nature genetics

# Genome-wide association study of B cell non-Hodgkin lymphoma identifies 3q27 as a susceptibility locus in the Chinese population

Dennis E K Tan[1,22], Jia Nee Foo[1,22], Jin-Xin Bei[1,2,22], Jiang Chang[3,4,22], Roujun Peng[2], Xiaohui Zheng[2], Lixuan Wei[3,4], Ying Huang[3,4], Wei Yen Lim[5], Juan Li[6], Qian Cui[2], Soo Hong Chew[7], Richard P Ebstein[8], Ponnudurai Kuperan[9], Soon Thye Lim[10], Miriam Tao[10], Suat Hoon Tan[11], Alvin Wong[12], Gee Chuan Wong[13], Soo Yong Tan[14], Siok Bian Ng[15], Yi-Xin Zeng[2,16], Chiea Chuen Khor[1,5,17–19], Dongxin Lin[3,4,23], Adeline L H Seow[5,23], Wei-Hua Jia[2,23] & Jianjun Liu[1,5,20,21,23]

© 2013 Nature America, Inc. All rights reserved.

**To identify genetic risk factors underlying non-Hodgkin lymphomas (NHLs) from the B cell lineage, we conducted a genome-wide association study (GWAS) of 253 Chinese individuals with B cell NHL (cases) and 1,438 controls and further validation in 1,175 cases and 5,492 controls. We identified a new susceptibility locus, rs6773854, located between *BCL6* (encoding B cell lymphoma protein 6) and *LPP* (encoding lipoma preferred partner) on oncogene-rich chromosome 3q27 that was significantly associated with increased risk of B cell NHL (meta-analysis $P = 3.36 \times 10^{-13}$, per-allele odds ratio (OR) = 1.44) and with diffuse large B cell lymphoma (DLBCL) in particular (meta-analysis $P = 1.14 \times 10^{-11}$, OR = 1.47). We found no evidence of association of rs6773854 with non–B cell NHLs (T cell and natural killer (NK) lineages) ($P = 0.17$, OR = 1.12) and observed significant heterogeneity between B cell and non–B cell subtypes ($P_{het} = 0.01$, $I^2 = 84\%$). Our results provide insight that germline variation in the intergenic region between *BCL6* and *LPP* has a role in risk of B cell lymphomagenesis.**

NHLs (MIM 605027), a complex group of malignant tumors arising from lymphoid tissue, are characterized by the clonal expansion of tumorigenic immune cells in various stages of differentiation[1]. NHLs from B cell lineages are more common, with DLBCL being the most common and aggressive histological subtype worldwide[1,2]. In support of a genetic basis for the disease, NHLs were observed to exhibit familial aggregation, and up to tenfold increased risk has been reported for the first-degree relatives of probands with DLBCL[3]. Although genetic risk factors have been hypothesized to be specific to each histological subtype of NHL, epidemiological studies suggest that some of these risk factors may be shared by different subtypes[4].

Several genetic association studies have been performed in search of common risk variants for NHL[5–8]; however, these studies were predominantly conducted in individuals of European ancestry, and investigation in Asians has been lacking. Besides robust associations reported between HLA (human leukocyte antigen) variants and risk of follicular lymphoma[5–7], no other genome-wide significant locus has been identified for the other NHL subtypes. Hence, to gain insights into genetic susceptibility factors for B cell NHLs in Asians, we conducted a GWAS of 274 Chinese individuals with B cell NHL and 1,500 controls recruited in Singapore. After applying stringent quality filters (Online Methods), there were 253 B cell NHLs (including 51 follicular lymphomas, 54 marginal zone lymphomas and 148 DLBCLs), 1,438 controls (**Supplementary Table 1a**) and 550,946 autosomal SNPs available for association analysis.

[1]Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore. [2]State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Guangzhou, China. [3]State Key Laboratory of Molecular Oncology, Cancer Institute and Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing, China. [4]Department of Etiology and Carcinogenesis, Cancer Institute and Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing, China. [5]Saw Swee Hock School of Public Health, National University Health System, National University of Singapore, Singapore. [6]Department of Hematology, the First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. [7]Department of Economics, National University of Singapore, Singapore. [8]Department of Psychology, National University of Singapore, Singapore. [9]Division of Haematology and Laboratory Pathology, Tan Tock Seng Hospital, Singapore. [10]Division of Medical Oncology, National Cancer Centre, Singapore. [11]Department of Dermatology, National Skin Centre, Singapore. [12]Department of Haematology-Oncology, National University Health System, National University of Singapore, Singapore. [13]Department of Haematology, Singapore General Hospital, Singapore. [14]Department of Pathology, Singapore General Hospital, Singapore. [15]Department of Pathology, National University Health System, National University of Singapore, Singapore. [16]Peking Union Medical College, Chinese Academy of Medical Science, Beijing, China. [17]Singapore Eye Research Institute, Singapore. [18]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. [19]Department of Pediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. [20]School of Biological Sciences, Anhui Medical University, Hefei, China. [21]Institute of Dermatology and Department of Dermatology, No.1 Hospital, Anhui Medical University, Hefei, China. [22]These authors contributed equally to this work. [23]These authors jointly directed this work. Correspondence should be addressed to J. Liu (liuj3@gis.a-star.edu.sg), W.-H.J. (jiawh@sysucc.org.cn), A.L.H.S. (adeline_seow@nus.edu.sg) or D.L. (lindx72@cicams.ac.cn).

Received 19 October 2012; accepted 13 May 2013; published online 9 June 2013; doi:10.1038/ng.2666

**Table 1 Summary statistics for rs6773854 in GWAS and all validation stages**

| Cases | Controls | Study | MAF | | Trend *P* | OR (95% CI) | $P_{meta}$ | $OR_{meta}$ (95% CI) | $P_{het}$ ($I^2$) |
|---|---|---|---|---|---|---|---|---|---|
| | | | Cases | Controls | | | | | |
| 253 | 1,438 | GWAS | 0.31 | 0.22 | $6.89 \times 10^{-6}$ | 1.64 (1.32–2.03) | – | – | – |
| 168 | 1,632 | Validation 1 (Guangzhou) | 0.26 | 0.18 | $4.47 \times 10^{-4}$ | 1.61 (1.2 –2.09) | – | – | – |
| 294 | 2,127 | Validation 2 (Guangzhou) | 0.26 | 0.21 | 0.003 | 1.34 (1.10–1.63) | – | – | – |
| 713 | 1,733 | Validation 3 (Beijing) | 0.21 | 0.17 | $1.00 \times 10^{-4}$ | 1.36 (1.16–1.58) | – | – | – |
| 1,175 | 5,492 | All validation | – | – | – | – | $4.11 \times 10^{-9}$ | 1.39 (1.24–1.55) | 0.50 (0) |
| 1,428 | 6,930 | All data | – | – | – | – | $3.36 \times 10^{-13}$ | 1.44 (1.31–1.59) | 0.37 (3) |

Association results are shown for the minor (C) and major (T) alleles of rs6773854. Only B cell NHLs were included in the single-SNP analysis, and the breakdown for selected subtypes is detailed in **Supplementary Table 1**. Trend *P*, Cochran-Armitage trend test *P* value.

We performed a standard GWAS statistical analysis (Online Methods) and identified a total of 59 SNPs with associations that exceeded discovery $P \leq 1 \times 10^{-4}$ in single SNP–based association analysis (**Supplementary Table 2a**). We were able to design assays for 52 of these 59 SNPs for validation on the Sequenom MassArray platform (**Supplementary Table 2b**). Forty-seven of 52 SNPs were successfully genotyped and passed quality control filters similar to those used in the GWAS stage (Online Methods and **Supplementary Table 2b**) in a first validation set collected from a Han Chinese population in Guangzhou, China, comprising 168 cases and 1,632 controls. We selected a total of five SNPs—four SNPs with association *P* values ranging from $7.58 \times 10^{-5}$ to 0.04 and one SNP with association $P = 0.061$ (ref. 9)—which showed similar effect sizes and had no significant heterogeneity ($P_{het} > 0.10$, $I^2 < 30\%$) in the preliminary meta-analysis (**Supplementary Table 2a**) for further validations in two additional sample collections recruited in Guangzhou (294 cases and 2,127 controls) and in Beijing (713 cases and 1,733 controls) (**Supplementary Table 1a**). Our GWAS approach is summarized in **Supplementary Figure 1**. Consistent evidence of association was observed between the minor allele (C) of rs6773854 and increased risk of B cell NHL across both validation studies (Guangzhou: validation $P = 0.003$, per-allele OR = 1.34, 95% confidence interval (CI) = 1.10–1.63; Beijing: validation $P = 1.0 \times 10^{-4}$, per-allele OR = 1.36, 95% CI = 1.16–1.58). The remaining four SNPs showed no evidence of association and were not analyzed further (**Supplementary Table 3**). Meta-analysis of all data sets in both the discovery and validation stages showed that rs6773854 was significantly associated with overall B cell NHL risk at genome-wide significance and without heterogeneity between the four independent sample collections ($P_{meta} = 3.36 \times 10^{-13}$, per-allele OR = 1.44, 95% CI = 1.31–1.59; $P_{het} = 0.37$, $I^2 = 3\%$; **Table 1**).

Minimal heterogeneity between sample collections was observed, despite the different proportions of B cell NHL subtypes in the discovery and validation sets, suggesting that the risk conferred by rs6773854 might be consistent across most B cell NHL subtypes. Nonetheless, we performed further analyses by including only B cell NHL subtypes that were consistently found in all GWAS and validation sample collections (follicular lymphoma and DLBCL). We observed consistent associations in the meta-analyses of all data sets for both single (follicular lymphoma or DLBCL alone) and combined subtypes (follicular lymphoma and DLBCL) without evidence of heterogeneity between studies (**Supplementary Table 4**). The association of rs6773854 with DLBCL alone also reached genome-wide significance in the meta-analysis ($P_{meta} = 1.14 \times 10^{-11}$, OR = 1.47, 95% CI = 1.32–1.65; $P_{het} = 0.21$, $I^2 = 33\%$; **Supplementary Table 4**). We further examined the association of rs6773854 with non–B cell NHLs (T cell and NK NHLs) in 462 cases and 6,930 controls and observed no evidence of association ($P = 0.17$, per-allele OR = 1.12, 95% CI = 0.94–1.33) and significant heterogeneity between overall B cell and non–B cell subtypes ($P_{het} = 0.01$, $I^2 = 84\%$), although a similar trend of association with weaker effect size was observed in non–B cell NHLs (**Fig. 1** and **Supplementary Table 4**). This finding further suggests that rs6773854 has a stronger genetic effect in B cell NHLs than in non–B cell subtypes; however, more work is required to dissect its role(s) in other B cell NHL subtypes that are rarer in Asians to support our hypothesis. Finally, we analyzed risk variants for DLBCL and follicular lymphoma that were previously reported by other genetic studies[5–8]. Although associations were not statistically significant in our discovery data set, similar trends of association (ORs in the same direction) were observed for some SNPs in our data set, particularly for HLA risk variants for follicular lymphoma, which showed consistent association in different populations of European ancestry[5,7] (**Supplementary Table 5**).

The minor allele frequency (MAF) for rs6773854 was observed to be slightly different between the southern (Singapore and Guangzhou) and northern (Beijing) Chinese controls (**Table 1**). However, in each of the three validation cohorts, the cases and controls were geographically matched, which is a good surrogate for genetic matching in the Han Chinese population[10–13]. No heterogeneity ($I^2 = 3\%$) was observed among all four independent sample collections, suggesting that population stratification, if any, was unlikely to have affected our results (**Table 1**). Notably, different MAFs might explain differences in the incidence rate of NHL across the southern, central and northern regions of China, with higher MAF among southern Chinese individuals possibly being correlated with the higher disease incidence rate in this region (**Supplementary Table 6**, Guangzhou versus Beijing).

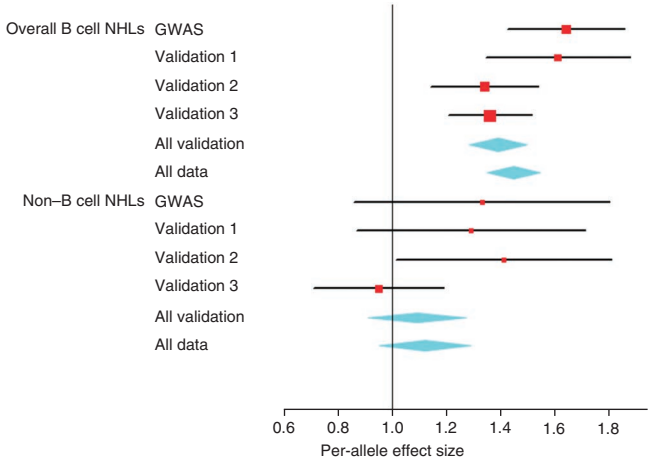


**Figure 1** Forest plot showing the per-allele ORs of rs6773854 for overall B cell NHLs compared to non–B cell NHLs in each stage of the study. Ranges represent the 95% CIs for the effects.

© 2013 Nature America, Inc. All rights reserved.

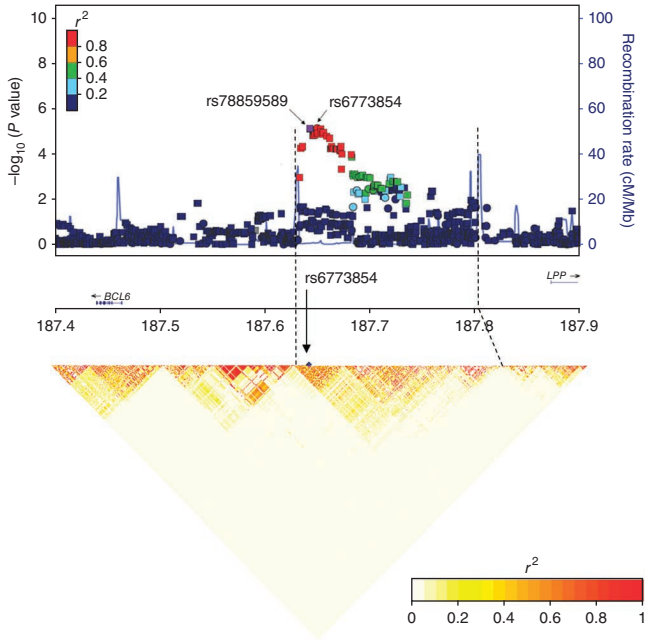


**Figure 2** Regional association plot on chromosome 3q27. Association evidence ($-\log_{10}$ *P* values, derived from trend tests) is shown for both genotyped (circles) and imputed (squares) SNPs plotted according to chromosomal position. Bottom, pairwise LD measures (in $r^2$) for SNPs within the region. rs6773854 appears to be confined within a unique LD block on its own, and imputation did not identify coding variants within *BCL6* or *LPP* in LD with rs6773854.

The associated risk variant rs6773854 maps to the intergenic region between *BCL6* and *LPP*. To maximize the coverage of genetic variants at this locus, we performed imputation analysis in the 5-Mb genomic region centered on rs6773854 using the GWAS data set (Online Methods). After applying stringent quality control filters for genotype imputation (Online Methods), a total of 12,364 SNPs (1,072 genotyped and 11,292 imputed) were available for association analysis. Despite the increased number of variants analyzed, we did not observe any stronger association with imputed SNPs over and above those with the directly genotyped SNPs. No residual association signal remained after conditioning on rs6773854, showing that this SNP alone accounts for the majority of the genetic association at this locus, and the associated SNP was not in linkage disequilibrium (LD) with any imputed coding variants in the exons of *LPP* or *BCL6* (**Fig. 2**).

Recently, this intergenic region (~400 kb in length) between *BCL6* and *LPP* was found to be highly conserved and to be the site of epigenetic modifications that were correlated with expression changes in some genes, including *BCL6* and *LPP*, suggesting the presence of strong regulatory elements within this region[14]. *BCL6* encodes a nuclear transcriptional repressor that is essential for germinal center formation and the regulation of several lymphocyte functions[1,15]. The other candidate, *LPP*, was first identified by its involvement in recurring t(3;12) translocation in lipomas[16]. Dysregulated expression of both *BCL6* and *LPP* has been documented to contribute to malignant transformation, especially for *BCL6* in B cell NHLs of germinal center origin, by inhibiting cellular differentiation and apoptosis or by enhancing cell proliferation[15,17,18]. As rs6773854 lies within the conserved region containing strong regulatory elements[14], we speculate that rs6773854 might tag polymorphisms in the regulatory elements that may modulate the expression of either *BCL6* or *LPP*, and we further propose that the association between rs6773854 and susceptibility to B cell NHL might be due to altered *BCL6* or *LPP* expression. We therefore screened rs6773854 and its tagging SNPs in available expression quantitative trait locus (eQTL) databases (MuTHER, Gencord and HapMap 3) using Genevar[19] but found no evidence ($P > 0.05$) supporting the association of these SNPs with *LPP* or *BCL6* expression in any data set, although none of these included data collected with germinal center B cells (the most relevant tissue type).

In conclusion, we have identified a new susceptibility locus that is strongly associated with increased risks of B cell NHL and DLBCL in the Chinese population. Further fine mapping of the intergenic region between *BCL6* and *LPP* is warranted to identify the causal variant. Our result supports the notion of shared genetic susceptibilities among NHL subtypes. Cross-validation of our data with European data sets[5–7] will be key in identifying additional genetic variants that might help to unravel the disease mechanisms involved in B cell NHL lymphomagenesis and to understand the differential risks for NHL subtypes between populations.

**URLs.** Sequenom, http://www.sequenom.com/; LocusZoom, http://csg.sph.umich.edu/locuszoom/; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; R, http://www.r-project.org/; Applied Biosystems, http://www.appliedbiosystems.com/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/.

## METHODS

Methods and any associated references are available in the online version of the paper.

*Note: Supplementary information is available in the online version of the paper.*

### ACKNOWLEDGMENTS

This work was supported by the Agency for Science, Technology and Research (A*STAR), Singapore, a grant from the Biomedical Research Council, A*STAR, Singapore (04/1/21/19/336), a grant from HSBC Trustee Ltd. (Singapore) as trustees of the Major John Long Trust Fund, the Chew Woon Poh Trust Fund, the National Key Scientific and Technological Project (W.-H.J., 2011ZX09307-001-04) and the Guangzhou Zhujiang nova program (J.-X.B.). Additional Singapore Lymphoma Project Investigators include B.-C. Tai, S.-E. Chia, K.-M. Lee, S. Loong, L.H.C. Tan and B. Mow. We also acknowledge the contributions of K.-Y. Wong and C. Ying to this work.

### AUTHOR CONTRIBUTIONS

D.E.K.T., J.N.F., J.-X.B. and J.C. contributed equally to the work and wrote the manuscript, with contributions from J. Liu, C.C.K., D.L., A.L.H.S. and W.-H.J. D.E.K.T., J.C., X.Z. and Q.C. performed genotyping. D.E.K.T., J.N.F., J.-X.B. and J.C. performed statistical analyses. R.P., X.Z., L.W., Y.H., W.Y.L., J. Li, Q.C., S.H.C., R.P.E., P.K., S.T.L., M.T., S.H.T., A.W., G.C.W., S.Y.T., S.B.N. and Y.-X.Z. provided samples and data for their respective sample collections. C.C.K., D.L., A.L.H.S., W.-H.J. and J. Liu jointly supervised the study. All authors contributed to and approved the final manuscript.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Lenz, G. & Staudt, L.M. Aggressive lymphomas. *N. Engl. J. Med.* **362**, 1417–1429 (2010).
2. Roman, E. & Smith, A.G. Epidemiology of lymphomas. *Histopathology* **58**, 4–14 (2011).
3. Goldin, L.R., Bjorkholm, M., Kristinsson, S.Y., Turesson, I. & Landgren, O. Highly increased familial risks for specific lymphoma subtypes. *Br. J. Haematol.* **146**, 91–94 (2009).
4. Kristinsson, S.Y. *et al.* Genetic and immune-related factors in the pathogenesis of lymphoproliferative and plasma cell malignancies. *Haematologica* **94**, 1581–1589 (2009).
5. Conde, L. *et al.* Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nat. Genet.* **42**, 661–664 (2010).

© 2013 Nature America, Inc. All rights reserved.

6. Skibola, C.F. *et al.* Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. *Nat. Genet.* **41**, 873–875 (2009).
7. Smedby, K.E. *et al.* GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma. *PLoS Genet.* **7**, e1001378 (2011).
8. Rothman, N. *et al.* Genetic variation in *TNF* and *IL10* and risk of non-Hodgkin lymphoma: a report from the InterLymph Consortium. *Lancet Oncol.* **7**, 27–38 (2006).
9. Thomas, G. *et al.* A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (*RAD51L1*). *Nat. Genet.* **41**, 579–584 (2009).
10. Chen, J. *et al.* Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *Am. J. Hum. Genet.* **85**, 775–785 (2009).
11. Xu, S. *et al.* Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am. J. Hum. Genet.* **85**, 762–774 (2009).
12. Chen, Z.J. *et al.* Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat. Genet.* **43**, 55–59 (2011).
13. Shi, Y. *et al.* Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat. Genet.* **43**, 1224–1227 (2011).
14. Ramachandrareddy, H. *et al. BCL6* promoter interacts with far upstream sequences with greatly enhanced activating histone modifications in germinal center B cells. *Proc. Natl. Acad. Sci. USA* **107**, 11930–11935 (2010).
15. Jardin, F., Ruminy, P., Bastard, C. & Tilly, H. The *BCL6* proto-oncogene: a leading role during germinal center development and lymphomagenesis. *Pathol. Biol.* (*Paris*) **55**, 73–83 (2007).
16. Petit, M.M., Mols, R., Schenmakers, E.F.P.M., Mandahl,, N. & Van, de Ven, W.J.M. *LPP*, the preferred fusion partner gene of *HMGIC* in lipomas, is a novel member of the LIM protein gene family. *Genomics* **36**, 118–129 (1996).
17. Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–346 (2001).
18. Wu, S. *et al.* Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3′ untranslated region. *Oncogene* **29**, 2302–2308 (2010).
19. Yang, T.P. *et al.* Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* **26**, 2474–2476 (2010).


© 2013 Nature America, Inc. All rights reserved.

## ONLINE METHODS

**Subject enrollment and diagnosis.** *Discovery set (Singapore).* The Singapore Lymphoma Study was a hospital-based case-control study carried out in three major public and two national referral centers for skin disease and cancer. The objectives and methodology are described elsewhere[20]. Briefly, eligible cases ($N$ = 324) were Singapore residents aged 18 years and older with newly diagnosed lymphoid neoplasm, including mature B and T cell neoplasms and Hodgkin lymphoma, determined on the basis of World Health Organization (WHO) classification, presenting to these hospitals and centers between February 2005 and December 2008. Controls were patients admitted to the orthopedic, internal medicine and general surgery departments in the same hospitals whose admission was not for any of the following: diagnosis or suspicion of malignancy, asthma, atopic eczema or allergy, immune-related disorders, peptic ulcer disease, viral hepatitis or tuberculosis. Controls ($N$ = 311) were frequency matched for age (± 5 years), sex, study center and month of diagnosis. Any individuals with lymphoma or hospital controls with a history of HIV/AIDS or organ transplant were excluded. The study was approved by the institutional review board or ethics committee at each participating hospital and institution. Written informed consent was obtained from all participants. All participants answered a structured questionnaire and provided a blood or saliva specimen. Genomic DNA was extracted from the buffy coat of 5-ml samples of whole blood using the FlexiGene DNA kit (Qiagen), and DNA concentrations were measured with the NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific). An additional 1,189 healthy undergraduate students of Singaporean Chinese ancestry were recruited at the National University of Singapore to participate in a study on the biological basis of decision making. They were genotyped on the Human OmniExpress BeadChip and were used as controls in the primary analysis. All subjects gave written informed consent, and the experiment was approved by the institutional review board at the National University of Singapore.

*Validation sets.* Validations 1 and 2: Guangzhou. The first Guangzhou validation set included 255 cases and 1,632 controls, and the second set included 459 cases and 2,162 controls of Han Chinese descent passing quality control. All of the cases were recruited from the Sun Yat-sen University Cancer Center and the First Affiliated Hospital of Sun Yat-sen University in Guangdong province between January 2002 and June 2011. All cases were histopathologically diagnosed as having NHL according to WHO classification. Healthy controls were recruited from physical examination centers at general hospitals in local communities in the same province and were frequency matched to the cases by age (± 5 years), sex and ancestry. Informed consent was obtained. Peripheral blood was collected from each individual, and DNA was extracted from whole-blood samples, peripheral blood mononuclear cells or the buffy coat using the QIAamp DNA Blood kit according to the manufacturer's instructions (Qiagen). DNA was stored at −80 °C before use.

Validation 3: Beijing. The validation samples from Beijing consisted of 1,030 individuals with NHL and 2,018 cancer-free controls. All subjects were self-reported Han Chinese. Subjects were consecutively recruited between January 1992 and December 2010 at the Cancer Hospital at the Chinese Academy of Medical Science. Cases were all diagnosed with histologically confirmed NHL, and subtypes of NHL were determined according to WHO classification. Controls were cancer-free individuals selected from a community nutritional survey of 6,450 individuals recruited from the same region and during the same collection period as for the cases. At the point of recruitment, informed consent was obtained from each subject, and personal data from each participant regarding demographic information and clinical characteristics were collected via clinical record or interview. This study was approved by the institutional review board of the Chinese Academy of Medical Science Cancer Institute.

**Use of population controls.** A total of 6,681 Chinese population controls (1,189 in the GWAS stage and 5,492 in the validation stage) were used in the statistical analyses. The incidence of NHL is relatively low in Asians (<1%), especially in Chinese (age-adjusted incidence: Singapore: 5 in 100,000; Shanghai, 5.5 in 100,000)[21]. Hence, assuming a prevalence of 1%, we estimate the number of potentially misclassified controls to be 12 (out of 1,189) in the GWAS stage and 54 (out of 5,492) in the validation stage. We expect the use of population controls to have had minimal impact on the downstream statistical analysis.

**Genotyping.** Case and control samples were randomized on plates and were genotyped on the Illumina Human OmniExpress BeadChip following the manufacturer's protocol in the GWAS stage. For the first validation, candidate SNPs were genotyped with the Sequenom (see URLs) MassArray iPlex Gold genotyping system. Subsequent validations were carried out using the Applied Biosystems (see URLs) TaqMan genotyping platform.

**Data quality control and statistical analyses.** Stringent quality controls were applied to remove poorly performing SNPs and samples using PLINK (version 1.07; see URLs). SNPs were excluded from downstream analysis if they had greater than 5% missing data, showed gross departure from Hardy-Weinberg equilibrium ($P < 1.0 \times 10^{-4}$) or had a MAF less than 5%. Accordingly, 180,588 SNPs (out of 731,534) were removed, of which 878 showed significant deviation from Hardy-Weinberg equilibrium, 193 had call rates below 95%, and 179,517 had a MAF less than 5%. A total of 550,946 autosomal SNPs were retained for analyses. For sample quality control, samples with an overall genotyping success rate below 95% were excluded from analysis. Principal-component analysis (PCA) was then applied to the remaining samples to detect outliers in terms of population stratification. PCA was performed using EIGENSTRAT, and the Singapore Genome Variation Project (SGVP) reference panel[22] was used (**Supplementary Fig. 2a**). A total of 107 individuals (10 cases and 97 controls; **Supplementary Fig. 2b**) were removed, and PCA plots were generated using the R statistical package (R Project; see URLs). To detect cryptic relatedness between the study samples, identity-by-state (IBS) information was derived using PLINK; no relatives were present in the data set.

For the discovery stage analysis, we excluded cases who had NHL subtypes that were present in small numbers ($N < 20$), whereas, for the validation analysis, all B cell NHL cases were included in the association test, with the exception of those with Burkitt's lymphoma, which is known to have an infectious etiology[23] (derived from Epstein-Barr virus (EBV) infection), as information regarding EBV infection status was lacking in the validation sample collections.

For both the GWAS and validation stages, analyses of SNP associations were performed using the Cochran-Armitage trend test. We generated a Manhattan plot with Haploview (**Supplementary Fig. 3**). The quantile-quantile plot of single-SNP analysis showed a modest excess of extreme $P$ values compared to the null distribution with minimal genome-wide inflation ($\lambda_{GC}$ = 1.003), suggesting the presence of minimal population stratification (**Supplementary Fig. 4**). The quantile-quantile plot was generated with R. Visual inspection of genotyping cluster plots for all 52 SNPs with associations exceeding $P < 1 \times 10^{-4}$ was performed before proceeding to the validation stages. Meta-analysis was performed under a fixed-effect model with pooling of weighted inverse variance using PLINK. The $I^2$ heterogeneity index and Cochran $Q$ statistics were computed to assess the overall effects of heterogeneity. The genotype cluster plot for rs6773854 in the GWAS stage is shown in **Supplementary Figure 5**.

**Genotype imputation.** We used IMPUTE2 for the imputation of SNPs at the associated locus with the 1000 Genomes Project Phase 1 reference panel (June 2011 release; see URLs). We set strict thresholds for imputation, using only SNPs with confidence score of ≥0.9, genotype call rate of ≥0.95, no deviation from Hardy-Weinberg equilibrium ($P > 0.001$), info score of > 0.9 and MAF of >0.01. Imputation was only performed on GWAS samples, and single-SNP associations were analyzed under the trend model. LocusZoom (see URLs) was used to visualize regional association signals. Pairwise LD between SNPs in the regional association plots was calculated using LDheatmap[24], a statistical package in R.


20. Wong, K.Y. *et al.* Sun exposure and risk of lymphoid neoplasms in Singapore. *Cancer Causes Control* **23**, 1055–1064 (2012).
21. Curado, M.P. *et al. Cancer Incidence in Five Continents Vol. IX* (IARC Scientific Publications, Lyon, France, 2007).
22. Teo, Y.Y. *et al.* Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. *Genome Res.* **19**, 2154–2162 (2009).
23. Rezk, S.A. & Weiss, L.M. Epstein-Barr virus–associated lymphoproliferative disorders. *Hum. Pathol.* **38**, 1293–1304 (2007).
24. Shin, J.-H., Blay, S., McNeney, B. & Graham, J. LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Soft.* **16**, Code Snippet 3 (2006).


© 2013 Nature America, Inc. All rights reserved.


doi:10.1038/ng.2666