A search for overlapping genetic susceptibility loci between non-Hodgkin lymphoma and autoimmune diseases

Lucia Conde a,⁎, Paige M. Bracci b, Eran Halperin c,d, Christine F. Skibola a,⁎

a Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA 94720, USA
b Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94143, USA
c Department of Microbiology and Biotechnology, and School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel
d International Computer Science Institute, 1947 Center Street, Suite 600, Berkeley, CA 94704, USA

A R T I C L E   I N F O

Article history:
Received 22 October 2010
Accepted 15 March 2011
Available online 23 March 2011

Keywords:
Non-Hodgkin lymphoma
Autoimmune diseases
Genome-wide association studies
Human leukocyte antigen

A B S T R A C T

Non-Hodgkin lymphoma (NHL) is a hematological malignancy of the immune system, and, as with autoimmune and inflammatory diseases (ADs), is influenced by genetic variation in the major histocompatibility complex (MHC). Persons with a history of specific ADs also have increased risk of NHL. As the coexistence of ADs and NHL could be caused by factors common to both diseases, here we examined whether some of the associated genetic signals are shared. Overlapping risk loci for NHL subtypes and several ADs were explored using data from genome-wide association studies. Several common genomic regions and susceptibility loci were identified, suggesting a potential shared genetic background. Two independent MHC regions showed the main overlap, with several alleles in the human leukocyte antigen (HLA) class II region exhibiting an opposite risk effect for follicular lymphoma and type I diabetes. These results support continued investigation to further elucidate the relationship between lymphoma and autoimmune diseases.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Recent genome-wide association studies (GWAS)① of the three most common subtypes of non-Hodgkin lymphoma (NHL) have detected multiple genetic factors in the major histocompatibility complex (MHC) region that are associated with follicular lymphoma (FL) [1,2]. Genetic variation in the MHC, which spans approximately 4 Mb on chromosome 6p21 and encodes >400 known protein-coding genes (www.ensembl.org, assembly GRCh37), also plays a crucial role in modulating risk of chronic autoimmune and pro-inflammatory diseases (ADs) such as rheumatoid arthritis (RA), Crohn's disease (CD) and type 1 diabetes (T1D). Several GWAS have successfully identified human leukocyte antigen (HLA) and non-HLA susceptibility alleles for these diseases [Supplementary Table 1], with some of these loci shared by more than one AD. Although most of the consistently associated regions have been found in the MHC, genetic variants have been identified outside the MHC [3] where some clustering of susceptibility alleles has been identified across autoimmune diseases [4,5].

The co-localization of susceptibility alleles in the MHC for NHL and ADs suggests the potential for shared risk loci. This hypothesis is supported by results from several epidemiological studies that have investigated whether a history of AD influences NHL risk [6,7] including a large pooled analysis of 29,423 participants in 12 case-control studies within the International Lymphoma Epidemiology Consortium (InterLymph). Increased risks were observed for self-reported history of Sjögren syndrome with all NHL, marginal zone lymphoma (MZL), diffuse large B-cell lymphoma (DLBCL) and FL; systemic lupus erythematosus (SLE) with NHL, DLBCL and MZL; and celiac disease and psoriasis with T-cell NHL [6]. Interestingly, neither a history of CD nor T1D was associated with NHL or its subtypes, whereas results for RA were heterogeneous among studies. The lack of consistency for RA may be due to misclassification of rheumatoid and non-rheumatoid arthritis based on self-report. The InterLymph findings were relatively consistent with those from a study of 44,350 lymphoid malignancy cases and 122,531 population-based controls from the U.S. Surveillance Epidemiology and End Results–Medicare database. Results indicated positive associations between a history of Sjögren syndrome and risk of DLBCL and MZL, for SLE with MZL, and for RA with DLBCL [7]. These results suggest that specific autoimmune disorders are associated with NHL risk and specific NHL subtypes.

① Corresponding authors at: Division of Environmental Health Sciences, School of Public Health, University of California, 237A Hildebrand Hall, Berkeley, CA 94720–7356, USA. Fax: + 1 510 642 0427.
E-mail addresses: lconde@berkeley.edu (L. Conde), chrisf@berkeley.edu (C.F. Skibola).
Some studies suggest that shared susceptibility between specific autoimmune diseases and lymphoma subtypes might be due to reverse causation and thus related to AD severity and related inflammatory processes or to disease treatment [8,9]. To investigate whether there is a genetic basis that may contribute to the positive associations observed between NHL risk and history of autoimmune disease, we compared our NHL GWAS data with publicly available genome-wide level data on RA, CD and T1D from the Wellcome Trust Case-Control Consortium (WTCCC) [10] to explore potential shared genetic susceptibility between NHL and these ADs. Our global genomic approach provides a comprehensive view of the genetic overlap of these diseases that expands upon previous epidemiologic studies of ADs and NHL risk, and aims to improve our understanding of mechanisms of lymphomagenesis. Analyses focused specifically on common genetic variants among the NHL subtypes, FL, chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL) and DLBCL, and the autoimmune disorders, RA, CD and T1D, for which data is publicly available.

2. Results

Genome-wide association results for ADs and the major NHL subtypes, FL, DLBCL and CLL/SLL, are shown in Fig. 1. A strong association signal can be observed in the MHC region for FL and all the ADs, especially for T1D and RA. Although several locations across the genome showed overlapping signals for NHL subtypes and ADs, only two specific locations showed statistically significant overlap (p < 0.05, see Materials and methods). These two 10 kb regions contained genetic variants that showed association with all ADs and with NHL subtypes, particularly FL. Due to the high LD structure in the MHC and because the associated signals could reflect LD with a causative marker located further away, we expanded these two regions to include neighboring SNPs in linkage disequilibrium (LD, r^2 > 0.6 in HapMap-CEU).

2.1. Region of overlap on chromosome 6: 31,215–31,671 Mb

The first region with overlapping signals for NHL subtypes and ADs was located in the MHC, from the class I psoriasis susceptibility locus (PSORS1) where a FL-associated locus (rs6457327, OR = 0.59, p = 4.7 \times 10^{-11}) was previously described [1], to the class III region that overlaps the NHL-associated genes TNF/LTA [11]. Several SNPs were observed in this region at the p < 0.001 level for all the ADs, FL and DLBCL (Fig. 2), but not for CLL/SLL. Some of these SNPs were associated with one disease only, whereas several loci were associated with multiple diseases (Supplementary Table 2). The overlapping loci mainly were found between FL and the ADs, particularly for T1D and RA. In general, these common susceptibility alleles were inversely associated with disease risk.

Among the shared associated genetic variants, we identified several tag SNPs for HLA-B and HLA-C alleles (Supplementary Fig. 1. Genotype results for non-Hodgkin lymphoma subtypes (FL, DLBCL, SLL/CLL) and autoimmune diseases (T1D, RA, CD). Chromosomes are plotted on the x-axis, and trend p-values (as -log_{10} adjusted using Genomic Control) are plotted on the y-axis. The T1D, RA and CD panels have been truncated at -log_{10} (p-value) = 15.

Fig. 2. The top panel shows, for each disease, all the genotyped or imputed SNPs that were ranked in the top-5000 and showed GC-adjusted p < 10^{-4} in the HLA region between positions 31215 and 31671 Mb. The lower panel shows the same plot after removing linked SNPs (r^2 > 0.2), with horizontal lines showing the extent of LD for each SNP (r^2 > 0.6 in each disease). The strongest associated SNP for each disease in the region is marked with a bigger point.
Table 1

<table>
<thead>
<tr>
<th>Allelotype</th>
<th>Tag SNPs</th>
<th>Tagging haplotype</th>
<th>FL Ca/Co freq</th>
<th>Chi-square p-val</th>
<th>RA Ca/Co Freq</th>
<th>Chi-square p-val</th>
<th>TID Ca/Co Freq</th>
<th>Chi-square p-val</th>
<th>CD Ca/Co Freq</th>
<th>Chi-square p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-B*0702</td>
<td>rs2596438, rs805288</td>
<td>GC</td>
<td>0.06/0.13</td>
<td>2.0E–4</td>
<td>0.12/0.14</td>
<td>2.0E–4</td>
<td>0.07/0.14</td>
<td>2.30E–2</td>
<td>0.13/0.14</td>
<td>0.0373</td>
</tr>
<tr>
<td>HLA-B*4402</td>
<td>rs3130501, rs2769, rs2844533</td>
<td>CTT</td>
<td>0.07/0.07</td>
<td>0.9405</td>
<td>0.14/0.11</td>
<td>6.48E–5</td>
<td>0.09/0.11</td>
<td>0.0058</td>
<td>0.12/0.11</td>
<td>0.0289</td>
</tr>
<tr>
<td>HLA-C*0701</td>
<td>rs2524074, rs7382297</td>
<td>CC</td>
<td>0.20/0.16</td>
<td>0.0903</td>
<td>0.18/0.18</td>
<td>0.3025</td>
<td>0.28/0.18</td>
<td>7.58E–3</td>
<td>0.17/0.18</td>
<td>0.292</td>
</tr>
<tr>
<td>HLA-DRB1*1301</td>
<td>rs2395173, rs2157051</td>
<td>GG</td>
<td>0.03/0.05</td>
<td>0.0407</td>
<td>0.04/0.07</td>
<td>2.96E–11</td>
<td>0.02/0.06</td>
<td>7.85E–17</td>
<td>0.07/0.07</td>
<td>0.5065</td>
</tr>
<tr>
<td>HLA-DQA1*0103</td>
<td>rs2273017, rs2157051</td>
<td>AG</td>
<td>0.04/0.07</td>
<td>0.0221</td>
<td>0.04/0.07</td>
<td>2.47E–13</td>
<td>0.02/0.07</td>
<td>5.70E–26</td>
<td>0.08/0.08</td>
<td>0.9551</td>
</tr>
<tr>
<td>HLA-DRB1*0101</td>
<td>rs4947332, rs6457614</td>
<td>GC</td>
<td>0.21/0.13</td>
<td>7.06E–5</td>
<td>0.16/0.11</td>
<td>1.15E–9</td>
<td>0.10/0.11</td>
<td>0.0418</td>
<td>0.12/0.11</td>
<td>0.4453</td>
</tr>
<tr>
<td>HLA-DQA1*0101</td>
<td>rs6457614, rs1794265</td>
<td>CG</td>
<td>0.21/0.13</td>
<td>7.06E–5</td>
<td>0.16/0.11</td>
<td>1.15E–9</td>
<td>0.10/0.11</td>
<td>0.0418</td>
<td>0.12/0.11</td>
<td>0.4453</td>
</tr>
</tbody>
</table>

The table shows the SNPs and their specific allele combination (haplotype) that tag each allelotype as described in [12]. The frequency of the tagging haplotype in cases and controls and the association p-value for the haplotype frequencies in cases vs. control are shown for each disease and were estimated using Haploview [21].

In particular, rs2596438, whose minor allele tags HLA-B*0702 together with rs805288 [12], was inversely associated with FL, T1D, and RA (ORs = 0.52 [0.37–0.74], 0.49 [0.43–0.55] and 0.80 [0.71–0.89], respectively). Accordingly, the frequency of the rs2596438–rs805288 haplotype-tagging HLA-B*0702 was in controls than in cases for each of these diseases (Table 1). We also found that rs9461684, which tags HLA-C*0401 [12], was positively associated with FL (OR = 1.72 [1.27–2.33]) and inversely associated with T1D (OR = 0.58 [0.49–0.69]).

Two other SNPs in this region, rs3130501 and rs7382297, which are part of haplotypes tagging, respectively, HLA-B*4402 and HLA-C*0701 [12], were inversely associated with FL and T1D (Supplementary Table 2). However, the haplotype analysis showed no association of these haplotypes to FL risk (Table 1), suggesting that linked loci other than the HLA-B*4402 and HLA-C*0701 allelotypes, might be the common susceptibility loci.

2.2. Region of overlap on chromosome 6: 32,320–33,112 Mb

Common associations for NHL subtypes and ADs also were found in HLA class II (Fig. 3). In this region that overlaps the HLA-DR and HLA-DQ genes, several shared associated loci were found at a p < 0.001 level (Supplementary Table 3). Thus, our results showed that SNP rs2157051, which together with rs2395173 tags HLA-DRB1*1301 [12], was inversely associated with FL (OR = 0.57 [0.41–0.79]) and RA (OR = 0.54 [0.47–0.61]). Haplotype analyses confirmed the inverse associations of HLA-DRB1*1301 with FL, RA, and T1D (Table 1). Similar results were found for a second haplotype tagged by the same SNP, HLA-DQA1*0103, that showed a reduced risk of FL, RA and T1D (Table 1).

We also observed that SNP rs6457614, which tags the HLA-DRB1*0101, HLA-DQA1*0101 and HLA-DQB1*0501 allelotypes [12], was associated with increased risk of FL (OR = 1.75 [1.32–2.31]) and RA (OR = 1.45 [1.29–1.64]). Analysis of the haplotypes containing rs6457614 confirmed the positive association of HLA-DRB1*0101 and HLA-DQA1*0101 with FL and RA, and interestingly, also showed an inverse association with T1D (Table 1). Opposing effects on risk estimates were repeatedly observed for SNPs associated with both FL and T1D in this region, with FL risk alleles being protective for T1D (Supplementary Table 3). For example, the rs9469220 A allele was associated with increased susceptibility to FL (OR = 1.47 [1.18–1.85]), CD (OR = 1.23 [1.14–1.35]), and RA (OR = 1.17 [1.08–1.27]), and decreased risk of T1D (OR = 0.37 [0.34–0.41]). Similarly, the rs10947332 (A), rs13218331 (C), rs3177928 (A), rs12529093 (G) and rs13209234 (A) alleles were positively associated with FL and RA, and inversely associated with T1D (Supplementary Table 3).

3. Discussion

Numerous epidemiological studies have shown a link between a personal history of autoimmune diseases and increased risk of specific lymphoma subtypes. Here, we compared the genetics of the NHL subtypes, FL, CLL/SLL and DLBCL, with the ADs, RA, T1D and CD using genome-wide association data to explore the hypothesis of a shared genetic background between lymphoma and ADs. Our analyses identified the presence of common genetic variants between NHL subtypes and ADs, suggesting a potential shared genetic mechanism. However, the co-localization of shared genetic variants was observed more often between FL and ADs than with the other NHL subtypes. Although several overlapping associated regions were found across the genome, only two that were located in the HLA class I–III and II regions were repeatedly observed in each of the diseases.
regions reached statistical significance. This non-random clustering in the HLA region supports the hypothesis that FL and autoimmune diseases might be influenced by a common set of immunoregulatory susceptibility genes.

Among the three ADs studied, CD exhibited the least genetic overlap with the three NHL subtypes studied. Although CD has been associated with risk of lymphoma [9], it may be that the association is more a consequence of disease-related inflammatory processes or related to the immunosuppressive therapies used to treat CD than to the presence of a shared genetic background.

We found several SNPs common to FL, RA and T1D, which were mainly located in intergenic or intronic regions. Among these, tag SNPs for HLA alleles were also identified, including tags for HLA-B*0702, HLA-DRB1*1301 and HLA-DQA1*0103, which showed protection for CD, RA and T1D. Whereas the use of tag SNPs to predict HLA allelotypes is acceptable for exploratory analyses and hypothesis generation, these findings will require validation and follow-up in HLA allelotype analyses. Further studies that include resequencing to exploit LD across all variants in the region are warranted to clarify and identify the causal variants and pathways that are unique and similar across these diseases.

Of particular interest was the observation that most of the SNPs in the HLA class II region tended to exert the same effect on disease risk for FL and RA, but opposite effects for FL and T1D. Specifically, we found that the HLA-DRBI*0101, HLA-DQA1*0103 showed increased risk of FL and RA and a reduced risk of T1D. Similarly, other SNPs not listed as haplotype-tagging SNPs in [12] had contrasting effects for FL and T1D. It could be that the alleles of these SNPs are tagging two different undetected allelotypes that are associated respectively with T1D and FL. It also might be that these alleles influence antigen presentation, and are important factors leading to different antigen-induced immune responses. Associations of opposite alleles to different autoimmune diseases have been observed in previous studies [13,14] and it has been suggested that risk alleles for one disease may confer selective advantage for another disease or resistance to infection. Opposite effects of haplotypes in the HLA class II region occur both across and within diseases, demonstrating the complexity of the HLA region. For example, there are several HLA class II haplotypes that lie within the same region and yet have been shown to confer increased and decreased risk for FL [15] and T1D [16]. Non-genetic factors associated with autoimmune disease such as chronic inflammation and treatment effects might also play a role in lymphomagenesis. Examples include the associations of Helicobacter pylori infection with gastric MALT lymphoma [17] and hepatitis C virus infection with B-cell NHL [18]. Further genotyping in additional cohorts will be necessary to validate these findings and to further explore the common genetic backgrounds between these two groups of diseases and their mechanisms of action.

4. Materials and methods

4.1. Study populations

4.1.1. Non-Hodgkin lymphoma (NHL) cohort

A population-based case-control study of NHL (2055 cases, 2081 controls) that included incident cases diagnosed from 2001 through 2006 was conducted in the San Francisco Bay Area. Details of the study design and methods have been described previously [1]. Briefly, eligible patients were identified through the cancer registry and met the following criteria at diagnosis: aged 20–85 years, resident of one of the six Bay Area counties and able to complete an in-person interview in English. Controls were identified by random digit dial and random sampling of Center for Medicare and Medicaid lists, met the same eligibility criteria as cases with the exception of NHL diagnosis, and were frequency-matched to patients by age in 5-year age groups, sex and county of residence. Blood and/or buccal specimens were collected from eligible cases and controls who participated in the laboratory portion of the study (participation rates, 87% and 89%, respectively). To confirm NHL diagnosis and for consistent classification of NHL subtypes using the WHO classification, the study’s expert hematopathologist re-reviewed patient diagnostic pathology materials (including diagnostic slides, pathology, immunohistochemistry and flow cytometry reports) for >98% of consenting cases, with review of diagnostic slides in addition to pathology reports conducted for 54% of cases. Approximately 23% of NHL subtypes were reclassified, and approximately 1% of cases were dropped as not NHL after expert re-review. Eighty-eight percent of FL and 92% of DLBCL case diagnoses did not change after re-review.

4.1.2. T1D, RA, CD cases and controls datasets

We used genotype data available from the Wellcome Trust Case-Control Consortium (WTCCC, http://www.wtccc.org.uk/). The WTCCC study populations used in this study consists of 1963 T1D, 1860 RA, and 1748 CD cases from Great Britain who mainly self-reported as white Europeans. A group of 2938 healthy individuals were used as shared controls for the T1D, RA and CD cases. These WTCCC controls came from two different sources, 1480 individuals from the 1958 British Birth Cohort and 1458 individuals from an UK national repository of anonymized DNA samples from 3622 controls (UK Blood Service sample) recruited as part of the WTCCC project. There was no evidence of systematic or marked differences in overall allele frequencies between the two control groups [10]. Phenotype description for the T1D, RA, CD cohorts and details about the control study populations have been extensively described elsewhere [10].

4.2. Genotyping and quality control

4.2.1. San Francisco Bay Area NHL study

Details of the genotyping and quality control have been published previously [2]. Briefly, DNA from 1577 study participants was genotyped using Illumina HumanCNV370-Duo BeadChip (Illumina, San Diego, CA), and genotype clustering was conducted with Illumina Beadstudio software from data files created by an Illumina BeadArray reader. Samples with call rates <95% were excluded from further analysis. SNPs with call rates <90%, minor allele frequency (MAF) < 0.05 and SNPs on sex chromosomes also were excluded. A total of 312,768 markers genotyped in 1568 individuals (236 FL, 221 CLL/SLL, 291 DLBCL, 9 other NHL and 811 controls) passed these quality control criteria and were used for analysis. Genotype data were used to search for individuals with evidence of non-European ancestry and closely related individuals, as described below. These individuals were excluded from further analysis.

4.2.2. T1D, RA, CD and WTCCC control datasets

Samples were genotyped on the Affymetrix GeneChip 500K Mapping Array Set at the Affymetrix Services Lab in California and genotypes were called using the CHIAMO algorithm as described elsewhere [10]. For each dataset, we applied multiple filters to exclude individuals and SNPs based on genotype data, using the same or similar thresholds and procedures as in the WTCCC study [10]. Thus, SNPs were excluded from each study if any of the following criteria was met: \( p < 5.7 \times 10^{-7} \) in test for deviation from Hardy–Weinberg equilibrium in controls, >5% of data missing for minor allele frequency (MAF) >0.05 or >1% of data missing for minor allele frequency (MAF) \( \leq 0.05 \), or allelic \( p = 5.7 \times 10^{-7} \) between the two WTCCC control groups. Individuals with \( >3\% \) of data missing also were excluded. A total of 474,016 markers genotyped in 1961 T1D cases, 474,680 markers in 1857 RA cases, 476,476 markers in 1741 CD cases and 462,424 markers in 2937 controls passed the quality control criteria and were included in this analysis. In each dataset, genotype data were used to search for individuals with evidence of non-European ancestry and closely related individuals, as described below.
4.3. Population structure

For each dataset (NHHL, T1D, RA, CD and WTCCC controls), we tested population stratification using the multidimensional scaling (MDS) procedure included in the software PLINK v1.04 [19]. To identify samples of non-European ancestry, we first merged the SNPs that remained after our quality controls with genotypes from 209 unrelated HapMap Phase II individuals from the CEU, YRI and JPT + CHB panels (JPT individual NA19012 was removed due to low call rate). In order to avoid strand issues when merging the datasets, we removed all the ambiguous (A/T and C/G) SNPs. We also removed non-biallelic SNPs, SNPs with different mapping positions in HapMap and SNPs with >5% missing data. From the remaining SNPs we selected a subset of unlinked SNPs by pruning those with $r^2$>0.1 using 50-SNP windows shifted at 5-SNP intervals. We ran the MDS analysis on the matrix of genome-wide identity-by-state pairwise distances and selected 100 as the number of dimensions to be extracted. Each dataset was projected over the first 2 MDS dimensions (which accounted for the 37.8%, 31.10%, 29.9%, 27.8% and 31.9% of the total variance in NHL, T1D, RA, CD and WTCCC controls, respectively) to identify clusters of European, Asian and African samples (Supplementary Fig. 1A–E). The cut-offs used to detect outliers in terms of population stratification were chosen based on inspection of the MDS plots. Thus, individuals that mapped outside the European cluster (117 in the NHL cohort [21 Fi, 32 DLBCL, 10 SLL/CLL and 54 controls], 21 in T1D, 27 in RA, 64 in CD and 27 in the WTCCC control datasets), regardless of their reported origin, were excluded from the association analysis. The final datasets were merged (NHL, T1D, RA, CD and WTCCC controls) and MDS analysis was conducted as described above. All the samples formed a single cluster together with the HapMap-CEU individuals (Supplementary Fig. 1F).

We also used genotype data to search for duplicates and closely related individuals. For each dataset, we used PLINK 1.04 to obtain identity-by-descent (IBD) estimates for all pairs of individuals using a subset of unlinked SNPs. Individuals with >86% IBD sharing (11 in NHL, 11 in T1D, 66 in RA, 156 in CD and 23 in the WTCCC control datasets) were excluded. After exclusion of individuals of non-European ancestry or with evidence of cryptic relatedness, we estimated an inflation factor $\lambda = 1.04$ in the NHL cohort ($\lambda = 1.04$, $\lambda = 1.01$ and $\lambda = 1.04$ in FL, DLBCL and SLL/SLC, respectively), and inflation factors $\lambda = 1.11$, $\lambda = 1.10$ and $\lambda = 1.17$ in the combined datasets of WTCCC controls with T1D, RA and CD, respectively. These inflation factors were used to adjust the $p$-values by Genomic Control in each cohort.

4.4. Statistical analysis

4.4.1. Imputations

Using the software package BEAGLE 3.0.3[20], we imputed all known SNPs that were not genotyped or that did not pass direct genotyping quality controls for each cohort (FL, DLBCL, SLL/CLL, T1D, RA and CD) separately. Genotypes were imputed based on directly genotyped SNPs and haplotype information from unrelated CEU samples in the HapMap Phase II data. When imputing genotypes in samples of unrelated individuals, BEAGLE produces posterior genotype probabilities for imputed genotypes. Those imputed SNPs with maximum posterior probability lower than 0.9 were set to missing. Using PLINK v1.04, we further removed imputed SNPs with >10% missing data or MAF<0.01. Finally, 1,621,903 SNPs were successfully imputed in FL, 1,627,752 in DLBCL, 1,624,197 in SLL/CLL, 1,383,155 in T1D, 1,383,984 in RA and 1,383,575 in CD. Imputed SNPs were tested in the same way as SNPs directly genotyped as described below.

4.4.2. Single SNP statistical analysis

For each NHL subtype (FL, DLBCL, SLL/CLL) and autoimmune disease (T1D, RA, CD), association tests were conducted using dominant and additive models (Cochran–Armitage trend test) as implemented in PLINK 1.04. ORs and 95% CI were computed for the dominant model (homozygous or heterozygous versus homozygous common allele carriers), by median-unbiased estimation using the mid-p method from the epitools R package (http://sites.google.com/site/medepi/epitools). Trend $p$-values in each cohort were adjusted using the Genomic Control procedure.

4.4.3. Search for overlapping signals in ADs and NHL

To look for overlapping regions among ADs and NHL subtypes, we first divided the genome into windows of 10 kb to fine-map autoimmune-associated regions. We considered a region to be autoimmune-associated if it contained at least one top 5000 SNP with an $GC$-adjusted $p<10^{-4}$ in all the three ADs studied. We found 19 windows that showed association to all the ADs. Some of these autoimmune-associated regions showed strong association signals in NHL subtypes, including two windows located in the HLA class I-III and II regions that showed FL-associated signals with unadjusted trend $p = 3.992 \times 10^{-5}$ and $3.593 \times 10^{-5}$, respectively. To address the possibility that the associations of NHL in autoimmune-associated regions were due to chance, we used a permutation approach to estimate the likelihood of finding an NHL signal at these $p$-value levels by chance in an autoimmune-associated region. Specifically, for each NHL subtype, we randomly permuted the case/control status and tested association in the 19 autoimmune-associated regions. After 1000 permutations, we found that the probability of finding an NHL signal in one of these autoimmune-associated regions with trend $p$-values of $3.992 \times 10^{-5}$ and $3.593 \times 10^{-5}$ is, respectively, $p = 0.029$ and 0.031. This suggests that these two regions are likely to contain real shared susceptibility loci. Due to the high linkage disequilibrium (LD) present in the HLA we expanded the associated 10 kb regions to include neighboring markers (up to 500 kb) in LD ($r^2>0.6$ in HapMap-CEU), allowing the windows to expand up to 456 kb in the HLA class I-III region, and 792 kb in the HLA class II region.

4.4.4. Haplotype analysis

We tested the association of HLA allele types by constructing haplotypes composed of the specific tag SNP alleles described in [12]. Estimation of haplotype frequencies in cases and controls and haplotype association tests (chi-square) were performed using Haplovie [21].

Conflict of interest statement

The authors declare no competing financial interest.

Acknowledgments

We want to thank Dr. Patrick Tressler from the University of California–San Francisco for leading and completing the pathology review. This work was supported by the National Cancer Institute, National Institutes of Health [grant numbers CA122663 and CA104682 to C.F.S. and CA45614 and CA09745 to P.M.B.]. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113 and 085475.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.ygeno.2011.03.007.
References


