Supplementary Methods

Patients and DNA samples

GWA studies

The two GWA studies on which the present analysis is based have been described in previously published literature[1]. Briefly, we analyzed the constitutional DNA of Caucasian patients diagnosed with childhood ALL from two patient cohorts: UK-GWA1 was based on 577 childhood ALL prevalent cases derived from the United Kingdom Childhood Cancer Study (UKCCS; 255 female, 322 male; mean age at diagnosis 5.5 years) an epidemiological study of childhood malignancies conducted between 1991 and 1998[2]; and UK-GWA2 on 392 childhood ALL incident cases derived from the UK Medical Research Council (MRC) ALL 97 (99) trial and 36 cases from Northern Institute for Cancer Research (NICR; 252 male; 176 female; mean age at diagnosis 5.8 years). Immunophenotyping and genotyping of diagnostic samples from patients was undertaken using standard methodologies. For controls we used genotype data from two publicly accessible data series for population SNP genotype frequencies; individuals from the 1958 Birth Cohort[3] and healthy individuals from a UK GWA study of colorectal cancer[4]. Genotyping of cases was conducted using Illumina Infinium HD Human370 Duo BeadChips according to the manufacturer's protocols (Illumina, San Diego, USA). A DNA sample was deemed to have failed if it generated genotypes at <95% of loci. Across both studies genotyped samples were excluded from analyses for the following reasons: failed genotyping (n=64), non-CEU ancestry (n=24) and cryptic relatedness (n=10). After removing failed samples UK-GWA1 consisted of 459 B-cell, 45 T-cell and UK-GWA2 of 365 B-cell and 38 T-cell cases; cytogenetic data were available on 632 individuals with B-cell ALL: hyperdiploid ALL (≥ 50 chromosomes - B-hyperdiploid, n = 289); B-cell lineage with the *ETV6/RUNX1* (alias *TEL/AML1*) fusion (n = 126) and B-cell other (n = 217).

Replication series

German case-control series: The German case-control study comprised 1,427 patients (822 male and 605 female; mean age 6.2 years) with childhood ALL (1,193 B-cell, 235 T-cell ALL) ascertained through the BFM trials between 1993 and 2004. 1,516 controls (762 male and 754 female; mean age 5.8 years) were ethnically matched healthy individuals of German origin recruited in 2004 at the Institute of Transfusion Medicine, Mannheim (Germany).

Canadian case-control series: Subjects from the Quebec Childhood ALL (QcALL) study (n = 260) were diagnosed in the Division of Hematology-Oncology of Ste-Justine Hospital, Montréal, between October 1985 and November 2006. The criteria for inclusion were: complete clinical history, whites of French-Canadian origin residing in the Province of Quebec as judged by their names, languages and places of birth and availability of biological material. The recruited patients comprised 155 males and 105 females between the ages of 1 and 16 years (mean age of 5.3). The distribution of ALL subtypes as determined by immunophenotyping was as follows: 1 T-cell, 251 B-cell and 8 with undetermined lineage. A general population control group composed of 266 (152 males, 114 females) was randomly selected from a large DNA institutional bank. The criteria for inclusion in the control group were: anonymous, healthy and unrelated individuals recruited from the population served by Ste-Justine Hospital; whites of French-Canadian origin residing in the Province of Quebec as judged by their language and place of birth.

Hungarian case-control series: DNA was obtained from a retrospective series of 550 childhood ALL cases (mean age at diagnosis 6.1 years). Patients were diagnosed with ALL between 1990 and 2002, aged 1–15 years at diagnosis and treated in 10 Hungarian centres (304 males, 246 females). The cohort represents 51.0% of all diagnosed childhood ALL. In the majority DNA was extracted from remission bloods (n = 485), while from patients who died, DNA was obtained from bone marrow smears (n = 65)[5]. The 450 controls (174 male and 276 females; average age 38 years) of the same ethnicity and from the same geographical region as the patients were randomly selected from healthy blood donors. None of the controls had a personal history of any malignancy.

Spanish case-control series:

The Spanish cohort consisted of 148 childhood ALL prevalent cases (82 male, 66 female; mean age 5.8 years; 127 B-cell and 21 T-cell). Samples were collected between 2006 and 2008 from two Spanish hospitals (Hospital la Paz and Hospital Doce de Octubre, both in Madrid). The 187 controls were unrelated Spanish individuals (94 male and 93 female).

Ethics

Collection of blood samples and clinico-pathological information from subjects was undertaken with informed consent and relevant ethical review board approval in accordance with the tenets of the Declaration of Helsinki.

Replication genotyping

Genotyping of SNPs was conducted using either competitive allele-specific PCR KASPar chemistry (KBiosciences Ltd, Hertfordshire, UK) or single-base primer extension chemistry

MALDI-TOF MS detection (Sequenom, San Diego, USA). All primers and probes used are available on request. A series of DNAs, whose genotype was established by sequencing, was used to validate genotyping assays in individual centers. Genotyping quality control was further evaluated through inclusion of duplicate DNA samples in SNP assays. For all SNP assays >99% concordant results were obtained. Samples having SNP call rates <90% were excluded from the analysis.

Statistical analysis

Statistical analyses were undertaken using R (v2.6) and STATA (v10; State College, Texas, US) software. Deviation of the genotype frequencies in the controls from those expected under HWE was assessed by a χ^2 test. The association between each SNP and risk of ALL was assessed by the Cochran-Armitage trend test. Odds ratios and associated 95% CIs were calculated by unconditional logistic regression. Relationships between multiple SNPs showing association with ALL risk in the 9p21.3 region were investigated using logistic regression analysis, and the impact of additional SNPs from the same region was assessed by a likelihood-ratio test. Meta-analysis was conducted using standard methods[6]. Cochran's Q statistic to test for heterogeneity and the I^2 statistic to quantify the proportion of the total variation due to heterogeneity were calculated[7]. Large heterogeneity is typically defined as $I^2 \geq 75\%$. The population attributable fraction was estimated from $1-\prod_i 1-(x_i-1)/x_i$ where $x_i = (1-p)^2 + 2p(1-p)OR_1 + p^2OR_2$, p is the population allele frequency, and OR_1 and OR_2 are the ORs associated with hetero- and homozygosity respectively. The relationship between SNP genotype and molecular and immunophenotypic subgroups was examined by case-only analysis using logistic regression.

Bioinformatics

LD metrics between SNPs reported in HapMap were based on Data Release 2/phase III Feb09 on NCBI B35 assembly, dbSNPb125. We used Haploview software (v3.2) for haplotype analysis and to infer the LD structure of the 174kb region at 9p21.3 (from 21,942,000 to 22,146,000 bps on NCBI B35). Prediction of the untyped SNPs in the UK GWA studies was carried out using IMPUTE (v2) based on phased haplotypes from HapMap phase III (January 2007 on NCBI B35 assembly, dbSNP build 125) and 1000Genomes data. Subsequent analysis was done using SNPTEST and LD metrics for SNPs not in HapMap were calculated from the imputed data.

Using the Margarita program we inferred ARGs for 27-SNP haplotypes mapping to the 174kb region of 9p21.3[8]. For every ARG, a putative risk mutation was placed on the marginal genealogy at each SNP position by maximizing the association between the mutation and disease status. We evaluated the significance of this observed association through 10^6 permutations on the phenotypes. To annotate potential regulatory sequences within the 174kb

region we implemented *in silico* searches using Transfac Matrix Database v7.2[9], PReMod[10] and EEL[11] software.

Relationship between rs3731217 genotype and *CDKN2A*, and *CDKN2B* expression

To examine for a relationship between SNP genotype and expression levels of *CDKN2A* and *CDKN2B* in lymphocytes, we made use of publicly available expression data generated from analysis of 90 Caucasian derived Epstein-Barr virus–transformed lymphoblastoid cell lines using Sentrix Human-6 Expression BeadChips (Illumina, San Diego, USA)[12]. Online recovery of data was performed using WGAViewer v1.25 Software. Differences in the distribution of levels of mRNA expression between SNP genotypes were compared using a Wilcoxon-type test for trend[13].

9p21.3 (*CDKN2A*) copy number analysis in ALL

To study the relationship between rs3731217 genotype and 9p21.3 monoallelic/bioallelic deletion in ALL we made use of previously generated data on: (1) a subset of 66 UK-GWA2 cases[14]. Genomic copy number for 9p21.3 was evaluated using arrayCGH, FISH or SNP arrays in diagnostic or relapse samples as previously described[14]; (2) 387 of the German cases - deletion status in leukemic clones was determined using 50k GeneChip Human mapping array (XbaI 240, Affymetrix) as previously reported[15]. Fisher's exact test was used to evaluate relationship between carrier status and *CDKN2A/CDKN2B* deletion.

Relationship between rs3731217 genotype and clinical outcome

Follow-up data were available for 378 ALL97(99) trial patients and 1,302 German BFM patients. Difference in patient outcome by genotype was assessed by Log rank comparison of Kaplan-Meier survival curves.

SUPPLEMENTARY MATERIAL

Supplementary Table 1: 34 SNPs genotyped in the German case-control series.

Supplementary Table 2: Details of directly typed SNPs and those imputed from HapMap and 1000Genomes data for UK-GWA1 and UK-GWA2. Only those SNPs with a high quality score as shown by the information score and having *P* value of < 0.05 are shown. The five imputed loci with $P<10^{-3}$ are bolded and rs3731217 is in blue.

Supplementary Table 3: Details of haplotype, logistic regression and ARG analysis of the 174kb region encompassing rs3731217. (a) Haplotypes analysis of rs3731239, rs2811709,

rs4074785, rs3731217, rs2811712, rs3218018. (b) Single SNP and conditional (on rs3731217) logistic regression analysis. (c) Results from ARG analysis using Margarita.

Supplementary Table 4: Correlation between carrier status for the rs3731217 (*CDKN2A*) risk allele and deletion of the locus in leukemic blasts in an available subset of the German and UK-GWA2 cases. Confidence intervals were calculated using Fisher's exact test.

Supplementary table 5: Pair-wise linkage disequilibrium between rs3731217 and SNPs within the region based on phased genotype data from HapMap. The amino acid change of each SNP is shown.

Supplementary Table 6: Details of transcription binding sites (TFBSs) as predicted by EEL (using binding profiles from the JASPAR2 database) and the Transfac Matrix Database. The positions of 3 transcriptional regulatory modules as predicted by PReMod are also shown. "Score" refers to the confidence value assigned to each predicted binding region by the three different programs. The predicted binding sites cluster in 3 regions (highlighted red, green and blue). For comparison, the observed and imputed SNPs and associated *P* values are shown.

Supplementary Figure 1: Details of the 174kb LD region at 9p21.3. (a) LD statistics (r^2) from HapMap phase II data. The darker shading indicates strong LD between SNPs. (b) Sequence conservation across the region in mammals. (c) Regions with high densities of TFBSs are shown: modules (predicted by PReMod) and groups (predicted by EEL and the Transfac Matrix Database). Also shown are the positions of the genes encoded by the region and the position of rs3731217.

Supplementary Figure 2: Relationship between event-free survival and rs3731217 genotype. Kaplan-Meier event-free survival curves are shown for 378 MRC97(99) patients from UK-GWA2 study.

Supplementary Figure 3: Relationship between rs3731217 genotypes and normalized lymphocyte *CDKN2A* and *CDKN2B* mRNA expression. Expression of genes (normalized – $\log_2$ levels) is based on data from analysis of 90 Epstein-Barr virus–transformed lymphoblastoid cell lines using Sentrix Human-6 Expression BeadChip (Illumina, San Diego, USA)[12,16]. Data was recovered using WGAViewer Version 1.25. Differences in the distribution of expression by SNP genotype were compared using a Wilcoxon-type test for trend[13].

REFERENCES

1. Papaemmanuil, E. et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet* **41**, 1006-10 (2009).
2. The United Kingdom Childhood Cancer Study: objectives, materials and methods. UK Childhood Cancer Study Investigators. *Br J Cancer 2000* **82**, 1073-102.
3. Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol* **35**, 34-41 (2006).
4. Tomlinson, I.P. et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* **40**, 623-30 (2008).
5. Semsei, A.F. et al. Association of some rare haplotypes and genotype combinations in the MDR1 gene with childhood acute lymphoblastic leukaemia. *Leuk Res* **32**, 1214-20 (2008).
6. Pettiti, D. Meta-analysis Decision Analysis and Cost-effectiveness Analysis. . *Oxford University Press, Oxford, New York* (1994).
7. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539-58 (2002).
8. Minichiello, M.J. & Durbin, R. Mapping trait loci by use of inferred ancestral recombination graphs. *Am J Hum Genet* **79**, 910-22 (2006).
9. Matys, V. et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34**, D108-10 (2006).
10. Ferretti, V. et al. PReMod: a database of genome-wide mammalian cis-regulatory module predictions. *Nucleic Acids Res* **35**, D122-6 (2007).
11. Hallikas, O. et al. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* **124**, 47-59 (2006).
12. Stranger, B.E. et al. Genome-wide associations of gene expression variation in humans. *PLoS Genet* **1**, e78 (2005).
13. Cuzick, J. A Wilcoxon-type test for trend. *Stat Med.* **4**, 87-90. (1985).
14. Sulong, S. et al. A comprehensive analysis of the CDKN2A gene in childhood acute lymphoblastic leukemia reveals genomic deletion, copy number neutral loss of heterozygosity, and association with specific cytogenetic subgroups. *Blood* **113**, 100-7 (2009).
15. Kawamata, N. et al. Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood* **111**, 776-84 (2008).
16. Stranger, B.E. et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848-53 (2007).

SNP	Chr.	Location (bp)	Nearby gene(s)	Minor allele	UK-GWA1 cases MAF (n=504)	UK-GWA1 control MAF (n=1438)	UK-GWA2 case MAF (n=403)	UK-GWA2 control MAF (n=960)	UK meta analysis *P* value	German case MAF (n=1,427)	German control MAF (n=1,516)	German *P* value
rs11799849	1	20,647,937	-	A	0.502	0.437	0.485	0.444	8.17×10^{-5}	0.464	0.468	0.727
rs17115122	1	96,641,799	-	A	0.095	0.125	0.089	0.123	4.08×10^{-4}	0.106	0.121	0.077
rs4853946	2	2,215,288	*MYT1L*	G	0.437	0.493	0.429	0.484	5.17×10^{-5}	0.483	0.473	0.464
rs896232	2	2,711,884	-	T	0.404	0.352	0.384	0.329	5.56×10^{-5}	0.356	0.368	0.362
rs2130904	4	31,874,929	-	G	0.345	0.297	0.351	0.704	6.16×10^{-5}	0.307	0.316	0.503
rs10002424	4	65,427,118	-	C	0.257	0.216	0.265	0.220	2.31×10^{-4}	0.225	0.210	0.160
rs405510	5	117,001,879	-	G	0.287	0.345	0.317	0.358	1.67×10^{-4}	0.347	0.343	0.763
rs10061417	5	120,793,117	-	G	0.161	0.202	0.151	0.185	4.19×10^{-4}	0.193	0.195	0.883
rs1800197	5	177,352,573	*PROP1*	A	0.263	0.303	0.254	0.305	4.07×10^{-4}	0.259	0.289	0.012
rs7448421	5	178,236,474	*ZNF354B; ZFP2*	C	0.205	0.247	0.224	0.266	4.50×10^{-4}	0.237	0.245	0.471
rs1336767	6	124,856,989	*TCBA1*	A	0.083	0.061	0.098	0.067	4.38×10^{-4}	0.062	0.060	0.738
rs11764793	7	115,618,496	*TES*	A	0.647	0.587	0.621	0.586	2.67×10^{-4}	0.416	0.405	0.401
rs7835507	8	10,212,547	*MSRA*	A	0.150	0.121	0.145	0.112	9.21×10^{-4}	0.131	0.133	0.825
rs6997224	8	10,213,773	*MSRA*	C	0.185	0.145	0.181	0.132	1.39×10^{-5}	0.152	0.156	0.670
rs10511573	9	12,194,748	-	C	0.335	0.284	0.351	0.301	5.62×10^{-5}	0.321	0.308	0.248
rs3731217	9	21,974,661	*CDKN2A; CDKN2B*	G	0.108	0.141	0.108	0.140	5.42×10^{-4}	0.101	0.147	1.15×10^{-7}
rs1331876	9	28,366,185	*LINGO2*	T	0.316	0.271	0.344	0.293	1.79×10^{-4}	0.326	0.297	0.014
rs1001919	9	108,331,482	-	C	0.172	0.214	0.188	0.240	4.57×10^{-5}	0.217	0.215	0.821
rs872863	9	125,194,175	*DENND1A; CRB2*	T	0.083	0.055	0.094	0.066	5.77×10^{-5}	0.083	0.075	0.225
rs11188661	10	97,950,983	*BLNK; ZNF518A*	A	0.315	0.346	0.273	0.336	5.46×10^{-4}	0.326	0.328	0.817
rs11188664	10	97,952,507	*BLNK; ZNF518A*	G	0.365	0.417	0.346	0.405	3.91×10^{-5}	0.407	0.392	0.249
rs7084370	10	113,126,069	-	G	0.549	0.498	0.530	0.466	3.94×10^{-4}	0.487	0.487	0.987
rs12582396	12	12,627,386	*DUSP16; CREBL2*	T	0.101	0.075	0.133	0.088	1.08×10^{-5}	0.092	0.103	0.129
rs10877094	12	56,984,007	-	C	0.220	0.179	0.231	0.810	1.41×10^{-4}	0.197	0.203	0.599
rs7971479	12	67,533,750	*CPM; MDM2*	T	0.228	0.194	0.239	0.184	1.20×10^{-4}	0.228	0.234	0.582
rs7317221	13	69,038,099	-	C	0.228	0.186	0.255	0.203	5.66×10^{-5}	0.209	0.210	0.899
rs1832050	13	104,303,225	-	A	0.105	0.139	0.103	0.144	1.03×10^{-4}	0.139	0.138	0.935
rs5021303	15	75,869,833	-	G	0.182	0.233	0.198	0.236	7.51×10^{-4}	0.224	0.228	0.771
rs567379	18	65,051,399	-	T	0.220	0.258	0.204	0.250	3.73×10^{-4}	0.247	0.226	0.060
rs2284378	20	32,051,756	*RALY; EIF2S2*	T	0.381	0.331	0.373	0.326	2.46×10^{-4}	0.293	0.317	0.051
rs4911414	20	32,193,105	*RALY; EIF2S2*	T	0.395	0.350	0.391	0.343	4.67×10^{-4}	0.304	0.331	0.027
rs2903908	20	44,127,354	*NCOA5; SLC12A5*	C	0.217	0.278	0.231	0.267	3.77×10^{-5}	0.275	0.286	0.360
rs6027571	20	58,395,771	-	C	0.354	0.409	0.362	0.430	1.02×10^{-5}	0.421	0.419	0.903
rs9613221	22	25,316,826	*CRYBB1; CRYBA4; TPST2*	G	0.211	0.178	0.239	0.183	9.13×10^{-5}	0.029	0.198	0.318

Supplementary Table 1: 34 SNPs genotyped in the German case-control series.

Dataset	SNP	Location (bp)	Information score	*P* value	r^2 with rs3731217	*D'* with r3731217
1000Genomes	rs56018935	21,956,527	0.94	1.13×10^{-3}	0.02	1.00
1000Genomes	**rs2518719**	**21,960,427**	**0.96**	**1.09×10^{-4}**	**0.81**	**0.97**
1000Genomes	rs3731246	21,961,989	0.95	2.46×10^{-3}	0.02	1.00
1000Genomes	rs2518721	21,969,204	0.97	4.39×10^{-3}	0.02	1.00
UK-GWA1&2	rs2811709	21,970,151	1.00	0.017	0.02	1.00
1000Genomes	**rs13297747**	**21,970,941**	**0.97**	**7.01×10^{-4}**	**0.99**	**1.00**
1000Genomes	**rs3731222**	**21,973,914**	**0.98**	**5.26×10^{-4}**	**1.00**	**1.00**
UK-GWA1&2	rs3731217	21,974,661	1.00	3.87×10^{-4}	1.00	1.00
1000Genomes	**rs3731204**	**21,977,584**	**0.98**	**5.52×10^{-4}**	**1.00**	**1.00**
1000Genomes	rs10757262	21,977,874	0.94	0.025	0.02	1.00
HapMap/1000Genomes	rs3731201	21,978,896	0.95	0.028	0.03	1.00
1000Genomes	**rs3731198**	**21,979,477**	**0.94**	**4.46×10^{-4}**	**0.96**	**1.00**
1000Genomes	9-21985941	21,985,941	0.95	1.61×10^{-3}	0.02	0.97
UK-GWA1&2	rs2811712	21,988,035	1.00	0.020	0.02	0.95
UK-GWA1&2	rs3218018	21,988,139	1.00	4.83×10^{-3}	0.02	0.96
1000Genomes	rs3218007	21,989,800	0.98	0.011	0.02	0.96
1000Genomes	rs3218005	21,990,247	0.98	0.011	0.02	0.96
UK-GWA1&2	rs2069426	21,996,273	1.00	5.26×10^{-4}	0.02	0.96
HapMap/1000Genomes	rs974336	21,996,348	0.99	0.020	0.02	0.96
UK-GWA1&2	rs2069422	21,998,026	1.00	9.87×10^{-3}	0.02	0.96
1000Genomes	9-22002441	22,002,441	0.98	0.013	0.02	0.96
HapMap/1000Genomes	rs10965212	22,013,795	0.98	0.032	0.01	0.28
1000Genomes	9-22015432	22,015,432	0.98	0.013	0.02	0.96
1000Genomes	9-22015885	22,015,885	0.98	0.013	0.02	0.96
1000Genomes	rs1591136	22,016,834	0.99	0.026	0.01	0.28
1000Genomes	rs598664	22,017,551	0.97	0.014	0.02	0.96
HapMap/1000Genomes	rs7049105	22,018,801	0.99	0.032	0.01	0.28
HapMap/1000Genomes	rs10965215	22,019,445	0.99	0.043	0.01	0.27
1000Genomes	rs4977753	22,020,027	0.99	0.028	0.01	0.28
HapMap/1000Genomes	rs662463	22,020,438	0.97	0.017	0.02	0.95
1000Genomes	rs10115049	22,022,119	0.99	0.031	0.01	0.28
1000Genomes	9-22023824	22,023,824	0.98	0.019	0.02	0.97
UK-GWA1&2	rs2151280	22,024,719	1.00	0.045	0.01	0.27
1000Genomes	9-22030839	22,030,839	0.98	0.022	0.02	0.96
1000Genomes	9-22031155	22,031,155	0.98	0.022	0.02	0.96
1000Genomes	rs1360590	22,031,443	0.99	0.043	0.01	0.27
1000Genomes	rs1333036	22,033,819	0.99	0.033	0.01	0.27
1000Genomes	rs1333035	22,034,059	0.98	0.018	0.02	0.96
UK-GWA1&2	rs1333034	22,034,122	1.00	0.028	0.02	0.96
1000Genomes	rs17694555	22,041,295	0.93	3.06×10^{-3}	0.01	0.96
HapMap/1000Genomes	rs17756311	22,043,895	0.94	0.016	0.01	0.96
1000Genomes	9-22044164	22,044,164	0.93	3.16×10^{-3}	0.01	0.96
HapMap/1000Genomes	rs17694572	22,044,356	0.94	0.014	0.01	0.96
UK-GWA1&2	rs10120688	22,046,499	1.00	0.035	0.01	0.22
1000Genomes	9-22047530	22,047,530	0.96	5.30×10^{-3}	0.01	0.96
1000Genomes	9-22049061	22,049,061	0.95	2.57×10^{-3}	0.01	0.92
1000Genomes	9-22050833	22,050,833	0.98	0.014	0.03	0.92
UK-GWA1&2	rs1011970	22,052,134	1.00	0.016	0.03	0.92
1000Genomes	rs10811647	22,055,002	0.98	0.016	0.00	0.01
HapMap/1000Genomes	rs10811650	22,057,593	0.98	0.014	0.00	0.01
UK-GWA1&2	rs1547705	22,072,375	1.00	1.73×10^{-3}	0.05	0.24

Supplementary Table 2: Details of directly typed SNPs and those imputed from HapMap phase III and 1000Genomes data for UK-GWA1 and UK-GWA2. Only those SNPs with a high quality score as shown by the information score and having $P<0.05$ are shown. The five imputed loci with $P<10^{-3}$ are bolded and rs3731217 is in blue.

(a)

Haplotype	Case frequency	Control frequency	*P* value
CGGTAA	0.399	0.379	0.1328
TGGTAA	0.26	0.266	0.571
TGGGAA	0.107	0.141	2.93×10^{-4}
TAGTGC	0.108	0.089	0.0188
TGATAA	0.081	0.082	0.9101
TAGTAA	0.024	0.022	0.5874

(b)

SNP	Location (bp)	Single SNP *P* value	*P* value conditional on rs3731217
rs10757260	21,943,137	0.1801	0.7265
rs7041637	21,951,866	0.6619	0.7561
rs2811708	21,963,422	0.1737	0.3010
rs3731239	21,964,218	0.1188	0.6457
rs2811709	21,970,151	0.0190	0.0768
rs4074785	21,971,583	0.8950	0.5453
rs2811712	21,988,035	0.0261	0.0949
rs3218018	21,988,139	0.0067	0.0288
rs3217992	21,993,223	0.0924	0.0945
rs1063192	21,993,367	0.6338	0.2897
rs2069426	21,996,273	0.0005	0.0290
rs2069422	21,998,026	0.0129	0.0518
rs2157719	22,023,366	0.4433	0.1674
rs1759417	22,023,389	0.0970	0.1390
rs2151280	22,024,719	0.0436	0.0218
rs1412829	22,033,926	0.4906	0.1858
rs1333034	22,034,122	0.0340	0.1134
rs10120688	22,046,499	0.0336	0.0183
rs1011970	22,052,134	0.0201	0.1020
rs4977756	22,058,652	0.4140	0.2131
rs1412832	22,067,543	0.1410	0.5545
rs10116277	22,071,397	0.2803	0.2622
rs1547705	22,072,375	0.0019	0.0263
rs1333040	22,073,404	0.1189	0.6565
rs2383207	22,105,959	0.2818	0.2236
rs1333050	22,115,913	0.6839	0.7448

(c)

SNP	Location (bp)	Score (χ^2)	*P* value
rs10757260	21,943,137	12.30	0.502
rs7041637	21,951,866	12.68	0.378
rs2811708	21,963,422	12.40	0.195
rs3731239	21,964,218	12.33	0.379
rs2811709	21,970,151	13.22	0.030
rs4074785	21,971,583	13.08	0.711
rs3731217	21,974,661	13.40	8.92×10^{-4}
rs2811712	21,988,035	12.92	0.017
rs3218018	21,988,139	12.82	3.05×10^{-3}
rs3217992	21,993,223	12.77	0.205

rs1063192	21,993,367	12.56	0.751
rs2069426	21,996,273	12.63	6.13×10^{-4}
rs2069422	21,998,026	12.61	9.65×10^{-3}
rs2157719	22,023,366	12.62	0.943
rs1759417	22,023,389	12.66	0.133
rs2151280	22,024,719	12.46	0.202
rs1412829	22,033,926	12.53	0.927
rs1333034	22,034,122	12.53	0.032
rs10120688	22,046,499	12.47	0.180
rs1011970	22,052,134	12.52	0.011
rs4977756	22,058,652	12.52	0.881
rs1412832	22,067,543	12.55	0.643
rs10116277	22,071,397	12.64	0.653
rs1547705	22,072,375	12.75	0.006
rs1333040	22,073,404	12.70	0.458
rs2383207	22,105,959	12.41	0.619
rs1333050	22,115,913	12.42	0.707

Supplementary Table 3: Details of haplotype, logistic regression and ARG analysis of the 174kb region encompassing rs3731217. (a) Haplotypes analysis of rs3731239, rs2811709, rs4074785, rs3731217, rs2811712, rs3218018. (b) Single SNP and conditional (on rs3731217) logistic regression analysis. (c) Results from ARG analysis using Margarita.

Study	Risk allele status	All cases		B cell only	
		No deletion	Deletion	No deletion	Deletion
German	Carrier	49	20	44	16
	Non-carrier	224	94	213	59
		OR: 0.97 (95%CI: 0.52-1.77) *P* = 0.93		OR: 1.31 (95%CI: 0.64-2.58) *P* = 0.40	
UK-GWA2	Carrier	10	1	8	1
	Non-carrier	34	20	26	8
		OR: 0.17 (95% CI: 0.004-1.39) *P* = 0.07		OR: 0.40 (95% CI: 0.008-3.99) *P* = 0.42	

Supplementary Table 4: Correlation between carrier status for the rs3731217 (*CDKN2A*) risk allele and deletion of the locus in leukemic blasts in an available subset of the German and UK-GWA2 cases. Confidence intervals were calculated using Fisher's exact test.

Gene	nsSNP	Change	D' v rs3731217	r^2 v rs3731217
CDKN2A - Isoform 1	rs45476696	N153D	na	na
	rs3731249	T148A	1.00	0.0029
	rs6413464	S127A	na	na
	rs34170727	C124R	na	na
	rs6413463	Q123H	na	na
	rs35741010	T102A	na	na
	rs34886500	W99R	na	na
	rs11552822	Y84D	na	na
	rs43968276	Q83H	na	na
	rs11552823	L81P	na	na
	rs36204594	V60A	na	na
CDKN2A - Isoform 3	rs75256133	L101P	na	na
	rs45456595	R63G	na	na
CDKN2A - Isoform 4	rs35741010	D157G	na	na
	rs34886500	L154P	na	na
	rs4987127	R147G	na	na
	rs11552822	L139R	na	na
	rs3731190	S58P	na	na
	rs1801022	L36P	na	na

Supplementary Table 5: Pair-wise LD statistics between rs3731217 and coding SNPs within the 9p21.3 region based data from HapMap and 1000Genomes. The amino acid change of each SNP is shown.
na = no data available

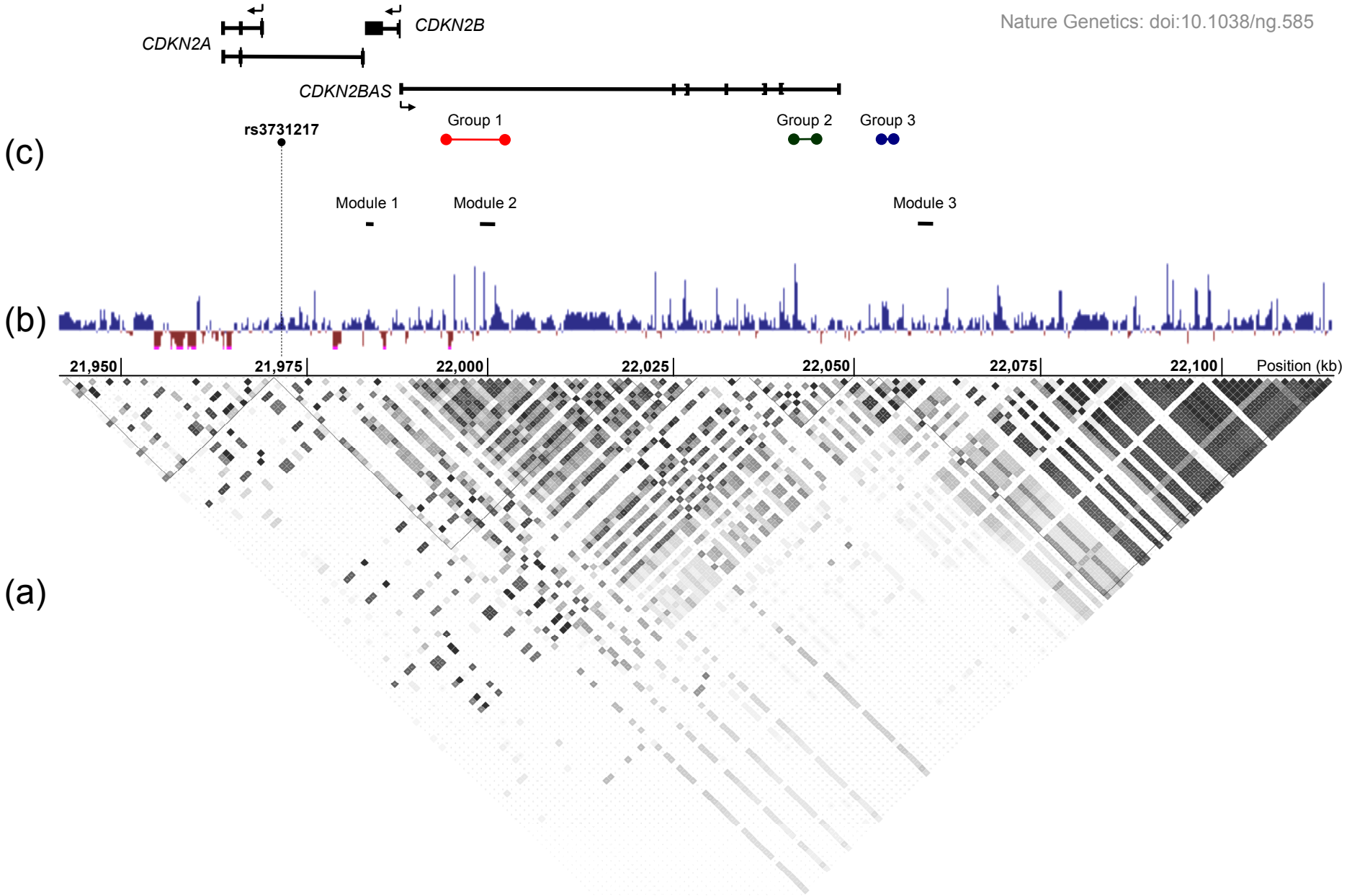
SNP	*P* value	SNP type	Predicted motif	Position	Program	Score
			V$IRF7_01	21,861,687 - 21,861,687	Transfac	2.91
			V$HNF3B_01	21,861,742 - 21,861,742	Transfac	2.12
			V$EVI1_02	21,863,209 - 21,863,209	Transfac	2.51
rs11515	0.0249	imputed		21,906,551		
			MA0010.pfm	21,936,028 - 21,936,028	EEL	83.69
			MA0010.pfm	21,940,754 - 21,940,754	EEL	90.24
			MA0001.pfm	21,947,559 - 21,947,559	EEL	86.68
			V$SOX5_01	21,947,637 - 21,947,637	Transfac	1.64
			V$EVI1_04	21,948,099 - 21,948,099	Transfac	2.1
			V$CEBPA_01	21,948,108 - 21,948,108	Transfac	1.84
			V$MEIS1BHOXA9_02	21,948,121 - 21,948,121	Transfac	2.35
			MA0010.pfm	21,951,062 - 21,951,062	EEL	124.59
			MA0010.pfm	21,953,911 - 21,953,911	EEL	86.13
			MA0010.pfm	21,954,908 - 21,954,908	EEL	84.43
			V$CDPCR3_01	21,960,036 - 21,960,036	Transfac	2.42
rs2518719	0.0005	imputed		21,960,427		
			V$PAX5_01	21,961,131 - 21,961,131	Transfac	1.66
rs3731246	0.0017	imputed		21,961,989		
rs2811708	0.1752	typed		21,963,422		
			MA0010.pfm	21,963,479 - 21,963,479	EEL	86
rs3731239	0.1342	typed		21,964,218		
rs3731238	0.1261	imputed		21,965,561		
			MA0008.pfm	21,968,311 - 21,968,311	EEL	102.34
rs2811709	0.0168	typed		21,970,151		
rs4074785	0.9205	typed		21,971,583		
rs3731222	0.0005	imputed		21,973,914		
rs3731221	0.1261	imputed		21,974,010		
rs3731217	0.0004	typed		21,974,661		
			V$OCT1_07	21,976,817 - 21,976,817	Transfac	1.68
rs3731211	0.1396	imputed		21,976,847		
			V$MEF2_04	21,976,924 - 21,976,924	Transfac	2.15
			V$NKX61_01	21,976,922 - 21,976,922	Transfac	1.92
			V$TCF11MAFG_01	21,976,954 - 21,976,954	Transfac	2.71
			V$SREBP1_01	21,976,955 - 21,976,955	Transfac	2.05
rs3731204	0.0005	imputed		21,977,584		
rs10757262	0.1153	imputed		21,977,874		
			V$ISRE_01	21,978,118 - 21,978,118	Transfac	1.71
			V$IRF2_01	21,978,117 - 21,978,117	Transfac	2.17
			V$IRF1_01	21,978,117 - 21,978,117	Transfac	2.11
			V$FREAC7_01	21,978,299 - 21,978,299	Transfac	2.11
rs3731201	0.1153	imputed		21,978,896		
			V$NFAT_Q6	21,979,462 - 21,979,462	Transfac	2.12
rs3731198	0.0005	imputed		21,979,477		
			V$IK1_01	21,980,345 - 21,980,345	Transfac	2.04
			V$IK2_01	21,980,345 - 21,980,345	Transfac	2.33
rs7036656	0.1496	imputed		21,980,457		
			MA0010.pfm	21,982,126 - 21,982,126	EEL	85.88
			V$MEF2_03	21,982,434 - 21,982,434	Transfac	1.9
			module1_hg17_mod114553	21,984,839 - 21,985,839	PReMod	170
			V$GATA1_03	21,984,000 - 21,985,000	Transfac	2.56
			V$NFY_Q6	21,984,997 - 21,984,997	Transfac	1.65
			V$IRF1_01	21,985,297 - 21,985,297	Transfac	1.67
rs3218020	0.0489	imputed		21,987,872		
rs2811712	0.0203	typed		21,988,035		
rs3218018	0.0048	typed		21,988,139		
			V$GATA1_05	21,988,546 - 21,988,546	Transfac	2.08
			V$AP1FJ_Q2	21,988,554 - 21,988,554	Transfac	1.81
			MA0010.pfm	21,988,920 - 21,988,920	EEL	84.42
rs3218005	0.0132	imputed		21,990,247		
rs3218002	0.0132	imputed		21,990,841		
			V$POU3F2_02	21,992,152 - 21,992,152	Transfac	2.99
rs3217992	0.0903	typed		21,993,223		
rs1063192	0.6759	typed		21,993,367		
			V$BACH1_01	21,994,707 - 21,994,707	Transfac	2.58
rs3217986	0.6851	imputed		21,995,330		
			V$NFKAPPAB50_01	21,995,756 - 21,995,756	Transfac	1.76
			V$STAT3_01	21,996,135 - 21,996,135	Transfac	1.67
			V$MYB_Q6	21,996,189 - 21,996,189	Transfac	2.08
			V$HEN1_02	21,996,224 - 21,996,224	Transfac	1.78
			V$COMP1_01	21,996,244 - 21,996,244	Transfac	2.76
			V$ER_Q6	21,996,250 - 21,996,250	Transfac	1.7
rs2069426	0.0003	typed		21,996,273		
			V$FOXJ2_01	21,996,741 - 21,996,741	Transfac	1.86
			V$HFH3_01	21,996,739 - 21,996,739	Transfac	1.76
			V$HNF3B_01	21,996,741 - 21,996,741	Transfac	2.34

			V$HFH1_01	21,996,739 - 21,996,739	Transfac	2.09
			V$FOXD3_01	21,996,739 - 21,996,739	Transfac	1.64
rs2069422	0.0099	typed		21,998,026		
			V$CP2_01	21,998,807 - 21,998,807	Transfac	1.76
			V$STAT1_01	21,998,862 - 21,998,862	Transfac	1.84
			V$STAT3_01	21,998,862 - 21,998,862	Transfac	1.72
			V$AP4_01	21,998,864 - 21,998,864	Transfac	1.7
			V$P53_02	21,998,935 - 21,998,935	Transfac	1.72
			module1_hg17_mod114554	21,999,209 - 22,000,209	PReMod	210
			V$IRF2_01	21,999,644 - 21,999,644	Transfac	2.39
			V$IRF1_01	21,999,644 - 21,999,644	Transfac	2.86
			V$BRN2_01	22,000,120 - 22,000,120	Transfac	1.64
			V$AML1_01	22,000,166 - 22,000,166	Transfac	1.68
			V$MEIS1BHOXA9_02	22,000,200 - 22,000,200	Transfac	2.3
			V$MEIS1_01	22,000,203 - 22,000,203	Transfac	1.69
			V$TGIF_01	22,000,203 - 22,000,203	Transfac	2.29
			V$GATA1_03	22,000,260 - 22,000,260	Transfac	2.4
			V$RFX1_01	22,000,369 - 22,000,369	Transfac	1.79
			V$CEBPA_01	22,000,366 - 22,000,366	Transfac	1.93
			V$NRSF_01	22,001,118 - 22,001,118	Transfac	4.61
			V$ZIC1_01	22,001,144 - 22,001,144	Transfac	2.91
			V$ZIC2_01	22,001,144 - 22,001,144	Transfac	2.47
			V$ZIC3_01	22,001,144 - 22,001,144	Transfac	2.96
			V$FOXJ2_02	22,001,290 - 22,001,290	Transfac	2.57
			V$EVI1_01	22,001,309 - 22,001,309	Transfac	2.48
rs13298881	0.0013	imputed		22,002,051		
			V$PAX6_01	22,002,206 - 22,002,206	Transfac	1.92
			V$USF_Q6	22,002,216 - 22,002,216	Transfac	1.88
			V$PAX2_01	22,002,253 - 22,002,253	Transfac	1.67
			V$RP58_01	22,002,305 - 22,002,305	Transfac	2.31
			V$YY1_02	22,002,317 - 22,002,317	Transfac	1.84
			V$GATA2_01	22,002,313 - 22,002,313	Transfac	1.76
			V$PPARA_01	22,002,366 - 22,002,366	Transfac	2.12
			V$E47_02	22,002,365 - 22,002,365	Transfac	1.65
			V$AREB6_03	22,002,363 - 22,002,363	Transfac	1.65
			V$EVI1_06	22,002,436 - 22,002,436	Transfac	2.24
			V$FAC1_01	22,002,518 - 22,002,518	Transfac	1.64
rs10811640	0.0209	imputed		22,003,411		
rs10811641	0.0304	imputed		22,004,137		
rs2106120	0.0199	imputed		22,007,101		
			V$EVI1_01	22,007,535 - 22,007,535	Transfac	2.33
			V$EVI1_04	22,007,540 - 22,007,540	Transfac	1.97
rs2106119	0.0197	imputed		22,007,550		
rs7044859	0.0208	imputed		22,008,781		
rs10757264	0.0353	imputed		22,009,732		
rs10965212	0.0297	imputed		22,013,795		
rs10811644	0.0114	imputed		22,015,067		
rs7035484	0.0119	imputed		22,015,240		
rs10738604	0.0365	imputed		22,015,493		
rs1591136	0.0274	imputed		22,016,834		
rs598664	0.0244	imputed		22,017,551		
rs7049105	0.0352	imputed		22,018,801		
rs10965215	0.0322	imputed		22,019,445		
rs662463	0.0223	imputed		22,020,438		
rs10115049	0.0323	imputed		22,022,119		
			V$GRE_C	22,022,268 - 22,022,268	Transfac	2.06
			V$NCX_01	22,023,331 - 22,023,331	Transfac	1.65
			V$MEIS1AHOXA9_01	22,023,518 - 22,023,518	Transfac	1.65
			V$FOXO4_01	22,023,551 - 22,023,551	Transfac	2.17
rs2151280	0.0450	typed		22,024,719		
			V$P53_01	22,027,416 - 22,027,416	Transfac	1.97
			V$FOXO4_01	22,027,470 - 22,027,470	Transfac	1.66
			V$HFH1_01	22,027,472 - 22,027,472	Transfac	1.76
			V$AML1_01	22,027,686 - 22,027,686	Transfac	1.68
			V$HAND1E47_01	22,027,870 - 22,027,870	Transfac	1.81
			V$CEBPB_02	22,027,010 - 22,028,010	Transfac	2.26
			V$BRACH_01	22,028,291 - 22,028,291	Transfac	3.68
			V$POU3F2_01	22,028,426 - 22,028,426	Transfac	1.68
			V$FOXJ2_01	22,029,039 - 22,029,039	Transfac	1.86
			V$SRY_02	22,031,110 - 22,031,110	Transfac	2.1
			MA0009.pfm	22,031,331 - 22,031,331	EEL	83.56
rs1360590	0.0377	imputed		22,031,443		
			V$P300_01	22,031,852 - 22,031,852	Transfac	1.87
			V$STAT_01	22,031,004 - 22,032,004	Transfac	2.54
			V$FOXO4_02	22,033,615 - 22,033,615	Transfac	2.47
			V$FOXO3_01	22,033,615 - 22,033,615	Transfac	1.81
rs1333035	0.0272	imputed		22,034,059		

rs1333034	0.0276	typed		22,034,122		
			V$HAND1E47_01	22,034,758 - 22,034,758	Transfac	2.31
			V$HNF1_01	22,034,801 - 22,034,801	Transfac	1.97
			MA0002.pfm	22,037,804 - 22,037,804	EEL	84.43
rs7028570	0.0235	imputed		22,038,683		
			V$NKX25_01	22,040,059 - 22,040,059	Transfac	1.81
			V$FAC1_01	22,041,151 - 22,041,151	Transfac	2.12
rs17694555	0.0126	imputed		22,041,295		
			V$OCT1_04	22,041,417 - 22,041,417	Transfac	1.69
			V$MEF2_01	22,041,412 - 22,041,412	Transfac	2.86
			V$NKX25_02	22,042,114 - 22,042,114	Transfac	2.04
			V$AP1_01	22,042,682 - 22,042,682	Transfac	1.86
			MA0002.pfm	22,042,708 - 22,042,708	EEL	83.8
			V$FOXJ2_01	22,042,710 - 22,042,710	Transfac	1.79
			V$FOXO3_01	22,042,707 - 22,042,707	Transfac	1.76
			V$FREAC7_01	22,042,709 - 22,042,709	Transfac	1.98
			V$FOXD3_01	22,042,706 - 22,042,706	Transfac	2.23
			V$HFH1_01	22,042,706 - 22,042,706	Transfac	1.97
			V$HFH3_01	22,042,707 - 22,042,707	Transfac	1.75
			V$FOXO1_01	22,042,705 - 22,042,705	Transfac	1.93
			V$FREAC3_01	22,042,713 - 22,042,713	Transfac	2.2
rs11790231	0.0003	imputed		22,043,591		
rs10965219	0.0189	imputed		22,043,687		
rs17756311	0.0129	imputed		22,043,895		
rs17694572	0.0140	imputed		22,044,356		
			MA0010.pfm	22,044,762 - 22,044,762	EEL	106.13
rs10120688	0.0346	typed		22,046,499		
			V$TATA_C	22,051,465 - 22,051,465	Transfac	1.89
rs1011970	0.0163	typed		22,052,134		
			V$EVI1_03	22,052,947 - 22,052,947	Transfac	3.09
			V$GATA_C	22,052,947 - 22,052,947	Transfac	1.66
			V$AML1_01	22,054,577 - 22,054,577	Transfac	1.68
			V$AP1_C	22,054,603 - 22,054,603	Transfac	1.72
rs10811647	0.0197	imputed		22,055,002		
			V$CART1_01	22,055,050 - 22,055,050	Transfac	2.13
			V$S8_01	22,055,053 - 22,055,053	Transfac	1.87
			V$CHX10_01	22,055,052 - 22,055,052	Transfac	2.34
			V$ZIC1_01	22,055,250 - 22,055,250	Transfac	2.24
			V$ZIC3_01	22,055,250 - 22,055,250	Transfac	2.41
			V$CHOP_01	22,055,413 - 22,055,413	Transfac	2.43
			V$IRF1_01	22,055,632 - 22,055,632	Transfac	2.2
			V$IRF2_01	22,055,632 - 22,055,632	Transfac	2.11
			MA0010.pfm	22,057,236 - 22,057,236	EEL	90.74
rs10811650	0.0192	imputed		22,057,593		
			MA0010.pfm	22,061,318 - 22,061,318	EEL	85.2
			module1_hg17_mod114556	22,063,274 - 22,070,274	PReMod	120
			MA0010.pfm	22,068,431 - 22,068,431	EEL	86.46
rs1547705	0.0017	typed		22,072,375		
			V$AP1_Q4	22,087,364 - 22,087,364	Transfac	2.05
			V$CEBPB_02	22,087,475 - 22,087,475	Transfac	1.71
rs6475607	0.0226	imputed		22,087,693		
			V$AML1_01	22,087,919 - 22,087,919	Transfac	1.68

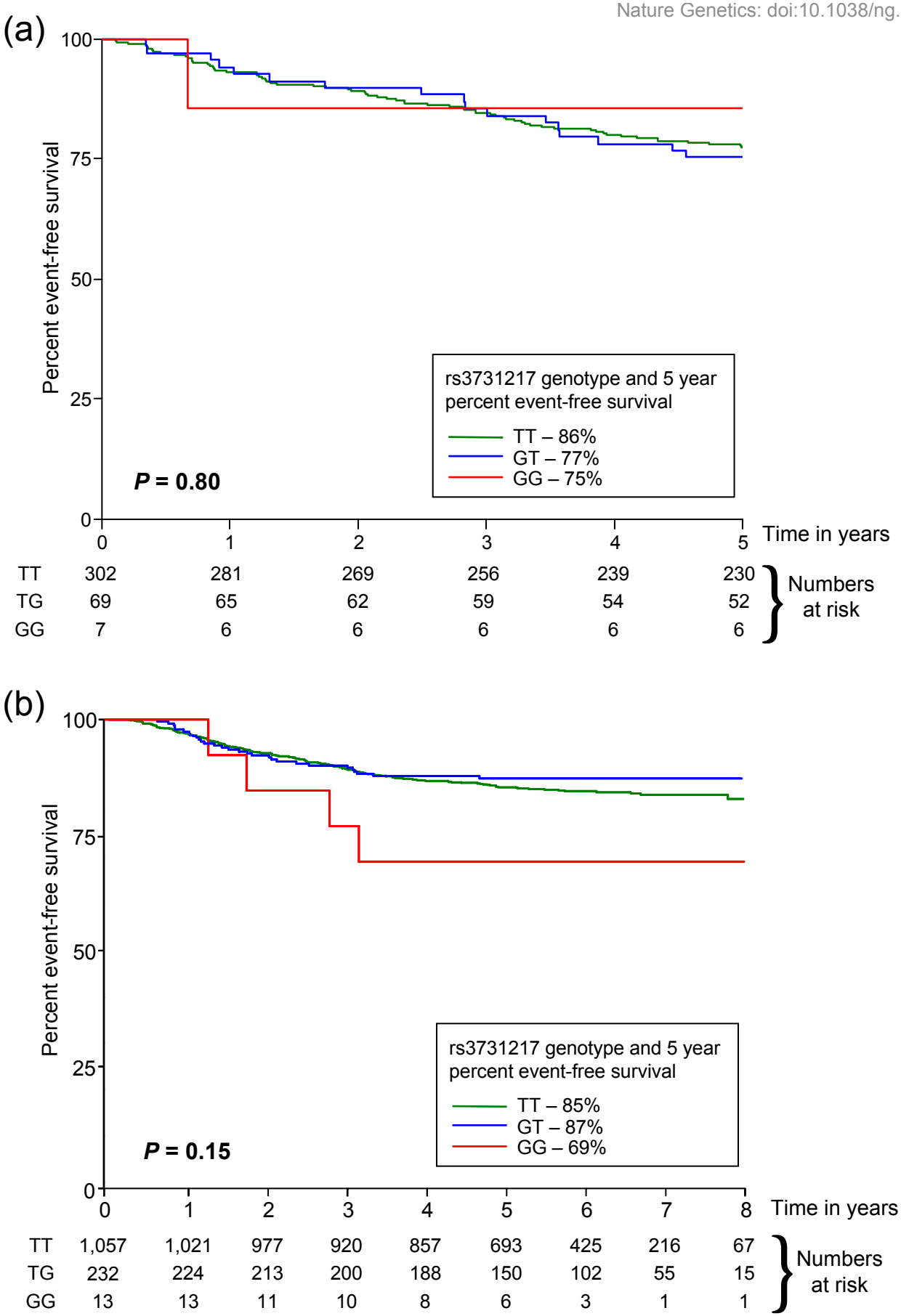
Supplementary Table 6: Details of transcription binding sites (TFBSs) as predicted by EEL (using binding profiles from the JASPAR2 database) and the Transfac Matrix Database. The positions of 3 transcriptional regulatory modules as predicted by PReMod are also shown. "Score" refers to the confidence value assigned to each predicted binding region by the three different programs. The predicted binding sites cluster in 3 regions (highlighted red, green and blue). For comparison, the observed and imputed SNPs and associated *P* values are shown.

Nature Genetics: doi:10.1038/ng.585

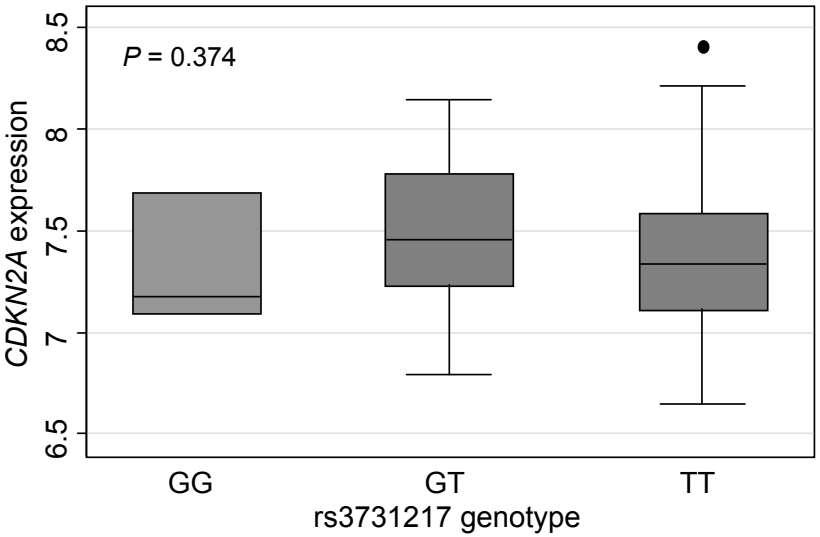


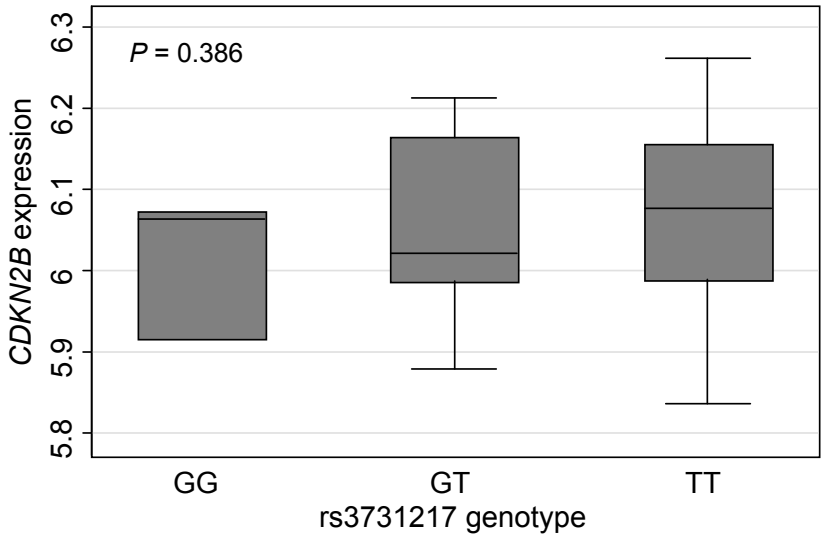
Supplementary Figure 1: Details of the 174kb LD region at 9p21.3. (a) LD statistics (r^2) from HapMap phase II data. The darker shading indicates strong LD between SNPs. (b) Sequence conservation across the region in mammals. (c) Regions with high densities of TFBSs are shown: modules (predicted by PReMod) and groups (predicted by EEL and the Transfac Matrix Database). Also shown are the positions of the genes encoded by the region and the position of rs3731217.

Nature Genetics: doi:10.1038/ng.585



Supplementary Figure 2: Relationship between event-free survival and rs3731217 genotype. Kaplan-Meier event-free survival curves are shown for (a) 378 MRC97(99) patients from UK-GWA2 study and (b) 1,302 BFM patients from the German series.





Supplementary Figure 3: Relationship between rs3731217 genotypes and normalized lymphocyte *CDKN2A* and *CDKN2B* mRNA expression. Expression of genes (normalized $-\log_2$ levels) is based on data from analysis of 90 Epstein-Barr virus–transformed lymphoblastoid cell lines using Sentrix Human-6 Expression BeadChip (Illumina, San Diego, USA). Data was recovered using WGAViewer Version 1.25. Differences in the distribution of expression by SNP genotype were compared using a Wilcoxon-type test for trend.