Supplementary Methods:

Discovery datasets

UK-GWAS: Briefly, 517 CLL cases (155 with one or more first-degree relatives affected with CLL or related B-cell lymphoproliferative disorders ascertained through haematologist in the International CLL Linkage Consortium between 1997 and 2007) and 362 cases ascertained through Leukemia Research UK CLL4 trial[1] between 1999 and 2004 were genotyped using HumanCNV370-Duo BeadChips (Illumina). For controls, we made use of publicly accessible Hap1.2M-Duo Custom array data generated on 2,930 individuals (1,509 men) from Wellcome Trust Case-Control Consortium 2 (WTCCC2) 1958 birth cohort (also known as the National Child Development Study)[2]. Quality control metrics included removal of samples with call rates <90% and SNP assays with call rates <95%. Participants with evidence of cryptic relatedness and non-European background were excluded from the analysis. We considered only the autosomal SNPs with MAF >1% in cases and controls, and SNPs with no extreme evidence of departure from Hardy-Weinberg equilibrium (HWE; $P>10^{-5}$) in cases or controls. After imposing these quality control metrics, the dataset provided 292,402 genotypes for 503 cases and 2,699 controls. Comparison of observed and expected distributions showed little evidence for an inflation of test statistics (inflation factor $\lambda = 1.04$). For imputation, we used IMPUTEv2/SNPTESTv2 software in conjunction with HapMap Phase II haplotypes release 2 (Jan 2008 on NCBI B36 assembly, dbSNP b126) and with HapMap Phase III haplotypes release 2 (HapMap Data Release 27/phase III Feb 2009 on NCBI B36 assembly, dbSNP b126).

SF-GWAS: Full details of the GWAS, including the process and criteria for participant selection, genotyping, quality control and statistical analysis have been previously described[3]. Briefly, a population-based case-control study of non-Hodgkin lymphoma (NHL) (2055 cases, 2081 controls) was conducted in the San Francisco Bay Area that included incident cases diagnosed from 2001 to 2006. Controls were frequency matched to patients by age in five-year groups, sex, and county of residence. Blood and/or buccal specimens were collected from eligible cases and controls who participated in the laboratory portion of the study (participation rates, 87% and 89%, respectively). DNA from 1,577 study participants was genotyped using Illumina HumanCNV370-Duo BeadChip. Samples with call rates <95%, and SNPs with call rates <90%, MAF <0.05 or on sex chromosomes were excluded from further analysis. Genotype data was used to test population stratification and cryptic relatedness as described previously, and closely related individuals with evidence of non-European ancestry were further excluded. A total of 312,555 markers genotyped in 211 CLL/SLL and 750 controls passed the quality control criteria.

For imputation, we used BEAGLE 3.0.3 software in conjunction with HapMap Phase II CEU population. Markers imputed with maximum posterior probability lower than 0.9 were set to missing, and those with >10% missing data or MAF<0.01 were further removed, resulting in an additional 1,624,197 imputed SNPs that were used for genome-wide association analysis.

GEC-GWAS: This GWAS has been previously described[4]. Briefly, peripheral blood samples were obtained from two ongoing studies: the Genetic Epidemiology of CLL (GEC) Consortium and the Mayo Clinic NHL/CLL study. The GEC Consortium started in 2004 and is an ongoing family-based study in which families with two or more members with CLL are recruited through hematology clinics or through the internet. The Mayo Clinic NHL/CLL case-control study started in 2002 and is an ongoing, clinic-based study of incident cases and frequency matched controls being conducted in Rochester, Minnesota. A total of 438 CLL cases (110 with one or more first-degree relatives affected with CLL, 328 sporadic CLL) and 328 controls were genotyped using the Affymetrix 6.0 SNP Array. Rigorous quality control measures were implemented, including excluding individuals or SNPs with call rates <95%, excluding related individuals, and excluding non-Caucasian individuals. We also excluded SNPs if call rates differed by 5% or more between cases and controls. A total of 827,777 autosomal SNPs passed quality control in 407 cases and 296 controls. Comparison of observed and expected distributions showed little evidence for an inflation of test statistics (λ = 1.003). For imputation, we used MACH1 software in conjunction with HapMap Phase II CEU population as a reference. Imputed SNPs with imputation $r^2<0.9$ were excluded from further analyses.

Mutational status

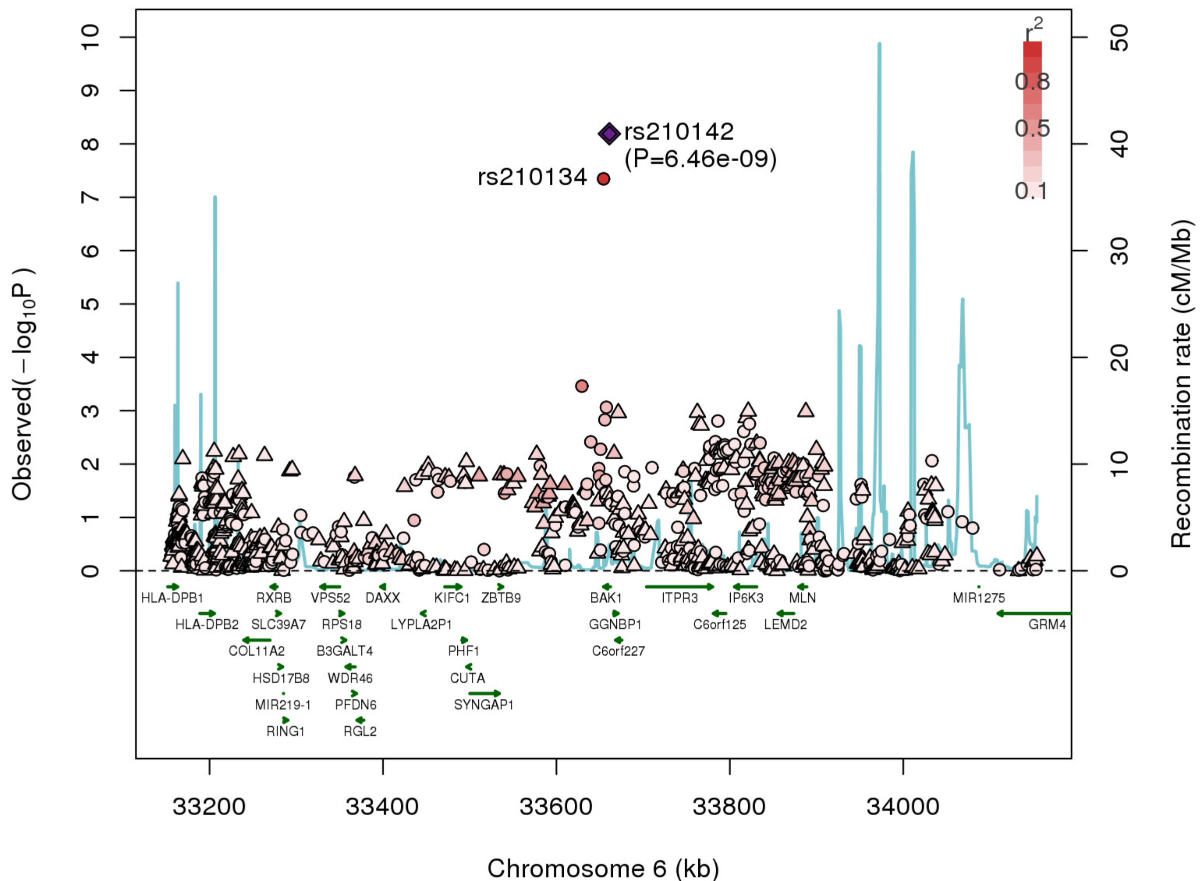
Immunoglobulin heavy-chain variable *(IGVH)* gene mutation status was determined according to BIOMED-2 protocols as described previously, using commercial reagents (InVivoScribe Technologies)[5]. Clonality was assessed by size discrimination of PCR products using semi-automated ABI3730xl sequencers in conjunction with Genescan software (Applied Biosystems). Sequence analysis was conducted using Chromas software version 2.23 (Applied Biosystems) and the international immunogenetics information system database. In accordance with published criteria, we classified *IGVH* sequences with a germline homology of 98% or greater as unmutated and those with homology of less than 98% as mutated.

Patient prognosis

To examine the relationship between SNP genotype and patient prognosis we used previously generated data on 356 patients entered into Leukemia Research CLL4 trial[6]. Briefly,

progression-free survival was defined as the time from the date of randomization to date of relapse needing further therapy, of progression, or of death from any cause. Cox-regression analysis was used to estimate genotype-specific hazard ratios (HR) adjusting for age at diagnosis, Binet stage, sex and treatment.

Supplementary Figure 1: Regional plot of association results and recombination rates for the 6p21.33 risk locus. Association results of both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates. $-\log_{10} P$ values (*y* axis) of the SNPs are shown according to their chromosomal positions (*x* axis). rs210142 in the combined GWAS analysis is denoted by a large diamond and is labelled by its rsID. The colour intensity of each symbol reflects the extent of LD with the top genotyped SNP: white (r^2=0) through to dark red (r^2=1.0). Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with a light blue line. Physical positions are based on NCBI build 36 of the human genome. Also shown are the relative positions of genes and transcripts mapping to each region of association. Genes have been redrawn to show the relative positions; therefore, maps are not to physical scale.



Supplementary Table 1: Odds Ratios (ORs) and 95% Confidence Intervals (CI) for previously reported CLL risk loci

Previously Reported		Risk	Number of Subjects		Allele Frequency					GEC-SF-UK
Locus	SNP	Allele	cases	controls	cases	controls	OR	95% CI	P-value	Direction*
2q13	rs17483466	G	1118	3741	0.26	0.20	0.70	(0.62 - 0.79)	4.96E-09	---
2q37.1	rs13397985	G	1113	3738	0.24	0.19	0.72	(0.63 - 0.81)	1.75E-07	---
2q37.3	rs757978	T	1121	3741	0.27	0.26	1.46	(1.25 - 1.72)	3.21E-06	+++
6p21.3	rs674313	T	400	292	0.31	0.26	1.30	(1.02 - 1.65)	3.30E-02	+??
6p25	rs872071	G	1117	3744	0.62	0.53	1.47	(1.33 - 1.63)	7.54E-14	---
	rs9378805	C	1119	3742	0.58	0.50	1.41	(1.27 - 1.55)	3.22E-11	---
8q24.21	rs2456449	G	703	3442	0.41	0.36	0.81	(0.72 - 0.91)	6.31E-04	?--
11q24	rs735665	A	1118	3742	0.28	0.20	1.52	(1.35 - 1.72)	3.36E-12	+++
15q21.3	rs7169431	A	1121	3742	0.11	0.08	1.45	(1.22 - 1.73)	2.65E-05	+++
15q23	rs7176508	A	1121	3741	0.46	0.37	1.42	(1.28 - 1.58)	2.87E-11	+++
15q25	rs783540	G	1114	3729	0.42	0.39	0.86	(0.78 - 0.95)	4.36E-03	---
16q24.1	rs305061	C	1121	3745	0.27	0.34	0.75	(0.67 - 0.83)	9.42E-08	+++
	rs391525	G	1110	3740	0.28	0.34	1.32	(1.19 - 1.48)	5.40E-07	+++
19q13	rs11083846	A	1121	3745	0.27	0.23	1.21	(1.07 - 1.36)	2.00E-03	+-+

* An indicator of the consistency and direction of the OR for each of the three studies. A ? Indicates that the SNP was not genotyped and poorly imputed.

Supplementary Table 2: Relationship between SNP genotypes and risk of CLL in the replication series of 861 CLL cases and 2,033 controls

				Cases				Controls						
SNP	**Locus**	**Allele A**	**Allele B**	**AA**	**AB**	**BB**	**MAF**	**AA**	**AB**	**BB**	**MAF**	***P***	**OR**	**95% CI**
rs4684085	3p25.2	T	C	255	413	171	0.45	603	1040	367	0.44	0.543	1.04	0.92-1.17
rs1395233	4q28.1	T	G	259	414	176	0.45	570	936	468	0.47	0.119	0.91	0.82-1.02
rs1395241	4q28.1	A	G	259	403	177	0.45	579	939	472	0.47	0.139	0.92	0.82-1.03
rs210134	6p21.31	G	A	453	333	52	0.26	906	902	191	0.32	4.87E-06	0.74	0.65-0.84
rs210142	6p21.31	C	T	485	307	41	0.23	947	889	169	0.31	2.41E-08	0.68	0.59-0.78
rs492638	8p21.1	G	T	411	362	71	0.30	936	861	202	0.32	0.184	0.92	0.81-1.04
rs484458	8p21.1	G	A	485	307	37	0.23	1165	698	120	0.24	0.589	0.96	0.84-1.10
rs6491466	13q32.2	G	C	339	360	121	0.37	741	962	260	0.38	0.461	0.96	0.85-1.08
rs2925269	15q11.2	A	G	395	348	77	0.31	907	871	193	0.32	0.347	0.94	0.83-1.07
rs9806169	15q11.2	G	T	422	342	63	0.28	1011	845	154	0.29	0.767	0.98	0.86-1.12

Supplementary Table 3: Relationship between rs210134 and rs210142 genotypes and sex, age at diagnosis, *IGHV* status.

		Age			**Sex**			***IGHV* Status**		
SNP	**genotype**	**< 63**	**>= 63**	***P***	**female**	**male**	***P***	**no**	**yes**	***P***
rs210134	AA	61	52	0.77	41	74	0.95	14	27	0.32
	AG	356	388		242	507		122	143	
	GG	542	535		362	724		182	209	
rs210142	TT	51	45	0.87	35	63	0.52	13	24	0.71
	TC	340	358		237	465		119	131	
	CC	562	557		371	758		182	216	

Age and sex based on all data. *IGHV* status based on UK-GWAS and GEC-GWAS data.

REFERENCES

1. Catovsky D, Richards S, Matutes E, 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):230-239. Prepublished on 2007/07/31 as DOI 10.1016/S0140-6736(07)61125-8.
2. Penegar S, Wood W, Lubbe S, et al. National study of colorectal cancer genetics. *British journal of cancer*. 2007;97(9):1305-1309. Prepublished on 2007/09/27 as DOI 10.1038/sj.bjc.6603997.
3. Conde L, Halperin E, Akers NK, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nature genetics*. 2010;42(8):661-664. Prepublished on 2010/07/20 as DOI 10.1038/ng.626.
4. Slager SL, Rabe KG, Achenbach SJ, et al. Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. *Blood*. 2011;117(6):1911-1916. Prepublished on 2010/12/07 as DOI 10.1182/blood-2010-09-308205.
5. van Krieken JH, Langerak AW, Macintyre EA, et al. Improved reliability of lymphoma diagnostics via PCR-based clonality testing: report of the BIOMED-2 Concerted Action BHM4-CT98-3936. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK*. 2007;21(2):201-206. Prepublished on 2006/12/16 as DOI 10.1038/sj.leu.2404467.
6. Wade R, Di Bernardo MC, Richards S, et al. 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. Prepublished on 2011/06/11 as DOI 10.3324/haematol.2011.043471.