grey box indicates that an SNP was associated with MS (at the P-value threshold shown in the top row), and was also associated with HL (FDR < 0.05; adjusted for the total number of SNPs that were tested in HL at the given MS threshold). Bottom: a grey box indicates that an SNP was associated with HL (at the P-value threshold shown in the top row), and was also associated with MS (FDR < 0.05; adjusted for the total number of SNPs that were tested in MS at the given HL threshold). Only independent SNPs are shown ($r^2 < 0.1$). The HLA region is omitted.

**CHR** chromosome.
similarity to other autoimmune diseases than to solid cancers \( (P = 3.5 \times 10^{-38}, \text{Figure 5A}) \), and analogously, the individual solid cancers showed more genetic similarity to other solid cancers than to autoimmune diseases \( (P = 5.4 \times 10^{-39}, \text{Figure 5C}) \). MS was closer to other autoimmune diseases than to solid cancers \( (P = 9.2 \times 10^{-6}, \text{Figure 5D}) \).

Haematological malignancies, as a group, displayed approximately equal proximity to solid cancers and autoimmune disorders \( (P = 0.49 \text{ for a difference, Figure 5B}) \). However, when the haematological malignancies were considered individually, HL was closer to autoimmune diseases than to solid cancers \( (P = 0.01, \text{Figure 5E}) \). Chronic
lymphocytic leukaemia (CLL) was also closer to autoimmune diseases ($P = 0.038$, Figure S2 available as Supplementary data at IJE online) but also shared some loci with solid cancers. In contrast, multiple myeloma (MM) was closer to solid cancers than to autoimmune diseases ($P = 0.08$, Figure S2). Figure 6 is a graphical representation of the proximity network between diseases.

**Discussion**

In this study, we undertook a series of analyses exploring genetic overlap between HL and MS. We found that top HL-associated SNPs were associated with MS, and conversely that top MS-associated SNPs were associated with HL. Overlap was particularly prominent in the HLA region of chromosome 6 but also applied to non-HLA loci. Genetic overlap between HL and MS was also observed in analyses of disease-specific polygenic risk scores (HLGB and MSGB) which included the HLA region. The HLGB explained 4.44% of the risk of HL and the MSGB explained 1.94% of the risk of HL. Similarly, MSGB explained 8.08% of the risk of MS and HLGB explained 2.36% of the risk for MS. Thus, the MSGB captured approximately 40% of the genetic susceptibility to HL measured by the HLGB and the HLGB captured approximately 30% of the genetic susceptibility to MS measured by the MSGB. Additionally, pathway analysis suggested shared biological pathways between HL and MS, involving a common theme of immune activation and cell proliferation, including lymphocyte-mediated immunity, positive regulation of JUN kinase activity (which plays roles in cellular response to stress, T cell differentiation, inflammation and apoptosis), peptidyl-tyrosine-phosphorylation (a non-specific intermediate step in multiple tyrosine pathways) and antigen processing and presentation.

The shared genetic element between HL and MS suggested by the present investigation is consistent with the original hypothesis of shared associations between the two conditions and with their previously observed mutual clustering within
families. Therefore further investigation of pathogenic pathways shared by these two clinically distinct conditions, similar to those being conducted for other neurodegenerative diseases associated with cancer risk, is warranted.

Both EBV-positive and EBV-negative HL and MS have been consistently and strongly associated with HLA alleles. Our study confirms that this locus confers the highest known effect for the three phenotypes. Interestingly, in stratified analyses the patterns of overlap with MS clearly differed between EBV-positive HL and EBV-negative HL. Risk alleles were shared between EBV-negative HL and MS, but not between EBV-positive HL and MS. These findings suggest that EBV-positive HL and EBV-negative HL may each be independently associated with MS, but via different genes.

The precise mechanisms underlying the HLA associations have not been firmly established either for HL or MS. In MS, one immune model holds that T cells, activated peripherally by an infectious agent, cross the blood-brain barrier and induce MS lesions upon reactivation by myelin fragment antigens presented in the context of HLA. For HL, speculation has centred primarily on its presumed infectious aetiology, with EBV present and expressing antigens with plausible pathogenic functions in all tumour cells and therefore likely playing a causal role in the EBV-positive subset of HL. Moreover, like MS, EBV-positive HL has been

---

**Table 3. Classification of immune and neoplastic diseases from the diseasome**

<table>
<thead>
<tr>
<th>Autoimmune diseases</th>
<th>Solid cancers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alopecia areata (AR)</td>
<td>Basal cell carcinoma (BCC)</td>
</tr>
<tr>
<td>Ankylosing spondylitis (AS)</td>
<td>Bladder carcinoma (BLC)</td>
</tr>
<tr>
<td>Behcet’s disease (Beh)</td>
<td>Breast carcinoma (BRCC)</td>
</tr>
<tr>
<td>Coeliac disease (CD)</td>
<td>Central nervous system cancer (CNS)</td>
</tr>
<tr>
<td>Crohn’s disease (CD)</td>
<td>Oesophageal carcinoma (OESCC)</td>
</tr>
<tr>
<td>Graves’ disease (GD)</td>
<td>Lung adenocarcinoma (LUA)</td>
</tr>
<tr>
<td>IGA glomerulonephritis (IGA)</td>
<td>Lung carcinoma (LUC)</td>
</tr>
<tr>
<td>Kawasaki disease (KAW)</td>
<td>Melanoma (ME)</td>
</tr>
<tr>
<td>Multiple sclerosis (MS)</td>
<td>Ovarian carcinoma (OVC)</td>
</tr>
<tr>
<td>Primary biliary cirrhosis (PBC)</td>
<td>Pancreatic carcinoma (PAC)</td>
</tr>
<tr>
<td>Psoriasis (PS)</td>
<td>Prostate carcinoma (PRC)</td>
</tr>
<tr>
<td>Psoriatic arthritis (PSA)</td>
<td>Renal cell carcinoma (RCC)</td>
</tr>
<tr>
<td>Rheumatoid arthritis (RA)</td>
<td>Squamous cell carcinoma (SCC)</td>
</tr>
<tr>
<td>Sclerosing cholangitis (PSC)</td>
<td>Stomach carcinoma (STC)</td>
</tr>
<tr>
<td>Systemic lupus erythematosus (SLE)</td>
<td>Thyroid carcinoma (THC)</td>
</tr>
<tr>
<td>Type 1 diabetes mellitus (T1D)</td>
<td></td>
</tr>
<tr>
<td>Ulcerative colitis (UC)</td>
<td></td>
</tr>
<tr>
<td>Vitiligo (Vit)</td>
<td></td>
</tr>
</tbody>
</table>

**Haematological cancers**

Chronic lymphocytic leukemia (CLL)
Hodgkin lymphoma (HL)
Multiple myeloma (MM)

---

**Figure 5. Diseasome analysis reveals that haematological malignancies lie somewhere between autoimmune diseases and solid cancer.**

A. Proximity of autoimmune diseases to other diseases. Density plots represent all possible pair-wise proximities between autoimmune diseases and solid cancers (orange), and all pair-wise proximities between autoimmune diseases and other autoimmune disease (purple). Higher degree of proximity (higher values on the x-axis) indicates more genetic similarity to autoimmune diseases. The \( P \)-value indicates that autoimmune diseases are closer to other autoimmune diseases than to solid cancers. B. Proximity of haematological malignancies to solid cancers. C. Proximity of solid cancers to other solid cancers (orange) and to autoimmune diseases (purple). Haematological malignancies show genetic overlap with both solid cancers and autoimmune diseases. C. Proximity of solid cancers to other solid cancers (orange) and to autoimmune diseases (purple). Solid cancers are closer to other solid cancers than to autoimmune diseases. D. Proximity of MS to all diseases. Each circle represents a disease in the diseasome. Higher degrees of proximity (higher values on x-axis) represent more genetic similarity with MS. Solid cancers are orange, autoimmune diseases are purple, HL is white. The \( P \)-value indicates MS is closer to autoimmune diseases than to solid cancers. E. Proximity of HL to all diseases. HL is closer to autoimmune diseases than to solid cancers.
associated with infectious mononucleosis caused by primary EBV infection\textsuperscript{20} and with aberrant anti-EBV nuclear antigen antibody patterns, albeit different from those associated with MS.\textsuperscript{19,21} Accordingly, HLA-specific variation in immune response to EBV infection may mediate the effects of the HLA associations shared by EBV-positive HL and MS. Although an analogous scenario involving an HLA-specific immune response directed against an infectious organism different from EBV may be envisioned for EBV-negative HL, no such agent has been linked to this HL subgroup\textsuperscript{50} or to MS as yet.

Construction of a diseasome based on disease-gene associations\textsuperscript{42} showed that the close genetic relationship between HL and the autoimmune disease MS applied to a broad spectrum of autoimmune conditions and that HL was in general closer to autoimmune diseases than to solid cancers. Importantly, the observation was not entirely explained by HLA, and the close relationship between HL and autoimmune conditions remained when the diseasome analysis was repeated after masking a seven mega-base region surrounding the HLA region. In line with this, there is evidence to suggest that the risk of HL is increased in patients with autoimmune diseases such as rheumatoid arthritis and systemic lupus erythematosus.\textsuperscript{51,52} In the absence of evidence of familial clustering of HL and autoimmune diseases,\textsuperscript{53} mechanisms such as chronic immune stimulation and/or immune-modulating treatment have been considered the most plausible explanations for the association. However, the present analyses suggest that shared genetic constitution may also contribute to the increased prevalence of HL among patients with autoimmune diseases (though an interaction between genetics and immune-modulating treatment is also a possibility). Indeed, our approach of combining GWAS data may prove more efficient in demonstrating overlapping pathogenic pathways between diseases than traditional epidemiological analytical designs, which may suffer from inadequate statistical power.

Besides HL, the diseasome analysis also included two other haematological malignancies. Among these, CLL was also closer to autoimmune diseases whereas multiple
myeloma showed similarity to solid cancers. CLL belongs to the group of non-Hodgkin lymphomas and, although strong associations between autoimmune disorders and CLL per se have not been noted, an increased risk of the combined group of non-Hodgkin lymphomas among patients with various autoimmune diseases is well documented in the literature.

The main limitation of the present study was the uneven distribution of HL and MS patients with GWAS data and the lack of independent validation datasets. These limitations likely do not affect the diseasome results which are based on multiple GWAS in each disease and are robust to the removal of any single GWAS. However, SNP-level summary data for other autoimmune diseases and other cancers would have allowed assessment of polygenic risk scores and specific genetic overlap between other pairs of diseases that were closely associated in the diseasome, i.e. in the same way that the genetic overlap between HL and MS was assessed. Perhaps the diseasome analysis will provide impetus for further collaborative meta-analyses of haematological malignancies and autoimmune diseases.

In summary, this study demonstrated commonalities in the genetic susceptibility to HL and MS as evidenced by analyses of individual SNPs, polygenic risk scores and protein-interaction networks. Diseasome analysis further suggested that HL shares a genetic architecture more similar to that of autoimmune diseases than to solid cancers. We speculate that autoimmune diseases and HL are different manifestations of a shared underlying genetic syndrome.

**Supplementary Data**

Supplementary data are available at IJE online.

**Funding**

Funding for this project was obtained from the Nordic Cancer Union (16-02-D to H.H.), the Lundbeck Foundation (R19-A2364 to H.H.), the Danish Cancer Research Foundation (41-08 to H.H.), the National Institutes of Health (CA110836 to W.C.; HD0433871, CA129045, and CA40046 to K.O.; CA55727 to L.L.R.; CA58839 to T.M.M.), the United States Army Medical Research and Materiel Command (Department of Defense PR054600 to W.C.), the American Cancer Society Illinois Division (to K.O.), the American Lebanese Syrian Associated Charities (to L.L.R.), the Leukemia & Lymphoma Society (TR6137-07 to W.C.), the Cancer Research Foundation (to K.O.), the National Cancer Institute Surveillance Epidemiology and End Results Population-based Registry Program at the National Institutes of Health in the Department of Health and Human Services (R01-NS-026799 to J.R.O., N01-PC-35139 to W.C. and N01-PC-35136 to the Cancer Prevention Institute of California), the National Cancer Institute (263-MQ-41755 to S.L.G.), the Leukemia Lymphoma Research and Kay Kendall Leukemia fund (grant 12022 to R.F.J.), the California Cancer Registry and the Dutch Cancer Society (KWF grants RUG 2009-4313 and RUG 2014-6698 to A.v.d.B.), the National Science Foundation (Graduate Research Fellowship Grant 1144247 to D.S.H.) and the Center for Neuroengineering and Therapeutics at the University of Pennsylvania (P.K.).

**Acknowledgments**

We thank the Childhood Cancer Survivor Study (National Cancer Institute grant U24 CA 55727) for the Hodgkin Lymphoma data, and the International Multiple Sclerosis Genetics Consortium and the Wellcome Trust Case-Control Consortium for the multiple sclerosis data. We would like to thank Hans-Olov Adami, David Conti, Arjan Diepstra, Bengt Glimelius, Annette Lake, Dorothy Montgomery, Les Robison, Malcolm Taylor, and Maria Timofeeva for data accrual, sample handling and analysis of Hodgkin lymphoma data. We would like to thank Stacy J. Caillier and Sarah Beth Hill for assistance with figure creation.

**Author contributions**


**Conflict of interest:** None declared.

**References**


