Association between single nucleotide polymorphism-genotype and outcome of patients with chronic lymphocytic leukemia in a randomized chemotherapy trial

Rachel Wade,[1] Maria Chiara Di Bernardo,[2] Sue Richards,[1] Davide Rossi,[3] Dalemari Crowther-Swanepoel,[2] Gianluca Gaidano,[3] David G. Oscier,[4] Daniel Catovsky,[5] and Richard S. Houlston[2]

[1]Clinical Trial Service Unit, University of Oxford, Oxford, UK; [2]Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, UK; [3]Division of Hematology, Department of Clinical and Experimental Medicine, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy; [4]Department of Haematology, Royal Bournemouth Hospital, Bournemouth, Dorset, UK, and [5]Section of Haemato-Oncology, Institute of Cancer Research, Sutton, Surrey, UK

ABSTRACT

Background
There is variability in the outcome of patients with chronic lymphocytic leukemia with apparently the same stage of disease. Identifying genetic variants that influence patients' outcome and response to treatment may provide important insights into the biology of the disease.

Design and Methods
We investigated the possibility that genetic variation influences outcome by conducting a genome-wide analysis of 346,831 single nucleotide polymorphisms in 356 patients entered into a phase III trial comparing the efficacy of fludarabine, chlorambucil, and fludarabine with cyclophosphamide as first-line treatment. Genotypes were linked to individual patients' outcome data and response to chemotherapy. The association between genotype and progression-free survival was assessed by Cox regression analysis adjusting for treatment and clinicopathology.

Results
The strongest associations were shown for rs1949733 (*ACOX3*; P=8.22x10-7), rs1342899 (P=7.72x10^{-7}) and rs11158493 (*PPP2R5E*; P=8.50×10^{-7}). In addition, the 52 single nucleotide polymorphisms associated at $P<10^{-4}$ included rs438034 (*CENPF*; P=4.86×10^{-6}), previously correlated with cancer progression, and rs2255235 (*B2M*; P=3.10×10^{-5}) and rs2064501 (*IL22RA2*; P=4.81×10^{-5}) which map to B-cell genes.

Conclusions
Our findings provide evidence that genetic variation is a determinant of progression-free survival of patients with chronic lymphocytic leukemia. Specific associations warrant further analyses. *(Clinicaltrials.gov identifier: NCT00004218)*

Key words: chronic lymphocytic leukemia, SNP, chlorambucil, fludarabine, cyclophosphamide.

Citation: Wade R, Di Bernardo MC, Richards S, Rossi D, Crowther-Swanepoel D, Gaidano G, Oscier DG, Catovsky D, and Houlston RS. Association between single nucleotide polymorphism-genotype and outcome of patients with chronic lymphocytic leukemia in a randomized chemotherapy trial. Haematologica 2011;96(10):1496-1503. doi:10.3324/haematol.2011.043471

©2011 Ferrata Storti Foundation. This is an open-access paper.

RW and MCDB contributed equally to this manuscript.

Acknowledgments: the authors would like to thank all patients and clinicians who participated in the LRF CLL-4 trial.

Funding: this work was primarily supported by a grant from Leukaemia Research and the MRC (G8223452). Additional funding was provided by the Arbib Foundation and Cancer Research UK.

Manuscript received on March 8, 2011. Revised version arrived on May 5, 2011. Manuscript accepted on May 31, 2011.

Correspondence:
Richard S. Houlston, Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK.
Phone: international +44.020.87224175.
Fax: international +44.020.87224059.
E-mail: richard.houlston@icr.ac.uk

The online version of this article has a Supplementary Appendix.

Introduction

Chronic lymphocytic leukemia (CLL) is the commonest lymphoid malignancy in Western populations.[1,2] While staging systems (Binet, Rai) are clinically useful for predicting prognosis and treatment requirements in CLL[3] there is variability in the outcome of patients who apparently have the same stage of disease. More precise assessment of prognosis would, therefore, be beneficial in the choice of specific therapeutic options, while assessment of the likelihood of response and adverse reactions to chemotherapeutic treatments would allow patient-tailored decisions on drug selection and consequent improvements in patients' outlook.

Molecular markers such as immunoglobulin heavy-chain variable region (*IGHV*) mutational status have been shown to greatly assist in defining patients' prognosis[4-6] In addition, genetic variability between patients may also contribute to outcome. To date most searches for polymorphic markers associated with patients' prognosis have been formulated on an "educated guess"; however, without a full understanding of the biology of CLL, definition of what constitutes a "suitable" candidate gene is problematic. This makes agnostic "genome-wide" searches an attractive proposition. Many of the previous candidate studies were based on small numbers of patients ascertained outside the context of any clinical trial. In addition to having limited power such studies are prone to bias as a consequence of survivorship and other confounders, especially when based on retrospectively ascertained patients who did not receive uniform treatment.

We have recently conducted a genome-wide association study of CLL searching for susceptibility alleles for the disease.[7,8] Three hundred and fifty-six of the patients genotyped in this study were participants in a phase III clinical trial. Linking genetic data to information on clinical outcome for these patients allowed us to search for novel genetic variants influencing the clinical behavior of CLL.

Design and Methods

Study population

We made use of a genome-wide association study based on CLL patients entered in the UK Leukemia Research Fund CLL-4 trial. Comprehensive details about the design and conduct of the trial, including drug dosages, have been published elsewhere.[9] Briefly, CLL-4 was a randomized phase III trial established to compare the efficacy of fludarabine, chlorambucil, and the combination of fludarabine plus cyclophosphamide as a first-line treatment for Binet stages B, C and A-progressive CLL. Age was not a criterion for entry into the study. Of the 777 patients entered into the trial the current analysis is based on a random subset of 356 Caucasians who had blood samples taken for clinical diagnostic purposes and cell marker studies at participating centers. *IGHV* mutational status was determined according to BIOMED-2 protocols, using commercial reagents (*InVivoScribe* Technologies, San Diego, USA), as described previously. Clonality was assessed by size discrimination of polymerase chain reaction (PCR) products using semi-automated ABI3730xl/ABI3700 sequencers in conjunction with Genescan software (Applied Biosystems, Foster City, USA). Sequences obtained were submitted to the IMGT/V-QUEST online database. In accordance with published criteria, sequences with a germline identity of 98% or more were classified as unmutated, and those displaying less than 98% identity as mutated.

Subjects for the replication phase

To validate selected associations of single nucleotide polymorphisms (SNP), we made use of outcome data on 380 patients attending the Division of Hematology of the Amedeo Avogadro University of Eastern Piedmont between January 1985 and September 2009.

Ethical approval for the study was obtained from the relevant Research Ethics Committees in each study center and the study was conducted in accordance with the tenets of the Declaration of Helsinki.

Genotyping

DNA was extracted by a standard salt-lysis protocol from venous blood samples collected into EDTA. Overall, 346,831 tagging SNP were genotyped using Illumina Human CNV370-Duo BeadChips (Illumina, San Diego, CA, USA) according to the manufacturer's protocols, as described previously.[7] Subsequent genotyping was conducted by competitive allele-specific PCR KASPar chemistry (KBiosciences Ltd, Hertfordshire, UK; *http://www.kbioscience.co.uk/*); primer sequences and conditions are available on request. As a quality control, we included samples of known genotype and replicates in each assay. A concordance of 99.9% was obtained for both sets of genotyping.

Statistical methods

Progression-free survival was the primary end-point of the analysis. The response recorded for each patient was the best achieved at any time due to first-line treatment. Progression-free survival was defined as the time from the date of randomization to relapse needing further therapy, progression or death from any cause. For non-responders and progressive disease, the date of progression was when no response or progressive disease was recorded. Data on age at diagnosis, Binet stage at diagnosis, sex and treatment were available for all patients, and *IGHV* mutation status was available for a subset of patients, enabling us to examine the significance of each SNP association conditional on these covariates. Cox-regression analysis was used to estimate genotype-specific hazard ratios (HR) and 95% confidence intervals (CI), adjusting for all covariates. For each SNP genotype the hazard ratios were generated using common allele homozygotes as the reference group. The *P*-values presented correspond to the significance of a test difference, trend test, across the three genotype groups (common allele homozygote, heterozygote, rare allele homozygote). For those SNP for which five or fewer minor allele homozygotes were observed, minor allele homozygote genotypes were combined with heterozygotes. If this combined frequency was still five or less then the SNP was removed from the analysis. Hazard ratios, 95% confidence intervals and associated *P* values under both dominant and recessive models were also generated. To assess the distribution of test statistics, we generated quantile-quantile (Q-Q) plots. Progression-free survival curves were plotted using Kaplan-Meier estimates. All statistical analyses were undertaken using R Version 2.7.0 and PLINK.

Bioinformatics

To predict the impact of missense variants on protein function, we applied two *in silico* algorithms, polymorphism phenotyping (PolyPhen)[10] and sifting intolerant from tolerant (SIFT).[11] PolyPhen predicts the functional impact of amino acid changes by considering evolutionary conservation, physicochemical differences, and the proximity of the substitution to predicted functional domains and/or structural features. SIFT predicts the functional importance of amino acid substitutions based on the alignment of orthologous and/or paralogous protein sequences. SIFT and PolyPhen scores were classified as intolerant, potentially intolerant, borderline, or

tolerant and probably damaging, possibly damaging, potentially damaging, borderline, or benign, respectively, according to the classifications proposed by Xi *et al.*[12] and Ng and Henikoff.[11]

Relationship between single nucleotide polymorphism genotype and expression levels

To examine for a relationship between SNP genotype and gene expression we made use of publicly available expression data generated from an analysis of 90 Caucasian-derived Epstein-Barr virus–transformed lymphoblastoid cell lines using Sentrix Human-6 Expression BeadChips (Illumina, San Diego, USA).[13,14] Online data were recovered using WGAViewer version 1.25 software. Differences in the distribution of levels of mRNA expression between SNP genotypes were compared using an extension of Wilcoxon's rank-sum test for trends across ordered groups.[15]

Results

Patients' characteristics

A third of the patients had been diagnosed with CLL before 60 years old and approximately two-thirds had presented with stage B or C disease. The median follow-up time for patients was 78 months with follow-up to October 31, 2008. Forty-nine percent of the cohort died. Only drug therapy significantly influenced progression-free survival. Specifically, the combination of fludarabine and cyclophosphamide as drug therapy was significantly associated with a more favorable prognosis (Table 1). For 305 (86%) of the 356 patients we had information on *IGHV* mutation status. As in the parent trial this covariate was prognostically important, being highly predictive of progression-free survival ($P<0.0001$). There was no difference in the demographics, treatment, and follow-up characteristics of the 356 cases genotyped in this study compared to those of the patients entered into CLL-4 but who were excluded from this analysis (*Online Supplementary Table S1*). It is, therefore, unlikely that any spurious biases have influenced our findings. The observation that conventional staging and other markers of prognosis were equally predictive of survival in the complete cohort of patients and in the subgroup analyzed provides further evidence that sub-selection of the genotyped cases is unlikely to have biased the overall findings of the study.

Table 1. Relationship between treatment, stage, age and sex, and progression-free survival in the patients genotyped.

		Progression-free survival		
	N.	Progression or death	O/E*	*P* value**
Treatment				
Chlorambucil	171	158	1.38	<0.0001
Fludarabine	91	83	1.19	
Fludarabine & cyclophosphamide	94	62	0.52	
Stage				
A	93	78	1.00	0.4
B	156	133	0.93	
C	107	92	1.13	
Age				
Less than 60 years	107	92	1.03	0.95
60 to 69 years	147	124	0.96	
Over 70 years	102	87	1.02	
Sex				
Male	265	230	1.05	0.2
Female	91	73	0.87	

*O/E indicates observed versus expected ratio. ** log-rank test for trend, except treatment (heterogeneity).

Genotyping quality control

SNP call rates per sample were greater than 97.8% in patients. Of the 346,831 SNP submitted for analysis, 345,665 were satisfactorily genotyped (99.7%), with mean individual SNP call rates of 99.6%. Of the 345,665 SNP loci that were satisfactorily genotyped, 7,165 were fixed in all patients' samples and 12,115 were either observed at sub-polymorphic frequencies (i.e. having a minor allele frequency <1%) or departed from Hardy-Weinberg equilibrium ($P\leq10^{-5}$), and were, therefore, excluded from all analyses. Of the 326,385 remaining SNP to be analyzed, cell counts of minor allele genotypes were sufficiently infrequent to exclude 65,909 from recessive model analyses.

Association between single nucleotide polymorphisms and progression-free survival

We examined the overall effect of patients' genotype on progression-free survival by generating a Q-Q plot of observed test statistics compared to the mean of the corresponding expected values and calculating an overdispersion factor, λ.[16] After adjustment for sex, age, stage of disease at presentation and drug regime, a Q-Q plot of test statistics of the association between SNP genotype and progression-free survival showed a significant deviation from the expected distribution with λ=1.071616 (95% CI: 1.071570-1.071661, Figure 1); this inflation reflects an over-representation of test statistics, indicative of an association between genotype and progression-free survival globally.

Fifty-two SNP (mapping to 38 genomic regions) showed an association with progression-free survival with a *P* value of 10^{-4} or less, significantly greater than that expected simply by chance ($P<0.05$; Table 2). Three of the 52 associations were of borderline statistical significance,

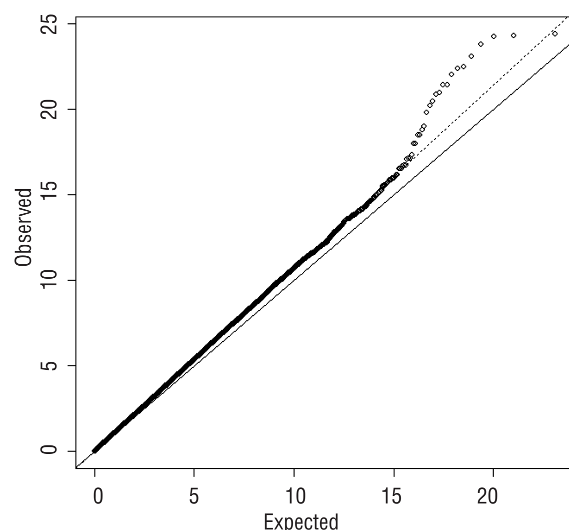

Figure 1. Quantile-quantile plot for progression-free survival from the genome-wide association study. The dotted line represents the observed *P* values. The solid line represents the expected line under the null distribution.

when Bonferroni's correction was applied (*i.e.* $P \leq 2.0\times10^{-7}$): rs1949733 ($P=8.22\times10^{-7}$), rs1342899 ($P=7.72\times10^{-7}$) and rs11158493 ($P=8.50\times10^{-7}$).

For 305 (86%) of the 356 patients, we had information on *IGHV* mutational status, allowing us to evaluate the impact of polymorphic variation adjusting for this clinically important prognostic marker. In a multivariate analysis 13 of the 52 SNP retained significance at the 10^{-4} threshold, and a further 26 showed evidence of an association with progression-free survival at the 10^{-3} threshold: these are annotated using asterisks in Table 2; all SNP retained conventional significance ($P<0.05$).

For purposes of clarity we have restricted our commentary here to the major loci defined by the 52 SNP. Of these 52 SNP, rs438034 has previously been reported to influence the prognosis of patients with breast cancer[17] and a further three SNP tag genes with a known or presumptive biological role in B-cell biology or tumor behavior (Table 2).

Centromeric protein F (*CENPF*, MIM 600236) is involved in the maintenance of chromosomal stability. Under the Cox proportional hazards model homozygosity for *CENPF* 2943T (rs438034) was associated with a poorer progression-free survival (HR=1.88; 95% CI: 1.44-2.46; $P=3.10\times10^{-6}$). The 5-year progression-free survival for T homozygotes was 8.9% (95% CI: 4.3-18.3%) compared to 19.9% (95% CI: 15.8-25.3%) for heterozygotes and C homozygotes (Figure 2A). An association between progression-free survival and the *CENPF* SNP rs3748697 (N2396D) was also shown (HR=0.64, 95% CI: 0.49-0.84; $P=1.46\times10^{-5}$; Table 2). rs438034 and rs3748697 are highly correlated, with LD metrics: $r^2=0.934$, D'=0.967.

rs2064501 maps to intron 4 of the gene encoding interleukin 22 receptor, alpha-2 (*IL22RA2*, MIM 606648). *IL22RA2* participates in the production of acute-phase reactants via interleukin-22, secreted by T cells. Under a dominant model, carrier status for the T allele of rs2064501 was associated with a more favorable prognosis (HR=0.56; 95% CI: 0.43-0.72; $P=1.00\times10^{-5}$). The 5-year progression-free survival rate for carriers of the T minor allele was 21.1% (95% CI: 16.6-26.8%) compared to 7.7% (95% CI: 3.7-15.7%) for non-carriers (Figure 2B).

Circulating levels of β_2 microglobulin (*B2M*, MIM 109700) have been reported to influence patients' outcome in CLL.[18] This is supported by an analysis of the data from the CLL-4 trial in which high levels of β_2 microglobulin (defined as ≥4 mg/L) were significantly associated with poorer progression-free survival ($P=0.0001$).[19] rs2255235 mapping 5' to *B2M* provided evidence of an association between genetic variation at this locus and progression-free survival ($P=3.10\times10^{-5}$). The hazard ratios for heterozygosity and TT homozygosity were 1.63 (95% CI: 1.23-2.16) and 3.07 (95% CI: 1.41-6.67), respectively ($P=3.10\times10^{-5}$). The 5-year progression-free survival associated with GG homozygosity was 18.7% (95% CI: 14.6-23.9%) compared to 13.2% (95% CI: 7.2-24.3%) for GT heterozygosity and 14.3% (95% CI: 2.3-87.7%) for TT homozygosity (Figure 2C). While a relationship between TT genotype and poor progression-free survival was shown, particularly striking in the early years, this observation was based on only seven patients. To investigate whether the rs2255235 genotype influenced β_2 microglobulin level we made use of data from 239 of the 356 genotyped individuals for whom β_2 microglobulin had been previously measured.[19] There was no difference in the β_2 microglobulin levels measured, or the quantity of missing β_2 microglobulin data between the genotyped and non-genotyped patients, and a relationship between β_2 microglobulin and progression-free survival was apparent in all groups. However, no statistically significant association between genotype and β_2 microglobulin level was observed ($P=0.6$).

An association was shown for three highly correlated SNP ($r^2>0.7$; rs3783741, rs1012145, rs11158493) mapping to protein phosphatase 2A, regulatory subunit B (B56), epsilon isoform (*PPP2R5E*, MIM 601647), the strongest evidence being provided by rs11158493 ($P=8.50\times10^{-7}$) which maps to intron 2 of *PPP2R5E*; moreover under a dominant model the association was statistically significant at the genome-wide threshold ($P=7.12\times10^{-8}$), with carrier status for the T minor allele conferring a more favorable progression-free survival (HR=0.50; 95% CI: 0.39-0.64). The 5-year progression-free survival rate for carriers of the T allele was 28.1% (95% CI: 21.1-37.4%) compared to 12.3% (95% CI: 8.7-17.3%) for non-carriers (Figure 2D). *PPP2R5E* has been implicated in a variety of regulatory processes including cell growth and gene transcription. Expression of *PPP2R5E* was correlated with rs3783741 and rs1012145 genotype ($P=0.015$) with increased expression being associated with risk genotype (Figure 3; *Online Supplementary Table S2*). Paradoxically, expression of *PPP2R5C* has been reported to be down-regulated in progressive CLL.[20]

Five correlated SNP mapping to 4p16.1 showed an association with progression-free survival (rs747580, rs2631731, rs940136, rs2631753, rs1949733), the strongest association being shown for rs1949733 ($P=8.22\times10^{-7}$). Hazard ratios for heterozygosity and AA homozygosity were 1.54 (95% CI: 1.21-1.97), and 2.36 (95% CI: 1.60-3.47), respectively; corresponding 5-year progression-free survival rates were 24.4% (95% CI: 18.9-31.5%), 11.9% (95% CI: 7.5-18.8%) and 3.1% (95% CI: 0.5-21.5%), respectively (Figure 2E). rs1949733 maps centromeric to a number of genes including acyl-coenzyme A oxidase 3, pristanoyl isoform b (*ACOX3*, MIM 603402) and predicted transcripts (e.g. C4orf23) which may be the basis for the observed association. This assertion is supported by the strong relationship found between rs747580 genotype and *ACOX3* expression in Epstein-Barr virus-transformed lymphocytes ($P=1.08\times10^{-6}$; *Online Supplementary Table S2*).

Adenosine deaminase, RNA-specific, b2 (*ADARB2*, MIM 602065) RNA-editing deaminase-2 (*RED2*, or *ADARB2*) is a member of the double-stranded RNA adenosine deaminases which play a role in RNA-editing, a process which has been implicated as an additional epigenetic mechanism relevant to cancer development and progression.[21] Two SNP, rs10903420 and rs1007147, which are in perfect LD and map to intron 3 of *ADARB2* provided evidence of an association ($P=3.00\times10^{-5}$; HR=0.58, 95% CI: 0.46-0.75). The 5-year progression-free survival associated with rs10903420 was 21.8% (95% CI: 17.3-27.4%) for carriers of the A minor allele, compared with 5.2% (95% CI: 2.0-13.5%) for non-carriers (Figure 2F).

Strong evidence for an association between progression-free survival and variation at 9q31.1 was provided by rs1342899 ($P=7.72\times10^{-7}$). Hazard ratios associated with heterozygosity and homozygosity for the A minor allele were 0.68 (95% CI: 0.54-0.86) and 0.32 (95% CI: 0.19-0.54) respectively, with 5-year progression-free survival rates of 12.4% (95% CI: 8.3-18.3%) for GG homozygotes,

Table 2. SNP showing significant (at a threshold of 10^{-4}) genotypic association with progression-free survival.

SNP	Minor allele	Gene	Chr	Position (bp)	*P* value trend test	HR_{het}	95% CI	HR	95% CI	N_{AA}	N_{Aa}	N_{aa}
rs686053	C		1	46,987,372	4.43E-05 *			1.69[d]	1.32-2.22	265	86	5
rs3748697	A	*CENPF*	1	212,886,722	1.46E-05 *†	0.64	0.49-0.84	0.49	0.36-0.68	100	171	85
rs438034	T	*CENPF*	1	212,897,240	4.86E-06 †			1.88[r]	1.44-2.46	109	170	77
rs335555	T		1	212,956,482	2.23E-05	1.29	0.99-1.69	2.00	1.46-2.72	109	166	81
rs2793086	C	*DISC1*	1	229,966,327	8.53E-06 †	1.57	1.23-2.01	3.01	1.51-6.01	241	106	9
rs1439887	G		2	188,781,591	5.82E-05 **†	1.39	1.09-1.77	2.04	1.39-2.98	167	152	37
rs6785504	T		3	154,155,469	6.95E-06 **†	1.55	1.18-2.05	2.11	1.51-2.94	101	179	75
rs9883654	T		3	154,204,665	1.54E-06 **†	1.54	1.19-1.98	2.20	1.57-3.08	132	168	55
rs747580	T	*ACOX3*	4	8,423,988	1.32E-05 *†	1.46	1.13-1.88	2.06	1.47-2.90	145	157	52
rs2631731	T	*C4orf23; ACOX3*	4	8,480,717	1.08E-06 **†	1.59	1.24-2.02	2.28	1.52-3.41	183	144	29
rs940136	G	*C4orf23; ACOX3*	4	8,499,892	6.01E-06 *†	1.47	1.15-1.88	2.21	1.53-3.20	153	164	38
rs2631753	T	*C4orf23; ACOX3*	4	8,506,145	3.52E-05 *†	0.65	0.49-0.85	0.50	0.37-0.70	93	171	87
rs1949733	A	*C4orf23; ACOX3*	4	8,554,259	8.22E-07 **†	1.54	1.21-1.97	2.36	1.60-3.47	184	140	32
rs3103078	T		4	8,575,667	7.90E-05 *†	1.41	1.08-1.86	1.87	1.36-2.55	107	168	81
rs1879724	G		4	190,948,130	3.51E-05 *	0.70	0.55-0.89	0.47	0.32-0.70	141	174	41
rs6844114	A		4	190,982,886	6.41E-05			1.75[d]	1.33-2.33	291	61	4
rs7676557	G		4	190,993,476	9.98E-05 *	1.46	1.10-1.94	1.92	1.38-2.67	90	186	79
rs773765	C	*TRIO*	5	14,344,614	6.38E-05 †			2.22[d]	1.52-3.33	326	30	0
rs4132580	G		5	123,969,149	5.84E-05 †	1.43	1.11-1.85	2.98	1.67-5.31	244	99	13
rs153753	T	*C5orf50*	5	171,117,576	4.31E-05 *	1.54	1.21-1.97	1.90	1.30-2.79	158	158	40
rs6457160	A	*NEDD9*	6	11,347,010	9.93E-05 †			1.72[r]	1.31-2.26	98	179	78
rs2064501	T	*IL22RA2*	6	137,519,516	4.81E-05			0.56[d]	0.43-0.72	98	177	78
rs215702	G		7	32,366,183	1.72E-05 **	1.43	1.12-1.84	2.11	1.47-3.02	150	160	46
rs10958369	G		8	54,572,761	2.11E-06 ††	0.71	0.55-0.93	0.43	0.3-0.61	106	184	65
rs10504154	T		8	54,591,549	2.22E-06 *††	0.76	0.58-0.99	0.43	0.31-0.61	98	182	76
rs13266634	T	*SLC30A8*	8	118,253,964	8.25E-05 *	1.39	1.09-1.76	2.22	1.46-3.38	156	170	30
rs10990381	G		9	104,725,086	2.69E-06 *	0.69	0.54-0.87	0.39	0.25-0.61	173	149	33
rs1342899	A		9	104,775,016	7.72E-07 *	0.68	0.54-0.86	0.32	0.19-0.54	177	151	27
rs1158809	G		9	120,849,439	6.85E-05 *	1.57	1.22-2.02	2.09	1.13-3.89	242	101	12
rs6478823	A	LCN2; *C9orf16*; CIZ1	9	129,946,669	6.50E-05 *	0.76	0.60-0.96	0.47	0.31-0.69	152	155	47
rs10903420	A	*ADARB2*	10	1,333,726	6.83E-05 **			0.58[d]	0.46-0.75	94	184	78
rs1007147	A	*ADARB2*	10	1,341,088	6.83E-05 **			0.58[d]	0.46-0.75	94	184	78
rs2935284	A		10	4,091,955	4.83E-05 *††			3.22[r]	1.98-5.24	251	86	19
rs3759225	C	*TMEM16B*	12	5,555,662	7.25E-05 *			1.72[d]	0.45-0.76	279	73	4
rs2252690	C		12	53,629,350	9.24E-05 †	1.40	1.10-1.78	2.24	1.43-3.50	189	141	23
rs2028809	T		13	43,421,136	7.66E-05 *†	0.64	0.49-0.84	0.44	0.25-0.80	243	94	17
rs4255691	T		14	31,053,122	8.20E-05	0.60	0.45-0.80	0.30	0.10-0.96	272	77	6
rs1168987	A		14	35,467,431	4.61E-06 **†			2.27[d]	0.31-0.63	316	39	1
rs3783741	A	*PPP2R5E*	14	62,991,425	3.64E-06 *			0.53[d]	0.41-0.68	240	108	8
rs1012145	A	*PPP2R5E*	14	63,009,835	3.64E-06 *			0.53[d]	0.41-0.68	240	108	8
rs11158493	T	*PPP2R5E*	14	63,023,275	8.50E-07 **			0.50[d]	0.39-0.64	235	110	11
rs7141841	C		14	98,621,051	4.83E-05 *	0.64	0.49-0.83	0.35	0.15-0.79	251	94	11
rs2255235	T	*B2M*	15	42,790,656	3.10E-05 *†	1.63	1.23-2.16	3.07	1.41-6.67	281	68	7
rs4417505	A	*FAM81A*	15	57,571,940	6.12E-05 *			1.64[d]	1.30-2.04	214	117	25
rs2194165	A	*GSG1L*	16	27,834,335	7.84E-05 *	0.75	0.58-0.96	0.53	0.38-0.73	132	156	68
rs2061450	T		18	58,911,792	2.19E-05	1.51	1.19-1.92	2.25	1.38-3.69	207	130	18
rs714789	G		18	69,712,477	4.33E-05 **	1.44	1.13-1.84	2.03	1.39-2.98	154	163	39
rs2228671	T	*LDLR*	19	11,071,912	1.71E-05 **			1.75[d]	1.35-2.27	269	82	5
rs4808047	C	*EPS15L1*	19	16,388,834	8.71E-05 *††	1.62	1.23-2.12	2.01	1.11-3.64	262	81	12
rs6056665	A	*PAK7*; *C20orf103*	20	9,452,468	9.33E-05 **†			1.64[d]	1.28-2.08	256	96	2
rs3746414	A	*ZFP64*	20	50,202,786	3.53E-05 *†			1.64[d]	1.32-2.08	212	129	14
rs1800522	C	*PFKL*; *AIRE*	21	44,541,978	7.52E-05	1.42	1.10-1.84	1.89	1.36-2.61	126	166	64

*Hazard ratios [HR] adjusted for age, sex, stage of disease and treatment in all analyses. **still significant at 10^{-4} threshold after adjustment for* IGHV *mutation status. *still significant at 10^{-3} threshold after adjustment for* IGHV *mutation status. ††significant association with overall survival at 10^{-3} threshold, trend test. †significant association with overall survival at 0.05 threshold, trend test. d,r indicates whether a dominant or recessive model was the most significant model rather than codominant*

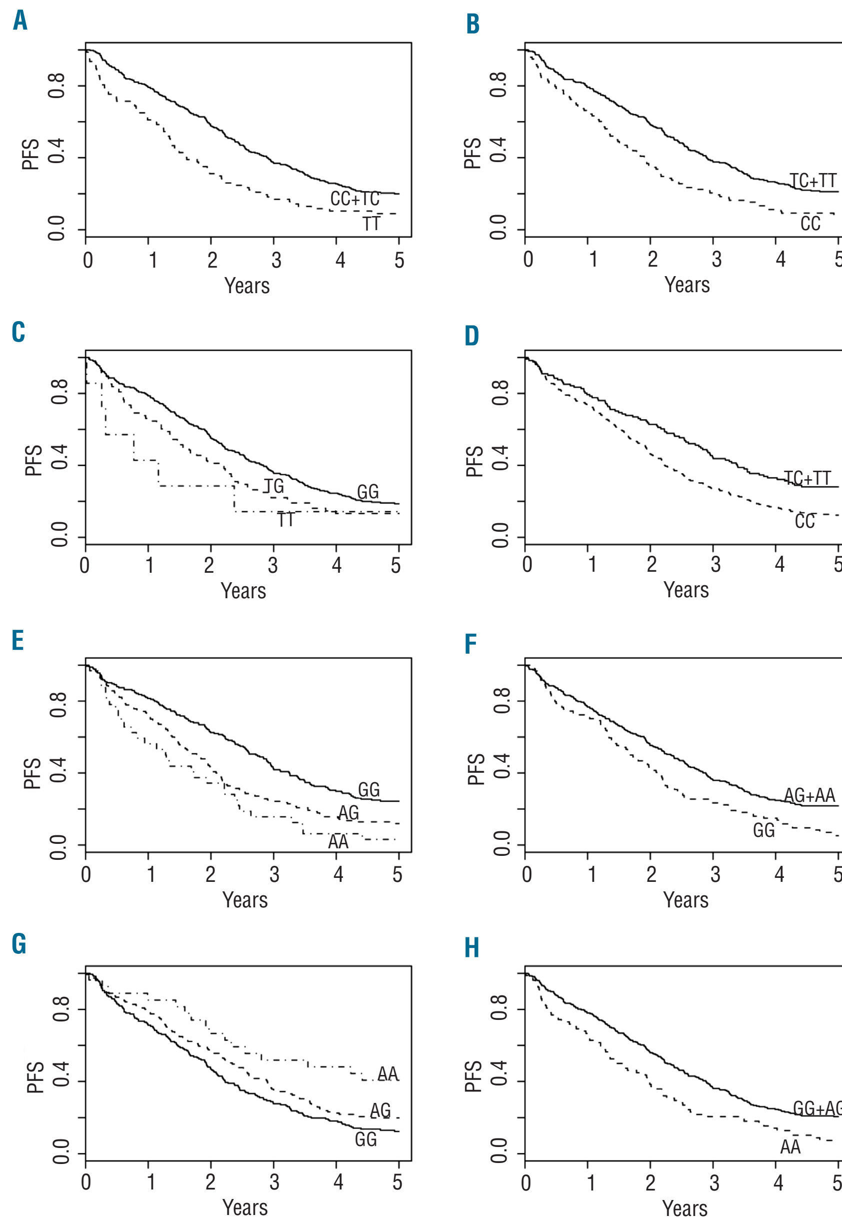

Figure 2. Kaplan-Meier curves. Progression-free survival (PFS) curves for (A) rs438034 (*CENPF* R2943G), (B) rs2064501 (*IL22RA2*), (C) rs2255235, (D) rs11158493 (*PPP2R5E*), (E) rs1949733, (F) rs10903420 (*ADARB2*), (G) rs1342899 and (H) rs6457160. (A) Progression-free survival curves associated with carriers of the C allele are shown as a solid line. The dashed line depicts the survival curve for those with the at risk variant genotype. (B) Progression-free survival curves associated with carriers of the T allele are shown as a solid line. The dashed line depicts the survival curve for those with the at risk variant genotype. (C) Progression-free survival curves associated with G homozygosity are shown as a solid line. The dashed and broken lines depict the survival curves for heterozygosity and A homozygosity respectively. (D) Lines are as defined in (B). (E) Progression-free survival curves associated with G homozygosity are shown as a solid line. The dashed and broken lines depict the survival curves for heterozygosity and A homozygosity, respectively. (F) Progression-free survival curves associated with carriers of the A allele are shown as a solid line. The dashed line depicts the survival curve for those with the at risk variant genotype. (G) Lines are as defined in (E). (H) Progression-free survival curves associated with carriers of the G allele are shown as a solid line. The dashed line depicts the survival curve for those with at risk variant genotype.

19.8% (95% CI: 14.3-27.3) for heterozygotes and 40.7% (95% CI: 25.9-64.2%) for AA homozygotes (Figure 2G).

Evidence for the role of genetic variation in the gene neural precursor cell expressed developmentally downregulated 9 (*NEDD9*, MIM 602265) was provided by rs6457160 (P=8.60×10^{-5}). *NEDD9* is an important component of a cytoskeleton-linked signaling pathway initiated by integrins regulating FAK and cell spreading, and has been shown to play a role in response to extracellular cues that drive invasion and metastasis.[22] Homozygosity for the A minor allele was associated with poorer progression-free survival (HR=1.72; 95%CI: 1.31-2.26) and a 5-year progression-free survival of 7.3% (95% CI: 3.3%-16.4%) compared to 20.5% (95% CI: 16.2%-25.9%) for heterozygotes and G homozygotes (Figure 2H).

We have previously shown that polymorphic variation at 2q13 (rs17483466), 2q37.1 (rs13397985, *SP140*), 6p25.3 (rs872071, *IRF4*), 11q24.1 (rs735665), 15q23 (rs7176508) and 19q13.32 (rs11083846, *PRKD2*) influences the risk of CLL.[7] Associations with progression-free survival (P=0.85, 0.74, 0.98, 0.94, 0.55 and 0.37, respectively) provided no evidence that these variants influence survival of patients with CLL.

Relationship between single nucleotide polymorphisms and chemotherapy

We examined for potential interactive effects between various SNP and response to chemotherapy on progression-free survival. Four of the 52 SNP showed evidence for an interaction at the 0.05 threshold: rs3103078 mapping to 4p16.1, rs6457160 mapping to *NEDD9*, rs215702 mapping to 7p14.3, and rs2061450 mapping to 18q21.33. *Online Supplementary Table S3* details the relationship between interactions for these four variants. For both rs6457160 and rs215702, the impact of genotype was greater in patients treated with fludarabine (HR=2.02, 95% CI: 1.44-2.84 and HR=1.79, 95% CI: 1.31-2.44, respectively) than in those treated with chlorambucil (HR=1.37, 95% CI: 1.09-1.71 and HR=1.53, 95% CI: 1.20-1.95, respectively) or fludarabine and cyclophosphamide (HR=0.96, 95% CI:

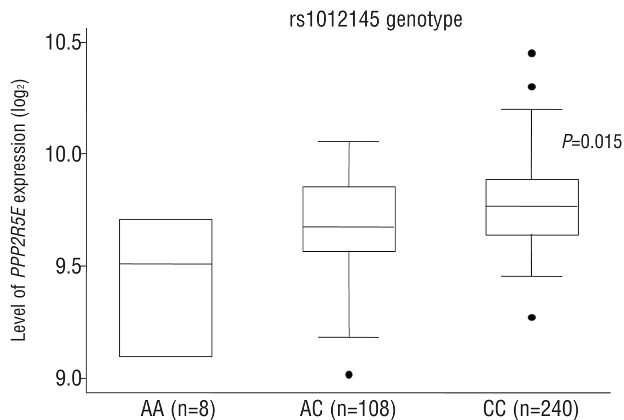

Figure 3. Relationship between lymphocyte mRNA expression levels of *PPP2R5E* and rs1012145 genotype.

0.66-1.40 and HR=1.01, 95% CI: 0.71-1.44, respectively). For rs2061450 the impact of genotype was greater in patients treated with the alkylating agent chlorambucil (HR=1.84, 95% CI: 1.42-2.38) than in those treated with purine analogs (HR for fludarabine=1.45, 95% CI: 1.02-2.08; HR for fludarabine + cyclophosphamide=0.98, 95% CI: 0.64-1.50). For rs3103078 the impact of genotype was greater in patients treated with the combination of fludarabine and cyclophosphamide (HR=1.90, 95% CI: 1.35-2.67), than in those treated with chlorambucil (HR=1.40, 95% CI: 1.13-1.73) or fludarabine monotherapy (HR=0.97, 95% CI: 0.70-1.33).

Replication results

We looked further at ten noteworthy SNP selected on the basis of *P* values and robustness of associations after adjustment for *IGHV* status (rs438034, rs9883654, rs1949733, rs10504154, rs10990381, rs1342899, rs1168987, rs11158493, rs2255235 and rs2228671), using an independent series of 380 CLL Italian patients, seeking to replicate findings. After a median follow-up of 67 months, 172 of the patients had required treatment and 90 had died. Statistical analysis of the relationship between SNP genotype and progression-free survival was made by multivariate analysis adjusting for age, sex, Binet stage, *IGHV* status, CD38 and *TP53* mutation. While for some SNP there was evidence for an association with progression-free survival consistent with that seen in the clinical trial, none of the associations was statistically significant (*i.e.* $P>0.05$; *Online Supplementary Table S4*).

Discussion

Genome-wide SNP genotype data generated from 356 patients within the CLL-4 trial have allowed us to gain insight into determinants influencing the biology of CLL and identify a number of genetic variants associated with progression-free survival. Notably a number of the SNP map to genes previously implicated as prognostic markers for malignancy, *CENPF*, and genes with a proven role in B-cell biology, *B2M* and *IL22RA2*, or with metastatic potential, *NEDD9*.

A major strength of our study is that it was based on a large series of patients entered into a phase III randomized trial thus minimizing biases. In addition, our analysis is unlikely to have been confounded by population stratification as we excluded non-Western European ancestral cases from the analysis. Furthermore we were able to adjust for the major molecular covariate, *IGHV* mutation status, and identify SNP independently predictive of progression-free survival.

It is intriguing that a number of the association signals map to genes with biological plausibility as determinants of prognosis. Furthermore, *CENPF* 2943T carrier status has previously been shown to be associated with a poorer prognosis in patients with breast cancer suggesting a generic effect on tumor prognosis.[17] As the SNP maps to exon 18 of the gene and the C>T change leads to substitution of glycine for arginine, the SNP may be directly functional. This does not, however, exclude the possibility that the association may be a consequence of LD with a causal variant, as the sequence change is predicted to be tolerated. Indeed excluding *CENPF* R2943G (rs438034), *SLC30A8* R325W (rs13266634), and *ZFP64* S451N (rs3746414), the majority of the SNP genotyped are generally not themselves candidates for causality. Accepting the caveat that HapMap is not comprehensive, we interrogated HapMap to identify non-synonymous SNP highly correlated ($r^2>0.8$) with the most strongly associated SNP. In this analysis we did not identify any correlated non-synonymous SNP within associated genes. A failure to demonstrate association suggests that many of the associations identified so far are mediated through LD with sequence changes that influence gene expression rather than protein sequences or through LD with low frequency variants (i.e. variants with minor allele frequencies of 0.01-0.05) that are not catalogued by HapMap. Such an assertion is supported by data showing an association between genotype and mRNA expression of the respective transcript in Epstein-Barr virus-lymphoblastoid cell lines.

Our study also provides insight into the genetic architecture of variants likely to influence patients' prognosis; specifically, our findings suggest that there are unlikely to be many common loci conferring hazard ratios of greater than 1.5 as the power of our analysis to detect this class of variant was high. Hence, if patients' prognosis is significantly influenced by genotype, by implication it is probable that a large number of variants influence the patients' outcome.

Although our study demonstrates strong associations between a number of SNP and progression-free survival, with biological plausibility, there are limitations to our findings. A major one is that replication analysis did not attain statistically significant levels. This failure of replication is not necessarily unexpected, given differences between the two case series. As previously discussed, replication studies face the risk of non-validation because of the lack of generalizability,[23,24] especially when there can be significant differences in individual patients' management. Our discovery cases were UK patients participating in a large phase III clinical trial: each patient had, therefore, received standardized management and treatment. In contrast, the cases in the replication phase were patients undergoing routine clinical care and thus likely to have had differing treatment and management over the period of ascertainment. This is reflected in the characteristics of the two groups of patients, who showed significant differences in survival time.

Furthermore, there is the issue of genetic heterogeneity between the two sets of patients which can erode the probability of demonstrating replication. Since most SNP associations are unlikely to be directly causal, replication of associations is contingent on patterns of LD being consistent between populations. Differences in LD structure between northern and southern European populations are well documented. Evidence for this and thus a possible reason for failure to replicate findings is given by the fact that allele frequencies for two of the ten SNP (P<0.05) were significantly different between the two sets of patients studied.

In conclusion, accepting the above caveats we have provided evidence implicating variations in a number of genes as a possible determinant of the outcome of patients with CLL. While any germline variant is unlikely to replace staging schemes and conventional molecular markers, they do have potential to assist in distinguishing different outcome patterns among patients with the same stage of disease, opening up the possibility of a rational, targeted approach to treatment based on a combination of the patients' genotype and tumor characteristics.

Authorship and Disclosures

The information provided by the authors about contributions from persons listed as authors and in acknowledgments is available with the full text of this paper at www.haematologica.org.

Financial and other disclosures provided by the authors using the ICMJE (www.icmje.org) Uniform Format for Disclosure of Competing Interests are also available at www.haematologica.org.

References

1. Stevenson FK, Caligaris-Cappio F. Chronic lymphocytic leukemia: revelations from the B-cell receptor. Blood. 2004;103(12): 4389-95.
2. Morton LM, Turner JJ, Cerhan JR, Linet MS, Treseler PA, Clarke CA, 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(2):695-708.
3. Binet JL. Prognostic factors in chronic lymphocytic leukaemia. Haematologica. 1999;84 (Suppl EHA-4):96-7.
4. Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. N Engl J Med. 2000;343(26):1910-6.
5. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood. 1999;94(6):1848-54.
6. Del Poeta G, Maurillo L, Venditti A, Buccisano F, Epiceno AM, Capelli G, et al. Clinical significance of CD38 expression in chronic lymphocytic leukemia. Blood. 2001;98(9):2633-9.
7. Di Bernardo MC, Crowther-Swanepoel D, Broderick P, Webb E, Sellick G, Wild R, et al. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. Nat Genet. 2008;40(10): 1204-10.
8. Crowther-Swanepoel D, Broderick P, Di Bernardo MC, Dobbins SE, Torres M, Mansouri M, et al. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. Nat Genet. 2010;42(2):132-6.
9. Catovsky D, Richards S, Matutes E, Oscier D, Dyer MJ, Bezares RF, et al. Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. Lancet. 2007;370 (9583):230-9.
10. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res. 2002;30(17):3894-900.
11. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001;11(5):863-74.
12. Xi T, Jones IM, Mohrenweiser HW. Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. Genomics. 2004;83(6):970-9.
13. Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, et al. Genome-wide associations of gene expression variation in humans. PLoS Genet. 2005;1(6):e78.
14. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science. 2007;315(5813):848-53.
15. Cuzick J. A Wilcoxon-type test for trend. Stat Med. 1985;4(1):87-90.
16. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet. 2005;37(11):1243-6.
17. Brendle A, Brandt A, Johansson R, Enquist K, Hallmans G, Hemminki K, et al. Single nucleotide polymorphisms in chromosomal instability genes and risk and clinical outcome of breast cancer: a Swedish prospective case-control study. Eur J Cancer. 2009;45(3):435-42.
18. Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do KA, et al. Characteristics associated with important clinical end points in patients with chronic lymphocytic leukemia at initial treatment. J Clin Oncol. 2009;27(10):1637-43.
19. Oscier D, Wade R, Davis Z, Morilla A, Best G, Richards S, et al. Prognostic factors identified three risk groups in the LRF CLL4 trial, independent of treatment allocation. Haematologica. 2010;95(10):1705-12.
20. Falt S, Merup M, Gahrton G, Lambert B, Wennborg A. Identification of progression markers in B-CLL by gene expression profiling. Exp Hematol. 2005;33(8):883-93.
21. Paz N, Levanon EY, Amariglio N, Heimberger AB, Ram Z, Constantini S, et al. Altered adenosine-to-inosine RNA editing in human cancer. Genome Res. 2007;17(11):1586-95.
22. Singh MK, Dadke D, Nicolas E, Serebriiskii IG, Apostolou S, Canutescu A, et al. A novel Cas family member, HEPL, regulates FAK and cell spreading. Mol Biol Cell. 2008;19(4):1627-36.
23. Gorroochurn P, Hodge SE, Heiman GA, Durner M, Greenberg DA. Non-replication of association studies: "pseudo-failures" to replicate? Genet Med. 2007;9(6):325-31.
24. Morgan TM, Krumholz HM, Lifton RP, Spertus JA. Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. JAMA. 2007;297(14):1551-61.