



This is a repository copy of *Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/107845/>

Version: Accepted Version

---

**Article:**

Berndt, S.I., Skibola, C.F., Joseph, V. et al. (127 more authors) (2013) Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nature Genetics*, 45 (8). pp. 868-876. ISSN 1061-4036

<https://doi.org/10.1038/ng.2652>

---

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>



Published in final edited form as:

Nat Genet. 2013 August ; 45(8): 868– 876. doi:10.1038/ng.2652.

## Genome-wide Association Study Identifies Multiple Risk Loci for Chronic Lymphocytic Leukemia

A full list of authors and affiliations appears at the end of the article.

Despite limited discovery stages (<1,125 cases), genome-wide association studies (GWAS) have successfully identified 13 loci associated with risk of chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL). To identify additional CLL susceptibility loci, we conducted the largest meta-analysis, to date, including four GWAS totaling 3,100 CLL cases and 7,667 controls with genotype data. In the meta-analysis, we discovered ten independent SNPs in nine novel loci at 10q23.31 (*ACTA2/FAS*;  $P=1.22\times 10^{-14}$ ), 18q21.33 (*BCL2*;  $P=7.76\times 10^{-11}$ ), 11p15.5 (*C11orf21*;  $P=2.15\times 10^{-10}$ ), 4q25 (*LEF1*;  $P=4.24\times 10^{-10}$ ), 2q33.1 (*CASP10/CASP8*;  $P=2.50\times 10^{-9}$ ), 9p21.3 (*CDKN2B-AS1*;  $P=1.27\times 10^{-8}$ ), 18q21.32 (*PMAIP1*;  $P=2.51\times 10^{-8}$ ), 15q15.1 (*BMF*;  $P=2.71\times 10^{-10}$ ), and 2p22.2 (*QPCT*;  $P=1.68\times 10^{-8}$ ) as well as an independent signal at an established locus (2q13, *ACOXL*,  $P=2.08\times 10^{-18}$ ). We also found evidence for two additional promising loci that reached marginal genome-wide significance ( $P<2.0\times 10^{-7}$ ) at 8q22.3 (*ODF1*;  $P=5.40\times 10^{-8}$ ) and 5p15.33 (*TERT*;  $P=1.92\times 10^{-7}$ ). Although further studies are required, proximity of several of these loci to genes involved in apoptosis suggests a plausible underlying biological mechanism.

CLL is a B-cell malignancy with a strong familial component<sup>1</sup> and an ~8.5-fold increased relative risk in first-degree relatives.<sup>2</sup> Previous CLL GWAS have identified 13 loci that

Users may view, print, copy, download and text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

Correspondence should be addressed to: Susan L. Slager, Ph.D., Mayo Clinic, 200 First Street SW, Rochester, MN 55905, Phone: 507.284.5965, Fax: 507.284.9542, [slager@mayo.edu](mailto:slager@mayo.edu).

<sup>90</sup>These authors contributed equally to this work.

<sup>91</sup>These authors jointly directed this work.

### AUTHORS' CONTRIBUTIONS

S.I.B., C.F.S., N.J.C., A.N., W.C., S.S.W., L.R.T., A.R.B.W., P.H., M.P.P., B.M.B., B.K.A., P.C., Y.Z., G.S., A.Z.J., C.L., K.E.S., J.M., P.V., J.J.S., A.K., S. S., H.H., J.R.C., S.J.C., N.R. and S.L.S. organized and designed the study. C.F.S., N.J.C., B.J.L.B., J.Y., A.H., L.C., P.M.B., E.A.H., J.M.C., J.R.C., S.J.C. and S.L.S. conducted and supervised the genotyping of samples. S.I.B., C.F.S., V.J., N.J.C., Z.W., N.C., C.C.C., M.Y., K.B.J., L.L., J.S., J.P., J.R.C., L.C., S.J.C., N.R. and S.L.S. contributed to the design and execution of statistical analysis. S.I.B., C.F.S., V.J., N.J.C., A.N., Z.W., W.C., A.M., R.S.K., N.C., C.C.C., M.Y., C.L., H.H., J.R.C., S.J.C., N.R. and S.L.S. wrote the first draft of the manuscript. S.I.B., C.F.S., V.J., N.J.C., A.N., W.C., A.M., S.S.W., R.S.K., Q.L., L.R.T., A.R.B.W., P.H., M.P.P., B.M.B., B.K.A., P.C., Y.Z., G.S., A.Z.J., T.G.C., T.D.S., A.J.N., N.E.K., M.L., A.H.W., K.E.S., H.O.A., M.M., B.G., E.T.C., M.G., K.C., L.A.C.A., B.J., W.R.D., B.K.L., G.J.W., L.C., P.M.B., J.R., E.A.H., M.T.S., R.D.J., L.F.T., S.D.S., Y.B., N.B., P.B., L.F., M.M., J.M., A.S., K.G.R., S.J.A., C.M.V., L.R.G., S.S.S., M.C.L., L.G.S., J.F.L., J.M.C., J.B.W., V.A.M., N.E.C., A.N., M.S.L., A.J.D.R., L.M.M., R.K.S., E.R., P.V., R.K., D.T., G.M., E.W., M.D.C., R.C.H.V., R.C.T., G.G.G., D.A., J.V., S.W., J.C., T.Z., T.R.H., K.O., A.Z., R.J.K., J.J.S., K.A.B., F.L., E.G., P.K., A.K., J.T., C.M.V., M.G.E., G.M.F., L.M., L.L., J.S., S.C., J.F.F., K.E.N., A.C., J.S., J.W., A.C., C.L.O., S.B., I.S., D.M., E.C., H.H., J.R.C., N.R. and S.L.S. conducted the epidemiological studies and contributed samples to the GWAS and/or follow-up genotyping. All authors contributed to the writing of the manuscript.

### COMPETING INTERESTS

The authors declare no competing financial interests

explain a portion of the familial risk,<sup>3–6</sup> suggesting that additional loci of modest effects can be found using a larger discovery sample size.<sup>7</sup>

As part of a larger initiative in non-Hodgkin lymphoma (NHL) (called the NHL-GWAS), we genotyped 2,343 CLL cases and 2,854 controls of European descent from 22 studies using the Illumina OmniExpress Beadchip (see **Online Methods** and Supplementary Table 1). Of those 5,197 subjects, 94% passed rigorous quality control criteria (see **Online Methods** and Supplementary Table 2) and 549,934 SNPs successfully passed quality control criteria with a median call rate >98%. We also utilized genotype data previously generated on the Illumina Omni2.5 from an additional 3,536 controls and one case from three studies<sup>8</sup> giving a total of 2,179 cases and 6,221 controls for the analysis of the NHL-GWAS (Supplementary Table 3).

In the NHL-GWAS (Stage 1) analysis, we observed an enrichment of SNPs with small *P*-values compared to the null distribution with a lambda of 1.026 in the Q-Q plot (Supplementary Figure 1). After exclusion of previously established loci, an excess of small *P*-values still remained suggesting additional novel loci were yet to be discovered. In our Stage 1 analyses, we observed SNPs from 10 unique loci (defined as separated by at least 500kb and linkage disequilibrium (LD)  $r^2 < 0.05$ ), which reached genome-wide significance ( $P < 5 \times 10^{-8}$ ), including eight established loci and two novel loci (Supplementary Figure 2).

We then performed a meta-analysis of the NHL-GWAS with three other independent CLL GWAS<sup>5,9</sup> that had a combined total of 921 CLL cases and 1,446 controls (Stage 2, Supplementary Tables 1 and 3). Because these other CLL GWAS studies were conducted on different commercial SNP microarrays, we imputed common SNPs from the 1000 Genomes Project<sup>10</sup> using IMPUTE2<sup>11</sup> (**Online Methods**, Supplementary Table 4). In the meta-analysis of stages 1 and 2 data, associations for all 13 established loci showed a consistent direction of effect with previously reported studies, and 10 loci achieved  $P < 5 \times 10^{-8}$  (Supplementary Table 5). However, two previously established loci, 15q25.2 and 19q13.3, were only nominally significant in the meta-analysis ( $P = 0.03$ , and  $P = 0.008$ , respectively), and no significant association was observed in stage 1 for the 15q25.2 locus ( $P = 0.10$ ). A suggestive locus on 18q21.1 that had not met genome-wide significance in prior studies<sup>12</sup> was also nominally significant ( $P = 5.06 \times 10^{-4}$ ) herein. From the meta-analysis of stages 1–2, we identified 10 promising SNPs in the eight novel loci and one promising SNP in an established locus that we carried forward for a *de novo* replication in stage 3: this included an additional 392 cases and 4561 controls and *in silico* replication in an independent CLL GWAS with 396 cases and 311 controls (see **Online Methods** and Supplementary Tables 1, 3, and 4).

Seven of the 10 SNPs in novel loci reached genome-wide significance in the meta-analysis of all three stages: 10q23.31 (*ACTA2/FAS*;  $P = 1.22 \times 10^{-14}$ ), 18q21.33 (*BCL2*;  $P = 2.66 \times 10^{-12}$ ), 11p15.5 (*C11orf21*;  $P = 2.15 \times 10^{-10}$ ), 4q25 (*LEF1*;  $P = 4.24 \times 10^{-10}$ ), 2q33.1 (*CASP10/CASP8*;  $P = 2.50 \times 10^{-9}$ ), 9p21.3 (*CDKN2B-AS1*;  $P = 1.27 \times 10^{-8}$ ), and 18q21.32 (*PMAIP1*;  $P = 2.51 \times 10^{-8}$ ) (Table 1, Figure 1). Further, within the 18q21.33 locus, a second SNP (rs4987852) in low LD ( $r^2 = 0.01$ ) with rs4987855 and located only 372 bp away, also

reached genome-wide significance (Table 1,  $P = 7.76 \times 10^{-11}$ ); this SNP was determined to be independent in conditional analyses ( $P_{\text{conditional}} = 3.87 \times 10^{-7}$ , Table 2).

To explore these regions in greater detail and identify additional loci that we may have missed using just the genotyped SNPs in Stage 1, we imputed Stage 1 of our NHL-GWAS using the 1000 Genomes Project<sup>10</sup> data (February 2012 release) and performed a meta-analysis of the results from stage 1 and stage 2. The most significant SNPs at three of our novel loci, 10q23.31 (rs2147420) 18q21.33 (rs4987856), and 4q25 (rs2003869), were highly correlated ( $r^2 = 0.95$ ) with our strongest genotyped SNPs, rs4406737, rs4987885, and rs898518, respectively (Supplementary Table 6). Only modest correlation ( $r^2$  range: 0.18–0.58) was observed for the most significant imputed SNPs at 11p15.5 (rs2521269), 2q33.1 (rs11688943), and 9p21.3 (rs1359742) and our strongest genotyped SNPs in each of the respective regions. The most significant of the imputed SNPs at 18q21.32 (rs35748167) appeared to be independent of our strongest genotyped SNP (rs4368253,  $r^2 = 0.003$ ,  $P_{\text{conditional}} < 7.89 \times 10^{-7}$  for both SNPs), suggesting a possible second, independent signal (Table 2).

Meta-analysis of our imputed scan data revealed two novel loci, 15q15.1 (*BMF*;  $P = 2.71 \times 10^{-10}$ ) and 2p22.2 (*QPCT*;  $P = 1.68 \times 10^{-8}$ ) (Table 1, Figure 1). In addition, although our genotyped SNP at 5p15.33 (*TERT*, rs10069690,  $P = 1.92 \times 10^{-7}$ ) (Supplementary Table 7) did not reach genome-wide significance, we did observe an imputed SNP in this region that reached genome-wide significance (rs7705526;  $P = 3.75 \times 10^{-8}$ ). Another promising locus was observed at 8q22.3 (*ODFI*;  $P = 5.40 \times 10^{-8}$ ) (Supplementary Table 7). Additional studies are needed to confirm these findings, particularly the signal on 5p15.33, which is already known to harbor risk variants for multiple cancers.<sup>13–20</sup>

An examination of established loci revealed a new SNP in 2q13 (*BCL2L11*, rs13401811,  $P = 6.09 \times 10^{-17}$ ; Table 1, Figure 2) that was independent of the previously reported SNP. After conditioning on the established 2q13 SNP (rs17483466,  $r^2 = 0.02$ ), the new SNP rs13401811 remained strongly associated with CLL risk ( $P_{\text{conditional}} = 1.60 \times 10^{-12}$ , Table 2). A putative second signal was observed at the established 2q37.3 locus (Supplementary Table 5, rs7578199,  $P = 5.39 \times 10^{-7}$ ) that was in low LD ( $r^2 = 0.01$ ) and independent of the previously reported rs757978 SNP ( $P_{\text{conditional}} = 6.10 \times 10^{-6}$ , Table 2), although rs7578199 was not genome-wide significant. Another possible second signal was observed on 6p21.32 (Supplementary Table 5, *HLA*, rs9273363,  $P = 2.24 \times 10^{-10}$ ). Rs9273363 showed some evidence of conditional independence with the originally reported SNPs ( $r^2 = 0.25$ ,  $P_{\text{conditional}} = 3.50 \times 10^{-9}$ , Table 2); however, it may be part of a shared HLA haplotype; thus accurate *HLA* typing is needed to further clarify its level of independence. Finally, we observed a SNP at 15q21.3 (Supplementary Table 5, rs11636802,  $P = 1.68 \times 10^{-13}$ ) that had stronger statistical significance than that of the previously reported SNP, rs7169431 ( $P = 1.72 \times 10^{-05}$ ). Although only modestly correlated ( $r^2 = 0.16$ ), rs11636802 explained all of the risk associated with rs7169431 in a conditional analysis (Table 2) suggesting that this SNP may be a better marker for the locus.

Heritability analysis indicated that the ten independent SNPs in our novel loci together with the new independent SNP at 2q13 (Table 1) explain approximately 5% more of the familial

risk in addition to ~12% for the established loci. When we explored the contribution of all common variants to the genetic heritability of CLL (using a method that estimates the variance explained by fitting all genotyped autosomal SNPs simultaneously<sup>21,22</sup>, **Online Methods**)<sup>21,22</sup> we estimate that common SNPs have the potential to explain up to ~46% of the familial risk, suggesting more common loci, likely of small effects, are still yet to be discovered. However, the analysis also implies that common SNPs probably do not explain all of the familial risk and other factors, such as uncommon SNPs with modest effects or rare highly penetrant variants, are likely to also play a role.

Five of the novel loci (10q23.31, 18q21.33, 2q33.1, 18q21.32, and 15q15.1) identified in this study as well as the new SNP at the established 2q13 locus are located in or near genes involved in apoptosis. Rs4406737 is located on 10q23.31 between the first and second exons of *FAS*, a member of the tumor necrosis factor receptor superfamily that has a crucial role in the initiation of the signaling cascade of the caspase family in apoptosis. Mutations in *FAS* leading to defective Fas-mediated apoptosis have been documented in inherited lymphoproliferative disorders associated with autoimmunity,<sup>23,24</sup> and families with germline *FAS* mutations have a substantially increased risk of other lymphoma subtypes.<sup>25</sup>

The two newly identified SNPs at 18q21.33 (rs4987855 and rs4987852) map to the 3'-UTR of B-cell CLL/lymphoma 2 (*BCL2*), which encodes an essential outer mitochondrial membrane protein that blocks lymphocyte apoptosis. Constitutive expression of *BCL2* through t(14;18) and other translocations is common in follicular lymphomas, but the translocation is also seen in CLL albeit rarely.<sup>26</sup> Both SNPs are located within a narrow region of *BCL2* where the majority of t(14;18) translocation breakpoints occur.<sup>27</sup> rs4987855 is in linkage disequilibrium with a SNP (rs4987856,  $r^2=1.0$ ) that is located within 200bp of a putative microRNA binding site for mir-195<sup>28</sup> and was found to be nominally correlated with *BCL2* expression (Supplementary Table 8,  $P=0.02$ )<sup>29</sup>. Forced overexpression of *BCL2* in mice leads to an increased incidence of B-cell lymphomas.<sup>30</sup>

The novel SNPs at 18q21.32 and 15q15.1 as well as the new SNP at the established 2q13 locus are located near Bcl-2 family member genes. Rs4368253 is located approximately 51kb downstream from phorbol-12-myristat-13-acetate-induced protein 1 (*PMAIP1*), which encodes the proapoptotic BCL2 protein, NOXA. Regulation of apoptosis through NOXA is critical for B-cell expansion after antigen triggering.<sup>31</sup> Down-regulation of NOXA contributes to the persistence of CLL B-cells in the lymph node environment.<sup>32</sup> Rs8024033 is located approximately 5.4kb upstream of Bcl-2 modifying factor (BMF), which encodes an apoptotic activator that binds to BCL2 proteins. BMF has been implicated in the survival of chronic lymphocytic leukemia cells<sup>33</sup>, and loss of BMF in mice leads to B-cell hyperplasia and an accelerated development of radiation-induced thymic lymphomas<sup>34</sup>. The new SNP (rs13401811) at 2q13, a locus previously implicated in risk of CLL<sup>3,35,36</sup> and more generally B-cell non-Hodgkin lymphomas,<sup>37</sup> is located approximately 262kb upstream of BCL2-like 11 (*BCL2L11*). *BCL2L11* encodes a pro-apoptotic member of the BCL2 family, BIM, which plays a key role in the regulation of apoptosis in T- and B-cell homeostasis. Loss of BIM accelerates Myc-induced leukemia in mice,<sup>38</sup> and this SNP has been previously reported to be nominally associated with CLL in a small candidate gene study.<sup>39</sup>

The novel 2q33.1 SNP (rs3769825) resides in intron 2 of caspase-8 (*CASP8*) and is in LD with a missense SNP (rs13006529,  $r^2=0.71$ ) in the nearby caspase-10 (*CASP10*) (Supplementary Table 9), both of which play a central role in cell apoptosis. SNPs within this region have been associated with breast cancer,<sup>40</sup> esophageal cancer,<sup>41</sup> and melanoma<sup>42</sup> susceptibility. SNPs in *CASP8/CASP10*, including one in moderate LD with ours (rs11674246,  $r^2=0.66$ ), were previously nominally associated with CLL risk in smaller case-control studies.<sup>43,44</sup>

The remaining four novel loci (11p15.5, 4q25, 9p21.3 and 2p22.2) map to other biologically interesting genes. The 4q25 SNP, rs898518, is located between the fourth and fifth exons of lymphoid enhancer-binding factor 1 (*LEF1*), which encodes a transcription factor involved in the Wnt signaling pathway, an essential component for the normal homeostasis of hematopoietic stem cells.<sup>45</sup> Aberrant protein expression of LEF1 has been observed in CLL cells as well as monoclonal B-cell lymphocytosis, suggesting that LEF1 plays an early role in CLL leukemogenesis.<sup>46</sup> Rs1679013 maps to an inter-genic region on 9p21.3, roughly 200kb upstream from *CDKN2B-AS1*, an antisense non-coding RNA implicated in the risk of acute lymphocytic leukemia.<sup>47</sup> The 2p22.2 SNP (rs3770745) is located approximately 52kb upstream of protein kinase D3 (*PRKD3*), which interacts with transcriptional repressor, B-cell lymphoma 6 (*BCL-6*). Lastly, the 11p15.5 region contains many imprinted genes and has been implicated in Beckwith-Wiedemann syndrome,<sup>48</sup> a disorder characterized by excessive growth and a high incidence of childhood tumors.<sup>49</sup>

In conclusion, our large GWAS of CLL identified ten SNPs in nine novel loci and one new independent SNP in a previously discovered locus. Together with the previously established loci, the cumulative set of SNPs correspond to an area-under-the-curve (AUC) of 0.73. Although further studies are required to fine-map the regions, the proximity of several of these loci to genes involved in apoptosis suggests a possible underlying mechanism of biological relevance. Our results further support a substantial contribution of common gene variants in the pathogenesis of CLL.

## ONLINE METHODS

### Stage 1: NHL-GWAS

As part of a larger initiative, we conducted a genome-wide association study (GWAS) of CLL using cases and controls of European descent from 22 studies of non-Hodgkin lymphoma (NHL) (Supplementary Table 1), including nine prospective cohort studies, eight population-based case-control studies, and five clinic or hospital-based case-control studies. All studies obtained informed consent from their participants and approval from their respective Institutional Review Boards for this study. As described in Supplementary Table 1, cases were ascertained from cancer registries, clinics or hospitals, or through self-report verified by medical and pathology reports. The phenotype information for all NHL cases was reviewed centrally at the International Lymphoma Epidemiology Consortium (InterLymph) Data Coordinating Center and harmonized according to the hierarchical classification proposed by the InterLymph Pathology Working Group based on the World Health Organization (WHO) classification (2008).<sup>50,51</sup>



All CLL cases with sufficient DNA ( $n=2,343$ ) and a subset of available controls frequency-matched by age and sex to cases ( $n=2,854$ ) including 4% quality control duplicates were genotyped on the Illumina OmniExpress at the NCI Cancer Genomic Research Laboratory (CGR). Genotypes were called using Illumina GenomeStudio software, and quality control duplicates showed >99% concordance. Extensive quality control metrics were applied to the data. Monomorphic SNPs and SNPs with a call rate <93% were excluded. Samples with a call rate <93%, mean heterozygosity <0.25 or >0.33 based on the autosomal SNPs, or gender discordance (>5% heterozygosity on X chromosome for males and <20% heterozygosity on the X chromosome for females) were excluded. Unexpected duplicates (>99.9% concordance) and first-degree relatives based on identity by descent (IBD) sharing with  $\text{Pi-hat} > 0.40$  were removed. Ancestry was assessed using the GLU *struct.admix* module based on the method proposed by Pritchard et al,<sup>52</sup> and participants with <80% European ancestry were excluded (Supplementary Figure 3). After exclusions, 2,178 (93%) cases and 2,685 (94%) controls remained (Supplementary Table 2). Genotype data previously generated on the Illumina Omni2.5 from additional 3,536 controls and 1 case from three of the studies (ATBC, CPSII, and PLCO) were also included,<sup>8</sup> resulting in a total of 2,179 cases and 6,221 controls for the stage 1 analysis. Of these additional controls, 703 (~235 from each study) were selected to be representative of their cohort and cancer-free<sup>8</sup>. The remaining 2,823 controls were cancer-free controls from an unpublished study of prostate cancer in PLCO. SNPs with call rate <99%, with Hardy-Weinberg equilibrium  $P$ -value <  $1 \times 10^{-6}$  or minor allele frequency <1% were excluded from analysis, leaving 549,934 SNPs for analysis. To evaluate population substructure, a principal components analysis (PCA) was performed using the Genotyping Library and Utilities (GLU), version 1.0, *struct.pca* module, which is similar to EIGENSTRAT.<sup>53</sup> Plots of the first ten principal components are shown in Supplementary Figure 4. Association testing was conducted assuming a log-additive genetic model, adjusting for age, sex, and significant principal components. All data analysis and management was conducted using GLU.

## Stage 2: Three Independent CLL GWAS

Three independent CLL GWAS provided genotype data for a meta-analysis (Supplementary Table 1). In all three studies, subjects with a genotyping call rate <95%, duplicates, related individuals, and SNPs with a call rate <95% were removed prior to imputation (Supplementary Table 4). Imputation was conducted separately for each study using IMPUTE2<sup>11</sup> and a hybrid of the 1000 Genomes Project version 2 (February 2012 release) and Division of Cancer Epidemiology and Genetics (DCEG) European reference panels.<sup>8,10</sup> SNPs were imputed for a total of 921 cases and 1446 controls. Association testing was conducted for each study using SNPTEST version 2, adjusting for age, sex, and significant principal components for GEC and UCSF2. No principal components were significant for the Utah study.

## Stage 3: Replication studies and technical validation

In stage 3, 10 SNPs in the most promising loci and one SNP from an established locus were taken forward for *de novo* replication in an additional 392 cases and 4561 controls from the NCI replication study (NCI Rep) and from the Utah/Sheffield Chronic Lymphocytic Leukemia study (Utah-Sheffield) (Supplementary Table 1). Additionally, these 10 SNPs

were also taken forward in an *in silico* replication in 396 CLL cases and 311 controls from the International Cancer Genome Consortium (ICGC) (Supplementary Table 1). Genotyping for the NCI Rep study was conducted using custom TaqMan genotyping assays (Applied Biosystems) at the NCI Core Genotyping Resource and genotyping for the Utah-Sheffield study was conducted at the Core Research Facilities at the University of Utah. Blind duplicates (~5%) yielded 100% concordance. The ICGC study provided results for eight SNPs (or proxies) that were genotyped on the Affymetrix 6.0 SNP microarray (Supplementary Table 4). Association results for the NCI Rep and Utah-Sheffield studies were adjusted for age and sex, and results from the ICGC were adjusted for age, sex, and significant principal components. A comparison of the genotyping calls from the OmniExpress microarray and confirmatory TaqMan assays (n=384) yielded 99.9% concordance.

### Meta analysis

Meta-analyses were performed using the fixed effects inverse variance method based on the beta estimates and standard errors from each study. For all SNPs in Tables 1 and 2, no substantial heterogeneity was observed among studies in stage 1 or among studies in stages 1– 3 combined after Bonferroni correction ( $P_{\text{heterogeneity}} < 0.02$  for all SNPs).

### Further follow-up analyses

Using 1000 Genomes data, we identified SNPs with  $r^2 > 0.7$  with our lead SNP that were reported to be non-synonymous or nonsense variants. We utilized HaploReg<sup>54</sup> which is a tool for exploring non-coding functional annotation using ENCODE data, to evaluate the genome surrounding our SNPs (Supplementary Table 9). In addition, we evaluated cis associations between all novel and promising SNPs discovered in this study and the expression of nearby genes in lymphoblastoid cell lines from subjects of European descent from three publically available datasets<sup>29,55,56</sup> (Supplementary Table 8).

### Heritability analyses

To evaluate the familial risk explained by the novel loci identified in this study, we estimated the contribution of each SNP to the heritability using the equation<sup>7</sup>,  $h^2_{\text{SNP}} = \beta^2 2f(1-f)$ , where  $\beta$  is the log-odds ratio per copy of the risk allele and  $f$  is the allele frequency, and then summed the contributions of all novel SNPs. Using the equation derived by Pharoah et al<sup>57</sup> to estimate the total heritability from the sibling relative risk (RR=8.5 from Goldin et al<sup>2</sup>), we then calculated the proportion of familial risk explained by dividing the summed contributions of the novel SNPs by the total heritability.

To estimate the contribution of all common SNPs to familial risk, we used the method proposed by Yang et al<sup>21</sup>, (which was extended to dichotomous traits<sup>22</sup> and implemented in the Genome-wide Complex Trait Analysis (GCTA) software.<sup>58</sup> The genetic similarity matrix was estimated from our discovery scan using all genotyped autosomal SNPs with a minor allele frequency  $> 0.01$ . We used restricted maximum likelihood (REML), the default option for GCTA, to fit the appropriate variance components model that included the top 10 eigenvectors as covariates. The final estimate of heritability on the underlying liability scale assumed that the lifetime risk of CLL was 0.005. From this estimate, we calculated the



proportion of familial risk explained based on a familial relative risk of 8.5. Details of fitting the variance components model and transforming from the observed to liability scale have been previously documented.<sup>22</sup>

### Estimate of recombination hotspots

To identify recombination hotspots in the region we used SequenceLDhot<sup>59</sup>, a program that uses the approximate marginal likelihood method<sup>60</sup> and calculates likelihood ratio statistics at a set of possible hotspots. We tested five unique sets of 100 control samples. PHASE v2.1 program was used to calculate background recombination rates<sup>61,62</sup> and LD heatmap was visualized in r2 using snp.plotter program.<sup>63</sup>

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Authors

Sonja I. Berndt<sup>1,90</sup>, Christine F. Skibola<sup>2,3,90</sup>, Vijai Joseph<sup>4,90</sup>, Nicola J. Camp<sup>5,90</sup>, Alexandra Nieters<sup>6,90</sup>, Zhaoming Wang<sup>7</sup>, Wendy Cozen<sup>8</sup>, Alain Monnereau<sup>9,10</sup>, Sophia S. Wang<sup>11</sup>, Rachel S. Kelly<sup>12</sup>, Qing Lan<sup>1</sup>, Lauren R. Teras<sup>13</sup>, Nilanjan Chatterjee<sup>1</sup>, Charles C. Chung<sup>1</sup>, Meredith Yeager<sup>7</sup>, Angela R. Brooks-Wilson<sup>14,15</sup>, Patricia Hartge<sup>1</sup>, Mark P. Purdue<sup>1</sup>, Brenda M. Birmann<sup>16</sup>, Bruce K. Armstrong<sup>17</sup>, Pierluigi Cocco<sup>18</sup>, Yawei Zhang<sup>19</sup>, Gianluca Severi<sup>20</sup>, Anne Zeleniuch-Jacquotte<sup>21</sup>, Charles Lawrence<sup>22</sup>, Laurie Burdette<sup>7</sup>, Jeffrey Yuenger<sup>7</sup>, Amy Hutchinson<sup>7</sup>, Kevin B. Jacobs<sup>7</sup>, Timothy G. Call<sup>23</sup>, Tait D. Shanafelt<sup>24</sup>, Anne J. Novak<sup>24</sup>, Neil E. Kay<sup>23</sup>, Mark Liebow<sup>25</sup>, Alice H. Wang<sup>26</sup>, Karin E Smedby<sup>27,28</sup>, Hans-Olov Adami<sup>29,30</sup>, Mads Melbye<sup>31</sup>, Bengt Glimelius<sup>28,32</sup>, Ellen T. Chang<sup>33,34</sup>, Martha Glenn<sup>35</sup>, Karen Curtin<sup>5</sup>, Lisa A. Cannon-Albright<sup>5,36</sup>, Brandt Jones<sup>5</sup>, W. Ryan Diver<sup>13</sup>, Brian K. Link<sup>37</sup>, George J. Weiner<sup>37</sup>, Lucia Conde<sup>2,3</sup>, Paige M. Bracci<sup>38</sup>, Jacques Riby<sup>2</sup>, Elizabeth A. Holly<sup>38</sup>, Martyn T. Smith<sup>2</sup>, Rebecca D. Jackson<sup>39</sup>, Lesley F. Tinker<sup>40</sup>, Yolanda Benavente<sup>41,42</sup>, Nikolaus Becker<sup>43</sup>, Paolo Boffetta<sup>44</sup>, Paul Brennan<sup>45</sup>, Lenka Foretova<sup>46</sup>, Marc Maynadie<sup>47</sup>, James McKay<sup>48</sup>, Anthony Staines<sup>49</sup>, Kari G. Rabe<sup>26</sup>, Sara J. Achenbach<sup>26</sup>, Celine M. Vachon<sup>26</sup>, Lynn R Goldin<sup>1</sup>, Sara S. Strom<sup>50</sup>, Mark C. Lanasa<sup>51</sup>, Logan G. Spector<sup>52</sup>, Jose F. Leis<sup>53</sup>, Julie M. Cunningham<sup>54</sup>, J. Brice Weinberg<sup>51</sup>, Vicki A. Morrison<sup>55</sup>, Neil E. Caporaso<sup>1</sup>, Aaron D. Norman<sup>26</sup>, Martha S. Linet<sup>1</sup>, Anneclaire J. De Roos<sup>40</sup>, Lindsay M. Morton<sup>1</sup>, Richard K. Severson<sup>56</sup>, Elio Riboli<sup>57</sup>, Paolo Vineis<sup>12,58</sup>, Rudolph Kaaks<sup>43</sup>, Dimitrios Trichopoulos<sup>30,59,60</sup>, Giovanna Masala<sup>61</sup>, Elisabete Weiderpass<sup>29,62,63,64</sup>, María-Dolores Chirlaque<sup>42,65</sup>, Roel C H Vermeulen<sup>66,67</sup>, Ruth C. Travis<sup>68</sup>, Graham G. Giles<sup>20</sup>, Demetrius Albanes<sup>1</sup>, Jarmo Virtamo<sup>69</sup>, Stephanie Weinstein<sup>1</sup>, Jacqueline Clavel<sup>9</sup>, Tongzhang Zheng<sup>19</sup>, Theodore R Holford<sup>70</sup>, Kenneth Offit<sup>4</sup>, Andrew Zelenetz<sup>4</sup>, Robert J. Klein<sup>4,71</sup>, John J. Spinelli<sup>72</sup>, Kimberly A. Bertrand<sup>16,30</sup>, Francine Laden<sup>16,30,73</sup>, Edward Giovannucci<sup>16,30,74</sup>, Peter Kraft<sup>30,75</sup>, Anne Krickler<sup>17</sup>, Jenny Turner<sup>76,77</sup>, Claire M. Vajdic<sup>78</sup>, Maria Grazia Ennas<sup>79</sup>, Giovanni M. Ferri<sup>80</sup>, Lucia Miligi<sup>81</sup>, Liming Liang<sup>30,75</sup>, Joshua Sampson<sup>1</sup>, Simon Crouch<sup>82</sup>, Juhyun Park<sup>83</sup>, Kari E. North<sup>84</sup>, Angela Cox<sup>85</sup>, John A. Snowden<sup>86</sup>, Josh Wright<sup>86</sup>,

Angel Carracedo<sup>87</sup>, Carlos Lopez-Otin<sup>88</sup>, Silvia Bea<sup>89</sup>, Itziar Salaverria<sup>89</sup>, David Martin<sup>89</sup>, Elias Campo<sup>89</sup>, Joseph F. Fraumeni Jr<sup>1</sup>, Silvia de Sanjose<sup>41,42,91</sup>, Henrik Hjalgrim<sup>31,91</sup>, James R. Cerhan<sup>26,91</sup>, Stephen J. Chanock<sup>1,91</sup>, Nathaniel Rothman<sup>1,91</sup>, and Susan L. Slager<sup>26,91</sup>

## Affiliations

<sup>1</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA <sup>2</sup>Division of Environmental Health Sciences, University of California Berkeley School of Public Health, Berkeley, California, USA <sup>3</sup>School of Public Health, University of Alabama, Birmingham, Alabama, USA <sup>4</sup>Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA <sup>5</sup>Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA <sup>6</sup>Center of Chronic Immunodeficiency, University Medical Center Freiburg, Freiburg, Hesse, Germany <sup>7</sup>Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA <sup>8</sup>Department of Preventive Medicine, USC Keck School of Medicine, University of Southern California, Los Angeles, California, USA <sup>9</sup>U1018 EQ6, Institut National de la Santé et de la Recherche Médicale (INSERM), Villejuif Cedex, France <sup>10</sup>Registre des hémopathies malignes de la Gironde, Institut Bergonié, Bordeaux Cedex, France <sup>11</sup>Division of Cancer Etiology, City of Hope Beckman Research Institute, Duarte, California, USA <sup>12</sup>MRC-HPA Centre for Environment and Health, School of Public Health, Imperial College London, London, United Kingdom <sup>13</sup>Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA <sup>14</sup>Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada <sup>15</sup>Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada <sup>16</sup>Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA <sup>17</sup>Sydney School of Public Health, The University of Sydney, Sydney, New South Wales, Australia <sup>18</sup>Department of Public Health, Clinical and Molecular Medicine, University of Cagliari, Monserrato, Cagliari, Italy <sup>19</sup>Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut, USA <sup>20</sup>Cancer Epidemiology Centre, Cancer Council Victoria, Carlton, Victoria, Australia <sup>21</sup>Department of Population Health, New York University School of Medicine, New York, New York, USA <sup>22</sup>Health Studies Sector, Westat, Rockville, Maryland, USA <sup>23</sup>Division of Hematology, College of Medicine, Mayo Clinic, Rochester, Minnesota, USA <sup>24</sup>Department of Medicine, College of Medicine, Mayo Clinic, Rochester, Minnesota, USA <sup>25</sup>Division of General Internal Medicine, College of Medicine, Mayo Clinic, Rochester, Minnesota, USA <sup>26</sup>Health Sciences Research, College of Medicine, Mayo Clinic, Rochester, Minnesota, USA <sup>27</sup>Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden <sup>28</sup>Department of Oncology and Pathology, Karolinska Institutet, Karolinska University Hospital Solna, Stockholm, Sweden <sup>29</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden <sup>30</sup>Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA <sup>31</sup>Department of Epidemiology

Research, Division of Health Surveillance and Research, Statens Serum Institut, Copenhagen, Denmark <sup>32</sup>Department of Radiology, Oncology and Radiation Science, Uppsala University, Uppsala, Sweden <sup>33</sup>Center for Epidemiology and Computational Biology, Health Sciences, Exponent, Inc., Menlo Park, California, USA <sup>34</sup>Division of Epidemiology, Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California, USA <sup>35</sup>Department of Internal Medicine, Huntsman Cancer Institute, Salt Lake City, Utah, USA <sup>36</sup>George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, Utah, USA <sup>37</sup>Department of Internal Medicine, Carver College of Medicine, The University of Iowa, Iowa City, Iowa, USA <sup>38</sup>Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, California, USA <sup>39</sup>Division of Endocrinology, Diabetes and Metabolism, The Ohio State University, Columbus, Ohio, USA <sup>40</sup>Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA <sup>41</sup>Unit of Infections and Cancer (UNIC), Cancer Epidemiology Research Programme, Institut Catala d' Oncologia, IDIBELL, Barcelona, Spain <sup>42</sup>Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain <sup>43</sup>Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Baden-Württemberg, Germany <sup>44</sup>The Tisch Cancer Institute, Mount Sinai School of Medicine, New York, New York, USA <sup>45</sup>Group of Genetic Epidemiology, Section of Genetics, International Agency for Research on Cancer, Lyon, France <sup>46</sup>Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic <sup>47</sup>EA 4184, Registre des Hémopathies Malignes de Côte d' Or, University of Burgundy and Dijon University Hospital, Dijon, France <sup>48</sup>Genetic Cancer susceptibility Group, Section of Genetics, International Agency for Research on Cancer, Lyon, France <sup>49</sup>School of Nursing and Human Sciences, Dublin City University, Dublin, Leinster, Ireland <sup>50</sup>Department of Epidemiology, M.D. Anderson Cancer Center, Houston, Texas, USA <sup>51</sup>Department of Medicine, Duke University and VA Medical Centers, Durham, North Carolina, USA <sup>52</sup>Division of Epidemiology/Clinical Research, University of Minnesota, Minneapolis, Minnesota, USA <sup>53</sup>Division of Hematology/Oncology, College of Medicine, Mayo Clinic, Phoenix, Arizona, USA <sup>54</sup>Department of Laboratory Medicine and Pathology, College of Medicine, Mayo Clinic, Rochester, Minnesota, USA <sup>55</sup>Minneapolis Veterans Affairs Medical Center, Minneapolis, Minnesota, USA <sup>56</sup>Department of Family Medicine and Public Health Sciences, Wayne State University, Detroit, Michigan, USA <sup>57</sup>School of Public Health, Imperial College London, London, United Kingdom <sup>58</sup>Human Genetics Foundation, Turin, Italy <sup>59</sup>Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece <sup>60</sup>Hellenic Health Foundation, Athens, Greece <sup>61</sup>Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute (ISPO), Florence, Italy <sup>62</sup>Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, Breivika, Norway <sup>63</sup>Cancer Registry of Norway, Oslo, Norway <sup>64</sup>Folkhalsan Research Center, Samfundet Folkhalsan, Helsinki, Finland <sup>65</sup>Department of Epidemiology, Murcia Regional Health Authority, Murcia, Spain <sup>66</sup>Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The

Netherlands <sup>67</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands <sup>68</sup>Cancer Epidemiology Unit, University of Oxford, Oxford, United Kingdom <sup>69</sup>Department of Chronic Disease and Prevention, National Institute for Health and Welfare, Helsinki, Finland <sup>70</sup>Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut, USA <sup>71</sup>Program in Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, New York, USA <sup>72</sup>Cancer Control Research, BC Cancer Agency, Vancouver, British Columbia, Canada <sup>73</sup>Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, USA <sup>74</sup>Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA <sup>75</sup>Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA <sup>76</sup>Pathology, Australian School of Advanced Medicine, Macquarie University, Sydney, New South Wales, Australia <sup>77</sup>Department of Histopathology, Douglass Hanly Moir Pathology, Macquarie Park, New South Wales, Australia <sup>78</sup>Prince of Wales Clinical School, University of New South Wales, Sydney, New South Wales, Australia <sup>79</sup>Department of Biomedical Science, University of Cagliari, Monserrato, Cagliari, Italy <sup>80</sup>Interdisciplinary Department of Medicine, University of Bari, Bari, Italy <sup>81</sup>Environmental and Occupational Epidemiology Unit, Cancer Prevention and Research Institute (ISPO), Florence, Italy <sup>82</sup>Epidemiology and Genetics Unit, Department of Health Sciences, University of York, York, United Kingdom <sup>83</sup>Dongguk University-Seoul, Seoul, South Korea <sup>84</sup>Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA <sup>85</sup>Department of Oncology, University of Sheffield, Sheffield, UK <sup>86</sup>Department of Oncology, University of Sheffield and Department of Haematology, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK <sup>87</sup>Genomic Medicine Group CIBERER, University of Santiago de Compostela, Santiago de Compostela, Spain <sup>88</sup>Department of Biochemistry and Molecular Biology, Institute of Oncology, University of Oviedo, Oviedo, Spain <sup>89</sup>August Pi I Sunyer Biomedical Research Institute (IDIBAPS), Hospital Clinic of Barcelona, University of Barcelona, Barcelona, Spain

## Acknowledgments

We thank C. Allmer, E. Angelucci, A. Bigelow, I. Brock, K. Butterbach, A. Chabrier, D. Chan-Lam, J.M. Connors, D. Connley, M. Cornelis, K. Corsano, C. Dalley, D. Cox, H. Cramp, R. Cutting, H. Dykes, L. Ershler, A. Gabbas, R.P. Gallagher, R.D. Gascoyne, P. Hui, L. Irish, L. Jacobus, S. Kaul, J. Lunde, M. McAdams, R. Montalvan, M. Rais, T. Rattle, L. Rigacci, K. Snyder, G. Specchia, M. Stagner, P. Taylor, G. Thomas, C. Tornow, G. Wood, M. Yang, M. Zucca. The overall GWAS project was supported by the intramural program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, U.S. National Institutes of Health. A full list of acknowledgements is provided in the Supplementary Note.

## References

1. Albright F, Teerlink C, Werner TL, Cannon-Albright LA. Significant evidence for a heritable contribution to cancer predisposition: a review of cancer familiality by site. *BMC cancer*. 2012; 12:138. [PubMed: 22471249]

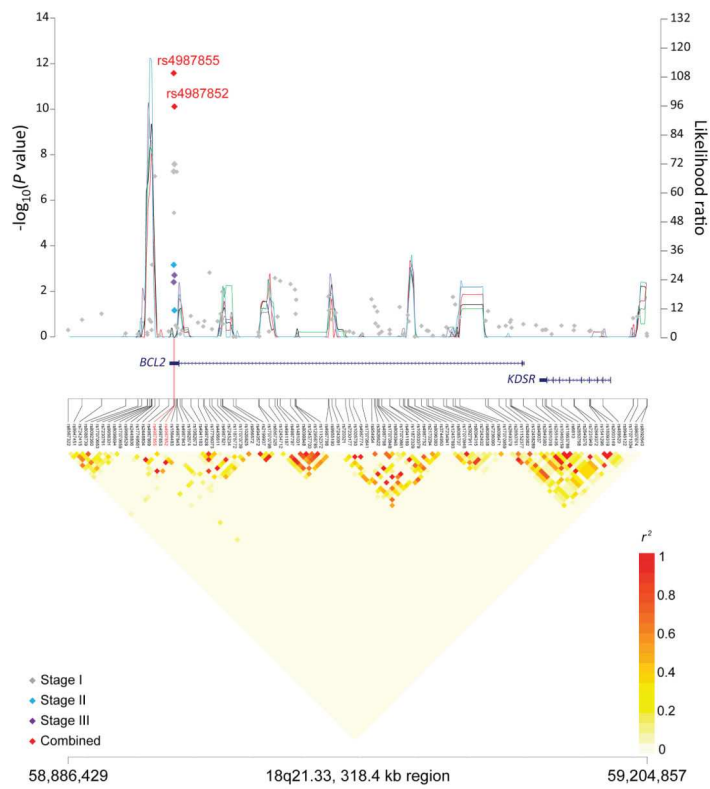
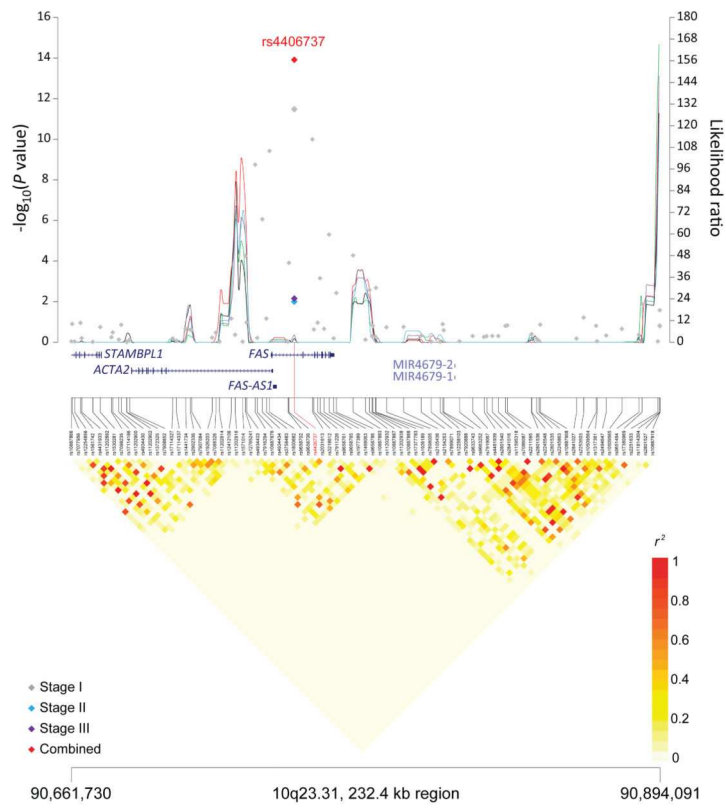
2. Goldin LR, Bjorkholm M, Kristinsson SY, Turesson I, Landgren O. Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia. *Haematologica*. 2009; 94:647– 53. [PubMed: 19286886]
3. Di Bernardo MC, et al. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nature genetics*. 2008; 40:1204– 10. [PubMed: 18758461]
4. Crowther-Swanepoel D, et al. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. *Nature genetics*. 2010; 42:132– 6. [PubMed: 20062064]
5. Slager SL, et al. Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. *Blood*. 2011; 117:1911– 6. [PubMed: 21131588]
6. Slager SL, et al. Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia. *Blood*. 2012
7. Park JH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature genetics*. 2010; 42:570– 5. [PubMed: 20562874]
8. Wang Z, et al. Improved imputation of common and uncommon SNPs with a new reference set. *Nature genetics*. 2012; 44:6– 7. [PubMed: 22200770]
9. Conde L, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nature genetics*. 2010; 42:661– 4. [PubMed: 20639881]
10. Consortium GP. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061– 1073. [PubMed: 20981092]
11. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics*. 2009; 5:e1000529. [PubMed: 19543373]
12. Crowther-Swanepoel D, et al. Common genetic variation at 15q25.2 impacts on chronic lymphocytic leukaemia risk. *British journal of haematology*. 2011; 154:229– 33. [PubMed: 21554262]
13. Rafnar T, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nature genetics*. 2009; 41:221– 7. [PubMed: 19151717]
14. Shete S, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nature genetics*. 2009; 41:899– 904. [PubMed: 19578367]
15. McKay JD, et al. Lung cancer susceptibility locus at 5p15.33. *Nature genetics*. 2008; 40:1404– 6. [PubMed: 18978790]
16. Wang Y, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nature genetics*. 2008; 40:1407– 9. [PubMed: 18978787]
17. Petersen GM, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nature genetics*. 2010; 42:224– 8. [PubMed: 20101243]
18. Haiman CA, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nature genetics*. 2011; 43:1210– 4. [PubMed: 22037553]
19. Kote-Jarai Z, et al. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nature genetics*. 2011; 43:785– 91. [PubMed: 21743467]
20. Sheng X, et al. TERT polymorphisms modify the risk of acute lymphoblastic leukemia in Chinese children. *Carcinogenesis*. 2012
21. Yang J, et al. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics*. 2010; 42:565– 9. [PubMed: 20562875]
22. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *American journal of human genetics*. 2011; 88:294– 305. [PubMed: 21376301]
23. Fisher GH, et al. Dominant interfering Fas gene mutations impair apoptosis in a human autoimmune lymphoproliferative syndrome. *Cell*. 1995; 81:935– 46. [PubMed: 7540117]
24. Drappa J, Vaishnav AK, Sullivan KE, Chu JL, Elkon KB. Fas gene mutations in the Canale-Smith syndrome, an inherited lymphoproliferative disorder associated with autoimmunity. *The New England journal of medicine*. 1996; 335:1643– 9. [PubMed: 8929361]

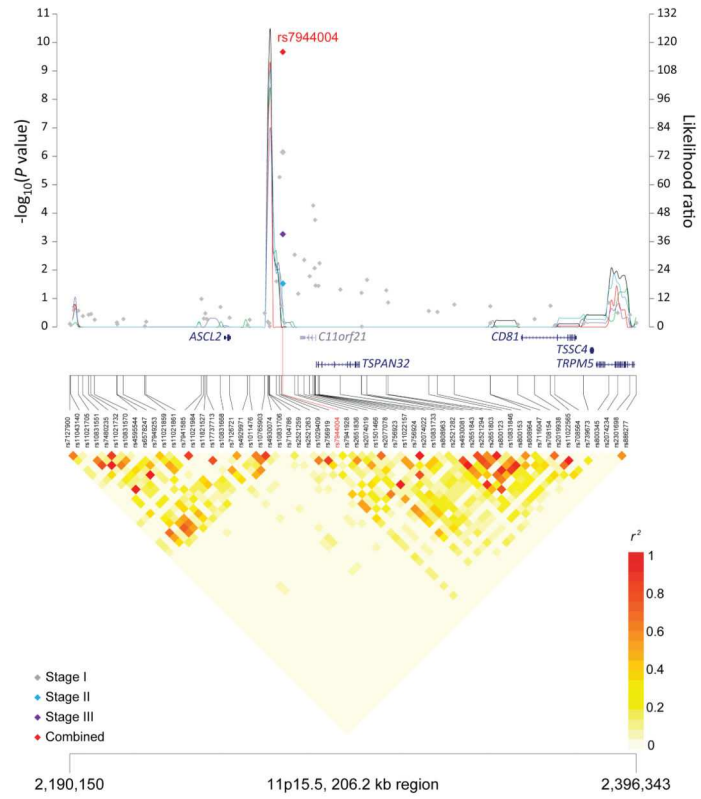


25. Straus SE, et al. The development of lymphomas in families with autoimmune lymphoproliferative syndrome with germline Fas mutations and defective lymphocyte apoptosis. *Blood*. 2001; 98:194–200. [PubMed: 11418480]
26. Baseggio L, et al. In non-follicular lymphoproliferative disorders, IGH/BCL2-fusion is not restricted to chronic lymphocytic leukaemia. *British journal of haematology*. 2012; 158:489– 498. [PubMed: 22686190]
27. Cleary ML, Smith SD, Sklar J. Cloning and structural analysis of cDNAs for bcl-2 and a hybrid bcl-2/immunoglobulin transcript resulting from the t(14;18) translocation. *Cell*. 1986; 47:19– 28. [PubMed: 2875799]
28. Reshmi G, et al. C-T variant in a miRNA target site of BCL2 is associated with increased risk of human papilloma virus related cervical cancer--an in silico approach. *Genomics*. 2011; 98:189– 93. [PubMed: 21704150]
29. Cheung VG, et al. Polymorphic cis- and trans-regulation of human gene expression. *PLoS biology*. 2010; 8
30. Strasser A, Harris AW, Cory S. E mu-bcl-2 transgene facilitates spontaneous transformation of early pre-B and immunoglobulin-secreting cells but not T cells. *Oncogene*. 1993; 8:1– 9. [PubMed: 8423986]
31. Wensveen FM, et al. BH3-only protein Noxa regulates apoptosis in activated B cells and controls high-affinity antibody formation. *Blood*. 2012; 119:1440– 9. [PubMed: 22144184]
32. Smit LA, et al. Differential Noxa/Mcl-1 balance in peripheral versus lymph node chronic lymphocytic leukemia cells correlates with survival capacity. *Blood*. 2007; 109:1660– 8. [PubMed: 17038534]
33. Morales AA, et al. Expression and transcriptional regulation of functionally distinct Bmf isoforms in B-chronic lymphocytic leukemia cells. *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U K*. 2004; 18:41– 7.
34. Labi V, et al. Loss of the BH3-only protein Bmf impairs B cell homeostasis and accelerates gamma irradiation-induced thymic lymphoma development. *The Journal of experimental medicine*. 2008; 205:641– 55. [PubMed: 18299399]
35. Crowther-Swanepoel D, et al. Verification that common variation at 2q37.1, 6p25.3, 11q24.1, 15q23, and 19q13.32 influences chronic lymphocytic leukaemia risk. *British journal of haematology*. 2010; 150:473– 9. [PubMed: 20553269]
36. Lan Q, et al. Genetic susceptibility for chronic lymphocytic leukemia among Chinese in Hong Kong. *European journal of haematology*. 2010; 85:492– 5. [PubMed: 20731705]
37. Nieters A, et al. PRRC2A and BCL2L11 gene variants influence risk of non-Hodgkin lymphoma: results from the InterLymph consortium. *Blood*. 2012
38. Egle A, Harris AW, Bouillet P, Cory S. Bim is a suppressor of Myc-induced mouse B cell leukemia. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101:6164– 9. [PubMed: 15079075]
39. Kelly JL, et al. Germline variation in apoptosis pathway genes and risk of non-Hodgkin' s lymphoma. *Cancer epidemiology, biomarkers & prevention*. 2010; 19:2847– 58.
40. Cox A, et al. A common coding variant in CASP8 is associated with breast cancer risk. *Nature genetics*. 2007; 39:352– 8. [PubMed: 17293864]
41. Abnet CC, et al. Genotypic variants at 2q33 and risk of esophageal squamous cell carcinoma in China: a meta-analysis of genome-wide association studies. *Human molecular genetics*. 2012; 21:2132– 41. [PubMed: 22323360]
42. Barrett JH, et al. Genome-wide association study identifies three new melanoma susceptibility loci. *Nature genetics*. 2011; 43:1108– 13. [PubMed: 21983787]
43. Enjuanes A, et al. Genetic variants in apoptosis and immunoregulation-related genes are associated with risk of chronic lymphocytic leukemia. *Cancer research*. 2008; 68:10178– 86. [PubMed: 19074885]
44. Lan Q, et al. Genetic variants in caspase genes and susceptibility to non-Hodgkin lymphoma. *Carcinogenesis*. 2007; 28:823– 7. [PubMed: 17071630]
45. Reya T, et al. A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature*. 2003; 423:409– 14. [PubMed: 12717450]



46. Gutierrez A Jr, et al. LEF-1 is a prosurvival factor in chronic lymphocytic leukemia and is expressed in the preleukemic state of monoclonal B-cell lymphocytosis. *Blood*. 2010; 116:2975–83. [PubMed: 20595513]
47. Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B. Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. *PLoS genetics*. 2010; 6:e1000899. [PubMed: 20386740]
48. Koufos A, et al. Familial Wiedemann-Beckwith syndrome and a second Wilms tumor locus both map to 11p15.5. *American journal of human genetics*. 1989; 44:711– 9. [PubMed: 2539717]
49. Weksberg R, Shuman C, Beckwith JB. Beckwith-Wiedemann syndrome. *European journal of human genetics*. 2010; 18:8– 14. [PubMed: 19550435]
50. Morton LM, et al. Proposed classification of lymphoid neoplasms for epidemiologic research from the Pathology Working Group of the International Lymphoma Epidemiology Consortium (InterLymph). *Blood*. 2007; 110:695– 708. [PubMed: 17389762]
51. Turner JJ, et al. InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): update and future directions. *Blood*. 2010; 116:e90– 8. [PubMed: 20699439]
52. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155:945– 59. [PubMed: 10835412]
53. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*. 2006; 38:904– 9. [PubMed: 16862161]
54. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research*. 2012; 40:D930– 4. [PubMed: 22064851]
55. Dixon AL, et al. A genome-wide association study of global gene expression. *Nature genetics*. 2007; 39:1202– 7. [PubMed: 17873877]
56. Stranger BE, et al. Population genomics of human gene expression. *Nature genetics*. 2007; 39:1217– 24. [PubMed: 17873874]
57. Pharoah PD, et al. Polygenic susceptibility to breast cancer and implications for prevention. *Nature genetics*. 2002; 31:33– 6. [PubMed: 11984562]
58. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics*. 2011; 88:76– 82. [PubMed: 21167468]
59. Fearnhead P. SequenceLDhot: detecting recombination hotspots. *Bioinformatics*. 2006; 22:3061– 6. [PubMed: 17060358]
60. Fearnhead P, Harding RM, Schneider JA, Myers S, Donnelly P. Application of coalescent methods to reveal fine-scale rate variation and recombination hotspots. *Genetics*. 2004; 167:2067– 81. [PubMed: 15342541]
61. Li N, Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*. 2003; 165:2213– 33. [PubMed: 14704198]
62. Crawford DC, et al. Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nature genetics*. 2004; 36:700– 6. [PubMed: 15184900]
63. Luna A, Nicodemus KK. snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. *Bioinformatics*. 2007; 23:774– 6. [PubMed: 17234637]



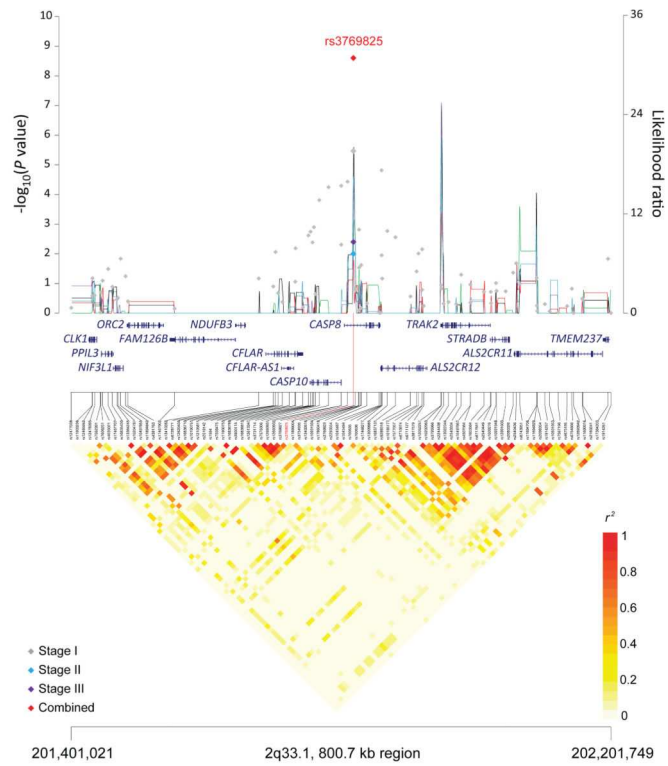
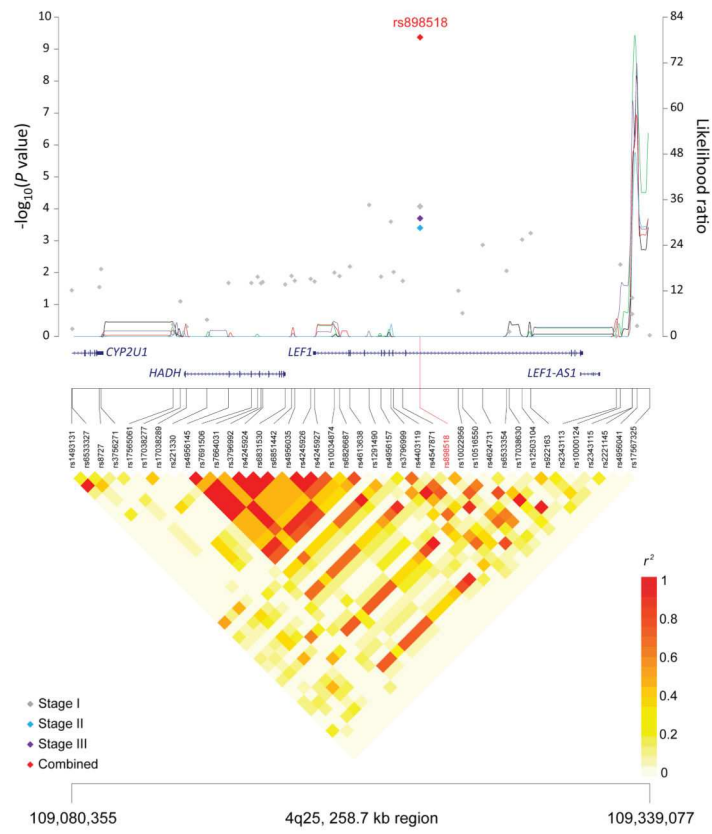


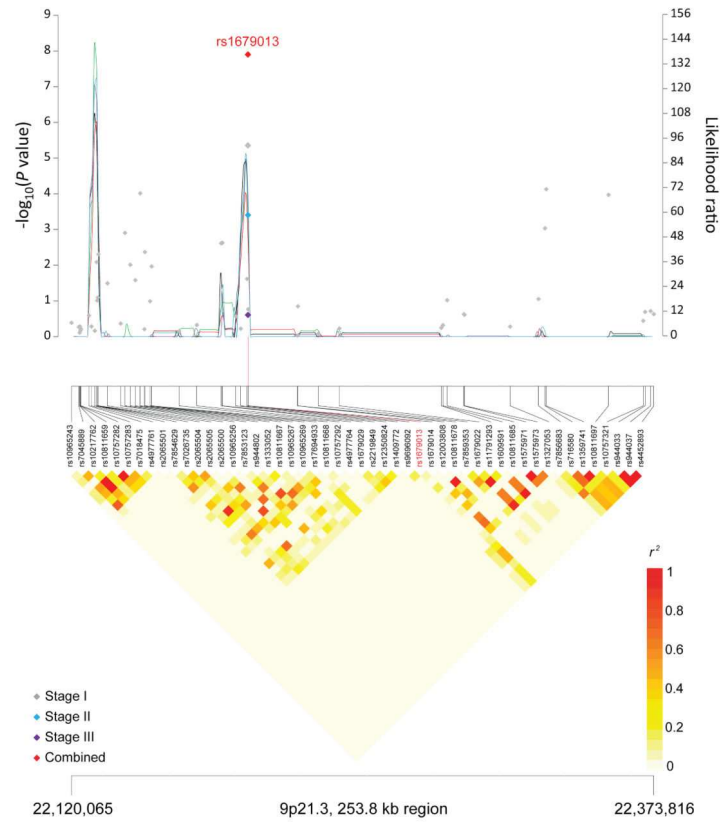
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



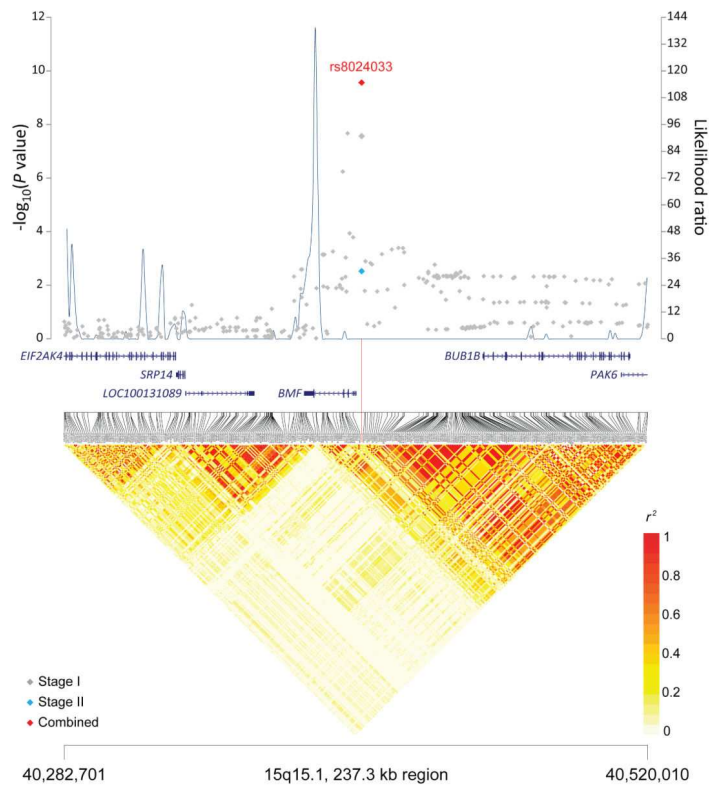
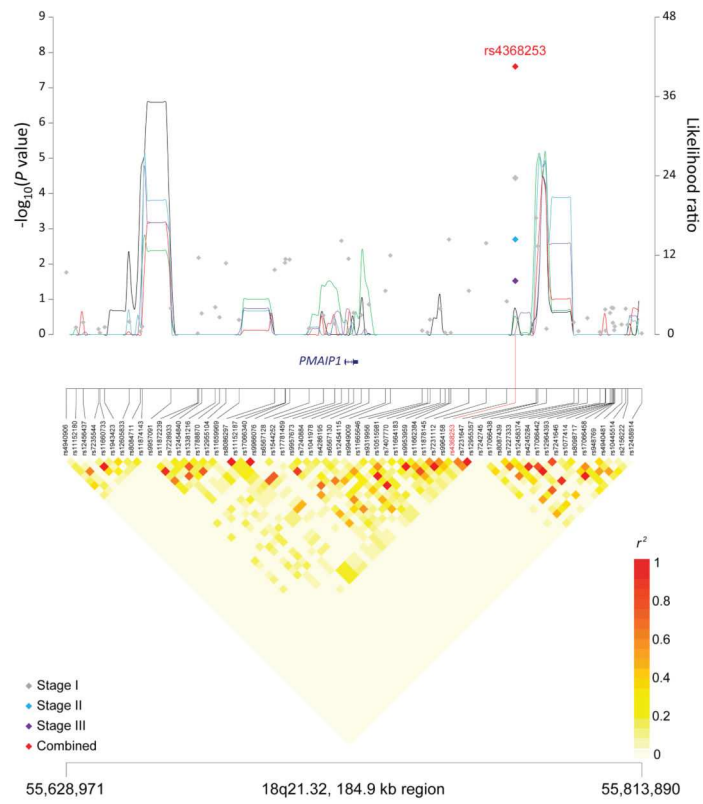


Author Manuscript

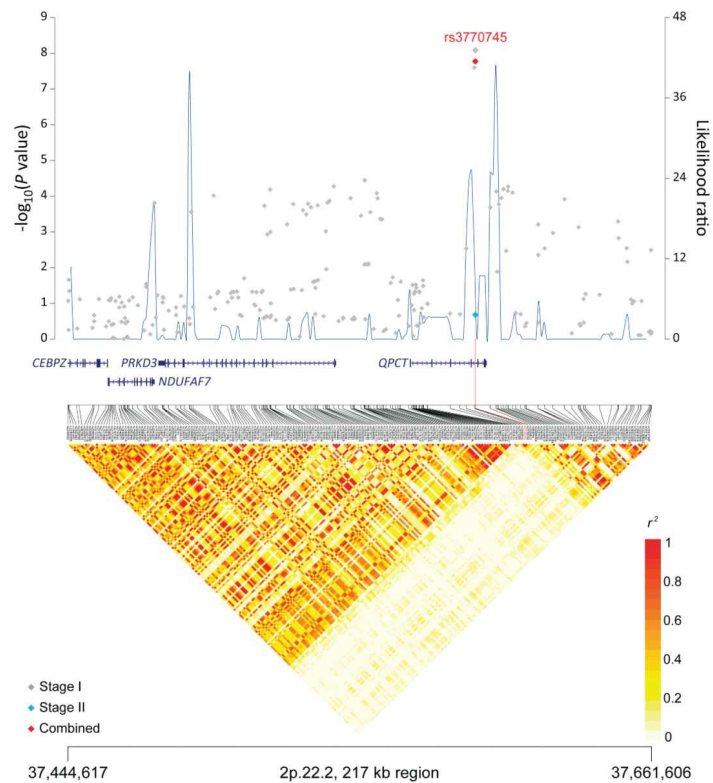
Author Manuscript

Author Manuscript

Author Manuscript

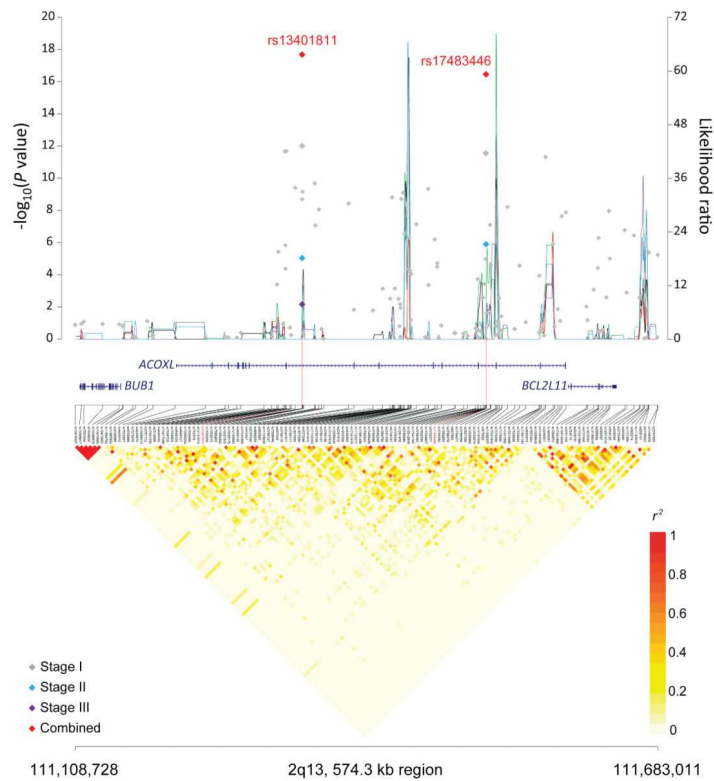






**Figure 1. Association results, recombination hot-spots, and linkage disequilibrium (LD) plots for the regions newly associated with CLL**

Top, association results of GWAS data from Stage 1 NHL-GWAS (grey diamonds), Stage 2 combined data (blue diamond), Stage 3 combined data (purple diamond), and Stages 1– 3 combined data (red diamond) are shown in the top panel with  $-\log_{10}(P)$  values (left y axis). Overlaid are the likelihood ratio statistics (right y axis) to estimate putative recombination hotspots across the region on the basis of 5 unique sets of 100 randomly selected control samples. Bottom, LD heatmap based on  $r^2$  values from total control populations for all SNPs included in the GWAS. (a) 10q23.31 region; (b) 18q21.33 region; (c) 11p15.5 region; (d) 4q25 region; (e) 2q33.1 region; (f) 9p21.3 region; (g) 18q21.32 region; (h) 15q15.1 region; (i) 2p22.2 region.



**Figure 2. Association results, recombination hot-spots, and linkage disequilibrium (LD) plot for the new independent CLL susceptibility SNP in the 2q13 established locus**  
 Top, association results of GWAS data from Stage 1 NHL-GWAS (grey diamonds), Stage 2 combined data (blue diamond), Stage 3 combined data (purple diamond), and Stages 1– 3 combined data (red diamond) are shown in the top panel with  $-\log_{10}(P)$  values (left y axis). Overlaid are the likelihood ratio statistics (right y axis) to estimate putative recombination hotspots across the region on the basis of 5 unique sets of 100 randomly selected control samples. Bottom, LD heatmap based on  $r^2$  values from total control populations for all SNPs included in the GWAS.

**Table 1**

Association results for novel loci and new independent SNPs

Chr	Nearest gene(s)	SNP	Position	Risk	Other	RAF	Stage	OR (95% CI)	p
				allele <sup>a</sup>	allele				
<b>Novel loci</b>									
10q23.31	<i>ACTA2, FAS</i>	rs 4406737	90,749,704	G	A	0.57	Stage 1	1.30 (1.21– 1.40)	$3.30 \times 10^{-12}$
							Stage 2	1.17 (1.03– 1.32)	0.01
							Stage 3	1.27 (1.06– 1.52)	0.007
							<b>Combined<sup>b</sup></b>	<b>1.27 (1.19–1.33)</b>	<b><math>1.22 \times 10^{-14}</math></b>
18q21.33	<i>BCL2</i>	rs4987855*	58,944,529	G	A	0.91	Stage 1	1.47 (1.28– 1.69)	$5.51 \times 10^{-8}$
							Stage 2	1.47 (1.18– 1.85)	0.0007
							Stage 3	1.43 (1.12– 1.82)	0.004
							<b>Combined<sup>b</sup></b>	<b>1.47 (1.32–1.61)</b>	<b><math>2.66 \times 10^{-12}</math></b>
		rs 4987852	58,944,901	G	A	0.06	Stage 1	1.43 (1.26– 1.63)	$2.67 \times 10^{-8}$
							Stage 2	1.24 (0.98– 1.56)	0.07
							Stage 3	1.52 (1.17– 1.97)	0.002
							<b>Combined<sup>b</sup></b>	<b>1.41 (1.27–1.56)</b>	<b><math>7.76 \times 10^{-11}</math></b>
11p15.5	<i>C11orf21, TSPAN32</i>	rs 7944004	2,267,728	T	G	0.49	Stage 1	1.19 (1.11– 1.28)	$7.20 \times 10^{-7}$
							Stage 2	1.15 (1.02– 1.32)	0.03
							Stage 3	1.27 (1.11– 1.45)	0.0006
							<b>Combined<sup>b</sup></b>	<b>1.20 (1.13–1.27)</b>	<b><math>2.15 \times 10^{-10}</math></b>
4q25	<i>LEF1</i>	rs898518*	109,236,273	A	C	0.59	Stage 1	1.16 (1.08– 1.24)	$8.47 \times 10^{-5}$
							Stage 2	1.26 (1.11– 1.43)	0.0004
							Stage 3	1.30 (1.14– 1.49)	0.0002
							<b>Combined<sup>b</sup></b>	<b>1.20 (1.14–1.27)</b>	<b><math>4.24 \times 10^{-10}</math></b>
2q33.1	<i>CASP10, CASP8</i>	rs 3769825	201,819,625	T	C	0.45	Stage 1	1.18 (1.10– 1.27)	$3.43 \times 10^{-6}$
							Stage 2	1.16 (1.03– 1.32)	0.01
							Stage 3	1.22 (1.07– 1.40)	0.004

Chr	Nearest gene(s)	SNP	Position	Risk	Other	RAF	Stage	OR (95% CI)	p
				allele <sup>a</sup>	allele				
							Combined <sup>b</sup>	<b>1.19 (1.12–1.25)</b>	<b>2.50 × 10<sup>-9</sup></b>
9p21.3	<i>CDKN2B-AS1</i>	rs 1679013	22,196,987	C	T	0.52	Stage 1	1.18 (1.10– 1.27)	4.47×10 <sup>-6</sup>
							Stage 2	1.32 (1.12– 1.52)	0.0004
							Stage 3	1.11 (0.93– 1.32)	0.25
							Combined <sup>b</sup>	<b>1.19 (1.12–1.27)</b>	<b>1.27 × 10<sup>-8</sup></b>
18q21.32	<i>PMAIP1</i>	rs 4368253	55,773,267	C	T	0.69	Stage 1	1.18 (1.09– 1.27)	3.65×10 <sup>-5</sup>
							Stage 2	1.24 (1.08– 1.41)	0.002
							Stage 3	1.18 (1.02– 1.37)	0.03
							Combined <sup>b</sup>	<b>1.19 (1.12–1.27)</b>	<b>2.51 × 10<sup>-8</sup></b>
15q15.1	<i>BMF</i>	rs8024033 <sup>†</sup>	38,190,949	C	G	0.51	Stage 1	1.22 (1.14– 1.32)	2.72×10 <sup>-8</sup>
							Stage 2	1.22 (1.08– 1.39)	0.003
							Stage 3	-	-
							Combined <sup>b</sup>	<b>1.22 (1.15–1.30)</b>	<b>2.71 × 10<sup>-10</sup></b>
2p22.2	<i>QPCT, PRKD3</i>	rs3770745 <sup>†</sup>	37,449,593	T	C	0.22	Stage 1	1.29 (1.18– 1.40)	8.23×10 <sup>-9</sup>
							Stage 2	1.10 (0.95– 1.28)	0.21
							Stage 3	-	-
							Combined <sup>b</sup>	<b>1.24 (1.15–1.33)</b>	<b>1.68 × 10<sup>-8</sup></b>
<b>New independent SNP in established locus</b>									
2q13	<i>ACOXL, BCL2L11</i>	rs13401811 <sup>*</sup>	111,332,575	G	A	0.81	Stage 1	1.43 (1.28– 1.56)	9.76×10 <sup>-13</sup>
							Stage 2	1.45 (1.23– 1.72)	9.39×10 <sup>-6</sup>
							Stage 3	1.32 (1.08– 1.59)	0.007
							Combined <sup>b</sup>	<b>1.41 (1.30–1.52)</b>	<b>2.08 × 10<sup>-18</sup></b>

<sup>a</sup>The risk allele is the allele corresponding to the estimated odds ratio; RAF= risk allele frequency in controls; OR= per allele odds ratio adjusted for age, sex and significant principal components.

<sup>b</sup>Number of cases and controls in the joint analysis of stage 1+stage2+stage3: rs4406737 (3,481/12,170), 20 rs4987855 (3,883/12,446), rs4987852 (3,880/12,497), rs7944004 (3,869/12,476), rs898518 (3,879/12,441), rs3769825 (3,885/12,471), rs1679013 (3,482/12,148), rs4368253 (3,882/12,473), rs8024033 (3096/7663), rs3770745 (3097/7663), rs13401811 (3,839/12,264).

<sup>\*</sup>For the ICGC study in stage 3, results for proxy SNPs were provided (rs4987856/rs4987855,  $r^2=1.0$ ; rs7698317/rs898518,  $r^2=1.0$ ; rs1554005/rs13401811,  $r^2=1.0$ ).

Author Manuscript

Author Manuscript

Author Manuscript

<sup>†</sup>Identified from the 1000 Genomes meta-analysis of stage 1 and stage 2 with imputation information >0.9 in the NHL-GWAS.

Author Manuscript

Table 2

Conditional analyses for select SNPs

New SNP	Chr	Position	Nearest gene	OR <sup>a</sup>	P <sup>a</sup>	Conditional	Conditional	Established	r <sup>2</sup> *	OR <sup>c</sup>	P <sup>c</sup>	Conditional	Conditional
						OR <sup>b</sup>	P <sup>b</sup>	SNP				OR <sup>d</sup>	P <sup>d</sup>
rs13401811	2q13	111,332,575	<i>ACOXL, BCL2L11</i>	1.43	6.09×10 <sup>-17</sup>	1.35	1.60×10 <sup>-12</sup>	rs17483466	0.02	1.37	3.53×10 <sup>-17</sup>	1.31	6.70×10 <sup>-13</sup>
rs7578199	2q37.3	241,841,521	<i>HDLBP, FARP2</i>	1.20	5.39×10 <sup>-7</sup>	1.19	6.10×10 <sup>-6</sup>	rs757978	0.01	1.29	1.35×10 <sup>-7</sup>	1.26	2.37×10 <sup>-6</sup>
rs9273363	6p21.32	32,734,250	<i>HLA</i>	1.24	2.24×10 <sup>-10</sup>	1.24	3.50×10 <sup>-9</sup>	rs674313	0.21	1.13	5.00×10 <sup>-4</sup>	1.06	0.11
rs9273363	6p21.32	32,734,250	<i>HLA</i>	1.24	2.24×10 <sup>-10</sup>	1.23	3.14×10 <sup>-9</sup>	rs9272535	0.11	1.18	7.60×10 <sup>-6</sup>	1.12	0.002
rs11636802	15q21.3	54,562,889	<i>MNS1</i>	1.41	1.68×10 <sup>-13</sup>	1.38	1.54×10 <sup>-9</sup>	rs7169431	0.16	1.27	1.72×10 <sup>-5</sup>	1.06	0.32
rs35748167	18q21.32	56,188,413	<i>PMAIP1, MC4R</i>	1.32	9.31×10 <sup>-9</sup>	1.25	7.89×10 <sup>-7</sup>	rs4368253 <sup>e</sup>	0.003	1.19	2.82×10 <sup>-7</sup>	1.18	5.76×10 <sup>-7</sup>
rs4987852	18q21.33	58,944,901	<i>BCL2</i>	1.41	7.76×10 <sup>-11</sup>	1.36	1.50×10 <sup>-8</sup>	rs4987855 <sup>e</sup>	0.01	1.47	2.66×10 <sup>-12</sup>	1.41	1.33×10 <sup>-10</sup>

\* r<sup>2</sup> linkage disequilibrium is based on 1000 Genomes Project and is between the new SNP and established SNP in the locus

<sup>a</sup> OR per allele odds ratio and *P* for the new SNP from the unconditional meta-analysis based on stage 1 + 2 for all loci, except 18q21.33. Data from stages 1– 3 was used for 18q21.33.

<sup>b</sup> OR and *P* for the new SNP from the conditional meta-analysis

<sup>c</sup> OR and *P* for the established SNP from the unconditional meta-analysis

<sup>d</sup> OR and *P* for the established SNP from the conditional meta-analysis

<sup>e</sup> SNP discovered and confirmed in the current study