Variation at 3p24.1 and 6q23.3 influences the risk of Hodgkin's Lymphoma

Matthew Frampton[1], **Miguel Inacio da Silva Filho**[2], **Peter Broderick**[1], **Hauke Thomsen**[2], **Asta Försti**[2,3], **Jayaram Vijayakrishnan**[1], **Rosie Cooke**[1], **Victor Enciso-Mora**[1], **Per Hoffmann**[4,5], **Markus M Nöthen**[4,6], **Amy Lloyd**[1], **Amy Holroyd**[1], **Lewin Eisele**[7], **Karl-Heinz Jöckel**[7], **Sabine Ponader**[8], **Elke Pogge von Strandmann**[8], **Tracy Lightfoot**[9], **Eve Roman**[9], **Annette Lake**[10], **Dorothy Montgomery**[10], **Ruth F Jarrett**[10], **Anthony J Swerdlow**[1,11], **Andreas Engert**[8], **Kari Hemminki**[2,3], and **Richard S Houlston**[1,+]

[1]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK

[2]Division of Molecular Genetic Epidemiology, German Cancer Research Centre, Heidelberg, Germany

[3]Center for Primary Health Care Research, Lund University, 205 02 Malmö, Sweden

[4]Institute of Human Genetics, University of Bonn, Bonn, Germany

[5]Genomics Research Group, Medical Genetics, University Hospital Basel

[6]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

[7]Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University of Duisburg–Essen, Essen, Germany

[8]University Hospital of Cologne, Department of Internal Medicine, Cologne, Germany

[9]Epidemiology and Cancer Statistics Group, Department of Health Sciences, University of York, York, UK

[10]MRC University of Glasgow Centre for Virus Research, Glasgow, UK

[11]Division of Breast Cancer Research, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK

[+]Corresponding author: Richard.houlston@icr.ac.uk, Tel: +44 (0) 208-722-4175, Fax: +44 (0) 208-722-4365.

Abstract

In addition to HLA, recent genome-wide association studies (GWASs) of Hodgkin's Lymphoma (HL) have identified susceptibility loci for HL at 2p16.1, 8q24.21 and 10p14. In this study, we perform a GWAS meta-analysis with published GWAS (totaling 1,465 cases and 6,417 controls of European background), and follow up the most significant association signals in 2,024 cases and 1,853 controls. A combined analysis identifies new HL susceptibility loci mapping to 3p24.1 (rs3806624; $P=1.14 \times 10^{-12}$, odds ratio [OR]=1.26) and 6q23.3 (rs7745098; $P=3.42 \times 10^{-9}$, OR=1.21). rs3806624 localizes 5' to the *EOMES* (eomesodermin) gene within a p53 response element affecting p53 binding. rs7745098 maps intergenic to *HBS1L* and *MYB*, a region previously associated with hematopoiesis. These findings provide further insight into the genetic and biological basis of inherited susceptibility to HL.

Hodgkin's lymphoma (HL) is a common lymph node cancer of germinal center B-cell origin, which is characterized by malignant Hodgkin and Reed-Sternberg (HRS) cells mixed with a dominant background population of reactive lymphocytes and other inflammatory cells1. Although Epstein-Barr virus (EBV) infection may be causally related to a number of cases, there is little evidence to support the involvement of other environmental risk factors2.

Evidence for inherited genetic influence on susceptibility is provided by the increased familial risk and very high concordance between monozygotic twins3. While the risk of HL is well recognized to be influenced by HLA genotype variation within the major histocompatibility complex (MHC), much of the familial risk is thought to be the consequence of non-HLA genotype variation. Support for this hypothesis has come from recent genome-wide association studies (GWASs) of HL which, in addition to demonstrating multiple HLA associations4–7 have identified non-HLA susceptibility loci at 2p16.1 (*REL*), 8q24.21 and 10p14 (*GATA3*)7.

In this study to identify additional susceptibility loci for HL, we conducted an independent primary scan of German patients with classical HL (cHL) and performed a genome-wide meta-analysis with one previously published GWAS of UK cHL patients. The most significant single nucleotide polymorphisms [SNPs] were analyzed in two additional series totaling 2,024 cases and 1,853 controls. A combined analysis identified new HL susceptibility loci mapping to 3p24.1 and 6q23.3.

Results

GWAS and meta-analysis

In the primary scan (German-GWAS), 1,001 cHL cases, ascertained by the German Hodgkin Study Group during 1998–2007, were genotyped using the Illumina Human OmniExpress-12 v1.0 arrays. For controls, we used genotype data on 1,226 individuals enrolled into the Heinz Nixdorf Recall (HNR) study genotyped using Illumina OmniExpress-12 v1.0. A total of 133 case samples were removed during quality control steps for reasons including a failure to genotype, duplicates, closely related individuals, or non-CEU ancestry (CEU - Utah residents with Northern and Western European ancestry; Figures 1 and 2).

The UK-GWAS has been previously reported[7]; briefly 622 cHL cases were genotyped 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) and 2,737 individuals from the UK Blood Service collections (UKBS) that had been genotyped using the Illumina Human1.2M-Duo Custom_v1 Array. There was no evidence of systematic bias between these two series, which were combined to provide genotype data for 5,667 controls[7]. Quality control steps for the UK GWAS have been previously reported[7].

After filtering of genotype data from each GWAS on the basis of pre-specified quality-control measures, 296,129 autosomal SNPs were common to both case-control series. Quantile-quantile plots of the genome-wide Armitage trend test chi-square values showed there was minimal inflation of the test statistics rendering substantial cryptic population substructure or differential genotype calling between cases and controls unlikely in either GWAS (genomic control inflation factor[8], λ_{gc}=1.03 and 1.09 in UK and German-GWAS respectively; Figure 3). In the combined analysis under a fixed effects model we identified 27 SNPs in 21 genomic regions not previously associated with cHL risk which showed good evidence for a relationship (*i.e. P*< 5.0x10^{-5}; Figure 4, Supplementary Table S1).

Replication genotyping

To validate these findings we genotyped the SNP showing the best evidence of association in each of the 21 regions in the UK-replication 1 series. In the combined analysis rs3806624 and rs7745098 showed consistent evidence for an association with HL risk applying a fixed effects model to Armitage trend test statistics (combined *P*-value <5.0x10^{-7}; Supplementary Table S2) and these SNPs were genotyped in the UK-replication 2 series. In a combined analysis of all datasets the rs3806624 and rs7745098 associations were statistically significant on a genomewide basis (*i.e. P*<5.0x10^{-8}; Figure 5; Supplementary Table S3).

rs3806624 localizes to 3p24.1 (27,764,623 base pairs [bps]; combined *P*=1.14 x10^{-12}, odds ratio [OR]=1.26) 5' to the *EOMES* (eomesodermin; MIM:604615) gene (Figure 6) within a 110kb region of linkage disequilibrium (LD). rs7745098 localises to 6q23.3 (135,415,004bps; combined *P*=3.42×10^{-9}; OR= 1.21) and maps intergenic to *HBS1L* (HBS1-like protein; MIM 612450) and *MYB* (V-MYB avian myeloblastosis viral oncogene homolog; MIM:189990) genes. To explore the 3p24.1 and 6q23.3 regions of association further we imputed unobserved genotypes in GWAS cases and controls using 1000genomes data. This analysis did not reveal a significantly stronger association at 3p24.1 and 6q23.3 to that provided by rs3806624 and rs7745098 respectively (Figure 6).

Subtype analysis

HL is biologically heterogeneous[2], something reflected in the histology of tumor subtypes. Additionally 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 lower prevalence of EBV in young cHL cases (*i.e.* <40 years), with ~70% of mixed cellularity (MC) cHL being EBV-positive compared with only ~20% of nodular sclerosis (NS) cHL. We assessed by case-only analysis the relationship between

cHL and sex, age, histology and EBV status and rs3806624, rs7745098 genotype (Supplementary Table S4). Using case-only logistic regression a strong association was seen between rs3806624 and age at diagnosis of cHL with an increased prevalence of the G risk allele in cases diagnosed before age 40 (P=8.35x10^{-4}). None of the other associations were statistically significant.

Impact of SNPs on heritability of HL

There was no evidence of significant interaction between either rs3806624 and rs7745098 and the previously identified risk loci at 2p16.1 (rs1432295), 6p21.32 (rs6903608), 8q24.21 (rs2608053), 8q24.21 (rs2019960) and 10p14 (rs501764), an observation compatible with each locus having an independent effect on HL risk. To quantify the impact of the known loci on the heritability associated with common variation at the non-HLA linked loci (3p24.1, 6q23.3, 2p16.1, 8q24.21, 8q24.21 and 10p14), using data from UK-GWAS and German-GWAS we computed the receiver operator characteristic associated with rs3806624, rs7745098, rs1432295, rs2608053, rs2019960 and rs501764 genotypes. The area under the curve corresponding to these variants was 0.63 translating into them collectively accounting for ~7% of the familial risk.

Deciphering association signals

The functional basis of many GWAS signals can be ascribed to sequence changes impacting on gene expression and sequence conservation in non-coding regions has been shown to be a good predictor of *cis*-regulatory sequences. Using publicly accessible expression quantitative trait loci (eQTL) data on lymphoblastoid cell lines (LCLs) and T-cells we examined whether either rs3806624 or rs7745098 genotype is associated with differential expression. While the associations identified did not show consistent statistically significant evidence of *cis*-acting regulatory effects in publicly accessible eQTL data 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 is small. Furthermore, it is likely that only a cumulative long-term imbalance in expression of target genes will influence cHL development, and expression differences may be relevant only to a specific subpopulation of B cells, which may not be well modeled by EBV-transformed lymphocytes.

To examine whether any directly typed or imputed SNPs lie within or very close to a putative transcription factor binding/enhancer element, we conducted a bioinformatic search of each region of association. rs3806624 maps within a predicted promoter with histone marks of regulatory elements associated with promoters shown in multiple cell types; rs7745098 is within a highly evolutionary conserved region with predicted weak enhancer activity (Figure 6; Supplementary Table S5).

Discussion

We have identified common variants on chromosome 3p24.1 and 6q23.3 that influence the risk of cHL. Some (HLA; 8q24.21 rs2608053; 5q31 rs20541) but not all of the previously identified risk loci for cHL have differential effects according to histology and EBV-

status5,7. Our study findings suggest that 3p24.1 and 6q23.3 have generic effects on the development of cHL akin to 10p14 variation7.

Since rs3806624 localizes 5' to the *EOMES* gene and there are no other known genes within this region of LD, there is a high likelihood that the functional basis of the 3p24.1 association is mediated through variation in this gene *a priori*. While we cannot exclude the possibility that rs3806624 is simply acting as a marker for the 3p24.1 association, interrogation of the association signal through imputation failed to recover a statistically stronger association. Moreover, since rs3806624 maps within a p53 response element affecting p53 binding, with the risk allele G displaying weaker binding than the A allele9 it is likely that the SNP is causal.

EOMES is a member of the TBR1 subfamily of T-box genes which play a critical role in embryogenesis and pluripotency. While *EOMES* appears important in extra-nodal NK/T-cell lymphoma10 evidence for a role of *EOMES* in cHL etiology is currently lacking. A high proportion of the reactive infiltrate in cHL tumors is however composed of $T_{helper}2$-like cells with a $T_{regulatory}$ phenotype and cross talk between these cells and the HRS cells appears essential for tumor growth11. Notably, a key characteristic of HRS cells is the production of cytokines and chemokines driven by *GATA3* and *T-bet* expression, and other T-cell transcription factors12. *EOMES* is necessary for full effector differentiation of CD8+ T-cells complementing the actions of *T-bet* and acting as a key regulatory gene in the development of cell-mediated immunity13; CD8+ T cells deficient in Eomes and T-bet fail to differentiate into functional killers required for viral defense14. Differential expression of Eomes and T-bet appears to facilitate the cooperative maintenance of the pool of antiviral CD8+ T-cells during chronic viral infection15, something highly pertinent to cHL.

CD44 -positive memory Th2 cells expressing Il5 have lower levels of Eomes than Th2 cells lacking Il5 expression16. We have previously shown that variation at *GATA3* is a determinant of cHL risk7. Down-regulation of Eomes is required for Il5 expression and Eomes suppresses Gata3 transcriptional activity by inhibiting Gata3 binding to the *Il5* promoter16. These findings support a role for an extended pathway involving *GATA3* and *EOMES* genes in the etiology of cHL.

rs170934 which maps 314Kb telomeric to *EOMES* (28,079,085bps) has previously been shown to be a risk factor for multiple sclerosis (MS)17. Intriguingly familial clustering of HL and MS is seen suggesting a common etiological basis to both diseases18. This coupled with the biology of *EOMES* strengthens the link between autoimmunity and HL and raises the possibility of a model by which T-cell activation plays a role in both diseases through aberrant antigenic response.

rs7745098 maps intergenic to *HBS1L* and *MYB,* a region of the genome which has previously been reported to play a substantial role in hematopoiesis with polymorphisms influencing platelet 19, white cell count 20, and hemoglobin levels 21. While evidence for a role for HBS1L in hematological malignancy is lacking, MYB, through its interaction with p300, controls the proliferation and differentiation of hematopoietic stem and progenitor cells.

While we do not find evidence of *cis*-acting regulatory effects in publicly accessible eQTL data from analysis of LCLs or T-cells, steady state levels of RNA in lymphocytes at a single time point and in cycling mature cells may not adequately capture the impact of differential expression in tumorigenesis. In summary, we have identified new risk loci at 3p24.1 and 6p22 for cHL which provide additional insight into the development of this B-cell malignancy.

Methods

Ethics

Collection of 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 (In the UK – the Royal Marsden Hospital NHS Trust and Multicentre Research Ethics Committee; In Germany, The Ethics committee of the University of Cologne).

Discovery subjects

The German-GWAS comprised 1,001 cHL cases (597 male; mean age at diagnosis = 34.8 years, s.d.= 12.3) ascertained by the German Hodgkin Study Group during 1998–2007. Cases were genotyped using Illumina Human OmniExpress-12 v1.0 arrays according to the manufacturer's protocols (Illumina, San Diego, USA). For controls, we used genotype data on 1,226 individuals enrolled into the Heinz Nixdorf Recall (HNR) study genotyped using Illumina OmniExpress-12 v1.0.

The UK-GWAS has been previously reported[7]; briefly 622 cHL cases (63 male, mean age at diagnosis=24.4 years, s.d.=9.6) were genotyped 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)[22] study of 2,930 individuals from the 1958 British Birth Cohort (58C)[23] and 2,737 individuals from the UK Blood Service collections (UKBS) that had been genotyped using the Illumina Human1.2M-Duo Custom_v1 Array. There was no evidence of systematic genotyping bias between these two series, which were combined to provide genotype data for 5,667 controls[7].

Quality control of GWAS datasets

DNA samples with GenCall scores <0.25 at any locus were considered "no calls". A SNP was deemed to have failed if <95% of DNA samples generated a genotype at the locus. A sample was deemed to have failed if <95% of SNPs were successfully genotyped. The same quality control metrics on the German-GWAS data were applied as in the UK-GWAS[7]. 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 0.80, thereby excluding first-degree relatives). For all identical pairs the sample having the highest call rate was retained. 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.

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 $P<1\times10^{-7}$ (χ^2 test with 1 degree of freedom or Fisher's exact test if a cell count was <5). After stringent quality control filtering (Figure 1) we analyzed 296,129 autosomal SNPs, common to all cases and controls, in 1,465 cHL cases and 6,417 controls. Cluster plots were manually inspected for all SNPs considered for replication.

Replication series and genotyping

UK-replication 1 comprised 1,071 UK patients diagnosed with cHL of either mixed-cellularity HL (MC-HL; 269 male; mean age at diagnosis 46.9 years, range 15-86) or nodular sclerosis HL (NS-HL; 306 male; mean age at diagnosis 31.4 years, range 15-49). The patients were ascertained through the National Study of Hodgkin's Lymphoma Genetics (NSHLG; http://public.ukcrn.org.uk), an ongoing study of HL established in 2008. Controls were healthy individuals recruited through the National Study of Colorectal Cancer Genetics (NSCCG, n=1,186, 367 male; mean age 58.6)24 and the Royal Marsden Hospital Trust/ Institute of Cancer Research Family History and DNA Registry (n=102, 17 male; mean age 59.3) with no personal history of malignancy. Both cases and controls were UK residents and had self-reported European ancestry.

UK-replication 2 comprised 953 cHL cases (defined according to ICD10 C81.0-3; 290 male, mean AOD = 38 years, SD=16 years) ascertained from the Scotland and Newcastle Epidemiological Study of Hodgkin Disease (SNEHD), the Young Adult Hodgkin Case-Control Study (YHCCS), and the Epidemiology and Cancer Statistics Group Lymphoma Case-Control Study (ELCCS; http://www.elccs.info). Full details of the SNEHD, YHCCS and ELCCS studies were provided previously. Briefly, SNEHD involved ascertainment of incident cases from Scotland and Northern England during 1993–1997. YHCCS was based on newly diagnosed cases aged 16–24 years from Northern England during 1991–1995. ELCCS comprised cases residing in the north of England aged 16–69 years, with newly diagnosed, non-HIV–related Hodgkin's lymphoma, during 1998–2003. UK population controls were obtained from SNEHD, YHCCS and ELCCS (n=565, 326 male, mean age 41 years, SD=17 years). The EBV status of cHL tumors was determined by immunohistochemical staining for EBV latent membrane antigen (LMP)-1 and/or EBV EBV-encoded RNA (EBER) *in situ* hybridization using sections of paraffin-embedded material.

Genotyping was performed using competitive allele-specific PCR KASP chemistry (LGC, Hertfordshire, UK). Primers used are listed in Supplementary Table S6. To ensure quality of genotyping in all assays, at least two negative controls and 6-7% duplicates (showing a concordance >99.9%) were genotyped. Call rates were >95% per 384-well plate for each SNP; cluster plots were visually examined by two researchers.

Statistical and bioinformatic analysis

Main analyses were undertaken using R (v2.10.1; http://www.r-project.org), Stata v.10 (State College, Texas, US) and PLINK (v1.07)[25] software. The association between each SNP and risk 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 λ was based on the 90% least significant SNPs[8]. Odds ratios (ORs) and associated 95% confidence intervals (CIs) were calculated by unconditional logistic regression. Meta-analysis was conducted using standard methods under a fixed effects model[26–27]. 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[28]. I^2 values 75% are considered characteristic of large heterogeneity[28]. Associations by age, sex, histology and EBV status were examined by case-only analyses.

Assuming a sibling relative risk of 3.93 [29] and a prevalence of 2.8×10^{-5} (Surveillance Epidemiology and End Results [SEER] data) we made use of receiver operator characteristic curve analysis[30] to estimate the proportion of the genetic variance on the liability scale attributable to variation at rs3806624, rs7745098, rs1432295, rs6903608, rs2608053, rs2019960 and rs501764.

Prediction of untyped SNPs was carried out using IMPUTEv2[31], based on the 1000 genomes phase 1 integrated variant set (b37) from March 2012. Imputed data were analysed using SNPTEST v2 to account for uncertainties in SNP prediction[32]. LD metrics were calculated in PLINK[25] using 1000genomes data and plotted using SNAP[33]. LD blocks were defined on the basis of HapMap recombination rate (cM/Mb) as defined using the Oxford recombination hotspots[34] and on the basis of distribution of confidence intervals defined by Gabriel *et al.*[35]

To explore the epigenetic profile of association signals we made use of chromatin state segmentation in lymphoblastoid cell lines data generated by the ENCODE Project[36]. The states were inferred from ENCODE Histone Modification data (H4K20me1, H3K9ac, H3K4me3, H3K4me2, H3K4me1, H3K36me3, H3K27me3, H3K27ac and CTCF) binarized using a multivariate Hidden Markov Model. We made use of RegulomeDB[37] and HaploReg[38] to examine if any of the SNPs or their proxies (*i.e.* $r^2 > 0.8$ in 1000genomes CEU reference panel) annotate putative transcription factor binding/enhancer elements.

Relationship between SNP genotype and mRNA expression

To examine for a relationship between SNP genotype and expression we made use of publicly available expression data generated on lymphoblastoid cell lines (LCLs) and T-cells from HapMap3, Geneva and the MuTHER pilot data using Sentrix Human-6 Expression BeadChips (Illumina, San Diego, USA)[39–41]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Kuppers R. The biology of Hodgkin's lymphoma. Nat Rev Cancer. 2009; 9:15–27. [PubMed: 19078975]

2. Hjalgrim H, Engels EA. Infectious aetiology of Hodgkin and non-Hodgkin lymphomas: a review of the epidemiological evidence. J Intern Med. 2008; 264:537–48. [PubMed: 19017178]

3. Mack TM, et al. Concordance for Hodgkin's disease in identical twins suggesting genetic susceptibility to the young-adult form of the disease. N Engl J Med. 1995; 332:413–8. [PubMed: 7824015]

4. Cozen W, et al. A genome-wide meta-analysis of nodular sclerosing Hodgkin lymphoma identifies risk loci at 6p21.32. Blood. 2011; 119:469–75. [PubMed: 22086417]

5. Urayama KY, et al. Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. J Natl Cancer Inst. 2012; 104:240–53. [PubMed: 22286212]

6. Moutsianas L, et al. Multiple Hodgkin lymphoma-associated loci within the HLA region at chromosome 6p21.3. Blood. 2012; 118:670–4.

7. Enciso-Mora V, et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). Nat Genet. 2011; 42:1126–30.

8. Clayton DG, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet. 2005; 37:1243–6. [PubMed: 16228001]

9. Bandele OJ, Wang X, Campbell MR, Pittman GS, Bell DA. Human single-nucleotide polymorphisms alter p53 sequence-specific binding at gene regulatory elements. Nucleic Acids Res. 2010; 39:178–89. [PubMed: 20817676]

10. Zhang S, Li T, Zhang B, Nong L, Aozasa K. Transcription factors engaged in development of NK cells are commonly expressed in nasal NK/T-cell lymphomas. Hum Pathol. 2011; 42:1319–28. [PubMed: 21329961]

11. Liu Y, Sattarzadeh A, Diepstra A, Visser L, van den Berg A. The microenvironment in classical Hodgkin lymphoma: An actively shaped and essential tumor component. Semin Cancer Biol. 2013; doi: 10.1016/j.semcancer.2013.07.002

12. Atayar C, et al. Expression of the T-cell transcription factors, GATA-3 and T-bet, in the neoplastic cells of Hodgkin lymphomas. Am J Pathol. 2005; 166:127–34. [PubMed: 15632006]

13. Pearce EL, et al. Control of effector CD8+ T cell function by the transcription factor Eomesodermin. Science. 2003; 302:1041–3. [PubMed: 14605368]

14. Intlekofer AM, et al. Anomalous type 17 response to viral infection by CD8+ T cells lacking T-bet and eomesodermin. Science. 2008; 321:408–11. [PubMed: 18635804]

15. Paley MA, et al. Progenitor and terminal subsets of CD8+ T cells cooperate to contain chronic viral infection. Science. 2012; 338:1220–5. [PubMed: 23197535]

16. Endo Y, et al. Eomesodermin controls interleukin-5 production in memory T helper 2 cells through inhibition of activity of the transcription factor GATA3. Immunity. 2011; 35:733–45. [PubMed: 22118525]

17. Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011; 476:214–9. [PubMed: 21833088]

18. Landgren O, et al. Re: Familial clustering of Hodgkin lymphoma and multiple sclerosis. J Natl Cancer Inst. 2005; 97:543–4. author reply 544–5. [PubMed: 15812083]

19. Gieger C, et al. New gene functions in megakaryopoiesis and platelet formation. Nature. 2011; 480:201–8. [PubMed: 22139419]

20. Okada Y, et al. Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. PLoS Genet. 2011; 7:e1002067. [PubMed: 21738478]

21. van der Harst P, et al. Seventy-five genetic loci influencing the human red blood cell. Nature. 2012; 492:369–75. [PubMed: 23222517]

22. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]

23. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). Int J Epidemiol. 2006; 35:34–41. [PubMed: 16155052]

24. Penegar S, et al. National study of colorectal cancer genetics. Br J Cancer. 2007; 97:1305–9. [PubMed: 17895893]

25. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. [PubMed: 17701901]

26. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst. 1959; 22:719–48. [PubMed: 13655060]

27. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986; 7:177–88. [PubMed: 3802833]

28. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002; 21:1539–58. [PubMed: 12111919]

29. Goldin LR, Bjorkholm M, Kristinsson SY, Turesson I, Landgren O. Highly increased familial risks for specific lymphoma subtypes. Br J Haematol. 2009; 146:91–4. [PubMed: 19438470]

30. Wray NR, Yang J, Goddard ME, Visscher PM. The genetic interpretation of area under the ROC curve in genomic profiling. PLoS Genet. 2010; 6:e1000864. [PubMed: 20195508]

31. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5:e1000529. [PubMed: 19543373]

32. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007; 39:906–13. [PubMed: 17572673]

33. Johnson AD, et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008; 24:2938–9. [PubMed: 18974171]

34. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. Science. 2005; 310:321–4. [PubMed: 16224025]

35. Gabriel SB, et al. The structure of haplotype blocks in the human genome. Science. 2002; 296:2225–9. [PubMed: 12029063]

36. Ernst J, K M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol. 2010; 28:817–825. [PubMed: 20657582]

37. Boyle AP, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012; 22:1790–7. [PubMed: 22955989]

38. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40:D930–4. [PubMed: 22064851]

39. Dimas AS, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science. 2009; 325:1246–50. [PubMed: 19644074]

40. Nica AC, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet. 2011; 7:e1002003. [PubMed: 21304890]

41. Stranger BE, et al. Patterns of cis regulatory variation in diverse human populations. PLoS Genet. 2012; 8:e1002639. [PubMed: 22532805]

42. Enciso-Mora V, et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). Nat Genet. 2010; 42:1126–30. [PubMed: 21037568]
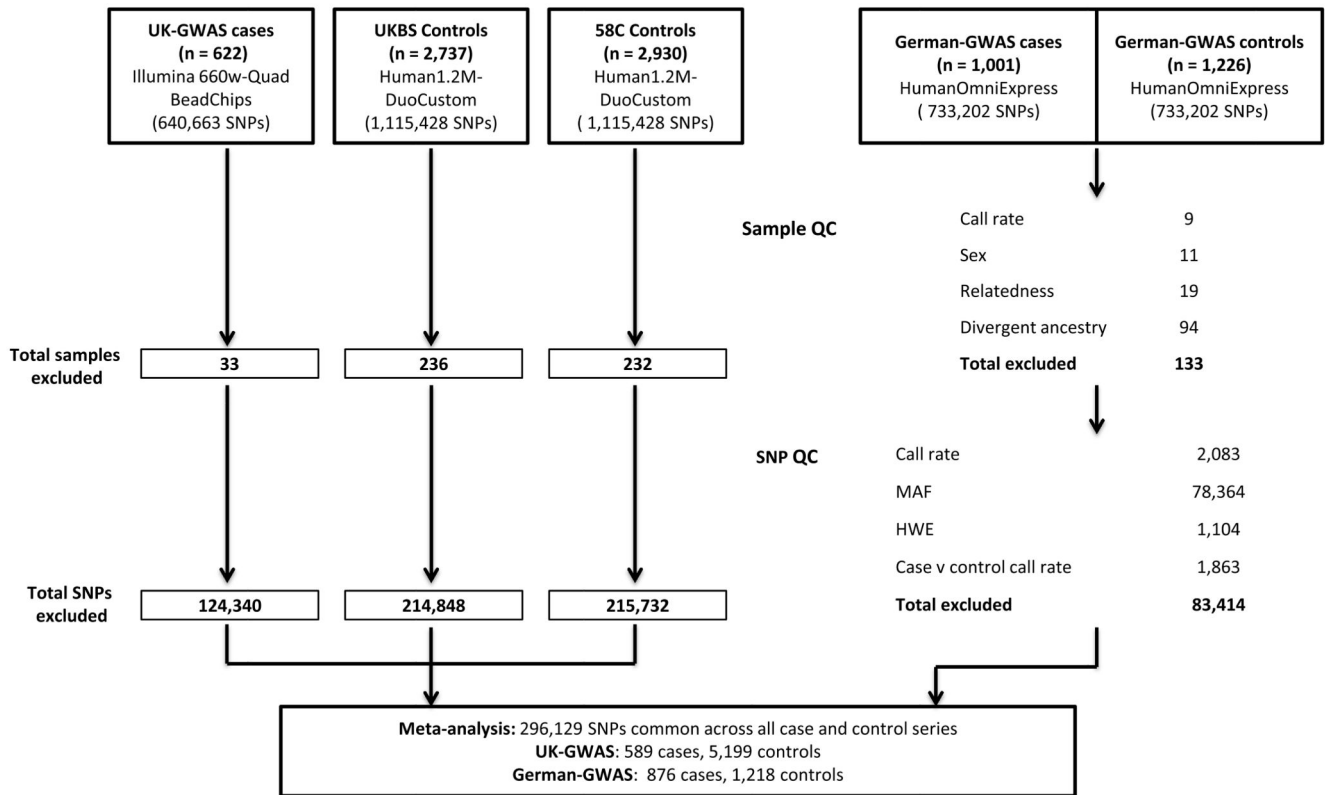
Figure 1. GWAS data quality control.
Details are provided of the samples and SNPs used in UK-GWAS and QC of German-GWAS. QC of UK-GWAS reported in previously published work42
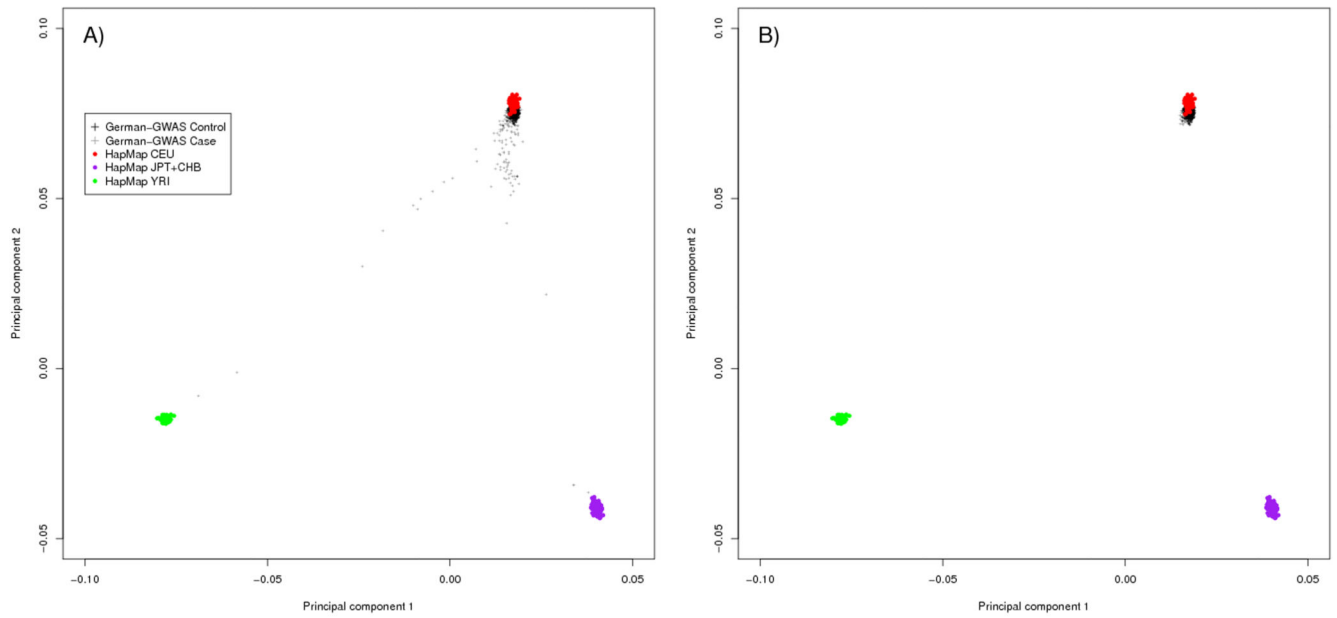
Figure 2. Identification of samples of non-European origin for the German GWAS.
The first two principal components of the analysis are plotted in each figure. (a) German controls (black crosses) and cases (gray crosses) are shown together with the HapMap CEU (Utah residents with Northern and Western European ancestry) individuals (red), CHB (Han Chinese in Beijing) +JPT (Japanese in Tokyo) individuals (purple) and YRI (Yoruba in Ibadan) individuals (green). (b) The same plot is shown after the removal of cases and controls of non-European origin. Respective data on UK-GWAS provided in previously published work42.

(a)
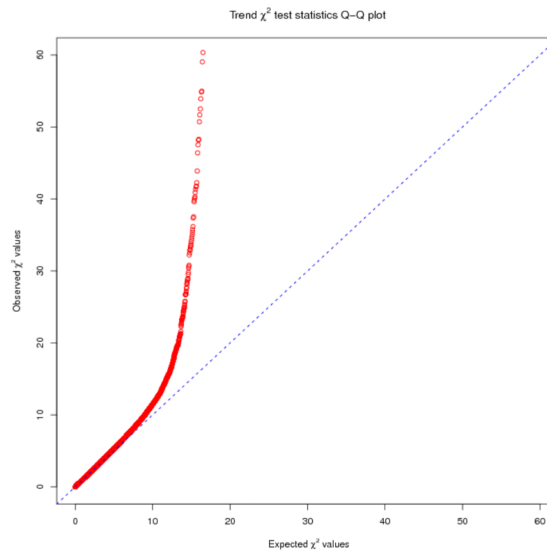
(b)
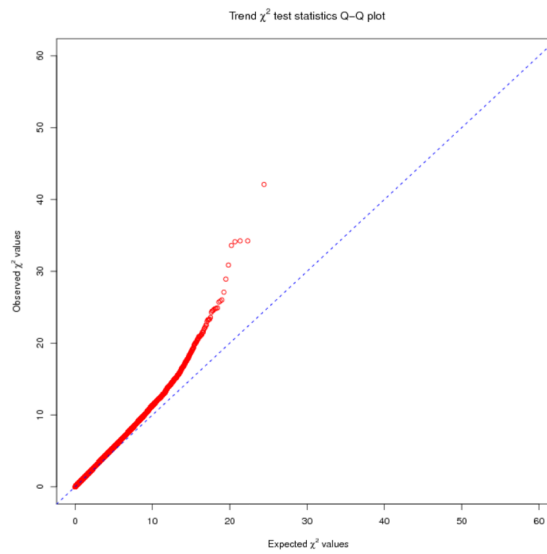
Figure 3. Q-Q plots of chi-square values for association.
(a) UK cases and UK controls (λ=1.03); (b) German cases and German controls (λ=1.09).
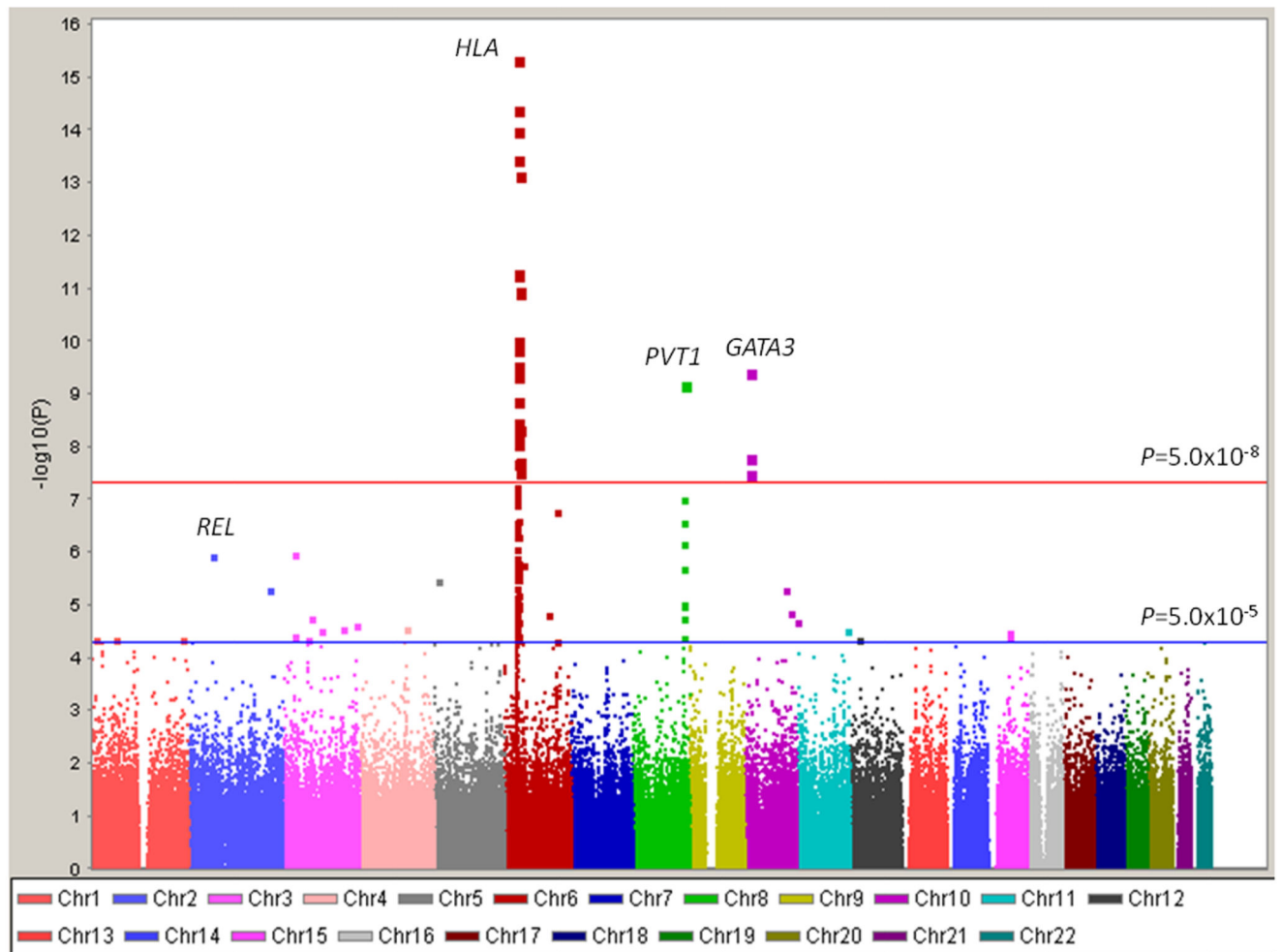
Figure 4. Genome-wide *P*-values plotted against their respective chromosomal positions.
Shown are the genome-wide *P*-values (two-sided) obtained using the Cochran-Armitage trend test from 296,129 autosomal SNPs in 1,465 Hodgkin's Lymphoma cases and 6,417 controls. The red horizontal line represents the genomewide significance threshold level (P=5.0x10^{-8}). The blue horizontal line represents the significance threshold at which SNPs were selected for genotyping in UK-replication 1 (P<5.0x10^{-5}). Previously identified risk loci at *REL, HLA, PVT1* and *GATA3* are labeled.

Dataset	Cases		Controls		OR (95% CI)
	RAF[a]	GG/AG/AA	RAF[a]	GG/AG/AA	
3p24.1 (rs3806624)					
UK-GWAS	0.50	146/298/145	0.45	1041/2604/1552	1.22 (1.08-1.38)
German-GWAS	0.55	253/450/173	0.49	272/646/300	1.26 (1.11-1.42)
UK-replication 1	0.51	274/521/258	0.43	242/614/412	1.35 (1.20-1.52)
UK-replication 2	0.50	234/472/239	0.46	115/286/159	1.16 (1.00-1.34)
Combined					1.26 (1.18-1.34)

$P = 1.14 \times 10^{-12}$
$P_{het} = 0.42$, $I^2 = 0\%$

Odds Ratio: 1.00 1.10 1.20 1.30 1.40 1.50 1.60

Dataset	Cases		Controls		OR (95% CI)
6q23.3 (rs7745098)					
UK-GWAS	0.53	160/301/128	0.48	1210/2603/1382	1.19 (1.06-1.34)
German-GWAS	0.51	225/447/202	0.44	248/574/390	1.33 (1.18-1.51)
UK-replication 1	0.50	271/530/261	0.48	302/612/356	1.11 (0.99-1.25)
UK-replication 2	0.52	253/479/208	0.47	129/272/160	1.23 (1.06-1.43)
Combined					1.21 (1.14-1.29)

$P = 3.42 \times 10^{-9}$
$P_{het} = 0.2$, $I^2 = 0\%$

Odds Ratio: 1.00 1.10 1.20 1.30 1.40 1.50 1.60

Figure 5. Plot of the odds ratios of Hodgkin's Lymphoma associated with rs3806624 and rs7745098.

Studies were weighted according to the inverse of the variance of the log of the OR calculated by unconditional logistic regression. Horizontal lines: 95% confidence intervals (95% CI). Box: OR point estimate; its area is proportional to the weight of the study. Diamond (and broken line): overall summary estimate under fixed effects model, with confidence interval given by its width. Unbroken vertical line: null value (OR=1.0).
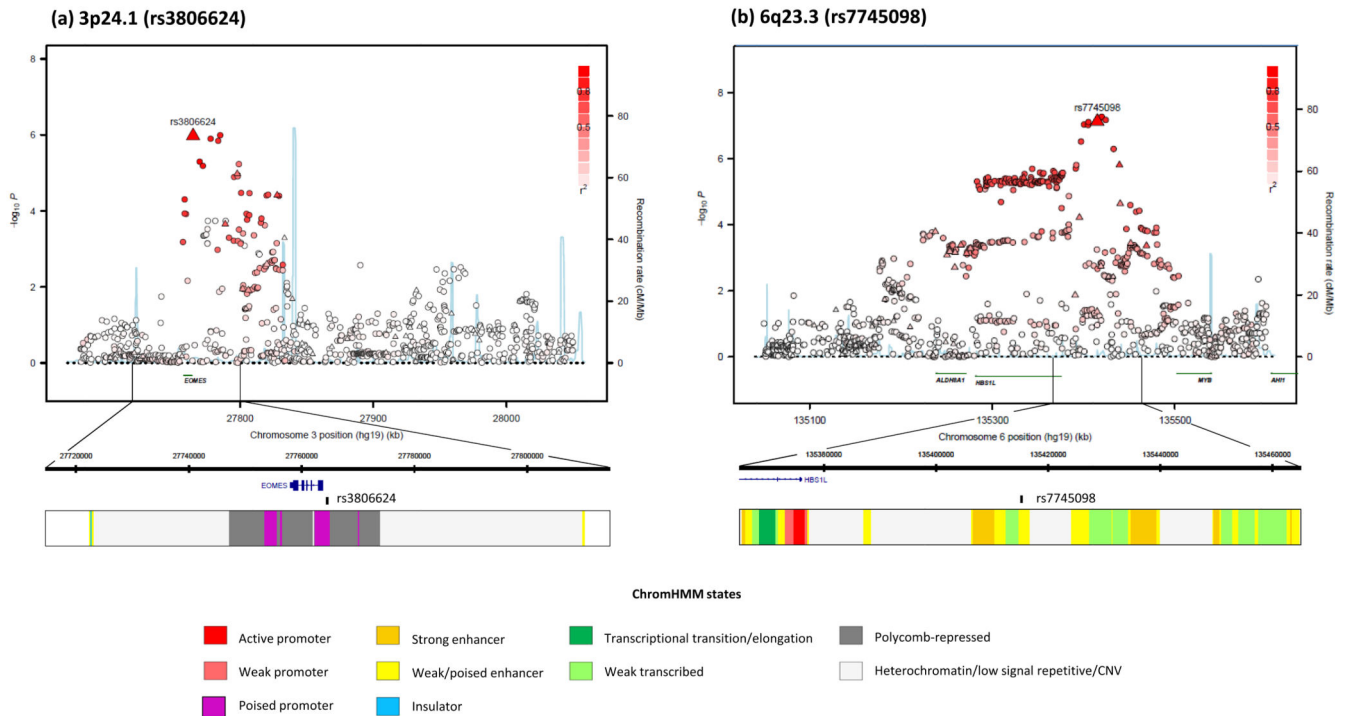
Figure 6. Regional plots of 3p24.1 and 6q23.3 associations with recombination rates and chromatin state segmentation tracks.

(a) 3p24.1 and (b) 6q23.3. Association results of both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates for rates. $-\log_{10} P$ values (y axis) of the SNPs are shown according to their chromosomal positions (x axis). The top genotyped SNP in each combined analysis is shown as a large triangle and is labeled by its rsID. Color intensity of each symbol reflects the extent of LD with the top genotyped SNP; white ($r^2=0$) through to dark red ($r^2=1.0$) Genetic recombination rates, estimated using HapMap Utah residents of Western and Northern European ancestry (CEU) samples, are shown with a light blue line. Physical positions are based on NCBI Build 37 of the human genome. Also shown are the relative positions of genes and transcripts mapping to the region of association. Genes have been redrawn to show the relative positions; therefore, maps are not to physical scale. The lower panel shows the exons and introns of the gene of interest; observed SNP and chromatin state segmentation track (ChromHMM).