nature genetics

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

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

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 another 456 FL cases and 2,785 controls (combined allelic $P = 4.7 \times 10^{-11}$). The region of strongest association overlapped *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]. NHL is the fifth most common among all cancers in the United States and has doubled in annual incidence 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: FL, diffuse large B-cell (DLBCL) and chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL)[1,2]. We therefore conducted genome-wide association studies (GWAS) using separate DNA pools from 189 individuals who had developed FL, 221 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 ancestry informative marker sets (AIMS) genotyping[4] to diminish potential underlying population stratification. Self-reported ethnicity and ancestry data were highly correlated (95%) and were used to construct homogeneous DNA pools of participants of European descent.

In the first phase, pools were hybridized to Human Hap550v.3 BeadChips (Illumina), 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. Eighty-seven percent of raw allelic *P*-values were <0.05

Table 1 Association of rs6457327 with follicular lymphoma and diffuse large B-cell lymphoma in the NC1 + NC2, NC3, G1 and C1 studies

	Number of cases			Allele frequency										
Study population	FL cases	DLBCL cases	Controls	FL cases	DLBCL cases	Controls	FL $OR_{allelic}$ (95% CI)	FL OR_{homoz} (95% CI)	FL $P_{allelic}$	FL P_{trend}	DLBCL $OR_{allelic}$ (95% CI)	DLBCL OR_{homoz} (95% CI)	DLBCL $P_{allelic}$	DLBCL P_{trend}
SF Bay Area (NC1 + NC2)	278	380	1,412	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	645	783	3,377	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}

Summary of the minor allele frequencies, allelic and homozygous variant genotype (homoz) odds ratios (OR) and *P*-values for rs6457327 and risk of follicular lymphoma (FL) and diffuse large B-cell lymphoma (DLBCL). The total allelic counts were used to generate a combined allelic *P*-value using a Pearson's chi-squared 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. Combined analyses includes data from all four preceding rows (NC1 + NC2, G1, NC3, C1). SF, San Francisco; CI, confidence interval.

[1]Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, California, USA. [2]Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA. [3]International Computer Science Institute, Berkeley, California, USA. [4]Translational Genomics Research Institute (TGen), Phoenix, Arizona, USA. [5]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany. [6]Canada's Michael Smith Genome Sciences Center and [7]Cancer Control Research, British Columbia Cancer Agency, Vancouver, British Columbia, Canada. [8]Molecular Epidemiology, Center of Chronic Immunodeficiency, University of Freiburg, Freiburg, Germany. Correspondence should be addressed to C.F.S. (chrisfs@berkeley.edu).

Received 29 January; accepted 18 June; published online 20 July 2009; doi:10.1038/ng.419

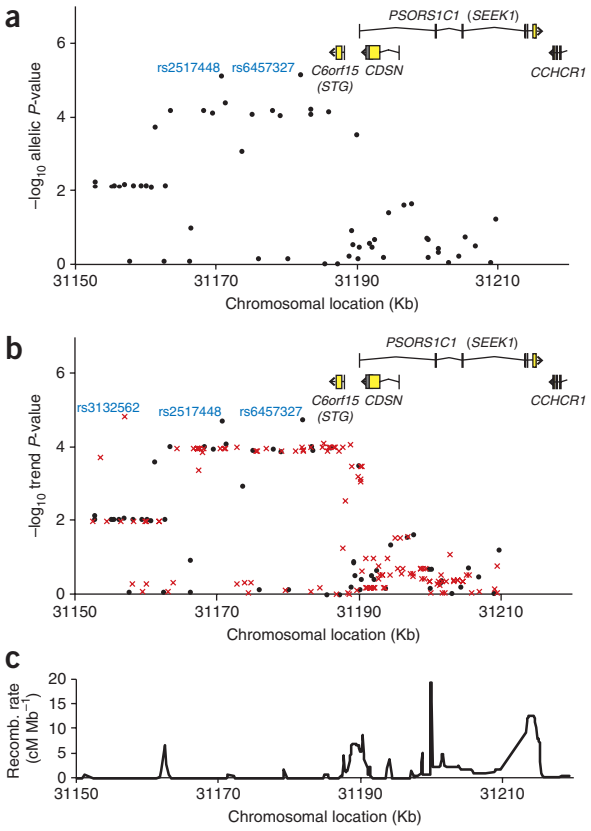

Figure 1 *P*-values for association testing of SNPs in the 6p21.33 region. (**a**) $-\log_{10}$ allelic *P*-values from the individual genotyping of the NC1 and NC2 sample sets. (**b**) $-\log_{10}$ trend *P*-values for both individually genotyped (•) and imputed (×) SNPs in the NC1 and NC2 sample sets. (**c**) Regional recombination (recomb.) rates (centimorgans per megabase) estimated from Phase II HapMap.

(**Supplementary Table 2a–c**), and genotype frequencies did not significantly differ from Hardy-Weinberg equilibrium. We subsequently genotyped 32 SNPs with subtype-specific allelic *q*-values (corrected *P*) <0.05 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 linkage disequilibrium, we excluded rs2517448 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 (odds ratio = 0.28, $P = 0.01$, **Table 1**). No other associations were validated. In the third phase, rs6457327 was validated in another two 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 $\alpha = 0.05$, 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 $P = 7.0 \times 10^{-5}$, trend $P = 1.1 \times 10^{-4}$; **Table 1**).

Respectively, rs6457327 and rs2517448 are 5 and 16 kb downstream of the 3′ UTR of the *C6orf15 (STG)* gene, telomeric to *HLA-C* on chromosome 6p21.33 in the major histocompatibility complex. This 300-kb region has been extensively evaluated because of its association with psoriasis, for which 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 to family history[10]. We found little linkage disequilibrium 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 more SNPs within 30 kb of rs6457327 in the NC1 and NC2 samples, where we found 13 more markers with $q < 0.05$ (**Fig. 1a** and **Supplementary Table 4**). The strongest signals remained for rs6457327 (allelic $P = 6.9 \times 10^{-6}$, $q = 5.0 \times 10^{-4}$) and rs2517448 (allelic $P = 7.7 \times 10^{-6}$, $q = 5.0 \times 10^{-4}$). 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 26-kb block of high linkage disequilibrium that covers *STG* and expands 23-kb 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 30 kb of rs6457327 in the NC1/2 sample, confirming that the main signal lies in the 26-kb block containing *STG*. Twelve imputed SNPs showed a trend $P < 10 \times 10^{-4}$ (**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 upon 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 linkage disequilibrium ($r^2 = 1$ in HapMap CEPH (Centre d'Etudes du Polymorphisme Humain) CEU individuals (Utah residents with northern and western European ancestry)) and recombination rates in this region are low relative to the genomic average[11], 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 been reported to be highly expressed in multiple hematopoietic tissues. We examined *STG* expression in human whole blood and lymphoblastoid cell lines, finding expression of an unspliced *STG* transcript, whereas a spliced *STG* transcript was found in tonsilar tissue (**Supplementary Table 6** and **Supplementary Fig. 3**). Eight

SNPs in the region, including six that we imputed or genotyped, were either nonsynonymous or could disrupt regulatory sequences (**Supplementary Methods**) and are thus functional candidates (**Supplementary Table 7**). Of particular interest was rs1265054, a nonsynonymous 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 to be in complete linkage disequilibrium 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, owing 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 chronic lymphocytic leukemia in a recent GWAS[12] 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 1,395, respectively, suggesting that these associations were detectable even with our modestly sized pooled study.

In summary, we have identified a new FL risk locus on chromosome band 6p21.33 near *PSORS1* with a combined allelic $P = 4.7 \times 10^{-11}$. 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), to evaluate whether risk alleles exist in common between FL and psoriasis and to fully dissect the association between the *PSORS1* locus and FL pathogenesis.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank S. Leach, J. Schuetz and A. MacArthur for assistance. This work was supported by US National Institutes of Health grants CA122663 (C.F.S.), CA104862 (M.T.S.), and CA45614, CA89745 and CA87014 (E.A.H.); the American Cancer Society (IRG-07-06401), the US National Institutes of Health (HL086528) and a charitable donation by Sylvia Chase (K.M.B.); the Stardust Foundation (D.W.C.); the Canadian Cancer Society and National Cancer Institute of Canada, and the Canadian Institutes of Health Research (J.J.S. and A.B.-W.); and the German José Carreras Leukemia Foundation (DJCLS R04/08 and R07/26f; A.N.).

AUTHOR CONTRIBUTIONS

C.F.S., J.J.S., A.B.-W., E.A.H. and N.B. are principal investigators for the participating studies; L.A. and J.R. did DNA extraction, normalization and quality control; K.M.B. and K.I. prepared DNA pools and performed the genome scan and analysis; D.W.C., C.F.S., K.M.B., M.T.S. and L.Z. consulted on study design; L.A. undertook genotyping; J.D.C. performed expression analysis; D.W.C., P.M.B., A.N. and L.C. performed the statistical analyses; L.C. and E.H. conducted bioinformatics analyses; C.S.F., L.C. and K.M.B. wrote the manuscript.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Morton, L.M. *et al.* Etiologic heterogeneity among non-Hodgkin lymphoma subtypes. *Blood* **112**, 5150–5160 (2008).
2. Skibola, C.F., Curry, J.D. & Nieters, A. *Haematologica* **92**, 960–969 (2007).
3. Skibola, C.F. *et al. PLoS ONE* **3**, e2816 (2008).
4. Salari, K. *et al. Genet. Epidemiol.* **29**, 76–86 (2005).
5. Homer, N. *et al. Bioinformatics* **24**, 1896–1902 (2008).
6. Becker, N., Deeg, E. & Nieters, A. *Leuk. Res.* **28**, 713–724 (2004).
7. Skibola, C.F. *et al. Cancer Epidemiol. Biomarkers Prev.* **13**, 779–786 (2004).
8. Spinelli, J.J. *et al. Int. J. Cancer* **121**, 2767–2775 (2007).
9. Nair, R.P. *et al. Am. J. Hum. Genet.* **78**, 827–851 (2006).
10. Nair, B., Raval, G. & Mehta, P. *Am. J. Hematol.* **82**, 1022–1024 (2007).
11. Su, A.I. *et al. Proc. Natl. Acad. Sci. USA* **99**, 4465–4470 (2002).
12. Di Bernardo, M.C. *et al. Nat. Genet.* **40**, 1204–1210 (2008).