Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma

Christine F. Skibola[1,9], **Paige M. Bracci**[2], **Eran Halperin**[3], **Lucia Conde**[1], **David W. Craig**[4], **Luz Agana**[1], **Kelly Iyadurai**[4], **Nikolaus Becker**[5], **Angela Brookes-Wilson**[6,7], **John D. Curry**[1], **John J. Spinelli**[6,7], **Elizabeth A. Holly**[2], **Jacques Riby**[1], **Luoping Zhang**[1], **Alexandra Nieters**[5,8], **Martyn T. Smith**[1], and **Kevin M. Brown**[4]

[1] Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA 94720

[2] Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94143

[3] International Computer Science Institute, 1947 Center Street, Suite 600, Berkeley, CA 94704

[4] Translational Genomics Research Institute (TGen), 444 North Fifth Street, Phoenix, AZ 85004

[5] Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany

[6] Canada's Michael Smith Genome Sciences Center, BC Cancer Agency

[7] Cancer Control Research, BC Cancer Agency

[8] Molecular Epidemiology, Center of Chronic Immunodeficiency, University of Freiburg, Germany

Abstract

We conducted genome-wide association studies of non-Hodgkin lymphoma using Illumina HumanHap550 BeadChips to identify subtype-specific associations in follicular, diffuse large B-cell and chronic lymphocytic leukemia/small lymphocytic lymphomas. We found that rs6457327 on 6p21.33 was associated with susceptibility to follicular lymphoma (FL, N=189 cases/592 controls) with validation in an additional 456 FL cases and 2,785 controls (combined allelic *p*-value=4.7×10^{-11}). The region of strongest association overlaps *C6orf15(STG)*, located near psoriasis susceptibility region 1(*PSORS1*).

Non-Hodgkin lymphoma (NHL) is a heterogeneous group of neoplasms of B- and T-cells that vary in their causes and molecular profiles[1]. With the fifth highest incidence amongst all cancers in the U.S., the annual incidence of NHL has doubled since the 1970s. With the increasing evidence supporting the importance of genetic determinants in lymphomagenesis[2], there is a strong impetus to identify genetic risk factors. Epidemiological and biological evidence suggest that environmental and genetic risk factors differ for the common NHL subtypes, follicular (FL), diffuse large B-cell (DLBCL) and chronic lymphocytic leukemia (CLL)/small lymphocytic lymphoma (SLL)[1]. We therefore conducted genome-wide association studies (GWAS) using separate DNA pools from 189 FL, 221

9Correspondance to: Dr. Christine Skibola, School of Public Health, 237A Hildebrand Hall, University of California, Berkeley, California 94720-7356, 510) 643-5041 tel/(510) 642-0427 fax/ chrisfs@berkeley.edu.

DLBCL, and 148 CLL/SLL cases and 592 controls (NC1 sample set) from a larger San Francisco Bay Area NHL case-control study[3] to identify subtype-specific NHL susceptibility genes (for study design see Supplementary Fig. 1; for description of study populations see Supplementary Table 1). We restricted genotyping to DNA collected from individuals with European ancestry as determined by AIMS genotyping[4] to diminish potential underlying population stratification. Self-reported ethnicity and ancestry data were highly correlated (95%) and used to construct homogeneous DNA pools of participants of European descent.

In the first phase, pools were hybridized to Human Hap550v.3 BeadChips (Illumina, San Diego, CA), and SNPs were ranked after adjusting for pooling error[5]. The top 30 ranked SNPs for each NHL subtype were subsequently individually genotyped across the NC1 sample set to confirm the accuracy of estimated allele frequency differences from the pooled data. 87% of raw allelic p-values were <0.05 (Supplementary Tables 2a–c), and genotype frequencies did not significantly differ from Hardy-Weinberg equilibrium. 32 SNPs with subtype-specific allelic q-values (corrected p) <0.05 were subsequently genotyped in an independent set of 89 FL, 159 DLBCL, 135 CLL/SLL and 363 other NHL cases and 820 controls from the same study population as NC1 (NC2 sample set). Joint analyses of the SNP data on all study participants revealed five SNPs associated with FL (rs6457327, rs2517448, rs13286028, rs11158098, rs16940565) and two with DLBCL (rs9936269, rs29605) at q<0.05 (Supplementary Table 3); no CLL/SLL SNPs were significant at q<0.05. As rs6457327 and rs2517448 were in complete LD, rs2517448 was excluded from further testing.

In the second phase, we performed validation genotyping for these six SNPs in a German case-control study[6] (G1) comprising 87 FL, 152 DLBCL, 102 CLL/SLL and 153 other NHL cases and 669 controls, where rs6457327 was associated with decreased FL risk (OR=0.28, p=0.01, Table 1). No other associations were validated. In the third phase, rs6457327 was validated in two additional independent NHL case-control studies (Table 1) that included 108 FL cases and 685 controls (NC3)[7] from the San Francisco Bay Area, and 172 FL cases and 611 controls from Canada (C1)[8]. The combined p-value for rs6457327 and risk of FL across all four studies was 4.7×10^{-11} for the allelic model and 1.9×10^{-10} for the Cochran-Armitage trend test. These p-values are lower than the threshold for genome-wide significance (Bonferroni corrected for 500,000 SNPs×4 genetic models×3 disease outcomes with α=0.05, 8.3×10^{-9}). Because some evidence suggests shared genetic factors among NHL subtypes[2], we also evaluated whether rs6457327 was associated with CLL/SLL or DLBCL in the combined sample from all study populations. No association was found for CLL/SLL, although a modest association was observed with DLBCL (allelic p-value=7.0×10^{-5}, trend p-value=1.1×10^{-4}; Table 1).

Respectively, rs6457327 and rs2517448 are 5 and 16kb downstream of the 3′ UTR of the *C6orf15 (STG)* gene, telomeric to *HLA-C* on chromosome 6p21.33 in the major histocompatibility complex. This 300kb region has been extensively evaluated due to its association with psoriasis, where the *HLA-C* gene was identified as a strong psoriasis susceptibility locus (*PSORS1*; psoriasis susceptibility region 1)[9]. History of psoriasis has been associated with increased risk of T-cell lymphoma; however, this association may be attributable to psoriasis treatment with immunosuppressive agents rather than family history[10]. Here, we found little LD between rs6457327 or rs2517448 and SNPs in *HLA-C* (r^2<0.35; Supplementary Fig. 2a). Although our data do not suggest that linked *HLA-C* SNPs are driving the association, we cannot conclusively rule out common genetic associations between the two diseases.

To explore the genetic interval containing the associated variants, we genotyped 52 additional SNPs within 30kb of rs6457327 in the NC1/2 sample, where we found 13 additional markers with q-values <0.05 (Fig. 1a, Supplementary Table 4). The strongest signals remained for rs6457327 (allelic p-value=6.9×10^{-6}, q-value=5.0×10^{-4}) and rs2517448 (allelic p-

value=7.7×10^{-6}, q-value=5.0×10^{-4}). Five neighboring associated SNPs were correlated with the two top SNPs ($r^2 > 0.6$, Supplementary Fig. 2b). All of the associated SNPs lie within a 26kb block of high LD that covers *STG* and expands 23kb downstream (Supplementary Fig. 2c). *STG* is the only gene overlapping this block, and no other associated SNPs were found outside this block.

We also imputed SNP genotypes within 30kb of rs6457327 in the NC1/2 sample, confirming that the main signal lies in the 26kb block containing *STG*. Twelve imputed SNPs showed trend p-values $< 10 \times 10^{-4}$ (Supplementary Table 5); 11 of these were located within the block (Fig. 1b) in a low recombination region (Fig. 1c). We used conditional logistic regression to evaluate whether a single SNP could account for all observed association signals in this region. Conditioning on any of the three most statistically significant polymorphisms (rs6457327, rs2517448 and the imputed SNP, rs3132562) in an additive test of association abolished the association signals from all other markers (Supplementary Table 5), suggesting that a single locus may be associated with FL in the *PSORS1* region. We found no evidence of an epistatic effect on FL risk, and the results of haplotype analyses did not provide additional information. Because these three SNPs are in complete LD ($r^2 = 1$ in HapMap-CEU), and recombination rates in this region are low relative to the genomic average[12], their effects could not be unambiguously separated, and we could not identify a more restricted associated region.

STG was originally described as a taste-bud specific gene in rhesus monkeys, though STG protein function in humans is unknown. *STG* has previously been reported to be highly expressed in multiple hematopoietic tissues[11]. We examined *STG* expression in human whole blood and lymphoblastoid cell lines revealing expression of an unspliced *STG* transcript, whereas a spliced *STG* transcript was found in tonsil tissue (Supplementary Fig. 3). Eight SNPs in the region, including six that we imputed or genotyped, were either non-synonymous or could disrupt regulatory sequences (Supplementary Methods) and are thus functional candidates (Supplementary Table 6). Of particular interest was rs1265054, a non-synonymous SNP located in an exonic splicing enhancer motif predicted to disrupt serine/arginine-rich protein binding and normal splicing. Upon genotyping rs1265054, we found it is in complete LD with rs6457327 ($r^2 = 1.0$). Future studies are needed to assess the potential relevance of *STG* splice variants as well as this candidate SNP to FL risk.

Our initial GWAS were limited in power, due to the relatively small sample sets included in the genome-wide genotyping phase, particularly for the CLL/SLL subtype. Though we found no statistically significant associations for CLL/SLL, three of seven SNPs highly associated with CLL in a recent GWAS[12] were ranked among the top 0.3% of SNPs associated with CLL/SLL. Specifically, rs735665 and rs13397985 in *SP140* and rs872071 in *IRF4* ranked 128, 503 and 1395, respectively, suggesting that these associations were detectable even with our modestly sized pooled study.

In summary, we have identified a novel FL risk locus on chromosome band 6p21.33 near *PSORS1* with a combined allelic p-value=4.7×10^{-11}. Although *STG* may be a plausible candidate FL susceptibility gene, we cannot exclude a potential role for other genes in this region. Further studies are required to identify the causal variant(s), evaluate whether common risk alleles exist between FL and psoriasis, and to fully dissect the association between the *PSORS1* locus and FL pathogenesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Morton LM, et al. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes. Blood. 2008

2. Skibola CF, Curry JD, Nieters A. Genetic susceptibility to lymphoma. Haematologica 2007;92:960–9. [PubMed: 17606447]

3. Skibola CF, et al. Polymorphisms in the estrogen receptor 1 and vitamin C and matrix metalloproteinase gene families are associated with susceptibility to lymphoma. PLoS ONE 2008;3:e2816. [PubMed: 18636124]

4. Salari K, et al. Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. Genet Epidemiol 2005;29:76–86. [PubMed: 15918156]

5. Homer N, et al. Multimarker analysis and imputation of multiple platform pooling-based genome-wide association studies. Bioinformatics 2008;24:1896–902. [PubMed: 18617537]

6. Becker N, Deeg E, Nieters A. Population-based study of lymphoma in Germany: rationale, study design and first results. Leuk Res 2004;28:713–24. [PubMed: 15158093]

7. Skibola CF, et al. Body mass index, leptin and leptin receptor polymorphisms, and non-hodgkin lymphoma. Cancer Epidemiol Biomarkers Prev 2004;13:779–86. [PubMed: 15159310]

8. Spinelli JJ, et al. Organochlorines and risk of non-Hodgkin lymphoma. Int J Cancer 2007;121:2767–75. [PubMed: 17722095]

9. Nair RP, et al. Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. Am J Hum Genet 2006;78:827–51. [PubMed: 16642438]

10. Nair B, Raval G, Mehta P. TNF-alpha inhibitor etanercept and hematologic malignancies: report of a case and review of the literature. Am J Hematol 2007;82:1022–4. [PubMed: 17654504]

11. Su AI, et al. Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A 2002;99:4465–70. [PubMed: 11904358]

12. Di Bernardo MC, et al. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. Nat Genet 2008;40:1204–10. [PubMed: 18758461]
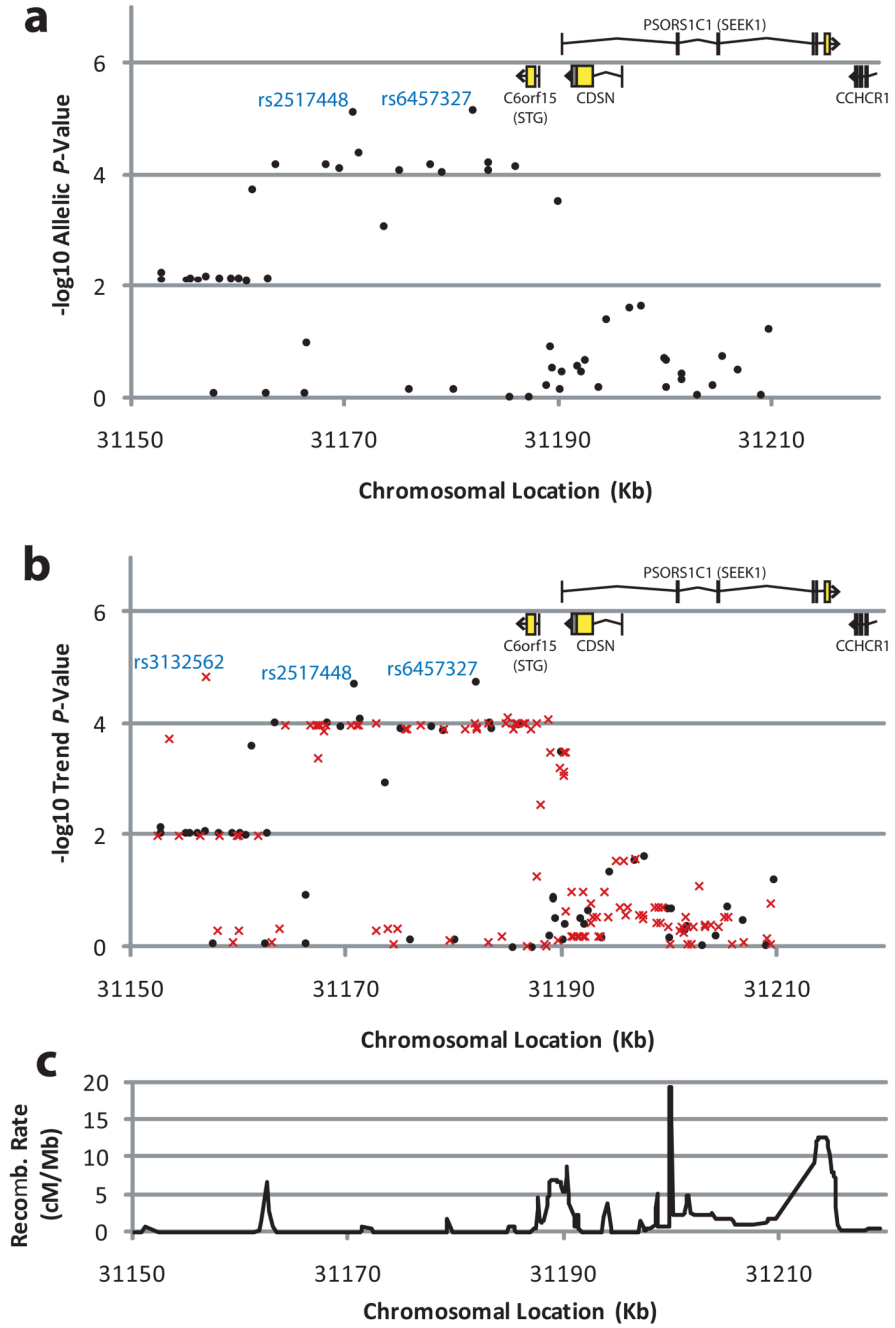
Figure 1. *P*-values for association testing of SNPs in the 6p21.33 region
(**a**) −log10 allelic *p*-values from the individual genotyping of the NC1 and NC2 sample sets.
(**b**) −log10 trend p-values for both individually genotyped (□) and imputed SNPs (**x**) in the
NC1 and NC2 sample sets. (**c**) Regional recombination rates (centimorgans per Mb) estimated
from Phase II HapMap are shown at the bottom.

Table 1

Association results for rs6457327 and follicular lymphoma (FL) and diffuse large B-cell lymphoma (DLBCL) in the San Francisco Bay Area (NC1/NC2, NC3), German (G1) and Canadian (C1) studies.

Study Population	Number of Cases			Allele Frequency			FL OR$_{allelic}$ (95% CI)	FL OR$_{homo}$ (95% CI)	FL $P_{allelic}$	FL P_{trend}	DLBCL OR$_{allelic}$ (95% CI)	DLBCL OR$_{homo}$ (95% CI)	DLBCL $P_{allelic}$	DLBCL P_{trend}
	FL Cases	DLBCL Cases	Controls	FL Cases	DLBCL Cases	Controls								
SF Bay Area (NC1&2)	278	380	1412	0.26	0.32	0.36	0.55 (0.43–0.72)	0.49 (0.30–0.76)	1.1×10^{-5}	1.8×10^{-5}	0.79 (0.63–0.99)	0.70 (0.47–1.0)	2.7×10^{-2}	3.0×10^{-2}
Germany (G1)	87	152	669	0.31	0.37	0.37	0.43 (0.22–0.84)	0.28 (0.10–0.79)	1.0×10^{-2}	1.0×10^{-2}	1.0 (0.67–1.7)	1.2 (0.59–2.5)	0.82	0.68
SF Bay Area (NC3)	108	98	685	0.26	0.28	0.38	0.59 (0.39–0.89)	0.26 (0.10–0.57)	6.7×10^{-4}	1.0×10^{-3}	0.63 (0.41–0.97)	0.34 (0.14–0.73)	5.1×10^{-3}	7.5×10^{-3}
Canada (C1)	172	153	611	0.30	0.32	0.41	0.62 (0.44–0.87)	0.41 (0.23–0.69)	3.9×10^{-4}	8.4×10^{-4}	0.75 (0.52–1.08)	0.39 (0.20–0.71)	5.0×10^{-3}	7.5×10^{-3}
Combined Analysis (NC1&2/G1/NC3/C1)	645	783	3377	0.28	0.32	0.38	0.59 (0.50–0.70)	0.45 (0.34–0.61)	4.7×10^{-11}	1.9×10^{-10}	0.78(0.67–0.92)	0.60(0.46–0.78)	7.0×10^{-5}	1.1×10^{-4}

Shown is a summary of the minor allele frequencies, odds ratios (OR) and p-values for rs6457327 and risk of follicular lymphoma (FL) and diffuse large B-cell lymphoma (DLBCL). The total allele counts were used to generate a combined allelic p-value using a Pearson's chi-square test, and the total genotypic counts were used to obtain a combined trend p-value using a two-sided asymptotically independence test. See Supplementary Methods for more detailed information on the statistical analyses.