



HAL
open science

Verification that common variation at 2q37.1, 6p25.3, 11q24.1, 15q23, and 19q13.32 influences chronic lymphocytic leukaemia risk

Dalemari Crowther-Swanepoel, Mahmoud Mansouri, Anna Enjuanes, Ana Vega, Karin Ekström Smedby, Clara Ruiz-Ponte, Jesper Jurlander, Gunnar Juliusson, Emili Montserrat, Daniel Catovsky, et al.

► To cite this version:

Dalemari Crowther-Swanepoel, Mahmoud Mansouri, Anna Enjuanes, Ana Vega, Karin Ekström Smedby, 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, 10.1111/j.1365-2141.2010.08270.x . hal-00552600

HAL Id: hal-00552600

<https://hal.science/hal-00552600>

Submitted on 6 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Verification that common variation at 2q37.1, 6p25.3, 11q24.1, 15q23, and 19q13.32 influences chronic lymphocytic leukaemia risk

Journal:	<i>British Journal of Haematology</i>
Manuscript ID:	BJH-2010-00494.R1
Manuscript Type:	Short Reports
Date Submitted by the Author:	30-Apr-2010
Complete List of Authors:	<p>Crowther-Swanepoel, Dalemari; Insitute of Cancer Research, Cancer Genetics Mansouri, Mahmoud; Uppsala University, Genetics and Pathology Enjuanes, Anna; University of Barcelona, Hematopathology Unit Vega, Ana; University of Santiago de Compostela, Genomic Medicine Group Smedby, Karin; Clinical Epidemiology, Department of Medicine Ruiz-Ponte, Clara; University of Santiago de Compostela, Genomic Medicine Group Jurlander, Jesper; Leukemia Laboratory, Department of Hematology Juliusson, Gunnar; Lund Strategic Research Center for Stem Cell Biology and Cell Therapy, Hematology and Transplantation Montserrat, Emili; University Barcelona, Department of Haematology Catovsky, Daniel; Institute of Cancer research, Haemato-oncology Campo, Elias; Hospital Clínic of Barcelona, Department of Pathology Carracedo, Angel; University of Santiago de Compostela, Genomic Medicine Group Rosenquist, Richard; Uppsala University, Dept of Genetics and Pathology Houlston, Richard; Haddow Laboratories, 15 Cotswold Road</p>
Key Words:	CLL, GENES, GENETICS, Genome wide association, risk

1
2
3 **Verification that common variation at 2q37.1, 6p25.3, 11q24.1, 15q23, and 19q13.32 influences**
4 **chronic lymphocytic leukaemia risk**
5
6
7
8

9
10 Dalemari Crowther-Swanepoel¹, Mahmoud Mansouri², Anna Enjuanes³, Ana Vega⁴, Karin E
11 Smedby⁵, Clara Ruiz-Ponte⁴, Jesper Jurlander⁶, Gunnar Juliusson⁷, Emilio Montserrat⁸, Daniel
12 Catovsky⁹, Elias Campo³, Angel Carracedo⁴, Richard Rosenquist², Richard S Houlston^{1*}
13
14

- 15
16 1. Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey. UK
17
18 2. Department of Genetics and Pathology, Uppsala University, Uppsala, Sweden
19
20 3. Hematopathology Unit, Department of Anatomic Pathology, Hospital Clinic, University of
21 Barcelona, Spain
22
23 4. Genomic Medicine Group, University of Santiago de Compostela and Galician Foundation of
24 Genomic Medicine, CIBERER, Santiago de Compostela, Spain
25
26 5. Unit of Clinical Epidemiology, Dept of Medicine, Karolinska Institutet, Stockholm, Sweden
27
28 6. Department of Hematology, Leukaemia Laboratory, Rigshospitalet, Copenhagen, Denmark
29
30 7. Lund Strategic Research Center for Stem Cell Biology and Cell Therapy, Hematology and
31 Transplantation, Lund University, Lund, Sweden
32
33 8. Department of Hematology, Institut d'Investigacions Biomediques August Pi i Sunyer, Hospital
34 Clinic, University of Barcelona, Spain
35
36 9. Section of Haemato-oncology, Institute of Cancer Research, Sutton, Surrey. UK
37

38
39 *Corresponding author

40 Richard Houlston

41 Institute of Cancer Research

42 15 Cotswold Rd, Sutton, Surrey SM2 5NG, UK

43
44
45 Tel: +44-(0)-208-722-4175; Fax: +44-(0)-208-722-4359; e-mail: richard.houlston@icr.ac.uk
46
47

48 **Key words:** CLL, risk, genotype, genome wide association

49 **Running title:** Verification of CLL risk loci
50
51

52
53 Word count: summary (115), text (1,494)
54
55
56
57
58
59
60

SUMMARY

A recent genome wide association study of chronic lymphocytic leukaemia (CLL) has provided evidence that common variation at 2q13 (rs17483466), 2q37.1 (rs13397985), 6p25.3 (rs872071), 11q24.1 (rs735665), 15q23 (rs7176508) and 19q13.32 (rs11083846) affects CLL risk. To verify and further explore the relationship between these variants and CLL risk we genotyped case-control datasets from Spain and Sweden totalling 824 cases and 850 controls. Combined data provided statistically significant support for an association between genotypes at rs13397985, rs872071, rs735665, rs7176508 and rs11083846 and CLL risk. CLL risk was shown to increase with increasing numbers of risk alleles ($P_{\text{trend}}=1.40 \times 10^{-15}$), consistent with a polygenic model of disease susceptibility. These data validate the relationship between common variation and risk of CLL.

INTRODUCTION

Chronic lymphocytic leukaemia (CLL; MIM 151400) accounts for ~25% of all leukaemia and is the most common form of lymphoid malignancy in Western countries (Stevenson and Caligaris-Cappio 2004). Despite a strong familial basis to CLL, (Sellick, *et al* 2006) with risks in first-degree relatives of cases being increased ~8-fold (Goldin, *et al* 2009), to date the inherited genetic basis of the disease is largely unknown.

A recent genome-wide association (GWA) study has provided evidence that common variation at six loci influences the risk of developing CLL. These loci map to 2q13 (rs17483466), 2q37.1 (*SP140*; rs13397985), 6p25.3 (*IRF4*; rs872071), 11q24.1 (rs735665), 15q23 (rs7176508) and 19q13.32 (*PRKD2*; rs11083846) (Di Bernardo, *et al* 2008). In addition to providing support for a model of CLL susceptibility based on the co-inheritance of multiple low-risk variants, this study provided evidence for the role of B-cell developmental genes in the aetiology of CLL development.

Association studies are often capricious and multiple replications are highly desirable for observations to be robust. Moreover, validation across different populations can augment generalization of findings. Estimates of the impact of risk loci on disease derived from GWA studies have a tendency to be inflated as a consequence of a “winner’s curse” especially as studies such as the GWA of CLL incorporated cases enriched for genetic predisposition by virtue of family history or early onset disease (Di Bernardo, *et al* 2008). To verify and further explore the relationship between the six variants and CLL risk we genotyped two case-control series from Sweden and Spain totalling 824 patients and 850 controls.

METHODS

The Spanish series of 424 CLL patients (241 males and 183 females) were ascertained through the Hospital Clinic of Barcelona between 1984 and 2005. Cases were defined as consecutive patients with diagnosis of CLL and with an available blood sample for DNA extraction before treatment. Mean age at diagnosis of CLL in patients was 64 years (SD±13). 450 healthy controls (253 males and 197 females, mean age 69 years, SD±12) were obtained from the MedXen Control population cohort (Galician Foundation of Genomic Medicine). Both cases and controls were Spanish residents and were of self-reported European ancestry. The Swedish case-control study comprised 400 CLL patients (249 males and 151 females) ascertained through the SCALE (Scandinavian Lymphoma Aetiology) (Smedby, *et al* 2005) study and the biobank at Uppsala University Hospital, Uppsala, Sweden. For the SCALE samples, blood samples were collected from 275 cases during 1999-2002, within a median of 4 months from diagnosis (0-29 months) while in 29 cases, follow-up samples ascertained in 2007-2008 were used. The 96 samples received from the biobank at Uppsala University Hospital were collected between 1982 and 2005. Mean age at diagnosis in patients was 63 years (SD ± 9). Blood samples collected from 400 healthy individuals (253 males and 147 females; mean age 61 years, SD±7) at the Karolinska Institute, Stockholm, Sweden served as a source of controls. In all patients the diagnosis of CLL was confirmed in accordance with the current WHO classification guidelines. Collection of blood samples and clinico-pathological information from patients and controls was undertaken with informed consent and relevant ethical review board approval in accordance with the tenets of the Declaration of Helsinki.

IGHV mutation status of Swedish cases was determined according to BIOMED-2 protocols and in accordance with published criteria sequences with a germline identity of 98% or greater were classified as unmutated, and those displaying identity of less than 98% as mutated (van Dongen, *et al* 2003, van Krieken, *et al* 2007). ZAP-70 expression was available for 291 Spanish cases. ZAP-70 expression levels were quantified as described previously (Bosch, *et al* 2006, Crespo, *et al* 2003).

DNA was extracted from EDTA venous blood samples using standard methodologies and quantified by Picogreen (Invitrogen, Paisley, United Kingdom). The Spanish case-control series was genotyped using Sequenom Mass Array (Sequenom Inc. San Diego, USA). Primer sequences and PCR conditions for assays are available on request. Genotyping of the Swedish case-control series was conducted using allele-specific PCR KASPar chemistry (KBiosciences Ltd, Hertfordshire, UK) and analyzed in an Applied Biosystem ABI 7900HT system (Foster City, USA).

Relative risk of CLL was quantified for the six SNPs by calculating odds ratios (ORs) with 95% confidence intervals (CIs) using the STATA statistical package (version 10.0; Stata Corporation,

1
2
3 College State, Tx, USA). All statistical tests were two-sided. To provide increased power to
4 demonstrate a relationship between SNP genotype and risk we pooled data from the two case-control
5 series. Meta-analysis was performed under a fixed effects model, estimating Cochran's Q statistic to
6 test for heterogeneity and the I^2 statistic to quantify the proportion of the total variation between studies
7 (Higgins and Thompson 2002, Pettiti 1994). Large heterogeneity is typically defined as $I^2 \geq 75\%$.
8 Combined effect of pairs of the six loci with risk was investigated by logistic regression modeling and
9 evidence for interactive effects between SNPs was assessed by a likelihood ratio test.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

RESULTS AND DISCUSSION

Genotypes were obtained for 415 out of 424 for DNA samples from the Spanish cases (97.9%) and 446 of 450 controls (99.1%). The call rate for Swedish series were 395 out of 400 (98.8%) in the cases and 398 of 400 (99.5%) in the controls. SNP call rates were not significantly different between all cases and controls or between each of the two case-control studies.

The genotype frequencies for five of the six SNPs in both case-control series showed a relationship with CLL risk consistent with previously published data (Table 1). The strongest associations were provided by rs7356665 and rs872071. Only for rs872071 in the Spanish series did the genotype distribution not meet the criterion for Hardy-Weinberg equilibrium in controls ($P < 0.01$). As a consequence of this, risk associated with heterozygosity was attenuated. The relationship between genotype and CLL risk was however not statistically significant for rs17483466 in either case-series (Table 1). This locus provided the fifth strongest evidence for an association in our previously published GWA study (Di Bernardo, *et al* 2008). **Failure to replicate this locus may be a consequence of study power. Although the risk allele frequency of rs17483466 in the Swedish series was similar to that documented in the UK and in HapMap CEU data, the frequency was markedly elevated in the Spanish series. This combined with the fact that the SNP itself is unlikely to be directly causal suggests the possibility of incomplete tagging in this population.**

In a pooled analysis of both Swedish and Spanish case-control series, per allele ORs associated with rs13397985, rs872071, rs735665, rs7176508 rs11083846 and rs17483466 and CLL risk were 1.28, ($P=4.22 \times 10^{-3}$; $P_{\text{het}}=0.35$, $I^2 = 0\%$), 1.42 ($P=9.96 \times 10^{-7}$; $P_{\text{het}}=0.37$, $I^2 = 0\%$), 1.59 ($P=1.23 \times 10^{-7}$; $P_{\text{het}}= 0.04$, $I^2 = 77\%$), 1.27 ($P=6.10 \times 10^{-4}$; $P_{\text{het}}= 0.47$, $I^2 = 0\%$), 1.24 ($P=9.50 \times 10^{-3}$; $P_{\text{het}}= 0.55$, $I^2 = 0\%$) and 1.10 ($P=0.21$; $P_{\text{het}}= 0.95$, $I^2 = 0\%$) respectively (Table 1). While there was evidence for between study heterogeneity for rs735665 (Table 1; Figure 1), ascribable to the strong effect of homozygosity for the AA genotype in the Swedish case series, the relationship between SNP genotype and CLL risk was consistent between the two studies. The association between five of the six loci and CLL risk were statistically significant even with adjustment for multiple testing. **Pooling with data from our recent GWA study provided unambiguous evidence for a relationship between variation at these five loci and CLL risk (Di Bernardo, *et al* 2008) with almost no evidence of between study heterogeneity (Figure 1). Furthermore, the data provide little evidence that the effect size of the six loci seen in the original GWA study were significantly over-inflated ($OR_{\text{per allele}}$ in GWA =1.42, $OR_{\text{per allele}}$ in replication=1.31).**

No evidence of interactive effects between any of the three pairs of loci ($P > 0.05$) was observed suggesting that each locus has an independent role in defining risk. While the risks of CLL associated with

1
2
3 the variants are individually modest the carrier frequencies of the risk alleles are high in the European
4 population and therefore the loci make a major contribution to the development of CLL. Using data from
5 both the Swedish and Spanish datasets the risk of CLL increased with increasing numbers of risk alleles
6 for the five loci (rs13397985, rs872071, rs735665, rs7176508 and rs11083846), counting two for a
7 homozygote and one for a heterozygote assuming equal weights ($OR_{\text{per-allele}}=1.31$, 95% CI: 1.23-1.40;
8 $P_{\text{trend}} = 1.40 \times 10^{-15}$). **The proportion of cases and controls grouped according to the number of risk**
9 **alleles carried is detailed in Supplementary Figure 1, which shows a shift toward a higher number**
10 **of risk alleles in cases.** Individuals with 8+ risk alleles had a >3-fold increase in risk compared to those
11 with a median number of risk alleles (**Supplementary Figure 1**). These ORs may be underestimates
12 because the additive model assumes equal weighting across SNPs. The present data only provide
13 estimates of the effect on susceptibility attributable to the variations. The effect of the actual causal variant
14 responsible for the association is likely to be larger.
15
16
17
18
19
20
21
22
23

24 CLL shows male predominance and can be classified on the basis of the presence or absence of somatic
25 hypermutations of the IGHV genes, with mutated CLL having a better prognosis. Likewise, patients with a
26 low percentage (*i.e.* <20%) of ZAP-70 positive B cells have a better prognosis than those with a higher
27 proportion (*i.e.* >20%) of ZAP-70 positive B cells (Bosch, *et al* 2006, Orchard, *et al* 2004). We assessed
28 the relationship between age, gender, IGHV mutation status, ZAP-70 status and SNP genotypes by case-
29 only logistic regression (Supplementary Table 1). None of the SNPs displayed evidence of a relationship
30 with gender or age at diagnosis in either series. The only evidence of a relationship between genotype
31 and IGHV mutation status (based solely on Swedish cases) was provided by rs7176508, with risk
32 genotypes being associated with mutated CLL ($P = 0.02$). rs735665, for which there was previously
33 evidence of a relationship between genotype and mutation status ($P = 0.027$, with risk genotypes being
34 associated with mutated CLL) (Di Bernardo, *et al* 2008) did show some evidence of a relationship, albeit
35 not statistically significant ($P = 0.08$).
36
37
38
39
40
41
42
43
44

45 Collectively these data provide further evidence for the role of common variation at 2q37.1, 6p25.3,
46 11q24.1, 15q23, and 19q13.32 in defining CLL risk. **The currently identified 10 loci for CLL**
47 **collectively account for only ~10% of the excess familial risk (Crowther-Swanepoel, *et al*).**
48 **Thus, it is likely that a large number of additional risk variants remain to be discovered.**
49
50
51
52
53
54

55 ACKNOWLEDGEMENTS

56 This work was supported by Leukaemia Research (UK). Support from Cancer Research UK
57 (C1298/A8362 supported by the Bobby Moore Fund) is also acknowledged. Dalemari Crowther-
58 Swanepoel is supported by a PhD studentship from the Institute of Cancer Research. EC and AE
59 acknowledge Fundacion Genoma España and CICYT funding (SAF 08/3630) and the RTICC from
60

1
2
3 the Instituto de Salud Carlos III. AC and MT acknowledge Xunta de Galicia funding for Spanish
4 control population (PGIDIT 08CSA005208PR) and the support of the Spanish National Genotyping
5 Center (USC Node). We would like to thank Dr Leonid Padyukov, Karolinska Institutet, Stockholm,
6 Sweden, and the EIRA group for providing control samples from the Swedish population for the
7 Swedish replication study. Sample collection in the SCALE study was supported by the National
8 Cancer Institute. Sample preparations in the SCALE study were supported by the Swedish Cancer
9 Society. We are grateful to all clinicians for their involvement in patient ascertainment. Finally we are
10 grateful to all the patients and other individuals for their participation.
11
12
13
14
15
16
17

18 **COMPETING INTERESTS STATEMENT**

19 The authors declare no competing financial interests
20
21

22 **FIGURE LEGEND**

23
24
25 **Figure 1: Forest plots of effect size and direction for the SNPs associated with CLL risk.** Boxes
26 denote per allele odds ratios (OR) point estimates, their areas being proportional to the inverse
27 variance weight of the estimate. Horizontal lines represent 95% confidence intervals. The diamond
28 (and broken line) represents the summary OR computed under a fixed effects model, with 95%
29 confidence interval given by its width. The unbroken vertical line is at the null value (OR = 1.0).
30 Phases 1, 2 and 3 refer to the original GWAS as described in Di Bernardo *et al.* (2008).
31
32
33
34
35

36 **Table 1: Association between SNPs and CLL risk**

37
38 * Large heterogeneity is typically defined as $I^2 \geq 75\%$.
39
40
41

42 **Supplementary Table 1: Clinico-pathological associations.** Age and gender based on data from
43 both case series; IGHV mutational status based on Swedish data; ZAP-70 status based on Spanish
44 data
45
46
47

48 **Supplementary Figure 1: Cumulative impact of the six variants on CLL risk.** (a) Distribution of
49 risk alleles in controls (blue) and cases (red) for the six CLL loci. (b) Plot of the increasing Ors for
50 CLL with increasing number of risk alleles. The ORs are relative to the median number of four risk
51 alleles; vertical bars correspond to 95% confidence intervals. The risk alleles follow a normal
52 distribution in both cases and controls, with a shift toward a higher number of risk alleles in the
53 cases. The horizontal line denotes the null value (OR=1.0)
54
55
56
57
58
59
60

REFERENCES

- Bosch, F., Muntanola, A., Gine, E., Carrio, A., Villamor, N., Moreno, C., Crespo, M. & Montserrat, E. (2006) Clinical implications of ZAP-70 expression in chronic lymphocytic leukemia. *Cytometry part B: Clinical Cytometry*, **70**, 214-217.
- Crespo, M., Bosch, F., Villamor, N., Bellosillo, B., Colomer, D., Rozman, M., Marce, S., Lopez-Guillermo, A., Campo, E. & Montserrat, E. (2003) ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *New England Journal of Medicine*, **348**, 1764-1775.
- Crowther-Swanepoel, D., Broderick, P., Di Bernardo, M.C., Dobbins, S.E., Torres, M., Mansouri, M., Ruiz-Ponte, C., Enjuanes, A., Rosenquist, R., Carracedo, A., Jurlander, J., Campo, E., Juliusson, G., Montserrat, E., Smedby, K.E., Dyer, M.J., Matutes, E., Dearden, C., Sunter, N.J., Hall, A.G., Mainou-Fowler, T., Jackson, G.H., Summerfield, G., Harris, R.J., Pettitt, A.R., Allsup, D.J., Bailey, J.R., Pratt, G., Pepper, C., Fegan, C., Parker, A., Oscier, D., Allan, J.M., Catovsky, D. & Houlston, R.S. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. *Nature Genetics*, **42**, 132-136.
- Di Bernardo, M.C., Crowther-Swanepoel, D., Broderick, P., Webb, E., Sellick, G., Wild, R., Sullivan, K., Vijayakrishnan, J., Wang, Y., Pittman, A.M., Sunter, N.J., Hall, A.G., Dyer, M.J., Matutes, E., Dearden, C., Mainou-Fowler, T., Jackson, G.H., Summerfield, G., Harris, R.J., Pettitt, A.R., Hillmen, P., Allsup, D.J., Bailey, J.R., Pratt, G., Pepper, C., Fegan, C., Allan, J.M., Catovsky, D. & Houlston, R.S. (2008) A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nature Genetics*, **40**, 1204-1210.
- Goldin, L.R., Bjorkholm, M., Kristinsson, S.Y., Turesson, I. & Landgren, O. (2009) Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia. *Haematologica*, **94**, 647-653.
- Higgins, J.P. & Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, **21**, 1539-1558.
- Orchard, J.A., Ibbotson, R.E., Davis, Z., Wiestner, A., Rosenwald, A., Thomas, P.W., Hamblin, T.J., Staudt, L.M. & Oscier, D.G. (2004) ZAP-70 expression and prognosis in chronic lymphocytic leukaemia. *Lancet*, **363**, 105-111.
- Pettiti, D. (1994) Meta-analysis Decision Analysis and Cost-effectiveness Analysis. *Oxford University Press, Oxford, New York*.
- Sellick, G.S., Catovsky, D. & Houlston, R.S. (2006) Familial chronic lymphocytic leukemia. *Seminars in Oncology*, **33**, 195-201.
- Smedby, K.E., Hjalgrim, H., Melbye, M., Torrang, A., Rostgaard, K., Munksgaard, L., Adami, J., Hansen, M., Porwit-MacDonald, A., Jensen, B.A., Roos, G., Pedersen, B.B., Sundstrom, C., Glimelius, B. & Adami, H.O. (2005) Ultraviolet radiation exposure and risk of malignant lymphomas. *Journal of the National Cancer Institute*, **97**, 199-209.
- Stevenson, F. & Caligaris-Cappio, F. (2004) Chronic lymphocytic leukemia: revelations from the B-cell receptor. *Blood*, 4389-4395.
- van Dongen, J.J., Langerak, A.W., Bruggemann, M., Evans, P.A., Hummel, M., Lavender, F.L., Delabesse, E., Davi, F., Schuurink, E., Garcia-Sanz, R., van Krieken, J.H., Droese, J., Gonzalez, D., Bastard, C., White, H.E., Spaargaren, M., Gonzalez, M., Parreira, A., Smith, J.L., Morgan, G.J., Kneba, M. & Macintyre, E.A. (2003) Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia*, **17**, 2257-2317.

1
2
3 van Krieken, J.H., Langerak, A.W., Macintyre, E.A., Kneba, M., Hodges, E., Sanz, R.G., Morgan,
4 G.J., Parreira, A., Molina, T.J., Cabecadas, J., Gaulard, P., Jasani, B., Garcia, J.F., Ott, M.,
5 Hannsmann, M.L., Berger, F., Hummel, M., Davi, F., Bruggemann, M., Lavender, F.L.,
6 Schuuring, E., Evans, P.A., White, H., Salles, G., Groenen, P.J., Gameiro, P., Pott, C. &
7 Dongen, J.J. (2007) Improved reliability of lymphoma diagnostics via PCR-based clonality
8 testing: report of the BIOMED-2 Concerted Action BHM4-CT98-3936. *Leukemia*, **21**, 201-
9 206.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

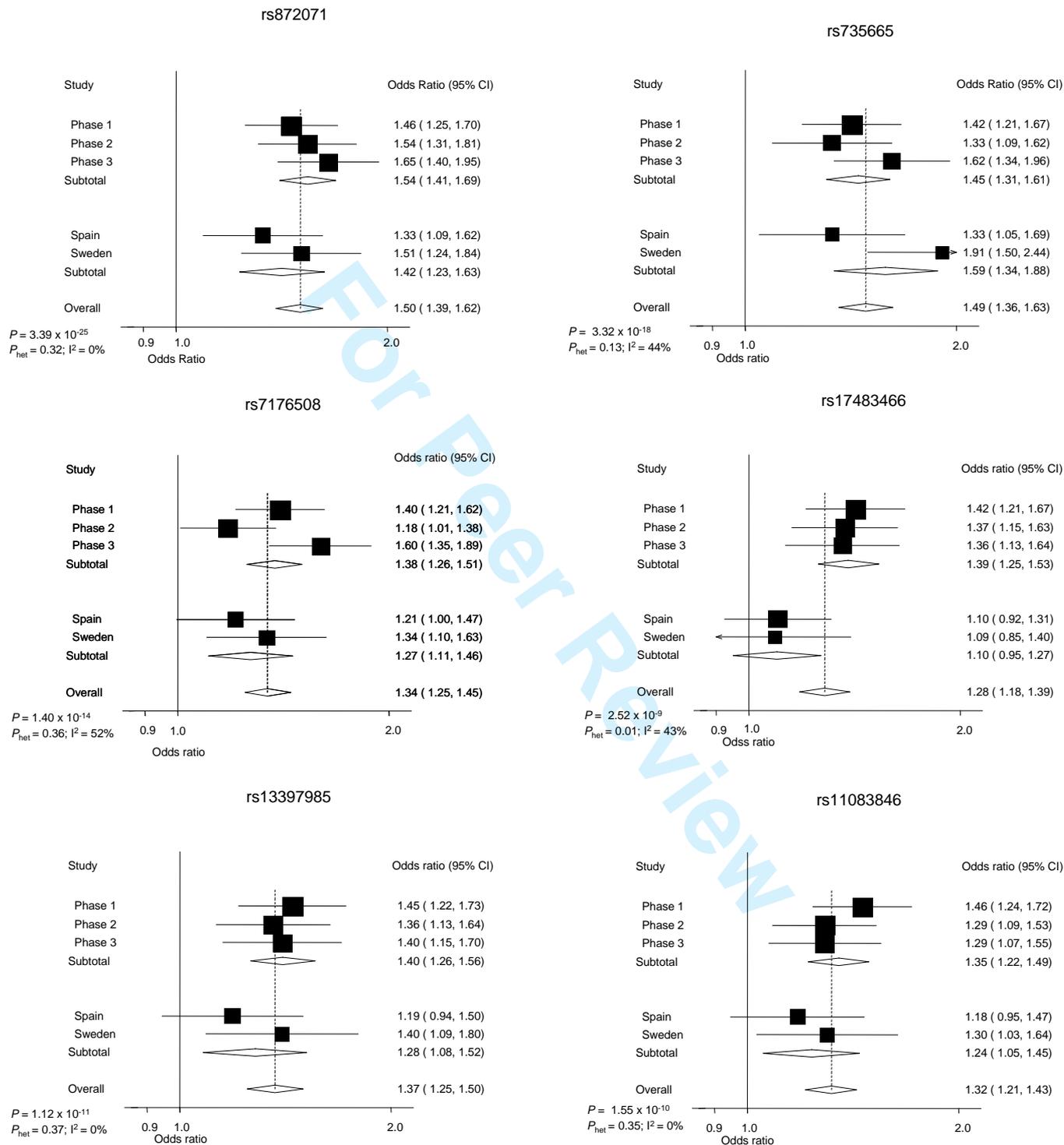


Figure 1: Forest plots of effect size and direction for the SNPs associated with CLL risk. Boxes denote allelic OR point estimates, their areas being proportional to the inverse variance weight of the estimate. Horizontal lines represent 95% confidence intervals. The diamond (and broken line) represents the summary OR computed under a fixed effects model, with 95% confidence interval given by its width. The unbroken vertical line is at the null value (OR = 1.0). Phases 1, 2 and 3 refer to the original GWAS as described in Di Bernardo *et al.* (2008).

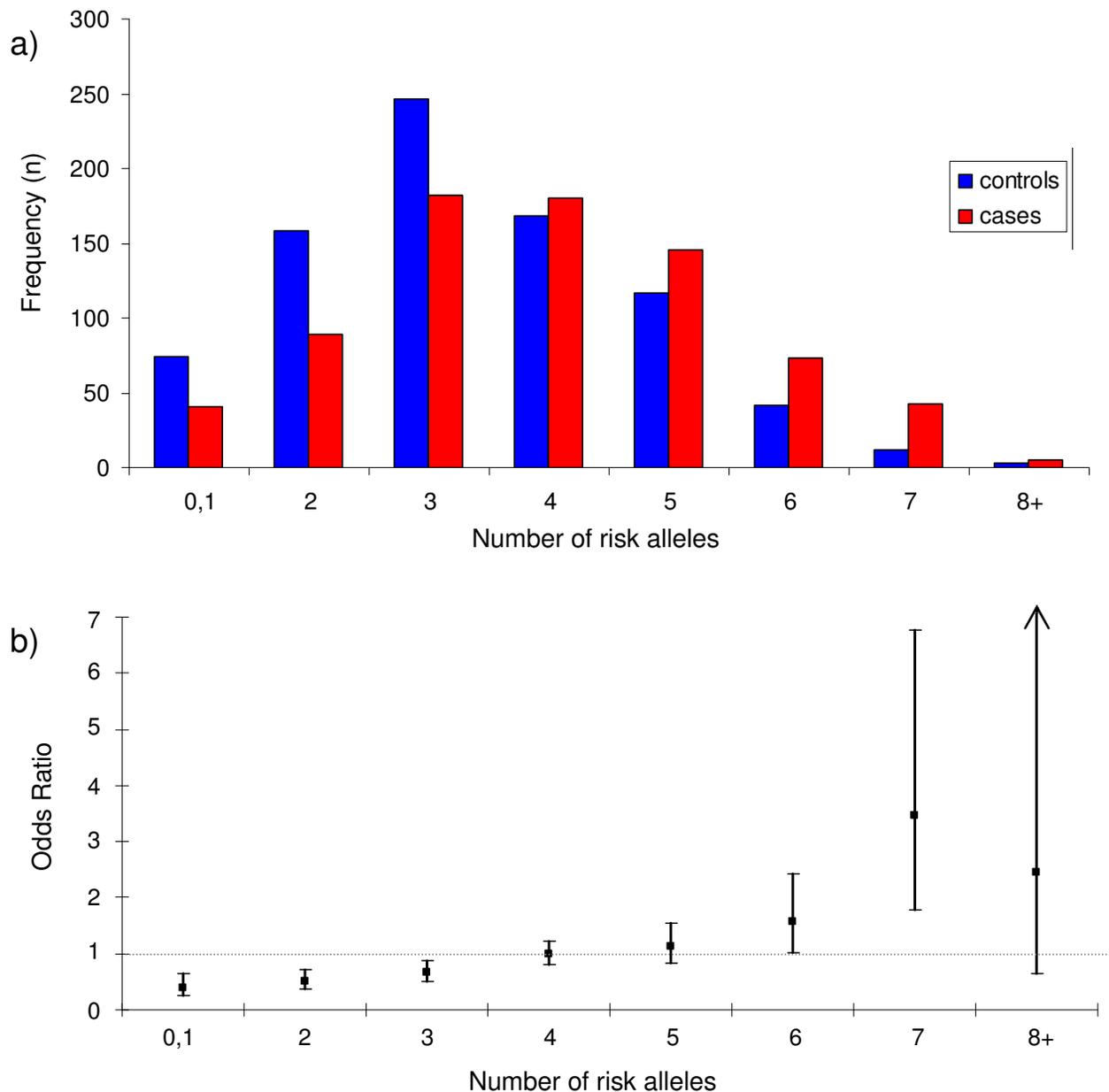
Table 1: Association between SNPs and CLL risk

SNP	Genotype	Spain						Sweden						Combined			
		Cases	Controls	OR	(95% CI)	P	Cases	Controls	OR	(95% CI)	P	OR	(95% CI)	P	P _{het}	I ²	
rs872071	AA	86	101	1.00	Ref		74	129	1.00	Ref							
6p25.3	AG	198	251	0.93	(0.65-1.33)	0.66	203	179	1.98	(1.37-2.85)	1.19 x 10 ⁻⁴	1.35	(1.06-1.72)	0.016	2.0 x 10 ⁻³	89%	
	GG	121	82	1.73	(1.14-2.64)	0.01	117	89	2.29	(1.51-3.48)	3.74 x 10 ⁻⁵	2.00	(1.51-2.65)	1.40 x 10 ⁻⁶	0.33	0%	
	Trend			1.33	(1.09-1.62)	0.01			1.51	(1.24-1.84)	3.80 x 10⁻⁵	1.42	(1.23-1.63)	9.96 x 10⁻⁷	0.37	0%	
rs735665	GG	239	298	1.00	Ref		181	248	1.00	Ref							
11q24.1	AG	157	130	1.51	(1.12-2.03)	0.01	178	135	1.81	(1.33-2.45)	7.80 x 10 ⁻⁵	1.65	(1.34-2.02)	1.96 x 10 ⁻⁶	0.39	0%	
	AA	18	18	1.25	(0.60-2.60)	0.52	32	10	4.38	(2.03-10.23)	2.39 x 10 ⁻⁵	2.31	(1.43-3.73)	4.37 x 10 ⁻⁴	0.01	84%	
	Trend			1.33	(1.05-1.69)	0.02			1.91	(1.50-2.44)	1.16 x 10⁻⁷	1.59	(1.34-1.88)	1.23 x 10⁻⁷	0.04	77%	
rs7176508	GG	127	159	1.00	Ref		114	147	1.00	Ref							
15q23	AG	208	222	1.17	(0.86-1.60)	0.30	181	184	1.27	(0.91-1.77)	0.14	1.22	(0.98-1.51)	0.079	0.73	0%	
	AA	76	64	1.49	(0.97-2.278)	0.06	95	67	1.83	(1.21-2.77)	3.00 x 10 ⁻³	1.65	(1.24-2.20)	5.02 x 10 ⁻⁴	0.48	0%	
	Trend			1.21	(1.00-1.47)	0.06			1.34	(1.10-1.63)	3.07 x 10⁻³	1.27	(1.11-1.46)	6.10 x 10⁻⁴	0.47	0%	
rs17483466	AA	223	239	1.00	Ref		254	267	1.00	Ref							
2q13	AG	151	176	0.92	(0.69-1.23)	0.56	118	114	1.09	(0.79-1.50)	0.59	0.99	(0.81-1.22)	0.947	0.43	0%	
	GG	40	31	1.38	(0.81-2.37)	0.21	16	14	1.20	(0.54-2.72)	0.63	1.32	(0.87-2.00)	0.187	0.76	0%	
	Trend			1.10	(0.86-1.31)	0.58			1.09	(0.85-1.40)	0.50	1.10	(0.95-1.27)	0.207	0.95	0%	
rs13397985	TT	262	301	1.00	Ref		240	281	1.00	Ref							
2q37.1	TG	126	125	1.16	(0.85-1.58)	0.33	136	104	1.53	(1.11-2.11)	0.01	1.33	(1.07-1.64)	0.010	0.20	39%	
	GG	25	19	1.51	(0.78-2.97)	0.19	16	13	1.44	(0.64-3.33)	0.34	1.48	(0.92-2.39)	0.106	0.92	0%	
	Trend			1.19	(0.95-1.50)	0.13			1.40	(1.08-1.80)	0.01	1.28	(1.08-1.52)	4.22 x 10⁻³	0.35	0%	
rs11083846	GG	234	272	1.00	Ref		218	250	1.00	Ref							
19q13.32	AG	147	147	1.16	(0.86-1.57)	0.31	141	129	1.25	(0.92-1.71)	0.14	1.21	(0.98-1.48)	0.078	0.72	0%	
	AA	32	26	1.43	(0.80-2.58)	0.20	30	19	1.81	(0.95-3.50)	0.05	1.59	(1.06-2.39)	0.024	0.57	0%	
	Trend			1.18	(0.95-1.47)	0.13			1.30	(1.03-1.64)	0.03	1.24	(1.05-1.45)	9.50 x 10⁻³	0.55	0%	

Large heterogeneity is typically defined as I² ≥ 75%.

Supplementary Table 1: Clinico-pathological association testing. Age and gender based on data from both populations; *IGHV* mutational status based on Swedish data; ZAP-70 based on Spanish data.

SNP	Locus	Genotype	Age (years)			Gender		IGHV mutation status			ZAP-70			
			≤63	>63	<i>P</i>	Male	Female	<i>P</i>	no	yes	<i>P</i>	no	yes	<i>P</i>
rs872071	6p25.3	AA	77	83	0.82	90	70	0.67	27	47	0.17	40	21	0.58
		AG	192	209		249	152		74	129		81	51	
		GG	117	121		141	97		53	64		48	31	
rs735665	11q24.1	GG	199	221	0.60	254	166	0.91	74	107	0.08	97	62	0.90
		AG	169	166		195	140		71	107		68	36	
		AA	24	26		32	18		6	26		8	7	
rs7176508	15q23	GG	117	124	0.75	142	99	0.77	54	60	0.02	59	29	0.25
		AG	193	196		234	155		68	113		86	56	
		AA	80	91		103	68		30	65		26	19	
rs17483466	2q13	AA	227	250	0.50	289	188	0.68	104	150	0.29	97	59	0.62
		AG	134	135		156	113		43	75		59	32	
		GG	28	28		34	22		5	11		16	14	
rs13397985	2q37.1	TT	248	254	0.86	295	207	0.59	94	146	0.70	105	65	0.46
		TG	122	140		162	100		53	83		53	34	
		GG	22	19		24	17		5	11		15	3	
rs11083846	19q13.32	GG	216	236	0.77	266	186	0.92	83	135	0.80	99	58	0.97
		AG	142	146		182	106		61	80		57	40	
		AA	30	32		33	29		8	22		16	7	



Supplementary Figure 1: Cumulative impact of the six variants on CLL risk. (a) Distribution of risk alleles in controls (blue) and cases (red) for the six CLL loci. (b) Plot of the increasing ORs for CLL with increasing number of risk alleles. The ORs are relative to the median number of four risk alleles; vertical bars correspond to 95% confidence intervals. The risk alleles follow a normal distribution in both cases and controls, with a shift toward a higher number of risk alleles in the cases. The horizontal line denotes the null value (OR=1.0)