



HAL
open science

Common genetic variation at 15q25.2 impacts on chronic lymphocytic leukaemia risk

Dalemari Crowther-Swanepoel, Maria Chiara Di Bernardo, Krzysztof Jamroziak, Lidia Karabon, Irena Frydecka, Silvia Deaglio, Giovanni d'Arena, Davide Rossi, Gianluca Gaidano, Bianca Olver, et al.

► To cite this version:

Dalemari Crowther-Swanepoel, Maria Chiara Di Bernardo, Krzysztof Jamroziak, Lidia Karabon, Irena Frydecka, et al.. Common genetic variation at 15q25.2 impacts on chronic lymphocytic leukaemia risk. *British Journal of Haematology*, 2011, 154 (2), pp.229. 10.1111/j.1365-2141.2011.08706.x. hal-00639394

HAL Id: hal-00639394

<https://hal.science/hal-00639394>

Submitted on 9 Nov 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.



Common genetic variation at 15q25.2 impacts on chronic lymphocytic leukaemia risk

Journal:	<i>British Journal of Haematology</i>
Manuscript ID:	BJH-2011-00113.R1
Manuscript Type:	Short Reports
Date Submitted by the Author:	09-Mar-2011
Complete List of Authors:	<p>Crowther-Swanepoel, Dalemari; Institute of Cancer Research, Cancer Genetics Di Bernardo, Maria Chiara; Institute of Cancer Research, Cancer Genetics Jamroziak, Krzysztof; Medical University of Lodz, Copernicus Memorial Hospital, Department of Hematology Karabon, Lidia; Polish Academy of Science, Institute of Immunology and Experimental Therapy Frydecka, Irena; Polish Academy of Science, Institute of Immunology and Experimental Therapy Deaglio, Silvia; University of Turin Medical School, Laboratory of Immunogenetics D'Arena, Giovanni; IRCCS, Casa Sollievo della Sofferenza Hospital Rossi, Davide; Division of Hematology, Department of Clinical and Experimental Medicine, Amedeo Avogadro University of Eastern Piedmont Gaidano, Gianluca; Division of Hematology, Department of Clinical and Experimental Medicine, Amedeo Avogadro University of Eastern Piedmont Olver, Bianca; Institute of Cancer Research, Cancer Genetics Lloyd, Amy; Institute of Cancer Research, Cancer Genetics Broderick, Peter; Institute of Cancer Research, Cancer Genetics Laurenti, Luca; Catholic University of the Sacred Heart, Institute of Haematology Szemraj-Rogucka, Zofia; Medical University of Lodz, Copernicus Memorial Hospital, Department of Hematology Robak, Tadeusz; Medical University of Lodz, Department of Hematology Catovsky, Daniel; Institute of Cancer Research, Haemato-oncology Houlston, Richard; Institute of Cancer Research, Cancer Genetics</p>
Key Words:	CLL, risk, common variant, CPEB1

1
2
3
4
5
6
7
8
9
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

SCHOLARONE™
Manuscripts

For Peer Review

Common genetic variation at 15q25.2 impacts on chronic lymphocytic leukaemia risk

Dalemari Crowther-Swanepoel¹, Maria Chiara Di Bernardo¹, Krzysztof Jamrozik², Lidia Karabon³, Irena Frydecka³, Silvia Deaglio⁴, Giovanni D'Arena⁵, Davide Rossi⁶, Gianluca Gaidano⁶, Bianca Olver¹, Amy Lloyd¹, Peter Broderick¹, Luca Laurenti⁷, Zofia Szemraj-Rogucka², Tadeusz Robak², Daniel Catovsky⁸ & Richard S Houlston¹

1. Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey. UK
2. Department of Hematology, Medical University of Lodz, Copernicus Memorial Hospital, Lodz, Poland
3. Institute of Immunology and Experimental Therapy, Polish Academy of Science, Wroclaw, Poland
4. Department of Genetics, Biology and Biochemistry, University of Turin and Human Genetics Foundation (HuGeF), Turin, Italy
5. IRCCS "Casa Sollievo della Sofferenza" Hospital, San Giovanni Rotondo, Italy
6. Division of Hematology, Department of Clinical and Experimental Medicine and IRCAD, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy
7. (Rome, Italy)
8. Section of Haemato-oncology, Institute of Cancer Research, Sutton, Surrey. UK

✉Corresponding author

Richard Houlston

Institute of Cancer Research

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

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

Key words: CLL, risk, common variant, CPEB1

Running title: Verification of CLL risk loci

SUMMARY

A genome-wide association study of chronic lymphocytic leukaemia (CLL) suggested that common variants at 15q25.2 (rs783540) and 18q21.1 (rs1036935) influence CLL. To validate these associations and explore their relationship with CLL risk we genotyped case-control datasets from Poland, UK and Italy totalling 1,428 cases and 1,920 controls. Combined data from these and previously genotyped series (2,503 cases and 5,789 controls) provided evidence for association between 15q25.2 and 18q21.1 loci and CLL risk ($P_{combined}=1.10 \times 10^{-7}$ and 1.30×10^{-5} respectively). These data provide further evidence for the involvement of common genetic variants in CLL risk and insight into the biological basis of disease development.

For Peer Review

INTRODUCTION

Chronic lymphocytic leukaemia (CLL) is the most common lymphoid malignancy in Western countries (Stevenson and Caligaris-Cappio 2004). While CLL shows an 8-fold increased risk in relatives of CLL cases (Goldin, *et al* 2009), the inherited basis of disease susceptibility is largely unknown.

We have recently conducted a genome-wide association study (GWAS) of CLL, identifying common variants at 10 loci influencing CLL risk (Di Bernardo, *et al* 2008)(Crowther-Swanepoel, *et al* 2010). In addition to these loci there was evidence for two risk loci for CLL at 15q25.2 (rs783540) and 18q21.1 (rs1036935), however associations were not genome-wide significant.

To verify and explore the relationship between these SNPs and CLL we have genotyped case-control series from Poland, the UK and Italy, totalling 1,428 patients and 1,920 controls. To increase power to demonstrate a relationship between SNP genotype and risk we pooled these data with our previously published data.

METHODS

Subjects

Poland 1: 302 Polish CLL patients from Lodz (177 male, mean age 60, SD±11) ascertained through the Department of Hematology, Medical University of Lodz, 1999-2008. Controls were 208 healthy volunteers (118 male, mean age 60 years, SD±12) from the Lodz region.

Poland 2: 193 Polish CLL patients from Wroclaw (100 male, mean age 64, SD±11) ascertained 1998-2008. Three hundred and fourteen healthy controls (137 male, mean age 31 years, SD±9) were obtained from the blood bank in Wroclaw and among employees of the Institute of Immunology and Experimental Therapy. Genetic homogeneity of the Polish population is observed, as reflected in virtually identical frequencies of H-Y polymorphisms in different regions of Poland (Ploski, *et al* 2002).

Italy 1: 227 Italian CLL patients (135 male, mean age 64, SD±10) collected through the Department of Genetics, Biology and Biochemistry, 2003-2008. Controls were 234 healthy volunteers (119 male, mean age 44, SD±16), resident in North West Italy.

Italy 2: 383 Italian CLL patients (206 male, mean age 67, SD±11) referred to the Division of Hematology, Amedeo Avogadro University of Eastern Piedmont, 1985-2009. Blood samples from 170 healthy individuals (95 male; mean age 68, SD±10) from the same geographical area with normal hematology served as controls.

UK 3: 446 CLL cases (288 male, mean age 62, SD±17) ascertained through the FCLL study, 1998-2006. Blood samples from 1,048 healthy individuals (389 male, mean age 59, SD±9) collected through NSCCG (Penegar, *et al* 2007) and GELCAPS studies served as controls.

In all patients the diagnosis of CLL was confirmed in accordance with the current IWCLL-NCI and WHO classification guidelines. Collection of blood samples and clinico-pathological information was undertaken with informed consent and ethical review board approval in accordance with the tenets of the Declaration of Helsinki.

DNA was extracted from EDTA-venous blood samples using standard methodologies and quantified by Picogreen (Invitrogen, Paisley, United Kingdom). Genotyping was conducted using allele-specific PCR KASPar chemistry (KBiosciences Ltd, Hertfordshire, UK) and analyzed using an Applied Biosystem ABI 7900HT system (Foster City, USA). To monitor quality control of genotyping, replicate samples of known genotype were included in each set of sample plates.

IGHV mutational and ZAP-70 expression status

IGHV mutation status of 369 Italian cases and 1,337 CLL cases from previously genotyped series was determined according to BIOMED-2 protocols. In accordance with published criteria sequences with a germline identity of $\geq 98\%$ were classified as unmutated, and those displaying identity $< 98\%$ as mutated (van Dongen, *et al* 2003, van Krieken, *et al* 2007). ZAP-70 expression information was available for 281 Spanish cases, 212 Italy 1 cases and 167 Italy 2 cases. ZAP-70 expression was quantified as described previously (Bosch, *et al* 2006, Crespo, *et al* 2003).

Statistical analysis

Statistical analyses were undertaken using STATA (v10.0). Deviation of the genotype frequencies in the controls from those expected under Hardy-Weinberg equilibrium (HWE) was assessed by the χ^2 test. Odds ratios (ORs) and 95% confidence intervals (CIs) were obtained by unconditional logistic regression. Data from Poland 1, Poland 2, Italy 1, Italy 2, UK 3 were combined with data from our previously published GWAS (Di Bernardo, *et al* 2008) and data from four previously published case-control series (UK 1, UK 2, Spain, Sweden) (Crowther-Swanepoel, *et al* 2010). Meta-analysis was performed under a fixed-effects model, estimating Cochran's Q and I^2 statistics (Higgins and Thompson 2002, Pettiti 1994). Evidence for interactive effects between SNPs was assessed by a likelihood ratio test. Associations by phenotype were examined by logistic regression in case-only analyses.

Relationship between SNP genotype and expression levels

To examine for a relationship between SNP genotype and expression of *CPEB1* in lymphocytes we used publicly available expression data on 90 Caucasian derived EBV-transformed lymphoblastoid cell lines. Online recovery of data was performed using WGAViewer Version 1.25 Software. Differences in levels of mRNA expression between genotypes were assessed using a Wilcoxon-type trend test.

RESULTS

Genotypes were obtained for >99.9% of DNA samples for both rs1036935 and rs783540; SNP call rates were not significantly different between all cases and controls or between each of the case-control series. The distribution of SNP genotypes in both controls and cases in each of the series showed no significant departure from HWE.

The genotype frequencies of rs783540 showed a relationship with CLL risk consistent with our previously published findings in four of the series (Figure 1; Supplementary Figure 1; Table 1). The strongest association was provided by UK 3. For rs1036935 three of the series provided support for a relationship with CLL risk consistent with published findings (Figure 1; Table 1). In a pooled analysis of these five case-control series, per allele ORs associated with rs1036935, and rs783540, and CLL risk were 1.15 ($P=0.009$; $P_{\text{het}}=0.56$, $I^2=0\%$), and 1.06 ($P=0.35$; $P_{\text{het}}=0.62$, $I^2=0\%$) respectively (Figure 1). Pooling with data from our previously published GWAS (Di Bernardo, *et al* 2008) and data from four previously published case-control series (UK 1, UK 2, Spain, Sweden) (Crowther-Swanepoel, *et al* 2010) provided unambiguous evidence for a relationship between variation at rs783540 and CLL risk ($P_{\text{combined}}=1.10 \times 10^{-7}$; OR =1.17) (Figure 1). For rs1036935 this combined analysis did not provide increased support for a relationship ($P_{\text{combined}}=1.30 \times 10^{-5}$). No evidence of interactive effects between these two loci ($P=0.69$) was shown compatible with each locus having an independent effect.

We searched for genotype-expression correlations in lymphoblastoid cell lines. *MBD1* expression was shown to correlate with genotype (Supplementary Figure 2); whereby reduced gene expression was associated with risk genotype.

CLL shows male predominance and can be classified on the basis of somatic hypermutation of *IGHV* genes, with mutated CLL having a better prognosis. Likewise, patients with $\leq 20\%$ of ZAP-70 positive B-cells have better prognosis than those with $>20\%$. Neither SNP displayed evidence of a relationship with gender, age at diagnosis, *IGHV* mutation or ZAP-70 status in either series (Supplementary Table 1).

DISCUSSION

While the risks of CLL associated with the rs1036935 and rs783540 variants are individually modest the carrier frequencies of the risk alleles are high in the European population and therefore the loci make a significant major contribution to CLL.

rs783540 localizes to intron 2 of *CPEB1* (cytoplasmic polyadenylation element binding protein 1; MIM 607342) which has an established role in the regulation of cyclin B1 during embryonic division-differentiation. *CPEB1* has been implicated in mediating both polyadenylation-dependent translation activation and CPE-directed translational repression (Mendez and Richter 2001, Welk, *et al* 2001). Furthermore, it has been shown that MYC mRNA is a target of CPEB1 regulation, where CPEB1 appears to act as a translational repressor protein to control MYC translation and its deregulation leads to senescence (Groisman, *et al* 2006). This is intriguing as variation at 8q24 influences CLL risk through probable effects on MYC expression (Crowther-Swanepoel, *et al* 2010). Although three of the five replication series showed support for an association between rs1036935 and CLL risk (Figure 1), overall we were not able to demonstrate enhanced evidence of an association between rs1036935 and CLL risk to that originally shown. This may be a consequence of LD heterogeneity between the different population series as evidenced by differences in allele frequencies between series. Accepting this caveat, rs1036935 still provides a reasonable basis for a CLL susceptibility locus. Moreover, the genes mapping adjacent to this SNP have biological plausibility for a role in CLL development. *CXXC1* (protein containing CXXC domain; MIM 609150) and *MBD1* (methyl-CpG-binding domain protein 1; MIM 156535) are both involved in gene regulation.

Accepting that HapMap is not comprehensive we did not identify any highly correlated (*i.e.* $r^2 > 0.5$) functional coding variants in *CPEB1* correlated with rs783540, or coding variants in *MBD1* or *CXXC1* correlated with rs1036935. These data suggest associations are mediated through LD with sequence changes influencing gene expression rather than protein sequence, or through LD with low frequency variants that are not catalogued. Although *MBD1* has no documented role in CLL, it has potential to affect CLL development through translational control of MYC via MDBP binding.

In conclusion this study provides further evidence for low risk variants predisposing to B-cell CLL and insight into the development of this haematological malignancy.

ACKNOWLEDGEMENTS

This work was supported by Leukaemia Research (UK). Support from Cancer Research UK (C1298/A8362 supported by the Bobby Moore Fund), from the Italian Association for Cancer Research and from the Italian Ministry of Health and Education is also acknowledged. Collection of blood samples from Polish patients and controls from Lodz area and DNA extraction was supported by the grant from Polish Ministry of Science and Higher Education no NN402178334. Dalemari Crowther-Swanepoel is supported by a PhD studentship from the Institute of Cancer Research.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests

For Peer Review

FIGURE LEGENDS

Figure 1: Forest plots of effect size and direction for rs783540 and rs1036935 with CLL risk. Boxes denote per allele odds ratios (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). Subjects in the GWAS, UK 1, UK 2, Spanish and Swedish datasets were described previously in (Crowther-Swanepoel, *et al* 2010, Di Bernardo, *et al* 2008).

Supplementary Figure 1: Regional plots of the rs783540 and rs1036935 risk loci. a) The 15q25.2 locus, with the local recombination rate plotted in light blue over this 340-kb chromosomal segment. Each square represents a SNP found in this locus and the most associated SNP in the combined analysis, rs783540, is marked by a blue diamond. The colour intensity of each square reflects the extent of LD with the top SNP - red ($r^2 > 0.8$) through to white ($r^2 < 0.3$). Physical positions are based on build 36 of the human genome. rs783540 is located in intron 2 of *CPEB1*.

b) The 18q21.1 locus, with the most associated SNP in this locus, rs1036935, highlighted by a blue diamond. All SNPs found within a 600-kb window centered on rs1036935 are indicated.

Supplementary Figure 2: Relationship between lymphocyte mRNA expression levels of a) *CPEB1* and rs783540 genotype and b) *MBD1* and rs1036935 genotype.

Adapted from (Crowther-Swanepoel 2010, *et al*)

Table 1: Association between SNPs and CLL risk

Supplementary Table 1: Clinico-pathological association testing. Age and gender are based on data from all ten series, IGHV mutational status is based on data from series GWAS, UK 1, Spain, Sweden, Italy-1 and Italy-2, ZAP-70 is based on series Spain, Italy-1 and Italy-2 data only.

REFERENCES

- 1
2
3
4
5 Bosch, F., et al. (2006) Clinical implications of ZAP-70 expression in chronic lymphocytic
6 leukemia. *Cytometry B Clin Cytom*, **70**, 214-217.
- 7
8 Crespo, M., et al. (2003) ZAP-70 expression as a surrogate for immunoglobulin-variable-region
9 mutations in chronic lymphocytic leukemia. *N Engl J Med*, **348**, 1764-1775.
- 10
11 Crowther-Swanepoel, D., et al. (2010) Common variants at 2q37.3, 8q24.21, 15q21.3 and
12 16q24.1 influence chronic lymphocytic leukemia risk. *Nat Genet*, **42**, 132-136.
- 13
14 Di Bernardo, M.C., et al. (2008) A genome-wide association study identifies six susceptibility
15 loci for chronic lymphocytic leukemia. *Nat Genet*, **40**, 1204-1210.
- 16
17 Goldin, L.R., et al. (2009) Elevated risk of chronic lymphocytic leukemia and other indolent non-
18 Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia.
19 *Haematologica*, **94**, 647-653.
- 20
21 Groisman, I., et al. (2006) Control of cellular senescence by CPEB. *Genes Dev*, **20**, 2701-2712.
- 22
23 Higgins, J.P. & Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Stat Med*,
24 **21**, 1539-1558.
- 25
26 Mendez, R. & Richter, J.D. (2001) Translational control by CPEB: a means to the end. *Nat Rev*
27 *Mol Cell Biol*, **2**, 521-529.
- 28
29 Penegar, S., et al. (2007) National study of colorectal cancer genetics. *Br J Cancer*, **97**, 1305-
30 1309.
- 31
32 Pettiti, D. (1994) Meta-analysis Decision Analysis and Cost-effectiveness Analysis.
33 . Oxford University Press, Oxford, New York.
- 34
35 Ploski, R., et al. (2002) Homogeneity and distinctiveness of Polish paternal lineages revealed by
36 Y chromosome microsatellite haplotype analysis. *Hum Genet*, **110**, 592-600.
- 37
38 Stevenson, F. & Caligaris-Cappio, F. (2004) Chronic lymphocytic leukemia: revelations from the
39 B-cell receptor. *Blood*, 4389-4395.
- 40
41 van Dongen, J.J., et al. (2003) Design and standardization of PCR primers and protocols for
42 detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect
43 lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936.
44 *Leukemia*, **17**, 2257-2317.
- 45
46 van Krieken, J.H., et al. (2007) Improved reliability of lymphoma diagnostics via PCR-based
47 clonality testing: report of the BIOMED-2 Concerted Action BHM4-CT98-3936.
48 *Leukemia*, **21**, 201-206.
- 49
50 Welk, J.F., et al. (2001) Identification and characterization of the gene encoding human
51 cytoplasmic polyadenylation element binding protein. *Gene*, **263**, 113-120.
- 52
53
54
55
56
57
58
59
60

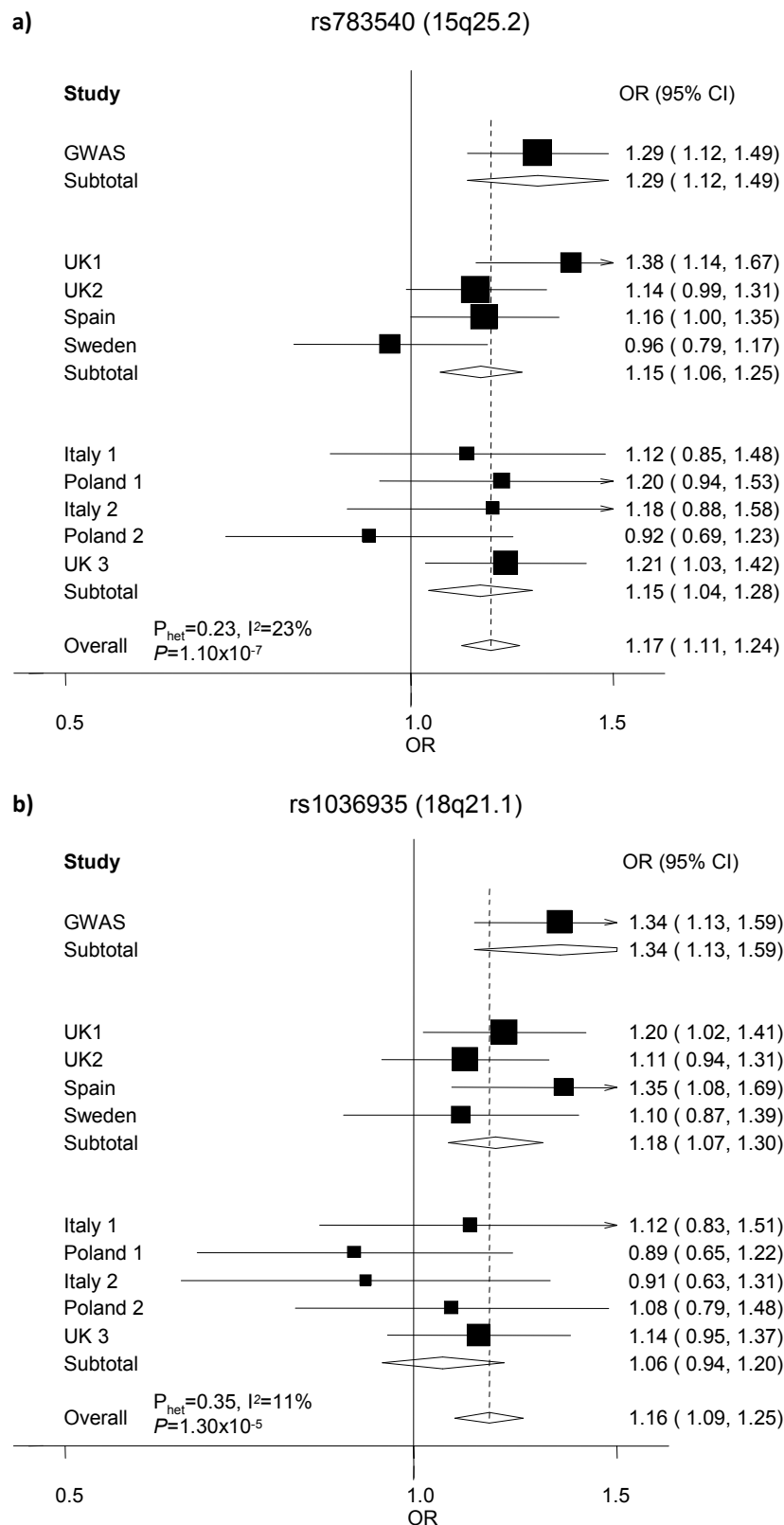


Figure 1: Forest plots of effect size and direction for rs783540 and rs1036935 with CLL risk. Boxes denote per allele odds ratios (OR) point estimates, their areas being proportional to the inverse variance weight of the estimate. Horizontal lines represent 95% confidence intervals. The diamond (and 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). Subjects in the GWAS, UK 1, UK 2, Spanish and Swedish datasets were described previously in (Crowther-Swanepoel, *et al* 2010, Di Bernardo, *et al* 2008).

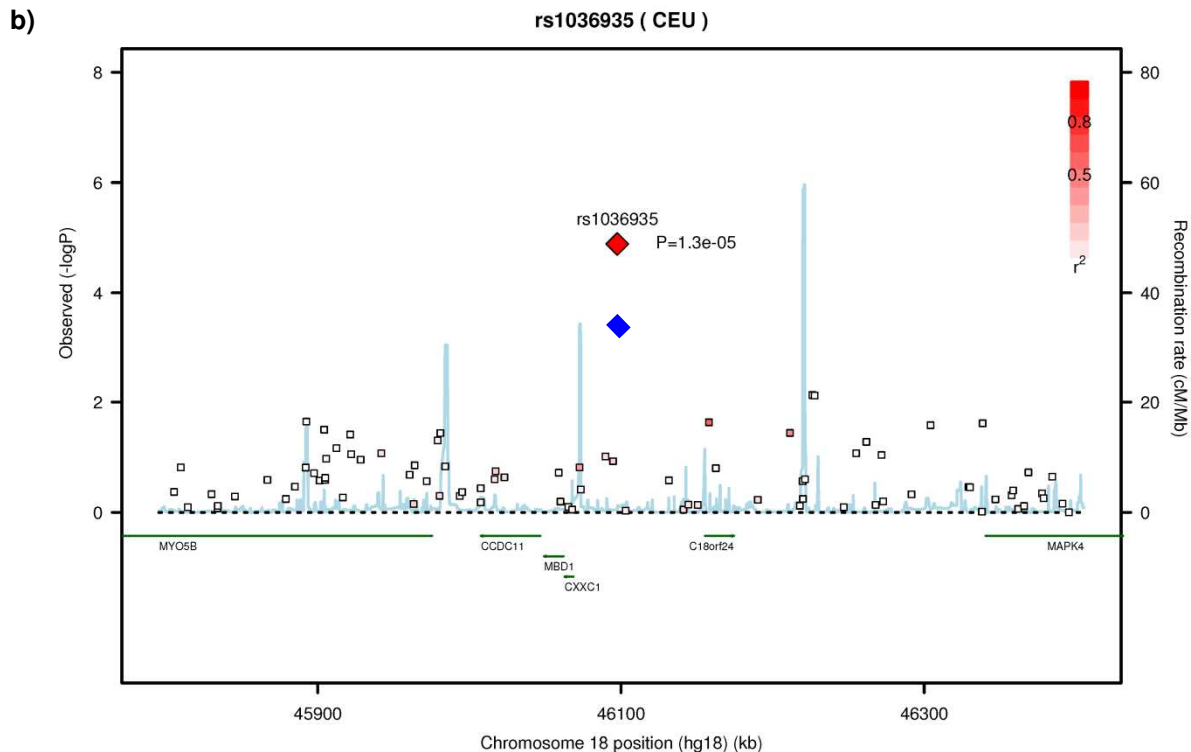
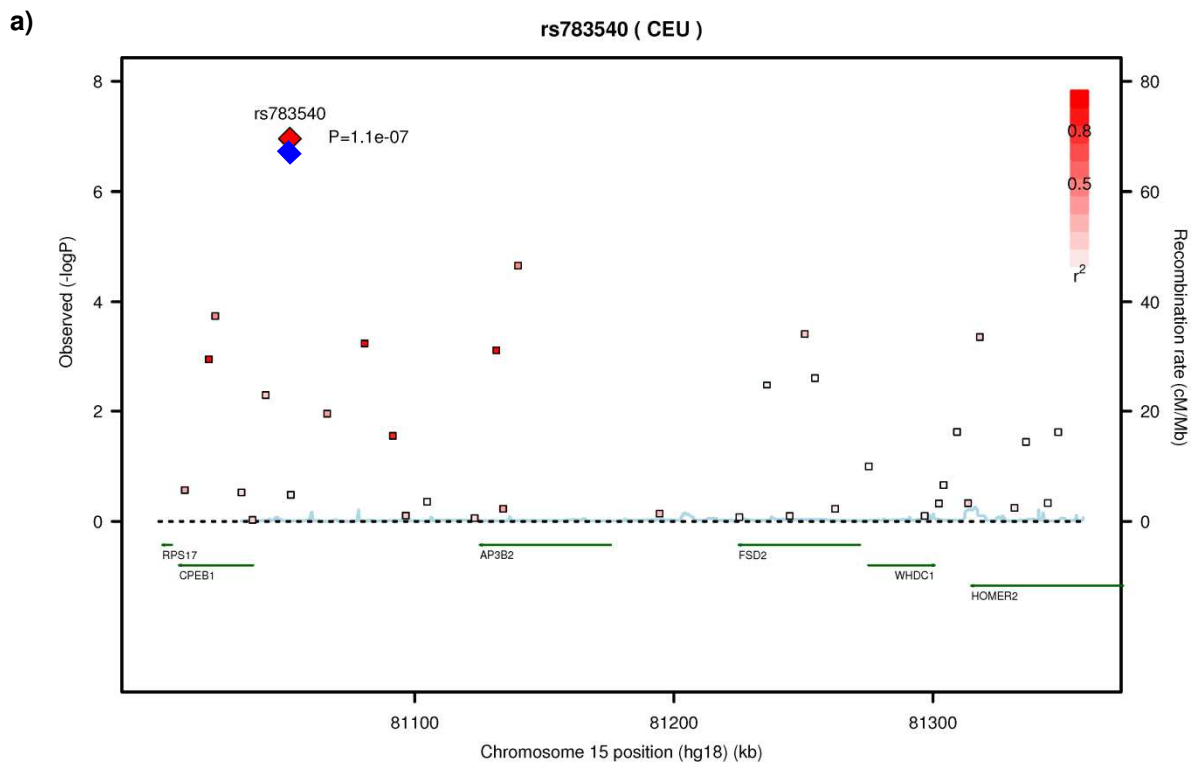
Table 1: Association between SNPs and CLL risk for replication series

Genotype rs783540	Poland 1				Italy 1				Poland 2				Italy 2				UK 3				Combined	
	Ca	Co	OR	(95% CI)	Ca	Co	OR	(95% CI)	Ca	Co	OR	(95% CI)	Ca	Co	OR	(95% CI)	Ca	Co	OR	(95% CI)	OR	(95% CI)
AA	118	70	1.00	Ref	69	79	1.00	Ref	67	134	1.00	Ref	65	64	1.00	Ref	157	411	1.00	Ref	1.00	Ref
AG	149	90	0.98	(0.66-1.46)	107	121	1.01	(0.67-1.53)	94	123	1.53	(1.03-2.26)	91	80	1.12	(0.71-1.77)	194	484	1.05	(0.82-1.35)	1.11	(0.94-1.30)
GG	33	23	0.85	(0.46-1.57)	40	35	1.31	(0.75-2.28)	32	51	1.26	(0.74-2.13)	36	24	1.48	(0.79-2.75)	86	145	1.55	(1.12-2.15)	1.35	(1.09-1.67)*
OR <i>trend</i>			0.92	(0.69-1.22)			1.12	(0.85-1.47)			1.2	(0.93-1.54)			1.18	(0.87-1.58)			1.21	(1.03-1.42)	1.15	(1.03-1.27)*
rs1036935																						
CC	179	116	1.00	Ref	120	136	1.00	Ref	109	158	1.00	Ref	114	105	1.00	Ref	250	651	1.00	Ref	1.00	Ref
CT	104	67	1.01	(0.68-1.48)	79	89	1.01	(0.68-1.49)	70	132	0.77	(0.53-1.23)	71	53	1.23	(0.79-1.92)	159	350	1.18	(0.93-1.50)	1.05	(0.91-1.23)
TT	18	9	1.30	(0.56-2.98)	17	12	1.61	(0.73-3.50)	13	17	1.11	(0.52-2.38)	5	11	0.42	(0.14-1.25)	21	47	1.16	(0.68-1.99)	1.12	(0.81-1.56)
OR <i>trend</i>			1.08	(0.78-1.47)			1.12	(0.83-1.52)			0.89	(0.66-1.21)			0.91	(0.63-1.31)			1.14	(0.94-1.37)	1.06	(0.94-1.19)

* P < 0.05

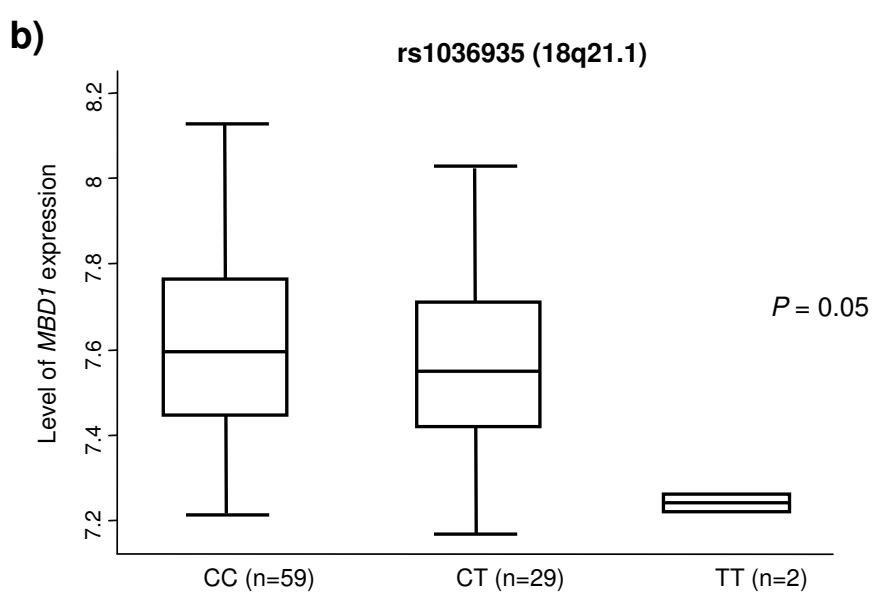
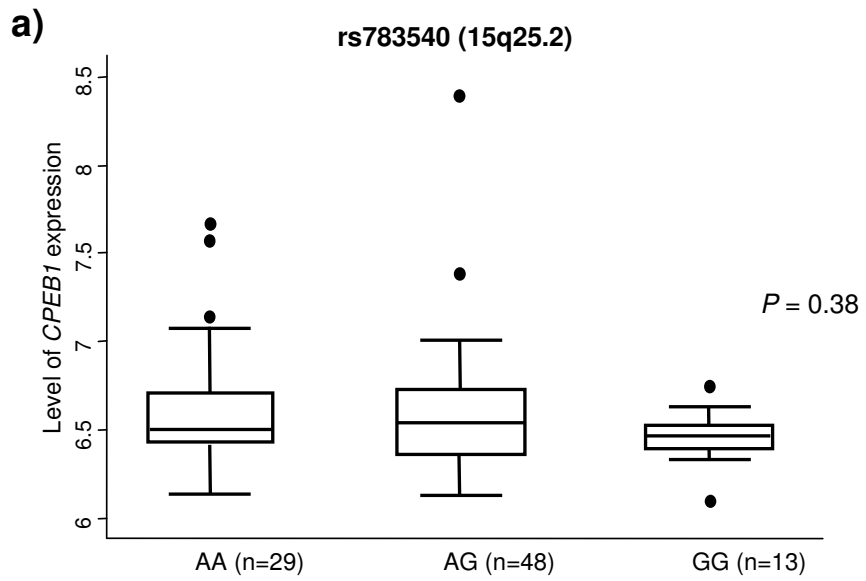
Ca = cases, Co = controls

Peer Review



Supplementary Figure 1: Regional plots of the rs783540 and rs1036935 risk loci. a) The 15q25.2 locus, with the local recombination rate plotted in light blue over this 340-kb chromosomal segment. Each square represents a SNP found in this locus and the most associated SNP in the combined analysis, rs783540, is marked by a blue diamond. The colour intensity of each square reflects the extent of LD with the top SNP - red ($r^2 > 0.8$) through to white ($r^2 < 0.3$). Physical positions are based on build 36 of the human genome. rs783540 is located in intron 2 of *CPEB1*. b) The 18q21.1 locus, with the most associated SNP in this locus, rs1036935, highlighted by a blue diamond. All SNPs found within a 600-kb window centered on rs1036935 are indicated.

1
2
3
4
5
6
7
8
9
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



Supplementary Figure 2: Relationship between lymphocyte mRNA expression levels of a) *CPEB1* and rs783540 genotype and b) *MBD1* and rs1036935 genotype.
Adapted from (Crowther-Swanepoel 2010, *et al*)

Supplementary Table 1: Clinico-pathological association testing. Age and gender are based on data from all ten series, IGHV mutational status is based on data from series GWAS, UK 1, Spain, Sweden, Italy-1 and Italy-2, ZAP-70 is based on series Spain, Italy-1 and Italy-2 data only.

SNP	Age				Gender			IGHV Mutation Status			ZAP-70		
	genotype	<= 63	> 63	<i>P</i>	female	male	<i>P</i>	no	yes	<i>P</i>	<= 20	>20	<i>P</i>
rs783540	AA	650	635	0.240	515	789	0.284	266	296	0.916	124	76	0.461
	AG	877	833		647	1,096		345	438		206	110	
	GG	313	354		252	424		159	163		80	43	
rs1036935	CC	1,068	1,028	0.219	806	1,317	0.477	437	519	0.999	236	122	0.612
	CT	685	678		536	855		287	324		141	97	
	TT	112	127		99	146		48	57		31	14	