A genome-wide association study of Hodgkin Lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21, and 10p14 (GATA3)

Victor Enciso-Mora1*, Peter Broderick1*, Yussanne Ma1*, Ruth F Jarrett2**, Henrik Hjalgrim3**, Kari Hemminki4, Anke van den Berg5, Bianca Olver1, Amy Lloyd1, Tracy Lightfoot6, Flora E van Leeuwen7, Asta Försti4, Arjan Diepstra5, Annegien Broeks8, Jayaram Vijayakrishnan1, Lesley Shield2, Annette Lake2, Dorothy Montgomery2, Eve Roman6, Andreas Engert9, Elke Pogge von Strandmann9, Katrin Reiners9, Ilja Nolte5, Karin E Smedby10, Hans-Olov Adami11, Nicola S Russell12, Bengt Glimelius13, Stephen Hamilton-Dutoit14, Marieke de Bruin7, Lars P Ryder15, Daniel Molin16, Karina Meden Sorensen17, Ellen T Chang18, Malcolm Taylor19, Rosie Cooke20, Robert Hofstra21, Helga Westers21, Tom van Wezel22, Ronald van Eijk22, Alan Ashworth23, Klaus Rostgaard3, Mads Melbye3, Anthony J Swerdlow20, Richard S Houlston1¥

1. Section of Cancer Genetics, Institute of Cancer Research, Sutton, SM2 5NG, UK
2. Leukaemia Research Fund Virus Centre, Institute of Comparative Medicine, University of Glasgow, Glasgow G61 1QH, UK
3. Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark
4. German Cancer Research Center Heidelberg, Germany and the Center for Primary Health Care Research, University of Lund, Malmö, Sweden
5. Department of Pathology & Medical Biology, University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands
6. Epidemiology & Genetics Unit, Department of Health Sciences, University of York, York, YO10 5DD, UK
7. Department of Epidemiology, The Netherlands Cancer Institute and Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands
8. Department of Experimental Therapy, The Netherlands Cancer Institute and Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands
9. University Hospital of Cologne, Department of Internal Medicine, Cologne, Germany
10. Unit of Clinical Epidemiology, Department of Medicine, Karolinska University Hospital, SE-171 76, Stockholm, Sweden
11. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-171 77, Stockholm, Sweden; Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA
12. Department of Radiotherapy, The Netherlands Cancer Institute and Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands
13. Department of Pathology and Oncology, Karolinska Institutet, SE-171 77, Stockholm, Sweden; Department of Oncology, Radiology and Clinical Immunology, Uppsala University, SE-751 85, Uppsala, Sweden
14. Institute of Pathology, Aarhus University Hospital, DK-8000, Aarhus, Denmark
15. Department of Clinical Immunology, University Hospital of Copenhagen, Rigshospitalet, DK-2100, Copenhagen, Denmark
16. Department of Oncology, Radiology and Clinical Immunology, Uppsala University, Upsalla, Sweden
17. Department of Clinical Biochemistry, Statens Serum Institut, Copenhagen, Denmark
18. Cancer Prevention Institute of California, Fremont, CA 94538; Division of Epidemiology, Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA 94305, USA
19. Cancer Immunogenetics Group, School of Cancer & Enabling Sciences, University of Manchester, Research Floor, St Mary's Hospital, Manchester, M13 9WL, UK
20. Section of Epidemiology, Institute of Cancer Research, Sutton, SM2 5NG, UK
21. Department of Genetics University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands
22. Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands
23. The Breakthrough Breast Cancer Research Centre, London SW3 6JB, UK

*These authors contributed equally to this work
**These authors should be considered to have equal positional status
¥ Corresponding author:
e-mail: Richard.houlston@icr.ac.uk
Tel: +44 (0) 208 722 4175
Fax: +44 (0) 208 722 4365
To identify predisposition loci for classical Hodgkin Lymphoma (cHL) we conducted a genome-wide association study of 589 cHL cases and 5,199 controls with validation in 4 independent samples totaling 2,057 cases and 3,416 controls. We identified three new susceptibility loci at 2p16.1 (rs1432295, REL; odds ratio [OR]=1.22, $P_{\text{combined}}=1.91\times10^{-8}$), 8q24.21 (rs2019960, PVT1; OR=1.33, $P_{\text{combined}}=1.26\times10^{-13}$) and 10p14 (rs501764, GATA3; OR=1.25, $P_{\text{combined}}=7.05\times10^{-8}$). Furthermore, we confirmed the role of the MHC in disease etiology by revealing a strong HLA association (rs6903608; OR=1.70, $P_{\text{combined}}=2.84\times10^{-50}$). These data provide new insight into the pathogenesis of cHL.

Classical Hodgkin Lymphoma (cHL) is a lymph node cancer of germinal center B-cell origin, characterized by malignant Hodgkin and Reed-Sternberg (HRS) cells mixed with a dominant background population of reactive lymphocytes and other inflammatory cells. cHL is one of the most common tumors in young adults in economically developed countries, with ~1,500 cases being diagnosed each year in the UK, and the disease accounts for ~1 in 3 of all lymphomas. While Epstein-Barr virus (EBV) infection may be causally related to a proportion of cases, the etiology of EBV-negative cHL remains largely unknown.

Evidence for inherited genetic predisposition to cHL is provided by the 3 to 9-fold increased risk of cHL in first-degree relatives of cHL patients. In the light of a possible viral basis to cHL it is interesting that cHL was the first disease to be associated with the HLA region. Subsequent studies have reported associations between various HLA class I and class II alleles and risk of cHL; specifically an association between the HLA-A*01 and A*02 for EBV-positive cHL. Genetic variation in HLA is, however, insufficient to account for the observed familial risk of cHL. To date no non-HLA genetic risk factors have been identified and convincingly replicated. Genome-wide linkage studies of cHL families have failed to demonstrate an additional major gene locus for cHL. This coupled with the very high concordance of Hodgkin Lymphoma in monozygotic compared with dizygotic twins is consistent with a genetic model of inherited susceptibility based on the co-inheritance of multiple low-risk variants.
Predicated on this hypothesis we conducted a genome-wide association study (GWAS) of 622 UK cHL cases using Illumina 660w Quad BeadChips. Genotype frequencies were compared with publicly accessible genotype data generated by the UK Wellcome Trust Case-Control Consortium 2 (WTCCC2) study of 2,930 individuals from the 1958 British Birth Cohort (58C)¹⁴ and 2,737 individuals from the UK Blood Service collections (UKBS), that had been genotyped using Illumina Human 1.2M-Duo Custom_v1 Array BeadChips (Supplementary Methods). There was no evidence of systematic bias between these two series (Supplementary Methods; Supplementary Figure 1), which were combined to provide genotype data on 5,667 controls. Data on 521,834 autosomal SNPs common to cases and controls were included in this analysis. After stringent quality control filtering (Supplementary Methods; Supplementary Table 1), we analyzed 504,374 SNPs in 589 cHL cases and 5,199 controls. Principal component analysis (PCA) demonstrated that these cases and controls were genetically well matched (Supplementary Figure 2). We therefore assessed the association between each SNP and cHL risk using the Cochran-Armitage trend test without PCA adjustment. The quantile-quantile (Q-Q) plots of the negative logarithm of genome-wide $P$-values showed a strong deviation from the null distribution (Figure 1), which could be ascribed to the strong association observed within the MHC region. After excluding 1,700 SNPs mapping to the major histocompatibility (MHC) region (6p21: 28-33Mb) there was only minimal inflation of test statistics, except at the upper tail of the distribution ($P < 10^{-4}$), thereby rendering cryptic population substructure or differential genotype calling between cases and controls unlikely (genomic control inflation factor $^{15}$, $\lambda_{gc}=1.04$; Figure 1). Using principal components analyses as implemented in Eigenstrat $^{16}$, correction for possible population substructure had no influence on findings for subsequently validated loci (Table 1). Furthermore, evidence for loci influencing cHL risk was provided by independent comparison with both 58C and UKBS control series (Supplementary Table 2).

This GWAS revealed multiple associations at chromosome 6, as well as suggestive associations on chromosomes 2, 5, 7, 8, 9, 10, 11 and 19 (Figure 1). To validate these associations we genotyped the HLA class II SNP rs6903608 and 10 SNPs from other regions showing an association, in the UK replication series (524 cases, 1,533 controls) (Supplementary Methods, Supplementary Table 1). In the combined analysis, associations for 6 of the SNPs were significant at $P_{combined} < 1.0 \times 10^{-4}$ (Supplementary Table 3). These 6 SNPs were successfully genotyped in 3
independent case-control replication series (Supplementary Methods, Supplementary Table 1) - SCALE (482 cases, 590 controls), Germany (498 cases, 655 controls) and Netherlands (553 cases, 638 controls). Combined analysis of all case-control series revealed genome-wide associations (i.e., \( P<5.0 \times 10^{-7} \)) at 2p16.1, 6p21, 8q24.21 and 10p14 (Table 1; Supplementary Table 3).

In our GWAS, 42 SNPs mapping to the 4.8Mb interval at 6p21, bordered by the TRIM27 and MLN genes (rs209130, 28,975,779bps and rs1547668, 33,883,424bps respectively) defining the classical MHC region, showed evidence of an association with cHL risk at \( P<5.0 \times 10^{-7} \) (Supplementary Figure 3). The most significant associations were with SNPs mapping to HLA class II; the strongest signal was attained at rs6903608 centromeric to HLA-DRA \( (P=8.12 \times 10^{-21}, 32,536,263\text{bps}; \text{Supplementary Figure 3}) \). The association between rs6903608 was consistently seen in each of the replication series, \( P_{\text{combined}}=2.84 \times 10^{-50} \) (Table 1, Supplementary Table 3).

The association with rs1432295 \( (P_{\text{combined}}=1.91 \times 10^{-8}, \text{OR}=1.22) \) on 2p16.1 (60,920,170bps) straddles a recombination hotspot between 2 regions of high linkage disequilibrium (LD) (Figure 2; Supplementary Figure 4). The 137Kb region defined by these two LD blocks encompasses the putative transcript FLJ16341 and REL (avian reticuloendotheliosis viral oncogene homolog). REL encodes c-Rel, a member of the Rel/NFκB family of transcription factors. Constitutive activity of NFκB transcription factors is a hallmark of cHL\(^1\) and inactivating somatic mutations of the NFκB signaling inhibitors play a major role in cHL pathogenesis\(^{18-20}\). Furthermore, studies have shown genomic amplifications of REL associated with increased c-Rel expression in cHL\(^{21-23}\).

We identified 2 SNPs on 8q24.21 associated with cHL risk, rs2019960 \( (P_{\text{combined}}=1.26 \times 10^{-13}, \text{OR}=1.33) \) and rs2608053 \( (P_{\text{combined}}=1.16 \times 10^{-7}, \text{OR}=1.20) \). rs2608053 mapping at 129,145,014bps localizes to a 56Kb region of LD that encompasses intron 6 of PVT1 (Figure 2, Supplementary Figure 4). rs2019960 mapping at 129,261,453bps localizes to a 82Kb region of LD telomeric to PVT1 (Figure 2, Supplementary Figure 4). The effects of rs2019960 and rs2608053 on cHL risk are maintained when adjusted for each other by logistic regression (\( \text{OR}=1.33, 95\% \text{ CI}:1.23-1.44, P=1.97 \times 10^{-13}; \) and \( \text{OR}=1.20, 95\% \text{ CI}:1.12-1.28, P=1.37 \times 10^{-7} \), respectively). Furthermore, correlation between rs2019960 and
rs2608053 is poor ($r^2=0.0$, $D'=0.01$ in HapMap CEU samples, $r^2=0.0$, $D'=0.03$ in our control data) and comparison of haplotype frequencies provided evidence of two haplotypes differing in frequency between cases and controls (Supplementary Table 4). Because rs2019960 or rs2608053 alone cannot fully account for the association between 8q24.21 and cHL, it is possible that a unique variant in LD with and capturing the effects of both SNPs may exist. We did not, however, identify a more significant association in LD with both SNPs through imputation, making it plausible that two independent signals exist at 8q24.21.

*PVT1* is frequently involved in translocations occurring in variant Burkitt’s lymphoma and murine plasmacytomas. The *PVT1* locus encodes several microRNAs thought to be as important as *MYC* in T-lymphomagenesis and T-cell activation. Co-activation of c-Myc and *PVT1* has been shown in a variety of human and animal tumors. The 128–130Mb genomic interval at 8q24.21 harbors multiple independent loci with different tumor specificities, including chronic lymphocytic leukemia (rs2456449; 128,262,163bps), prostate (rs16901979; 128,194,098bps), breast (rs13281615; 128,424,800bps), colorectal and prostate (rs6983267; 128,482,487bps), prostate (rs1447295; 128,554,220bps) and bladder (rs9642880; 128,787,250bps) cancer. The LD blocks defining these loci are distinct from the 8q24.21 cHL association signal ($r^2<0.03$; Supplementary Table 5). The colorectal cancer SNP rs6983267 shows differential binding of TCF4 to an enhancer element that physically interacts with the *MYC* promoter. A similar allele-specific cis-effect either on *MYC* or through *PVT1* impacting on *MYC* expression provides an attractive mechanistic basis for the 8q24.21 association with cHL risk. If the 8q24.21 locus influences risk through differential *MYC* expression, the association is intriguing since c-Myc and Rel/NFκB are the two master transcriptional systems activated in the latency III program of EBV-immortalized B-cells, which are responsible for the phenotype, growth pattern, and biological properties of cells driven into proliferation by EBV.

The two SNPs showing an association with cHL mapping to 10p14, rs501764 ($P_{combined}=7.05\times10^{-8}$, OR=1.25) and rs485411 ($P_{combined}=1.29\times10^{-7}$, OR=1.22) are in strong LD ($r^2=0.71$, $D'=0.95$ in HapMap CEU samples, $r^2=0.69$, $D'=1.00$ in our control data) and map to a 40Kb region of LD encompassing the transcription factor and putative tumor suppressor gene, *GATA3* (GATA binding protein 3 isoform 2) (Figure 2, Supplementary Figure 4). The expression of *GATA3* is important in
hematopoietic and lymphoid-cell development, acting as a master transcription factor for differentiation of T\(_h\)2 cells\(^{39}\). A high proportion of the reactive infiltrate in cHL tumors is composed of T\(_h\)2-like cells with T\(_{\text{reg}}\) phenotype which can influence EBV-positive cHL cell growth, depending on EBV antigenic presentation by MHC molecules\(^{40}\). Notably, a key characteristic of HRS cells is the production of cytokines and chemokines driven by GATA3 expression and other T-cell transcription factors \(^{41}\). Evidence for a biological relationship between the 2p16.1, 8q24.21 and 10p14 loci is that members of the Rel-family have differential effects on the MYC promoter\(^{42}\) and GATA3 is a target for c-Myc\(^{43}\).

Elucidation of the basis of each of the associations at 2p16.1, 8q24.21 and 10p14 will require fine-mapping and functional analyses. To examine if any directly typed or imputed SNPs annotate a putative transcription factor (TF) binding/enhancer element, we conducted a bioinformatic search of each of the regions of association using Transfac Matrix Database, PReMod and EEL software. At 10p14 an imputed SNP rs369421 provides the best evidence for the association signal \((P=6.20\times10^{-7})\) mapping within module 011553 (Supplementary Table 6, Supplementary Figure 4). Intriguingly, this module includes binding sites for ARID5B and E2F TFs. ARID5B has been previously implicated in development of acute leukemia\(^{44}\), and loss of PU.1, an E2F TF, has been associated with defective immunoglobulin expression in HRS cells\(^{45}\).

A hallmark of cHL epidemiology is the bimodal age specific incidence and it has been argued that the disease in young adults and older adults are etiologically different; in particular there is a low prevalence of EBV in younger cHL patients\(^{46}\). We assessed the relationship between cHL and EBV-status, age and sex at the 6p21, 2p16.1, 8q24.21 and 10p14 loci (defined by rs6903608, rs1432295, rs2019960, rs2608053, and rs501764 genotypes) by case-only analysis using data from SCALE, UK and Netherlands replication series (1,100 cases; Supplementary Table 7). Associations at all loci were not influenced by sex after adjustment for age and EBV-status. The rs501764 association with cHL was not related to age or EBV-status (Supplementary Table 7). The HLA class II association at 6p21 was primarily driven by EBV-negative cHL after adjustment for age and sex \((P_{\text{adjusted}}=1.63\times10^{-11})\). Similarly, rs1432295 (2p16.1) risk alleles were significantly enriched in EBV-negative cHL \((P_{\text{adjusted}}=0.01)\). At 8q24.21, while rs2608053 was associated with EBV-negative cHL \((P_{\text{adjusted}}=0.01)\), rs2019960 showed a relationship with early-onset cHL,
independent of EBV-status or sex \( (P_{\text{adjusted}}=0.002) \) (Supplementary Table 7). These phenotypic differences provide further support for two independent cHL risk loci at 8q24.21.

To explore whether any of the associations at 2p16.1, 8q24.21 and 10p14 reflect cis-acting regulatory effects on a nearby gene we searched for genotype-expression correlations in 90 EBV-transformed lymphoblastoid cell lines using previously described data\(^{47,48}\). We did not find any significant relationship between SNP genotype and gene expression, after adjustment for multiple testing (Supplementary Figure 5). This does not preclude the possibility that the causal variants at these disease loci have subtle effects on expression as the dynamic range of transcripts, such as \textit{MYC}, is small. Furthermore, it is likely that only a cumulative long-term imbalance in expression in target genes will influence cHL development and expression differences may only be relevant to a specific subpopulation of B-cells, which may not be well modelled by EBV-transformed lymphocytes.

While the HLA association with cHL is a very strong genetic effect, the identification of risk variants at 2p16.1, 8q24.21 and 10q14 implicates important roles for networks involving \textit{MYC}, \textit{GATA3} and the NF\(\kappa\)B pathway in cHL disease etiology. In the combined dataset there was some evidence for interactions between HLA (rs6903608) and 2p16.1 (rs1432295; \( P=0.05 \)) and between 8q24.21 (rs2608053) and 10p14 (rs501764 and rs485411; \( P=0.01 \)), albeit non-significant after correction for multiple testing (Supplementary Table 8). Further studies are needed to investigate possible interactions between these susceptibility loci and their interplay with EBV infection. Finally, the modest size of our study makes it likely that further risk variants for cHL can be identified through additional studies.
Note: Supplementary information is available on the Nature Genetics website

URLs

The R suite can be found at http://www.r-project.org/
Detailed information on the tag SNP panel can be found at http://www.illumina.com/
HapMap: http://www.hapmap.org/
1000Genomes: http://www.1000genomes.org/
KBioscience: http://kbioscience.co.uk/
WGAViewer: http://www.genome.duke.edu/centers/pg2/downloads/wgaviewer.php
SNAP http://www.broadinstitute.org/mpg/snap/
IMPUTE: https://mathgen.stats.ox.ac.uk/impute/impute.html
SNPTEST: http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html
EEL: http://www.cs.helsinki.fi/research/algodan/EEL/
PReMod: http://genomequebec.mcgill.ca/PReMod/welcome.do
JASPAR2 database: http://jaspar.cgb.ki.se/
EIGENSTRAT: http://genepath.med.harvard.edu/~reich/Software.htm
Wellcome Trust Case Control Consortium: www.wtccc.org.uk
METAL: www.sph.umich.edu/csg/abecasis/metal

ACKNOWLEDGEMENTS

Leukemia Lymphoma Research (UK) provided principal funding for the study. Additional support was provided by Cancer Research UK (C1298/A8362 supported by the Bobby Moore Fund). This study made use of control genotyping data generated by the Wellcome Trust Case-Control Consortium. We acknowledge use of genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. 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. At the Institute of Cancer Research sample and data acquisition was supported by Breakthrough Breast Cancer and the European Union and we acknowledge NHS funding to the NIHR Biomedical Research Centre. We are grateful to the patients and their clinicians who participated in this collection (see Supplementary Note). Work at the LRF Virus Centre was funded by Leukaemia and Lymphoma Research. Sample and data acquisition for the UK replication series was also supported by the Kay Kendall Leukaemia Fund. ELCCS was funded by Leukaemia & Lymphoma Research. Grant support to the German Study Group was through Deutsche Krebshilfe and the EU, HEALTH-F4-2007-200767. The SCALE study is supported by the Lundbeck Foundation R19 A2364; Danish Cancer Research Foundation grant 41-08 and Danish Cancer Society grant DP 08155. At the Department of Pathology & Medical Biology, University of Groningen, sample and data acquisition was supported by two grants from the Dutch Cancer Society (RUG 200-2315 and RUG 2009-4313). The Dutch NKI study was supported by the Dutch Cancer Society (Grants No. NKI 98-1833, NKI 04-3068, NKI 08-3994) and the EU6th project GRR (Project no 012926). We thank Ausra Kesminiene for coordinating the EU-GRR project, Michael Schaapveld and Anja Eggermond for data management, Linde Braaf and Izabela Mikolajewska for lab assistance. We are indebted to the patients and physicians who participated in this collection (see Supplementary Note).

**AUTHOR CONTRIBUTIONS**

RSH designed the study and obtained financial support. RSH drafted the manuscript with contributions from PB, VE, and YM. YM and VE performed statistical and bioinformatic analyses; PB sample coordination and laboratory analyses; BO and AL performed genotyping. AJS, AA and RC provided samples and data from a study conducted at the ICR. ER initiated ELCCS; TJ managed and prepared Epidemiology & Genetics Lymphoma Case-Control Study samples. RFJ, FEA and GMT designed and conducted studies contributing to the UK replication series and RFJ, LS, AL and DM prepared samples and collated data. FEvL designed the Dutch NKI study and obtained financial support; NSR and MdB, were involved in identification and inclusion of Dutch cases, study design, review board approval and clinical implementation; AB coordinated collection and preparation of NKI samples. AF, KH, AE and KR provided samples and data from German cases and controls. AD and
AvdB, collection of samples and data from cHL cases ascertained through Groningen. RH, HW, TvW and RvE, ascertainment and collection of control samples from the Netherlands. HH, MM, KR, LPR, KES, HOA, BG, DM, SHD, KMS, and ET provided samples and data from the SCALE study in Denmark and Sweden. SHD analyzed samples and provided data from Danish cHL cases. All authors contributed to the final paper.

**COMPETING INTERESTS STATEMENT**

The authors declare no competing financial interests
METHODS

Patients and samples

Genome-wide association study
We analyzed the constitutional DNA of 622 patients diagnosed with cHL (International Classification of Diseases [ICD] 10 codes C81.0-3) who were ascertained through the Royal Marsden Hospitals NHS Trust Family History study, between 2004 and 2008 (n=104, 63 male; mean age at diagnosis 38 years, SD 16) and an ongoing national study of cHL in females (n= 518, mean age at diagnosis 23 years, SD 6) being conducted by the Institute of Cancer Research (ICR). 146 of the cases had been diagnosed with breast cancer subsequent to diagnosis of cHL. All cases were British residents and were self-reported to be of European Ancestry.

For controls we made use of publicly accessible data generated by the UK Wellcome Trust Case-Control Consortium 2 (WTCCC2) study on 5,667 individuals from two sources. Firstly, 2,930 individuals ascertained through the British 1958 Birth Cohort (58C; also known as the National Child Development Study) which includes all births in England, Wales and Scotland14. Secondly, 2,737 UK Blood Services Controls (UKBS) aged 18-69 years, sex- and geographically matched to reproduce the distribution of samples within the 1958 Birth Cohort.

Replication series
The UK-replication series comprised 524 cHL cases (ICD10 C81.0-3; 290 male, mean age at diagnosis 38 years, SD 16) ascertained from the Scotland and Newcastle Epidemiological Study of Hodgkin Disease (SNEHD), the Young adult Hodgkin Case-Control Study (YHCCS) and the Epidemiology & Genetics Lymphoma Case-Control Study (ELCCS; www.elccs.info). Full details of the SNEHD, YHCCS and ELCCS studies are provided in previously published work49-51. Briefly, SNEHD involved the ascertainment of incident cases from Scotland and Northern England between 1993 and 1997. YHCCS was based on the newly diagnosed patients aged 16-24 years from Northern England between 1991 and 1995. ELCCS comprised patients residing in the north of England between the ages of 16 and 69 years, with newly diagnosed, non-HIV-related HL, during 1998-2003. UK population controls were obtained from SNEHD and YHCCS (n=495, 268 male, mean age at sampling
41, SD=17) and ongoing epidemiological studies of cancer being conducted at the ICR (n=1,038, 524 male, mean age at sampling 60, SD=9)\textsuperscript{52}.

The Scandinavian Lymphoma Etiology (SCALE) study has been described previously\textsuperscript{53,54}. Briefly, SCALE is a population-based case-control study of HL and non-Hodgkin lymphomas carried out in Denmark and Sweden during 1999 to 2002. The study population encompassed Danish and Swedish speaking residents aged 18 to 74 with no history of HIV infection, solid organ transplantation or previous hematopoietic malignancy in Denmark from June 1, 2000 to August 30, 2002, and in Sweden from October 1, 1999, to April 15, 2002. Participants recruited in a Danish regional pilot phase starting November 1, 1999, were also included, as were prevalent cases of HL diagnosed since January 1, 1999 in both countries. A total of 586 patients diagnosed with cHL according to the WHO classification in the study period and 3,187 controls representing 91\% and 71\% of eligible cases and controls, respectively, participated in the study, which included telephone interview and blood sampling. For the present investigation, DNA from 482 cases (82\% of all SCALE cHL cases, 282 male, mean age at diagnosis 40 years, SD 16) and from 255 Danish controls was extracted from dried filter paper blood spots with Extract-N-Amp\textsuperscript{T} as per the manufacturer’s instructions (Sigma-Aldrich, St. Louis, MO, USA) and subjected to whole genome amplification with AmpliQ Genomic Amplifier Kit (Ampliqon, Denmark)\textsuperscript{55}. In addition, germline DNA extracted from buffy coat for 335 Danish SCALE controls (randomly selected from 590 controls) was also included. (Mean age at sampling for combined SCALE controls 59 years, SD 13).

The Netherlands replication series comprised: (i) 281 cHL patients (149 male, mean age at diagnosis 36 years, SD 15) collected from the northern part of the Netherlands diagnosed between 1997 and 2000 as part of an ascertainment made by the University Medical Centre Groningen; (ii) 272 cHL cases, 97 of whom had been diagnosed with breast cancer subsequent to cHL (mean age at diagnosis 24 years, SD 6). These patients were selected in the framework of an ongoing case-control study of risk factors for breast cancer after HL being conducted by the Netherlands Cancer Institute, Amsterdam, within a larger cohort study of women who had been treated for cHL before age 60, between 1965 and 1995 and survived at least 5 years. Patient selection, methods of data/blood collection and DNA isolation have been described previously\textsuperscript{56-58}. Samples from healthy blood donors, aged 19-69 years, ascertained through medical centers in Groningen (mean age 52
years, SD 11) and Leiden (mean age 47 years, SD 12), served as a source of controls.

The German replication series comprised 498 cHL patients ascertained by the German Hodgkin Study Group between years 1998 and 2007 (292 male, mean age at diagnosis 34 years, SD 12). Controls were 655 healthy blood donors from Mannheim, located 200 km from Cologne (381 male, mean age at sampling 36, SD 13).

**EBV status of tumors**

The EBV status of cHL tumors was determined by immunohistochemical staining for EBV latent membrane antigen (LMP)-1 and or EBV EBER *in situ* hybridization using sections of paraffin-embedded material\(^53,59\).

**Ethics**

Collection of blood samples and clinico-pathological information from subjects was undertaken with informed consent and relevant ethical review board approval in accordance with the tenets of the Declaration of Helsinki.

**Genotyping**

DNA was extracted from samples using conventional methodologies and quantified using PicoGreen (Invitrogen, Carlsbad, USA).

Genotyping of cHL cases in the GWAS was conducted using Illumina Infinium HD Human660-Quad BeadChips according to the manufacturer's protocols (Illumina, San Diego, USA). DNA samples with GenCall scores <0.25 at any locus were considered “no calls”. A SNP was deemed to have failed if fewer than 95% of DNA samples generated a genotype at the locus. Cluster plots were manually inspected for all SNPs considered for replication.

We used data on controls from the 1958 Birth Cohort (58C) and National Blood Service (UKBS) which had been generated by the Wellcome Trust Case Control Consortium. Genotyping of both sets of controls was conducted using Illumina Human 1.2M-Duo Custon_v1 Array chips. SNP calling was performed using the Illuminus Software. Full details of genotyping, SNP calling and QC have been previously reported (www.wtccc.org.uk). Concordant with previous findings\(^17\).
comparison of the two control series showed little evidence for systematic bias (inflation factor $\lambda=1.022$; Supplementary Figure 1).

Validation and replication of associations were performed using competitive allele-specific PCR KASPar chemistry (KBiosciences Ltd, Hertfordshire, UK). All primers and probes used are available on request. Samples having SNP call rates of <90% were excluded from the analysis. To ensure quality of genotyping in all assays, at least two negative controls and 1-2% duplicates (showing a concordance >99.99%) were genotyped. To exclude technical artifact in genotyping we performed cross-platform validation and sequenced a random series of 96 samples to confirm genotyping accuracy (concordance >99.9%).

**Statistical and bioinformatic analysis**

We applied pre-determined quality control metrics to the GWAS data. We restricted analyses to samples for whom >95% of SNPs were successfully genotyped, thus eliminating 12 cases. We computed identity-by-state (IBS) probabilities for all pairs (cases and controls) to search for duplicates and closely related individuals amongst samples (defined as IBS $\geq 0.80$, thereby excluding first-degree relatives). For all identical pairs the sample having the highest call rate was retained, eliminating 2 cHL cases. To identify individuals who might have non-Western European ancestry, we merged our case and control data with phase II HapMap samples (60 western European [CEU], 60 Nigerian [YRI], 90 Japanese [JPT] and 90 Han Chinese [CHB]). For each pair of individuals we calculated genome-wide IBS distances on markers shared between HapMap and our SNP panel, and used these as dissimilarity measures upon which to perform principal component analysis. The first two principal components for each individual were plotted and any individual not present in the main CEU cluster (i.e., 5% furthest from cluster centroids) was excluded from analyses. We removed 30 cases with non-CEU ancestry (some of which had poor call rates) and 1 WTCCC2 control which was a duplicate case. We filtered out SNPs having a minor allele frequency [MAF] <1%, and a call rate <95% in cases or controls. We also excluded SNPs showing departure from Hardy-Weinberg equilibrium (HWE) at $P<10^{-5}$ in controls. For replication and validation analysis call rates were >95% per 384-well plate for each SNP; cluster plots were visually examined by two researchers.
Main analyses were undertaken using R (v2.6), Stata10 (State College, Texas, US) and PLINK (v1.06) software. The association between each SNP and risk of cHL was assessed by the Cochran-Armitage trend test. The adequacy of the case-control matching and possibility of differential genotyping of cases and controls were formally evaluated using quantile-quantile (Q-Q) plots of test statistics. The inflation factor $\lambda$ was based on the 90% least significant SNPs. We undertook adjustment for possible population substructure using Eigenstrat software. Odds ratios (ORs) and associated 95% confidence intervals (CIs) were calculated by unconditional logistic regression. Meta-analysis was conducted using standard methods. Cochran’s Q statistic to test for heterogeneity and the $I^2$ statistic to quantify the proportion of the total variation due to heterogeneity were calculated. Large heterogeneity is typically defined as $I^2 \geq 75\%$. To conduct a pooled analysis incorporating Eigenstrat adjusted $P$-values from the GWAS we used the weighted Z-method implemented in the program METAL. We examined each SNP for dose response by comparing 1-d.f. and 2-d.f. logistic regression models, adjusting for stage using a likelihood ratio test, and examined the combined effects of multiple SNPs by evaluating the effect of adding an interaction term on the model by using a likelihood ratio test and adjusting for stage. Associations by sex, age and EBV-status were examined by logistic regression in case-only analyses.

Prediction of the untyped SNPs was carried out using IMPUTEv2, based on HapMap Phase III haplotypes release 2 (HapMap Data Release 27/phase III Feb 2009 on NCBI B36 assembly, dbSNP26) and 1000genomes. Imputed data were analysed using SNPTEST v2 to account for uncertainties in SNP prediction. LD metrics between HapMap SNPs were based on Data Release 27/phase III (Feb 2009) on NCBI B36 assembly, dbSNP26, viewed using Haploview software (v4.2) and plotted using SNAP. LD blocks were defined on the basis of HapMap recombination rate (cM/Mb) as defined using the Oxford recombination hotspots and on the basis of distribution of confidence intervals defined by Gabriel et al.

To annotate potential regulatory sequences within disease loci we implemented *in silico* searches using Transfac Matrix Database v7.29, PReMod10 and EEL software.
**Relationship between SNP genotypes and expression levels**

To examine for a relationship between SNP genotype and expression levels of *GATA3*, *REL*, and *MYC* in lymphocytes we made use of publicly available expression data generated from analysis of 90 Caucasian derived Epstein-Barr virus–transformed lymphoblastoid cell lines using Sentrix Human-6 Expression BeadChips (Illumina, San Diego, USA)\(^{47,48}\). Online recovery of data was performed using WGAViewer Version 1.25 Software. Differences in the distribution of levels of mRNA expression between SNP genotypes were compared using a Wilcoxon-type test for trend\(^{57}\).
TABLE AND FIGURE LEGENDS

TABLES

Table 1: Summary results for the 6 SNPs associated with classical Hodgkin Lymphoma risk. Detailed data including genotype counts are shown in Supplementary Table 3. *Chromosome location based on NCBI Human Genome Build 36 coordinates. *Putative candidate genes mapping within 50kb to respective SNPs. †Risk Allele Frequency. ‡Odds ratio with 95% Confidence Interval. *Eigenstrat adjusted $P$-values: rs1432295, $P=8.87\times10^{-6}$; rs6903608, $P=2.93\times10^{-17}$; rs2608053, $P=4.20\times10^{-6}$; rs2019960, $P=9.14\times10^{-7}$; rs501764, $P=1.67\times10^{-6}$; rs485411, $P=1.25\times10^{-5}$; †Combined $P$-values using adjusted data rs1432295, $P=5.02\times10^{-9}$; rs6903608, $P=1.86\times10^{-46}$; rs2608053, $P=1.84\times10^{-8}$; rs2019960, $P=4.01\times10^{-14}$; rs501764, $P=1.80\times10^{-8}$; rs485411, $P=4.51\times10^{-8}$.

FIGURE LEGENDS

Figure 1: Genome-wide association results from the initial GWAS. Shown are the genome-wide $P$-values obtained using the Cochran-Armitage trend test from 504,374 autosomal SNPs in 589 HL cases and 5,199 controls. (a) Q-Q plots of observed $P$-values (-log10$P$) for association. The plot in blue shows $P$-values for all 504,374 SNPs, whereas the plot in green shows $P$-values excluding 1,700 SNPs mapping to the MHC region (chr 6: 28-33Mb). The black line represents the null hypothesis of no true association. (b). Genome-wide association $P$-values (-log10$P$, y axis) plotted against their respective chromosomal positions (x axis). Each chromosome is depicted in a different color. The points with $P<10^{-10}$ were truncated; the smallest $P$ value is 8.12 $\times 10^{-21}$.

Figure 2: Regional plots of association results and recombination rates for 2p16.1, 8q24.21, and 10p14 susceptibility loci. (a-c) Association results of both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates within the three loci: (a) 2p16.1; (b) 8q24.21; (c) 10p14. For each plot, -log10$P$ values (y-axis) of the SNPs are shown according to their chromosomal positions (x-axis). The top genotyped SNP in the combined analysis is
labeled by rs ID. The color intensity of each symbol reflects the extent of LD with the top genotyped SNP – red/blue ($r^2>0.8$) through to white ($r^2<0.2$). Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with a light blue line. Physical positions are based on build 36 (NCBI) of the human genome. Also shown are the relative positions of genes and transcripts mapping to each region of association. Genes and miRNAs have been redrawn to show the relative positions; therefore, maps are not to physical scale.

**SUPPLEMENTARY FIGURES AND TABLES**

**Supplementary Figure 1:** Q-Q plots of observed $P$-values ($-\log_{10}P$) for association between (a) 58C and UKBS controls ($\lambda=1.02$); (b) cHL cases and 58C ($\lambda=1.04$); (c) cHL cases and UKBS ($\lambda=1.04$); (d) cHL cases and combined controls (58C + UKBS, $\lambda=1.04$). The plot in blue shows $P$-values for all 504,374 SNPs, whereas the plot in green shows $P$-values excluding 1,700 SNPs mapping to the MHC region (chr 6: 28-33Mb). The black line represents the null hypothesis of no true association.

**Supplementary Figure 2:** Identification of individuals in the GWAS of non-European ancestry in cases and controls. The first two principal components of the analysis were plotted. HapMap CEU individuals are plotted in blue; CHB+JPT individuals are plotted in green; YRI individuals are plotted in red; GWAS cases are plotted in pink before (a) and after (c) removal, GWAS controls are plotted pink before (b) and after (d) removal.

**Supplementary Figure 3:** cHL associations in the MHC region. Major HLA genes are shown.

**Supplementary Figure 4:** Plots of linkage disequilibrium and transcription factor binding sites at 2p16.1, 8q24.21, and 10p14 loci. Upper panel shows: the positions of genes and transcripts encoded by the region and positions of relevant SNPs; regions with high densities of TFBSs, modules predicted by PReMod are shown in orange and groups predicted by EEL and the Transfac Matrix Database are shown in blue; sequence conservation across the region in mammals. Middle and lower panels show $r^2$ and $D'$ LD statistics from HapMap phase II data. The darker shading indicates strong LD between SNPs.
Supplementary Figure 5: Relationship between lymphocyte mRNA expression levels of: MYC and rs2608053, rs2019960 genotype; GATA3 and rs501764, rs485411 genotype; REL and rs1432295 genotype. Expression of genes (normalized –log2 levels) is based on data from analysis of 90 Epstein-Barr virus–transformed lymphoblastoid cell lines using Sentrix Human-6 Expression BeadChip (Illumina, San Diego, USA)47,48. Data was recovered using WGAViewer Version 1.25. Differences in the distribution of expression by SNP genotype were compared using a Wilcoxon-type test for trend67.

Supplementary Table 1: a) Basic characteristics b) exclusions, genotyping, quality control in each of the case-control series.

Supplementary Table 2: Association between 2p16.1, 6p21.32, 8q24.21, 10p14 genotypes and cHL in GWAS using individual 58C and UKBS controls sets

Supplementary Table 3: SNP genotype frequencies in cHL cases and controls in each case-control series.

Supplementary Table 4: Risk of cHL by rs2608053-rs2019960 8q24.21 haplotypes.

Supplementary Table 5: Linkage disequilibrium metrics between 8q24.21 SNPs.

Supplementary Table 6: Details of transcription factor binding sites (TFBSs) as predicted by EEL (using binding profiles from the JASPAR2 database), the Transfac Matrix Database and PReMod. "Score" refers to the confidence value assigned to each predicted binding region by the three different programs. For comparison, the observed and imputed SNPs and associated P-values are shown.

Supplementary Table 7: Relationship between 6p21.32, 2p16.1, 8q24.21 and 10p14 genotype and a) cHL EBV-subtype b) age and c) sex. RAF, risk allele frequency in series.
**Supplementary Table 8: Pairwise analysis of all SNPs associated with cHL risk.** For each row-column combination, numbers show the $P$-value for inclusion of an interaction term between the two SNPs. Numbers in parentheses are the number of samples from which statistics are calculated.
SUPPLEMENTARY NOTE

UK clinicians and study centers

G. Anghel, Lincoln Hospital; B. Attock, North Devon District Hospital, Barnstaple; S. Awward, Royal Shrewsbury Hospital; J. Barrett, Royal Berkshire Hospital, Reading; A. Bates, Southampton General Hospital; A. Bell, Poole Hospital; K. Benstead, Cheltenham General Hospital; E.M. Bissell, Nottingham University Hospital; A. Biswas, Royal Preston Hospital; N. Blesing, Great Western Hospital, Swindon; M. Brada, Royal Marsden Hospital; C. Brammer, New Cross Hospital, Northampton; J.E.S. Brock, Clatterbridge Centre for Oncology; A.M. Brunt, University Hospital of North Staffs; P. Coates, Queen Elizabeth Hospital, King’s Lynn; M.P. Collinson, Royal Cornwall Hospital, Tresliske; D. Cunningham, Royal Marsden Hospital, N. Davidson, Essex County Hospital, Colchester; S. Davies, North Middlesex; S. Falk, Bristol Haematology & Oncology Centre; I Fentiman, Guy’s Hospital; E. Gallop-Evans, Velindre Hospital, Cardiff; A. Garcia; Glen Clwyd Hospital, Bodelwyddan, Rhyl; D. Gilson; Cookridge Hospital, Leeds; A. Goodman, Royal Devon & Exeter Hospital; Mr. G. Gui, Royal Marsden Hospital; B.W. Hancock, Weston Park Hospital, Sheffield; A. Harnett, Norfolk & Norwich Hospital; S.J. Harris, St. Thomas’ Hospital, London; A. Horwich, Royal Marsden Hospital; P. Hoskin, Mount Vernon Hospital; C. Irwin, University Hospital, Coventry; S.A. Kelly, Derriford Hospital, Plymouth; J. Kingston, Great Ormond Street Hospital; C. Knechtl, Royal United Hospital, Bath; G. Levitt, Great Ormond Street Hospital; D. Linch, University College London Hospitals; A. Lister, St. Bartholomew’s; H. Luraft, Northern Centre for Cancer Care, Freeman Hospital; Z. Maung, James Cook Hospital; C. McNamara, Royal Free Hospital; J. Morgan, Ipwic; P. Murray, Colchester General; H. O’Brien, Cumberland Royal Infirmary; A. O’Callaghan, St. Mary’s Hospital, Portsmouth; N. O’Connor, Shrewsbury Hospital; R. Patmore, Hull Royal Infirmary; M. Persic, Derby City General Hospital; C. Pocock, Kent & Canterbury Hospital; H. Price, Taunton; K. Pritchard-Jones, Royal Marsden Hospital; J. Radford, Christie Hospital; G. Ross, Royal Marsden Hospital; JRY Ross, Kettering General Hospital; A Salisbury, Churchill Hospital, Oxford; M. Savage, Charing Cross Hospital/Hammersmith Hospital; J. Simpson, Royal Sussex County Hospital, Brighton; I Smith, Royal Marsden Hospital; A Stevens, Queen Elizabeth Hospital, Birmingham; P.S. Rassam, Kent & Canterbury Hospital; R. E. Taylor, Singleton Hospital, Swansea; G. Thomas, Leicester Royal Infirmary; C. Trask, Southend Hospital; J. Treleaven, Royal Marsden Hospital; Z Volkova, North Devon District Hospital, Barnstaple; N West, Whitehaven Hospital; S. Whitaker, Royal Surrey County Hospital; M. Williams, Addenbrooke’s Hospital, Cambridge; J. Yarnold, Royal Marsden Hospital

Dutch study centres: the Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital [NKI/AVL] in Amsterdam, Erasmus MC/Daniel den Hoed Kliniek [EMC/DdHK] in Rotterdam, Vrije Universiteit Medical Center [VUMC] in Amsterdam, Leiden University Medical Center [LUMC] in Leiden, the Emma Children’s hospital/Academic Medical Center [EKZ] in Amsterdam) and in hospitals in the region of the Eindhoven Cancer Registry (IKZ).
REFERENCES

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP</th>
<th>Location (bps)</th>
<th>Gene</th>
<th>Risk allele</th>
<th>RAF control</th>
<th>GWAS</th>
<th>Replication studies</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>2p16.1</td>
<td>rs1432295</td>
<td>60,920,170</td>
<td>REL</td>
<td>G</td>
<td>0.40</td>
<td>1.33 (1.18-1.51)</td>
<td>4.69x10^{-6}</td>
<td>1.17 (1.08-1.27)</td>
</tr>
<tr>
<td>6p21.32</td>
<td>rs6903608</td>
<td>32,536,263</td>
<td>HLA-DRA</td>
<td>G</td>
<td>0.27</td>
<td>1.81 (1.60-2.05)</td>
<td>8.12x10^{-21}</td>
<td>1.65 (1.52-1.80)</td>
</tr>
<tr>
<td>8q24.21</td>
<td>rs2608053</td>
<td>129,145,014</td>
<td>PVT1</td>
<td>G</td>
<td>0.52</td>
<td>1.33 (1.18-1.50)</td>
<td>4.06x10^{-6}</td>
<td>1.15 (1.06-1.24)</td>
</tr>
<tr>
<td>8q24.21</td>
<td>rs2019960</td>
<td>129,261,453</td>
<td>PVT1</td>
<td>G</td>
<td>0.23</td>
<td>1.38 (1.21-1.58)</td>
<td>2.14x10^{-6}</td>
<td>1.31 (1.19-1.44)</td>
</tr>
<tr>
<td>10p14</td>
<td>rs501764</td>
<td>8,133,040</td>
<td>GATA3</td>
<td>C</td>
<td>0.19</td>
<td>1.42 (1.23-1.63)</td>
<td>1.33x10^{-6}</td>
<td>1.18 (1.07-1.30)</td>
</tr>
<tr>
<td>10p14</td>
<td>rs485411</td>
<td>8,133,191</td>
<td>GATA3</td>
<td>A</td>
<td>0.25</td>
<td>1.35 (1.18-1.54)</td>
<td>6.83x10^{-6}</td>
<td>1.17 (1.07-1.28)</td>
</tr>
</tbody>
</table>

Table 1: Summary results for 6 SNPs associated with classical Hodgkin lymphoma risk. Detailed data including genotype counts are shown in Supplementary Table 3. \(^a\)Chromosome location based on NCBI Human Genome Build 36 coordinates. \(^b\)Putative candidate genes mapping within 50kb to respective SNPs. \(^c\)Risk Allele Frequency. \(^d\)Odds ratio with 95% Confidence Interval. *Eigenstrat adjusted P-values: rs1432295, P=8.87x10^{-6}; rs6903608, P=2.93x10^{-17}; rs2608053, P=4.20x10^{-6}; rs2019960, P=9.14x10^{-7}; rs501764, P=1.67x10^{-6}; rs485411, P=1.25x10^{-5}. †Combined P-values using adjusted data rs1432295, P=5.02x10^{-9}; rs6903608, P=1.86x10^{-46}; rs2608053, P=1.84x10^{-8}; rs2019960, P=4.01x10^{-14}; rs501764, P=1.80x10^{-8}; rs485411, P=4.51x10^{-6}. 
$\lambda = 1.04$
Figure 1: Genome-wide association results from the initial GWA study. Shown are the genome-wide $P$-values obtained using the Cochran-Armitage trend test from 504,374 autosomal SNPs in 589 HL cases and 5,199 controls. (a) Q-Q plots of observed $P$-values ($\log_{10}P$) for association. The plot in blue shows $P$-values for all 504,374 SNPs, whereas the plot in green shows $P$-values excluding 1,700 SNPs mapping to the MHC region (chr 6: 28-33Mb). The black line represents the null hypothesis of no true association. (b). Genome-wide association $P$-values ($\log_{10}P$, y axis) plotted against their respective chromosomal positions (x axis). Each chromosome is depicted in a different color. The points with $P < 10^{-10}$ were truncated; the smallest $P$ value is $8.12 \times 10^{-21}$. 
rs1432295
PAPOLG
REL
PUS10
PEX13
(c)
**Figure 2: Regional plots of association results and recombination rates for 2p16.1, 8q24.21, and 10p14 susceptibility loci.** (a-c) Association results of both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates within the three loci: (a) 2p16.1; (b) 8q24.21; (c) 10p14. For each plot, \(-\log_{10} P\) (y-axis) of the SNPs are shown according to their chromosomal positions (x-axis). The top genotyped SNP in the combined analysis is labeled by rs ID. The color intensity of each symbol reflects the extent of LD with the top genotyped SNP – red/blue \((r^2 > 0.8)\) through to white \((r^2 < 0.2)\). Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with a light blue line. Physical positions are based on build 36 (NCBI) of the human genome. Also shown are the relative positions of genes and transcripts mapping to each region of association. Genes and miRNAs have been redrawn to show the relative positions, therefore maps are not to physical scale.