Supplemental Information

Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia

Jun Yang[1] PhD, Cheng Cheng[1] PhD, Wenjian Yang[1] PhD, Deqing Pei[1] MS, Xueyuan Cao[1] MS, Yiping Fan[1] PhD, Stan Pounds[1] PhD, Lisa R. Treviño[1] PhD, Deborah French[1] PhD, Dario Campana[1] MD, PhD, James R. Downing[1] MD, William E. Evans[1] Pharm D, Ching-Hon Pui[1] MD, Meenakshi Devidas[2] PhD, W.P. Bowman[3] MD, Bruce M. Camitta[4] MD, Cheryl Willman[5] MD, Stella M. Davies[6] MBBS, PhD, Michael J. Borowitz[7] MD, PhD, William L. Carroll[8] MD, Stephen P. Hunger[9] MD, Mary V. Relling[1] Pharm D

[1]St. Jude Children's Research Hosp., Memphis, TN
[2]University of Florida, Gainesville, FL
[3]Cook Children's Medical Center, Ft. Worth, TX
[4]Medical College of Wisconsin, Milwaukee, WI
[5]University of New Mexico Cancer Center, Albuquerque, NM
[6]Cincinnati Children's Hospital and Medical Center, Cincinnati, OH
[7]Johns Hopkins Medical Institute, Baltimore, MD
[8]New York University Medical Center, NY, NY
[9]The Children's Hospital and the University of Colorado Cancer Center, Aurora, CO

Supplemental Methods

Genotyping calls and imputation: Genotypes were assigned using the BRLMM calling algorithm[1] and coded as 0, 10, or 20 based on the number of B alleles. In total, there was an average of ~9,500 SNPs per patient not called by BRLMM. Missing genotypes were imputed using linkage disequilibrium (LD) between missing and typed SNPs in the appropriate race/ancestry groups for each patient, with LD patterns established using the HapMap data sets release 22, when LD was informative (median of ~3,500 SNPs per patient). However, if the imputed genotype was discordant with the direction of the raw intensity signals (e.g. if raw data indicated AA or AB but not BB, but LD patterns indicated BB), then the genotype was not imputed and left as missing (median of < 400 SNPs per patient). Alternatively, if LD information was unavailable for a SNP with missing genotype, an intermediate value (between 0 to10 or 10 to 20) was assigned based on the actual signal intensity distribution observed from the chip (median of ~6,000 SNPs per patient).

Quality control (SNP and patient exclusion): Patient-specific quality control included accuracy of sex assignment, ruling out duplication of samples, lack of conflict for self-declared vs. genotype-based race/ancestry, and lack of DNA degradation by agarose gel electrophoresis. Also, a minimum of 95% of SNPs (markers interrogated by the Nsp and Sty 250K arrays) with successful genotyping call using the BRLMM software was required, and an acceptable proportion of heterozygote calls (within three standard deviations of the average proportion of heterozygote calls). Of 371 St. Jude patients with germline DNA and end-of-induction MRD, 25 failed to achieve 95% call rates and were

therefore removed from the cohort. Likewise, 20 of 227 COG patients with germline DNA and MRD status had call rates <95%, and were excluded.

To minimize confounding effects by subtype, patients with *BCR-ABL*, *E2A-PBX1*, and *MLL*-rearranged subtypes of ALL (n=28 in St. Jude and 38 in COG) were also excluded from the subsequent genome-wide association analyses. These 3 ALL molecular subtypes were significantly associated ($P<0.05$) with end-of-induction MRD and differed in frequency between the St. Jude and COG cohorts, as indicated by Chi-square tests (Supplemental Table 1S). Therefore, the final population for the genome-wide association analyses of SNPs vs. MRD consisted of 318 St. Jude patients and 169 COG patients.

SNP-specific quality control included removing SNPs with minor allele frequency (MAF) <1%, and/or call rate<95%. In total, 35,965 SNPs were removed due to rarity, and 73,828 were removed because of poor genotyping quality. Deviations from Hardy Weinberg equilibrium were assessed using a Chi-square test for each race/ancestry group; P values were assigned to each SNP but not used to exclude consideration of genotyping data. In total, 476,796 SNPs were included in the final genome wide association analyses. For each top ranked SNP, raw signal intensity plots (A vs B allele signal intensity) for the patients were analyzed for any deviations from the expected trimodal clustering of AA, AB, and BB genotypes; all 102 top ranked overlapping SNPs gave high quality cluster plots.

SNP annotation: A SNP was annotated to a gene if it was located in the gene, within 5 Kb upstream/downstream of the gene, or in linkage disequilibrium with SNPs that met the above criteria.

Race: Race was inferred based on hierarchical clustering by including genotypes from patients with those available from the HapMap resource (n=90 CEU (Europeans), 90 YRI (Africans), 30 CHB (Chinese), and 30 JPT (Japanese)), using all SNPs on the SNP array with 100% call rates. Good agreement was observed between self-declared race of patients and that determined by germline SNP genotypes: 99% and 93% accuracy of self-declared race in the St. Jude and COG cohorts, respectively.

Permutation-asymptotic hybrid P values for association with MRD: The end-of-induction MRD phenotype was regarded as an ordinal variable with 1, 2, and 3 coding for negative, positive, and high-positive (defined in the Methods), respectively. The genotype of each SNP was coded based on the number of B alleles in the genotype call (see above). For a given SNP, association between its genotype and the phenotype was measured by the rank (Spearman's) correlation coefficient, where average ranks were used for ties. More specifically, let $R_1, \ldots, R_n$ denote the average ranks of the observed phenotype values, and let $S_1, \ldots, S_n$ denote the average ranks of the observed genotype values. Then the rank correlation C_R is defined as the sample correlation of the average ranks:

$$C_R = \frac{n^{-1}\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{(n-1)^{-1}\sum_{i=1}^{n}(R_i - \bar{R})^2}\sqrt{(n-1)^{-1}\sum_{i=1}^{n}(S_i - \bar{S})^2}}. \qquad (1)$$

Because C_R is essentially in a form of a sample average, the Central Limit Theorem implies that, under the null hypothesis of no association between the SNP genotype and the phenotype, C_R asymptotically (as the sample size $n \to \infty$) follows a normal distribution with mean zero and some variance σ^2. We estimated this variance by 4,000 permutations. In each permutation round the first component in the n pairs (R_i, S_i), $i=1,\ldots, n$ were randomly permuted and C_R was recalculated. The sample variance of the 4,000 C_R values computed form the permutated data was used to estimate σ^2. The two-sided permutation-asymptotic P value was calculated as

$$P = 2\left(1 - \Phi\left(\frac{|C_R|}{\hat{\sigma}^2}\right)\right) \qquad (2)$$

where $\hat{\sigma}^2$ is the permutation estimate of σ^2, and Φ denotes the cumulative distribution function of the standard normal (i.e., $N(0,1)$) distribution.

Note that the above description is for a single SNP. The asymptotic variance of C_R under the null hypothesis was estimated for each SNP, and different SNPs usually have different estimates for the asymptotic variance of C_R under the null hypothesis.

By computing P values this way we overcame the granularity of purely permutation-based P values and substantially reduced the computing demand in the internal validation inference described below.

False discovery rate (FDR) estimation and internal validation: Using the estimator we developed,[2] we estimated the FDR at the P value cutoff (per-test significance level) γ =0.0005, 0.00075, 0.001, 0.0025, 0.005, 0.0075, 0.01, 0.0125, 0.015, 0.0175, 0.02 (Fig.

2S). For example, there were 8,635 SNPs with P≤0.0125 in the St. Jude cohort, with the FDR level 67.2%; there were 7,881 SNPs with P≤0.0125 on the COG cohort, with the FDR level 73.4%. Hence at the 0.0125 per-test significance level we roughly expected to capture on the St. Jude (COG) cohort 2,832 (2096) SNPs that are truly associated with the phenotype and make 5803 (5,785) false discoveries. The remaining issue then was: if in fact none of SNPs was associated with the phenotype, for a given per-test significance level, how often we could observe the evidence for SNP-phenotype association that is at least as strong as that expected by chance. To resolve this issue we performed an internal validation inference on each cohort as follows. In one cohort, the P value of each SNP carried the weight of evidence for that SNP's association with the phenotype; thus for a given P value cutoff γ, the aggregate weight of evidence in the SNPs with P≤γ can be described by the score statistic

$$S(\gamma) = -\sum_{j=1}^{m} I(P_j \le \gamma)\log(P_j),$$

where m is the total number of SNPs and I denotes the indicator function, i.e., $I(A)=1$ if the statement A is true and $I(A)=0$ otherwise. The validation inference then consisted of a permutation test based on this statistic. In each permutation round, the ranks of the phenotype values were randomly permuted, the rank correlation C_R was recalculated for each SNP using the permuted data, and the P value was recalculated according to Equation (2); then the score statistic $S(\gamma)$ was recalculated. The *profile significance* of the set of SNPs with P≤γ was then defined as

$$PS(\gamma) = \frac{1}{N}\sum_{j=1}^{N} I(S_j(\gamma) \ge S(\gamma)),$$

where N is the number of permutation rounds, $S_j(\gamma)$ is the recalculated score statistic in the jth permutation round, and $S(\gamma)$ is the score statistic calculated on the unpermuted (original) data. Profile significance for the above sequence of per-test significance levels was produced in each cohort (Fig. 2S). The profile significance at γ =0.0125 was 0.0650 (0.0545) on the St. Jude (COG) cohort, based on 2,000 permutations.

As the P value cutoff γ increases, the estimated FDR steadily increases, whereas the profile significance did not vary much (Fig. 2S). Considering the number of SNPs captured, the level of FDR, the expected number of false positives, the expected number of true positives, and the profile significance, we chose 0.0125 for the P value cutoff to determine a specific set of SNPs on each cohort for further investigation.

Bi-directional validation: We sought to address whether the evidence for genotype-phenotype associations among the SNPs captured by P≤0.0125 could be "validated" in some way externally, i.e., independent of the cohort in which they were discovered. Toward this end, we used the COG cohort to validate the evidence of genotype-phenotype association contained in the SNPs discovered in the St. Jude cohort, and vice versa.

Simon[3] pointed out that in the context of classifier validation, it is inappropriate to insist that each gene expression feature achieves the same or higher observed significance level (P value) on the validation set, because the observed P values of each feature can be unstable from one dataset to another; it is the classifier as a whole that should be

validated. The same principle may be applied to genomic association studies. One should not insist that each SNP reaches the same or higher statistical significance on the validation set, rather it is the total evidence for genotype-phenotype association contained in the set of captured SNPs that should be validated. For this purpose we developed a rank test procedure as follows.

A SNP was considered concordant on the St. Jude and COG cohorts if its genotype-phenotype association had the same direction on both cohorts. Let D_{SJ} denote the sign function of the rank correlation between the genotype and MRD levels on the St. Jude cohort (D_{SJ} =−1, 0, 1 for negative, zero, and positive rank correlation respectively); likewise define D_{COG}. Then a SNP was concordant if and only if $D_{SJ}\, D_{COG}>0$. The rank test procedure to validate the St. Jude SNPs on the COG cohort consists of the following steps.

(1) Sort the SNPs according to their P value in increasing order on the COG cohort.

(2) Suppose there are M St. Jude SNPs to validate (M=8,635 in this case). Let $R_1, \ldots, R_M$ be their ranks on the rank order obtained in (1). Define the "unitized" ranks as $W_i=R_i/m$, i=1,…,M, where m is the total number of SNPs (m=476,796).

(3) If none of the SNPs is associated in the COG cohort and the SNP-phenotype associations are observed purely by chance, then given M, the ranks obtained in (2) are approximately a simple random sample out of the set {1, 2, .., m}; with m being large, the unitized ranks (W's) form approximately a random sample from the *Uniform*(0,1) distribution; furthermore, for each of these SNPs, the probability

of concordance (i.e., $D_{SJ}\, D_{COG}>0$) is 0.5; this provides the null model. The SNP-phenotype association among these SNPs on the validation set (the COG cohort) can then be measured by the statistic

$$T = -\sum_{i=1}^{M} I\left(D_{SJ,i} D_{COG,i} > 0\right)\log\left(W_i\right).$$

The higher the concordant SNPs are ranked (the smaller are their rank numbers), the larger is T. Under the null model, T has the cumulative distribution function (cdf) $F(t) = 0.5^M I(t \geq 0) + \sum_{j=1}^{M} B(j; M, 0.5) G(t; j, 1), -\infty \leq t \leq \infty,$ where $B(\cdot\,; M, 0.5)$ is the *Binomial*(M,0.5) cdf and $G(\cdot\,; j, 1)$ is the *Gamma*(j,1) cdf. Thus the statistical significance (P value) of the validation inference is $P = 1 - F(T)$.

The COG SNPs were validated in the St. Jude cohort likewise, with the roles of St. Jude (COG) SNPs and St. Jude (COG) cohorts interchanged.

The validation P value of the St. Jude SNPs on the COG cohort was 2.2046×10^{-6}, and the validation P value of the COG SNPs on the St. Jude cohort reached a higher statistical significance of $P<10^{-11}$.

Combined P value: To account for an assessment of significance of a SNP in both cohorts, we estimated a "combined" P value for each SNP. For a SNP, we let P_{SJ} and P_{COG} be the permutation-asymptotic hybrid P value as described above for the St. Jude

and COG cohorts, respectively. Then a statistic combining evidence for the SNP-phenotype association on both cohorts is

$$S = -I(D_{SJ} D_{COG} > 0)\left[\log(P_{SJ}) + \log(P_{COG})\right].$$

The two P values are independent, and under the null hypothesis that the SNP is not associated with the phenotype in either cohort, the two P values follow approximately the *Uniform*(0,1) distribution and the probability of concordance is 0.5; thus *S* has the following cdf under the null hypothesis: $F_S(x) = 0.5I(x \geq 0) + 0.5G(x;2,1), -\infty < x < \infty,$ where G is the Gamma(2,1) cdf. Then the combined P value for the SNP was calculated as $P = 1 - F_S(S)$.

***IL15* gene expression:** Expression of *IL15* (probe set 205992_s_at) in the diagnostic leukemia blasts was assessed by Affymetrix U133A microarray in a subset of St. Jude and COG MRD cohorts (n=194 in St. Jude and 148 in COG), as described.[4] Gene expression values were generated and normalized within the St. Jude and COG cohorts, using MAS5.0 software. Association with MRD was analyzed using linear regression. Relationship between *IL15* SNP (rs17007695) genotype (CC vs. TC/TT) and *IL15* gene expression was assessed in the combined St. Jude and COG cohort using a t-test.

Supplemental Reference List

(1) Rabbee N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 2006 January 1;22(1):7-12.

(2) Cheng C, Pounds SB, Boyett JM, Pei D, Kuo ML, Roussel MF. Statistical significance threshold criteria for analysis of microarray gene expression data. *Stat Appl Genet Mol Biol* 2004;3:Article36.

(3) Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 2005 October 10;23(29):7332-41.

(4) Flotho C, Coustan-Smith E, Pei D et al. Genes contributing to minimal residual disease in childhood acute lymphoblastic leukemia: prognostic significance of CASP8AP2. *Blood* 2006 August 1;108(3):1050-7.

Table 1S. Patient characteristics and relation to minimal residual disease (MRD).

		St. Jude		COG**	
		Patients (%) n=371	Relation to MRD (P value)	Patients (%) n=227	Relation to MRD (P value)
Race	Caucasian	282 (76%)		132 (58%)	
	African	58 (16%)		12 (5%)	
	Other	21 (8%)	0.5187	63 (37%)	0.1862
Gender	Male	213 (57%)		154 (68%)	
	Female	158 (43%)	0.2481	73 (32%)	0.2793
Age at Diagnosis	<1 year	6 (1%)		0	
	1-10 years	262 (71%)		73 (32%)	
	>10 years	103 (28%)	0.0849	154 (68%)	0.0439*
WBC at Diagnosis	<50k cells/mm^3	269 (73%)		124 (55%)	
	>50k cells/mm^3	102 (27%)	0.03682*	103 (45%)	0.1889
Lineage	B-lineage	299 (81%)		227 (100%)	NA
	T-cell	72 (19%)	0.4467	0 (0%)	
Molecular Subtype	TEL-AML1	65 (17%)	0.1251	3 (1%)	0.3769
	BCR-ABL	12 (3%)	<0.0001*	0	NA
	E2A-PBX1	13 (3%)	0.7338	23 (10%)	0.0012*
	MLL rearrangements	7 (2%)	0.241	18 (8%)	0.0027*
	No common translocations	274 (74%)	NA	183 (81%)	NA
End-of-induction MRD	<0.01%	293 (79%)		149 (66%)	
	0.01---1%	63 (17%)	NA	52 (23%)	NA
	>1%	15 (4%)		26 (11%)	

WBC: white blood cell count; MRD: minimal residual disease; NA: non-applicable.

Association between MRD (categorical variables, as defined in the Methods) and patient characteristics was assessed by using a Chi-square test.
*Older age(>10 years), high WBC (>50k cells/mm^3), presence of BCR-ABL, MLL rearrangements or absence of E2A-PBX1were associated with MRD positivity.
**While St. Jude trials included all newly-diagnosed ALL patients, the COG P9906 study was intended for children with high risk B-precursor ALL.

Table 2S. Association of the 102 overlapping SNPs with end-of-induction MRD..

SNP	Chr	Location	RS_ID	Gene	MAF	Risk Allele	P_SJ	P_COG	P_Comb	P_Adjusted	OR
SNP_A-1709241	7	8970101	rs10486275		0.32	T	0.0041	0.00931	2.13E-04	3.08E-04	2.05 (1.25, 3.36)
SNP_A-4291612	14	48095268	rs4522336		0.16	C	0.00622	0.0105	3.47E-04	5.87E-03	2.22 (1.12, 4.39)
SNP_A-2155892	13	100595841	rs7992226	NALCN	0.25	A	0.00448	0.00485	1.28E-04	2.38E-05	2.05 (1.2, 3.51)
SNP_A-2292283	2	100961907	rs1542178	NPAS2	0.28	G	0.00137	0.0115	9.50E-05	1.18E-04	0.464 (0.276, 0.779)
SNP_A-2239298	8	18673653	rs11203995	PSD3	0.49	C	0.00887	0.00932	4.30E-04	5.99E-04	1.68 (1.13, 2.49)
SNP_A-4236270	3	170154427	rs9871556		0.45	C	6.87E-04	0.00335	1.61E-05	7.91E-06	1.95 (1.3, 2.93)
SNP_A-2191667	1	215239015	rs1339219	ESRRG	0.45	C	0.00868	0.00589	2.78E-04	4.60E-04	0.576 (0.384, 0.865)
SNP_A-1740495	3	7887606	rs7652838		0.12	G	0.0114	0.00598	3.61E-04	2.79E-04	2.88 (1.2, 6.89)
SNP_A-1673416	6	55918441	rs4312989		0.34	C	0.00557	0.00819	2.51E-04	5.38E-04	0.503 (0.305, 0.827)
SNP_A-2297092	1	43493581	rs3862227		0.32	G	7.88E-04	0.0111	5.53E-05	1.22E-04	1.86 (1.24, 2.79)
SNP_A-2240822	19	39306923	rs11673011		0.31	G	0.00777	0.00613	2.61E-04	8.70E-04	1.92 (1.18, 3.11)
SNP_A-2277218	19	39303589	rs11666638		0.31	G	0.00531	0.00851	2.49E-04	8.73E-04	0.508 (0.312, 0.828)
SNP_A-4264692	5	3469372	rs17683231		0.07	A	0.0115	0.00179	1.21E-04	1.03E-04	0.419 (0.226, 0.777)
SNP_A-1980351	5	3448915	rs17682547		0.05	C	0.0024	0.0077	1.10E-04	3.39E-05	2.77 (1.4, 5.47)
SNP_A-2216221	5	3470577	rs17632091		0.05	A	0.00557	0.00363	1.19E-04	3.04E-05	3.2 (1.43, 7.13)
SNP_A-2172039	4	33719815	rs13106616		0.15	A	0.00598	0.00128	4.89E-05	1.70E-04	2.01 (1.27, 3.18)
SNP_A-1976736	4	33700472	rs12644671		0.16	G	0.00474	0.00403	1.13E-04	3.14E-04	1.94 (1.25, 2.99)
SNP_A-2009694	4	33405626	rs6844110		0.16	G	0.00224	0.00692	9.36E-05	1.12E-04	2.22 (1.39, 3.54)
SNP_A-1930229	4	33424921	rs1373494		0.15	C	0.00394	0.0113	2.45E-04	9.20E-05	2.19 (1.38, 3.47)
SNP_A-1907313	4	33406171	rs7691996		0.15	G	0.0037	0.00692	1.48E-04	2.29E-04	0.462 (0.289, 0.739)
SNP_A-2009695	4	33406328	rs7664751		0.15	A	0.00447	0.00692	1.76E-04	2.69E-04	2.14 (1.34, 3.41)
SNP_A-4194613	10	126638997	rs17152408	CTBP2	0.16	A	0.00512	0.00842	2.38E-04	1.10E-04	2.3 (1.33, 3.96)
SNP_A-1913543	10	126639506	rs17152411	CTBP2	0.15	G	0.0078	0.00766	3.20E-04	2.43E-04	2.05 (1.2, 3.51)
SNP_A-2265775	10	126574989	rs7900929	CTBP2	0.15	C	0.00495	0.00422	1.23E-04	2.01E-04	2.08 (1.22, 3.54)
SNP_A-2218245	10	126577363	rs7086128	CTBP2	0.15	G	0.00526	0.00422	1.30E-04	2.74E-04	2.03 (1.22, 3.4)
SNP_A-1682324	17	28297991	rs2521984		0.29	T	0.00996	0.0117	5.86E-04	1.02E-04	1.92 (1.16, 3.18)
SNP_A-4218391	1	176044231	rs4650977		0.36	G	0.00399	0.0096	2.14E-04	5.25E-04	0.543 (0.364, 0.81)
SNP_A-1829813	1	67037519	rs497710	INSL5	0.15	A	0.0081	0.00522	2.34E-04	1.52E-03	1.94 (1.16, 3.26)
SNP_A-4244750	11	115818726	rs11215936		0.06	A	0.00259	0.00151	2.63E-05	2.18E-04	2.82 (1.36, 5.84)
SNP_A-1853385	5	120633604	rs17147528		0.05	C	0.0109	0.0032	1.96E-04	4.14E-04	0.363 (0.167, 0.789)
SNP_A-4199696	14	99145812	rs4905865		0.29	T	0.00799	0.00586	2.57E-04	7.20E-04	1.63 (1.07, 2.48)
SNP_A-2241347	14	99145526	rs11160533		0.29	C	0.00799	0.0036	1.65E-04	5.46E-04	0.613 (0.403, 0.933)
SNP_A-1786996	10	108035647	rs1359645		0.08	G	0.00229	0.0123	1.62E-04	1.20E-04	2.17 (1.24, 3.78)

SNP	Chr	Location	RS_ID	Gene	MAF	Risk Allele	P_SJ	P_COG	P_Comb	P_Adjusted	OR
SNP_A-2261153	9	8751778	rs10116682	PTPRD	0.04	A	0.00252	9.24E-04	1.63E-05	2.64E-04	0.344 (0.161, 0.733)
SNP_A-1796782	9	8739291	rs6477346	PTPRD	0.04	G	0.00267	0.00812	1.27E-04	1.23E-03	2.59 (1.26, 5.34)
SNP_A-1675970	8	18691618	rs335251	PSD3	0.40	C	0.0113	0.00289	1.85E-04	1.13E-04	0.582 (0.386, 0.879)
SNP_A-1812242	2	173455958	rs2676527	RAPGEF4	0.03	C	0.0109	0.00826	4.64E-04	1.24E-02	3.03 (1.19, 7.71)
SNP_A-1849963	22	47310862	rs16999349	FAM19A5	0.14	A	0.00672	0.0107	3.79E-04	1.31E-03	0.478 (0.284, 0.804)
SNP_A-2146428	14	47157164	rs8004608		0.02	A	0.00434	0.00832	2.03E-04	4.43E-04	0.186 (0.0597, 0.576)
SNP_A-2084593	X	27766314	rs16988013		0.01	C	0.00211	0.011	1.35E-04	2.18E-04	0.132 (0.025, 0.695)
SNP_A-1663705	4	41406619	rs6823379		0.11	C	0.01	0.0125	6.24E-04	5.97E-05	2.01 (1.2, 3.36)
SNP_A-1953524	2	200752142	rs1881539	C2orf47	0.08	G	0.00728	0.00455	1.87E-04	9.92E-03	2.35 (1.19, 4.63)
SNP_A-1966464	2	200606142	rs10497845	C2orf47	0.08	C	0.0119	0.00438	2.83E-04	2.95E-03	2.52 (1.34, 4.75)
SNP_A-1807959	2	200440235	rs769951	C2orf47	0.09	G	0.00579	4.69E-04	1.88E-05	6.19E-04	0.467 (0.257, 0.847)
SNP_A-1794325	2	200703932	rs4673727	C2orf47	0.10	C	8.39E-04	1.55E-04	1.10E-06	1.90E-04	2.85 (1.56, 5.19)
SNP_A-2264953	2	200730199	rs1569175	C2orf47	0.11	T	0.00116	9.56E-05	9.43E-07	1.69E-04	0.366 (0.203, 0.658)
SNP_A-1717194	8	62310549	rs1486649		0.13	C	0.00222	0.0038	5.35E-05	1.47E-03	2.16 (1.28, 3.63)
SNP_A-2044445	8	62315228	rs10435604		0.13	C	0.00327	0.00208	4.39E-05	4.22E-03	0.498 (0.292, 0.851)
SNP_A-1785028	8	62315455	rs3864670		0.13	A	0.00303	0.004	7.47E-05	7.86E-03	0.478 (0.277, 0.824)
SNP_A-2207718	8	62315603	rs3864671		0.13	C	0.00303	0.00247	4.79E-05	4.01E-03	0.478 (0.277, 0.824)
SNP_A-2047557	6	118634581	rs1413846	SLC35F1	0.39	A	0.00566	0.0107	3.24E-04	7.98E-04	1.8 (1.2, 2.7)
SNP_A-2061448	6	118634711	rs1413845	SLC35F1	0.39	C	0.00796	0.00566	2.48E-04	6.57E-04	1.77 (1.18, 2.65)
SNP_A-2295593	4	142948222	rs17015014	IL15	0.19	C	0.00425	0.0044	1.11E-04	1.11E-04	2.2 (1.42, 3.41)
SNP_A-1676613	4	142873151	rs10519612	IL15	0.13	C	0.0117	0.0106	6.20E-04	1.03E-03	2.42 (1.41, 4.15)
SNP_A-4264519	4	142873534	rs10519613	IL15	0.13	A	0.00769	0.00291	1.31E-04	8.62E-04	2.18 (1.3, 3.68)
SNP_A-2062945	4	142929173	rs17007695	IL15	0.10	C	4.43E-04	2.34E-04	8.85E-07	1.31E-04	2.67 (1.53, 4.68)
SNP_A-1958136	4	142899038	rs35964658	IL15	0.12	G	0.00225	0.00274	4.01E-05	1.27E-03	2.38 (1.38, 4.1)
SNP_A-4272012	22	20139185	rs463426	HIC2	0.39	T	0.0052	0.0109	3.05E-04	2.71E-03	1.73 (1.14, 2.62)
SNP_A-4259020	22	20136401	rs460106	HIC2	0.40	T	0.0119	0.00935	5.62E-04	1.97E-03	0.585 (0.382, 0.896)
SNP_A-4271870	2	157986459	rs16841722	PSCDBP	0.03	C	0.00258	0.0101	1.51E-04	2.37E-03	4.16 (1.53, 11.3)
SNP_A-1888709	2	157986491	rs3769376	PSCDBP	0.03	C	0.00258	0.0101	1.51E-04	2.37E-03	0.24 (0.0883, 0.654)
SNP_A-1787461	10	120417302	rs1312895	C1orf46	0.17	T	0.00241	0.00949	1.34E-04	4.09E-03	2.04 (1.21, 3.42)
SNP_A-2039544	10	120418749	rs1247118	C1orf46	0.16	A	0.00308	0.0125	2.15E-04	4.04E-03	2.01 (1.19, 3.37)
SNP_A-2171639	5	148869211	rs752822	CSNK1A1	0.27	T	0.0069	0.0044	1.73E-04	3.69E-04	1.75 (1.11, 2.76)
SNP_A-2069005	16	77996875	rs4888024		0.44	G	0.00321	0.0116	2.08E-04	6.75E-04	1.84 (1.21, 2.79)
SNP_A-4272973	11	95639748	rs7115578	MAML2	0.36	A	0.00205	5.03E-04	7.62E-06	2.36E-05	0.539 (0.359, 0.811)
SNP_A-1919387	22	18703211	rs6518604		0.02	G	0.00233	0.00475	6.87E-05	1.64E-04	0.201 (0.0699, 0.575)
SNP_A-1988256	6	155971493	rs35229355		0.03	T	7.24E-05	0.00899	4.96E-06	4.84E-04	7.24 (2.46, 21.3)

SNP	Chr	Location	RS_ID	Gene	MAF	Risk Allele	P_SJ	P_COG	P_Comb	P_Adjusted	OR
SNP_A-2139851	10	122890613	rs2901286		0.03	A	4.05E-04	0.00123	3.86E-06	6.28E-06	4.66 (1.58, 13.7)
SNP_A-1676306	10	122898209	rs10510089		0.04	C	0.0117	0.00114	8.15E-05	3.99E-04	0.362 (0.146, 0.9)
SNP_A-1980357	5	3587426			0.04	G	0.00549	9.51E-04	3.44E-05	6.69E-06	3.18 (1.31, 7.71)
SNP_A-1969697	21	27030694	rs2409191		0.11	A	0.00145	0.0106	9.29E-05	6.36E-04	0.502 (0.294, 0.857)
SNP_A-1930568	1	47515776	rs11211503	STIL	0.02	A	0.00654	0.00889	3.13E-04	2.51E-03	0.219 (0.0678, 0.708)
SNP_A-1754353	1	207185511	rs7517671		0.28	A	0.0112	0.00883	5.05E-04	4.54E-04	1.78 (1.16, 2.74)
SNP_A-1683695	14	36348715	rs1631933	SLC25A21	0.39	G	0.0123	0.00645	4.14E-04	1.13E-03	0.554 (0.356, 0.861)
SNP_A-1990753	7	103331139	rs6951651	RELN	0.41	C	0.00661	0.00626	2.30E-04	4.52E-04	1.85 (1.19, 2.86)
SNP_A-1937715	1	238934099	rs16840493		0.08	T	0.00341	0.00267	5.74E-05	2.58E-03	2.8 (1.4, 5.58)
SNP_A-1667637	7	19876055	rs10264784		0.17	C	0.00699	0.0103	3.79E-04	1.68E-03	1.86 (1.19, 2.91)
SNP_A-4285668	2	75564253	rs1158392	TMEM166	0.22	A	0.00965	0.00795	4.02E-04	1.99E-03	1.78 (1.16, 2.74)
SNP_A-1918014	5	36173377	rs267759	LMBRD2	0.05	A	2.88E-04	0.00316	6.78E-06	2.43E-04	3.23 (1.52, 6.87)
SNP_A-4234252	7	15757351	rs17169056		0.02	G	0.00396	5.19E-04	1.45E-05	1.82E-03	4.28 (1.3, 14)
SNP_A-1958126	6	94497325	rs9345389		0.05	G	0.00201	0.0094	1.12E-04	5.16E-03	0.312 (0.137, 0.708)
SNP_A-2152188	6	53564641	rs486060		0.09	C	0.00376	0.0118	2.45E-04	4.58E-03	2.59 (1.42, 4.73)
SNP_A-4281758	12	114362160	rs11067600		0.07	C	0.00391	0.00225	5.56E-05	7.66E-03	0.352 (0.164, 0.754)
SNP_A-2041620	6	131511929	rs6917207	AKAP7	0.16	T	0.0111	0.00118	8.02E-05	3.53E-04	0.538 (0.321, 0.902)
SNP_A-2174556	1	54847650	rs2289015	ACOT11	0.05	T	2.38E-04	0.0105	1.74E-05	4.65E-06	3.67 (1.76, 7.65)
SNP_A-1727090	11	106935264	rs10502094	ALKBH8	0.04	G	0.00252	0.00515	7.95E-05	2.60E-04	0.319 (0.144, 0.707)
SNP_A-1793591	11	20519095	rs7128311		0.03	C	5.90E-05	0.011	4.95E-06	1.32E-03	13.9 (2.72, 71.1)
SNP_A-1718563	2	17888754	rs10495669		0.46	C	0.0063	0.00919	3.11E-04	3.57E-04	0.56 (0.371, 0.846)
SNP_A-2184177	5	27993195	rs389719		0.25	A	6.83E-04	0.00428	2.01E-05	5.97E-04	1.89 (1.23, 2.89)
SNP_A-1750447	9	23765449	rs959091	ELAVL2	0.12	G	0.0124	0.011	6.75E-04	2.35E-03	2.12 (1.21, 3.72)
SNP_A-4233826	7	37232877	rs4723619	ELMO1	0.07	C	8.70E-04	4.45E-04	3.05E-06	4.59E-05	3.01 (1.5, 6.03)
SNP_A-4285077	16	20227588	rs4078023	GP2	0.04	G	0.0094	0.00546	2.79E-04	1.50E-02	0.436 (0.204, 0.931)
SNP_A-1709114	10	8190719	rs10508343		0.04	A	6.00E-04	0.00169	7.50E-06	1.52E-03	3.81 (1.4, 10.4)
SNP_A-4249789	20	45642012	rs6125048	NCOA3	0.04	T	0.00867	3.37E-05	2.34E-06	1.55E-03	2.73 (1.08, 6.88)
SNP_A-1713408	2	40842565	rs1012620		0.04	G	0.0085	0.00831	3.73E-04	9.09E-04	0.227 (0.0751, 0.688)
SNP_A-1892341	10	17428149	rs359312	ST8SIA6	0.04	T	0.00103	9.57E-06	9.58E-08	1.53E-06	3.91 (1.52, 10.1)
SNP_A-1916779	11	102137403	rs17099545		0.03	G	0.0112	0.00942	5.36E-04	3.20E-03	3.9 (1.25, 12.1)
SNP_A-2158344	4	26665455	rs6851702		0.07	G	0.00586	0.0107	3.35E-04	2.07E-03	2.17 (1.15, 4.08)
SNP_A-2105458	7	14412442	rs6971925	DGKB	0.02	T	5.44E-05	0.00777	3.31E-06	1.88E-04	0.0719 (0.0141, 0.367)
SNP_A-4277336	20	45061176	rs12481102	EYA2	0.03	T	0.00605	0.00863	2.84E-04	1.24E-02	0.319 (0.1, 1.01)
SNP_A-2242289	20	45077823	rs3827063	EYA2	0.03	A	0.0038	0.0117	2.45E-04	1.08E-02	0.248 (0.0858, 0.716)

SNP indicates the Affymetrix ID of a single nucleotide polymorphism (SNP) marker; Chr: chromosome; Location: physical location of a SNP based on Mar. 2006 human genome assembly (hg18); RS_ID: SNP ID according to the dbSNP database; Gene: gene annotation of the SNP (detailed in the supplemental information); MAF: minor allele frequency in the patients included in the genome-wide association analyses; Risk allele: allele that conferred risk of MRD-positive status; P_SJ: P value for association with MRD in the St. Jude cohort; P_COG: P value for association with MRD in the COG cohort; P_comb: P value for association with MRD in the combined analysis; P_Adjusted: P value for association with MRD after adjusting for race, age, and molecular ALL subtype; OR: odds ratio for MRD-positive status (values in parenthesis indicate the 95% confidence interval). SNPs in strong linkage disequilibrium (r^2>0.5) are grouped together, and each group is indicated by a distinct color.

Table 3S. Relation of MRD-associated SNPs to other phenotypes.

SNP	Chr	Location	RS_ID	Gene	MAF	MTX_CL	MTX_PG	VP_CL	Relapse_Crr	Relapse_Cin	Super_MRD_SJ	Super_MRD_COG
SNP_A-1709241	7	8970101	rs10486275		0.32	0.2015	0.4868	0.9939	>0.5	>0.5	0.127	0.00213
SNP_A-4291612	14	48095268	rs4522336		0.16	0.1672	0.2845	0.0271	>0.5	>0.5	0.191	0.12
SNP_A-2155892	13	100595841	rs7992226	NALCN	0.25	0.0107	0.5688	0.9710	0.0645	0.0288	0.0234	0.0000416
SNP_A-2292283	2	100961907	rs1542178	NPAS2	0.28	0.8772	0.2724	0.3157	>0.5	>0.5	0.0316	0.111
SNP_A-2239298	8	18673653	rs11203995	PSD3	0.49	0.2299	0.2207	0.4312	0.1111	0.0259	0.000736	0.011
SNP_A-4236270	3	170154427	rs9871556		0.45	0.0969	0.0322	0.8351	>0.5	>0.5	0.0144	0.0435
SNP_A-2191667	1	215239015	rs1339219	ESRRG	0.45	0.3197	0.7334	0.8202	>0.5	>0.5	0.0749	0.00873
SNP_A-1740495	3	7887606	rs7652838		0.12	0.1112	0.6120	0.2652	0.0371	0.3444	0.148	0.00338
SNP_A-1673416	6	55918441	rs4312989		0.34	0.6307	0.8674	0.7702	>0.5	>0.5	0.0164	0.0162
SNP_A-2297092	1	43493581	rs3862227		0.32	0.0326	0.0340	0.4461	>0.5	>0.5	0.00337	0.0993
SNP_A-2240822	19	39306923	rs11673011		0.31	0.4930	0.1939	0.2842	>0.5	>0.5	0.257	0.171
SNP_A-2277218	19	39303589	rs11666638		0.31	0.5229	0.4068	0.2850	>0.5	>0.5	0.383	0.187
SNP_A-4264692	5	3469372	rs17683231		0.07	0.5087	0.6992	0.2916	0.0699	0.0930	0.00682	0.0122
SNP_A-1980351	5	3448915	rs17682547		0.05	0.5999	0.9808	0.8392	>0.5	>0.5	0.00213	0.0923
SNP_A-2216221	5	3470577	rs17632091		0.05	0.6199	0.6855	0.6474	>0.5	>0.5	0.00223	0.0282
SNP_A-2172039	4	33719815	rs13106616		0.15	0.6923	0.2720	0.0031	0.0810	0.0170	0.0486	0.0127
SNP_A-1976736	4	33700472	rs12644671		0.16	0.4930	0.3352	0.0013	>0.5	>0.5	0.0244	0.0587
SNP_A-2009694	4	33405626	rs6844110		0.16	0.5156	0.7019	0.0080	>0.5	>0.5	0.00606	0.0783
SNP_A-1930229	4	33424921	rs1373494		0.15	0.6200	0.4105	0.0214	>0.5	>0.5	0.0123	0.128
SNP_A-1907313	4	33406171	rs7691996		0.15	0.6615	0.5074	0.0214	>0.5	>0.5	0.0163	0.0783
SNP_A-2009695	4	33406328	rs7664751		0.15	0.6980	0.5561	0.0214	>0.5	>0.5	0.0138	0.0783
SNP_A-4194613	10	126638997	rs17152408	CTBP2	0.16	0.9001	0.4255	0.3874	>0.5	>0.5	0.0736	0.000344
SNP_A-1913543	10	126639506	rs17152411	CTBP2	0.15	0.9434	0.3253	0.4292	>0.5	>0.5	0.121	0.000458
SNP_A-2265775	10	126574989	rs7900929	CTBP2	0.15	0.9701	0.6613	0.3353	>0.5	>0.5	0.119	0.000311
SNP_A-2218245	10	126577363	rs7086128	CTBP2	0.15	0.9576	0.6113	0.3010	>0.5	>0.5	0.104	0.000311
SNP_A-1682324	17	28297991	rs2521984		0.29	0.1653	0.9453	0.2028	>0.5	>0.5	0.137	0.0027
SNP_A-4218391	1	176044231	rs4650977		0.36	0.1272	0.3960	0.9254	>0.5	>0.5	0.0163	0.031
SNP_A-1829813	1	67037519	rs497710	INSL5	0.15	0.2148	0.3959	0.8976	>0.5	>0.5	0.349	0.368
SNP_A-4244750	11	115818726	rs11215936		0.06	0.7779	0.8878	0.7372	>0.5	>0.5	0.0177	0.0582
SNP_A-1853385	5	120633604	rs17147528		0.05	0.3643	-	0.8412	>0.5	>0.5	0.0388	0.0495
SNP_A-4199696	14	99145812	rs4905865		0.29	0.0052	0.2860	0.8084	>0.5	>0.5	0.0704	0.0466
SNP_A-2241347	14	99145526	rs11160533		0.29	0.0055	0.2972	0.7598	>0.5	>0.5	0.0704	0.0384
SNP_A-1786996	10	108035647	rs1359645		0.08	0.0496	0.1184	0.4356	0.0937	0.0472	0.007	0.00666

SNP	Chr	Location	RS_ID	Gene	MAF	MTX_CL	MTX_PG	VP_CL	Relapse_Crr	Relapse_Cin	Super_MRD_SJ	Super_MRD_COG
SNP_A-2261153	9	8751778	rs10116682	PTPRD	0.04	0.4367	0.0233	0.2568	>0.5	>0.5	0.00165	0.0121
SNP_A-1796782	9	8739291	rs6477346	PTPRD	0.04	0.9091	0.0633	0.8535	>0.5	>0.5	0.00166	0.0343
SNP_A-1675970	8	18691618	rs335251	PSD3	0.40	0.0678	0.7568	0.5179	>0.5	>0.5	0.00778	0.00129
SNP_A-1812242	2	173455958	rs2676527	RAPGEF4	0.03	0.9034	-	0.7783	>0.5	>0.5	0.123	0.0313
SNP_A-1849963	22	47310862	rs16999349	FAM19A5	0.14	0.9215	0.8270	0.9678	>0.5	>0.5	0.0211	0.00382
SNP_A-2146428	14	47157164	rs8004608		0.02	0.5118	0.0176	0.1855	>0.5	>0.5	0.000613	0.0308
SNP_A-2084593	X	27766314	rs16988013		0.01	0.3033	-	0.3909	>0.5	>0.5	NA	NA
SNP_A-1663705	4	41406619	rs6823379		0.11	0.3114	0.0844	0.9853	>0.5	>0.5	0.00318	0.0162
SNP_A-1953524	2	200752142	rs1881539	C2orf47	0.08	0.8912	0.6110	0.3591	0.0890	0.0061	0.0317	0.0324
SNP_A-1966464	2	200606142	rs10497845	C2orf47	0.08	0.2825	0.4286	0.5898	>0.5	>0.5	0.0449	0.0152
SNP_A-1807959	2	200440235	rs769951	C2orf47	0.09	0.6740	0.0831	0.5437	0.0101	0.0004	0.0194	0.00248
SNP_A-1794325	2	200703932	rs4673727	C2orf47	0.10	0.6626	0.8518	0.4509	0.0264	0.0011	0.0058	0.000777
SNP_A-2264953	2	200730199	rs1569175	C2orf47	0.11	0.5494	0.6819	0.7942	0.0455	0.0013	0.0105	0.000253
SNP_A-1717194	8	62310549	rs1486649		0.13	0.1142	0.3372	0.5838	0.0003	0.0001	0.0637	0.011
SNP_A-2044445	8	62315228	rs10435604		0.13	0.2061	0.2490	0.7501	0.0009	0.0001	0.0841	0.00662
SNP_A-1785028	8	62315455	rs3864670		0.13	0.1846	0.2213	0.7464	0.0028	0.0003	0.0858	0.0124
SNP_A-2207718	8	62315603	rs3864671		0.13	0.1846	0.2213	0.7464	0.0032	0.0003	0.0858	0.00792
SNP_A-2047557	6	118634581	rs1413846	SLC35F1	0.39	0.4978	0.2951	0.6173	0.1107	0.0353	0.153	0.373
SNP_A-2061448	6	118634711	rs1413845	SLC35F1	0.39	0.5280	0.3949	0.6173	0.0582	0.0203	0.14	0.284
SNP_A-2295593	4	142948222	rs17015014	IL15	0.19	0.6149	0.6441	0.0532	0.0504	0.0059	0.0765	0.0299
SNP_A-1676613	4	142873151	rs10519612	IL15	0.13	0.3771	0.2475	0.7285	>0.5	>0.5	0.181	0.141
SNP_A-4264519	4	142873534	rs10519613	IL15	0.13	0.1104	0.3842	0.4157	>0.5	>0.5	0.0918	0.0437
SNP_A-2062945	4	142929173	rs17007695	IL15	0.10	0.6519	0.3365	0.0218	>0.5	>0.5	0.00286	0.00277
SNP_A-1958136	4	142899038	rs35964658	IL15	0.12	0.8522	0.1878	0.0878	>0.5	>0.5	0.00968	0.0281
SNP_A-4272012	22	20139185	rs463426	HIC2	0.39	0.2998	0.7333	0.7689	>0.5	>0.5	0.124	0.0558
SNP_A-4259020	22	20136401	rs460106	HIC2	0.40	0.2446	0.9825	0.5574	>0.5	>0.5	0.244	0.0568
SNP_A-4271870	2	157986459	rs16841722	PSCDBP	0.03	0.2509	0.4009	0.7206	0.0031	0.0027	0.00785	0.0688
SNP_A-1888709	2	157986491	rs3769376	PSCDBP	0.03	0.2509	0.4009	0.7206	0.0031	0.0027	0.00785	0.0688
SNP_A-1787461	10	120417302	rs1312895	C1orf46	0.17	0.3179	0.5381	0.1805	>0.5	>0.5	0.231	0.0106
SNP_A-2039544	10	120418749	rs1247118	C1orf46	0.16	0.2761	0.4942	0.2381	>0.5	>0.5	0.286	0.00705
SNP_A-2171639	5	148869211	rs752822	CSNK1A1	0.27	0.1743	0.6729	0.6661	>0.5	>0.5	0.0368	0.0108
SNP_A-2069005	16	77996875	rs4888024		0.44	0.0312	0.2022	0.0496	>0.5	>0.5	0.206	0.0661
SNP_A-4272973	11	95639748	rs7115578	MAML2	0.36	0.4961	0.7099	0.7714	0.0920	0.0445	0.608	0.00232
SNP_A-1919387	22	18703211	rs6518604		0.02	0.4620	0.7465	0.1677	>0.5	>0.5	0.115	0.014
SNP_A-1988256	6	155971493	rs35229355		0.03	0.1893	0.2276	0.4553	>0.5	>0.5	0.00113	0.0314

SNP	Chr	Location	RS_ID	Gene	MAF	MTX_CL	MTX_PG	VP_CL	Relapse_Crr	Relapse_Cin	Super_MRD_SJ	Super_MRD_COG
SNP_A-2139851	10	122890613	rs2901286		0.03	0.8291	-	0.9788	>0.5	>0.5	0.00878	0.0146
SNP_A-1676306	10	122898209	rs10510089		0.04	0.2788	-	0.5226	>0.5	>0.5	0.0779	0.0124
SNP_A-1980357	5	3587426			0.04	0.8215	0.4855	0.3041	>0.5	>0.5	0.266	0.00745
SNP_A-1969697	21	27030694	rs2409191		0.11	0.6411	0.0318	0.1473	0.0092	0.0037	0.0119	0.135
SNP_A-1930568	1	47515776	rs11211503	STIL	0.02	0.7855	0.7056	0.8514	>0.5	>0.5	0.00817	0.00745
SNP_A-1754353	1	207185511	rs7517671		0.28	0.1225	0.2016	0.2314	>0.5	>0.5	0.239	0.0122
SNP_A-1683695	14	36348715	rs1631933	SLC25A21	0.39	0.1950	0.8370	0.8865	0.0717	0.0847	0.245	0.00847
SNP_A-1990753	7	103331139	rs6951651	RELN	0.41	0.6271	0.3331	0.5181	0.0939	0.0613	0.0253	0.00446
SNP_A-1937715	1	238934099	rs16840493		0.08	0.7309	0.4053	0.4634	0.0088	0.0016	0.112	0.00784
SNP_A-1667637	7	19876055	rs10264784		0.17	0.6729	0.0772	0.0912	>0.5	>0.5	0.116	0.0818
SNP_A-4285668	2	75564253	rs1158392	TMEM166	0.22	0.3522	0.6224	0.6976	0.0008	0.0184	0.0617	0.0528
SNP_A-1918014	5	36173377	rs267759	LMBRD2	0.05	0.5282	0.6447	0.3447	>0.5	>0.5	0.000835	0.0324
SNP_A-4234252	7	15757351	rs17169056		0.02	0.9997	0.5130	0.9304	>0.5	>0.5	0.0217	0.00897
SNP_A-1958126	6	94497325	rs9345389		0.05	0.0163	-	0.6142	>0.5	>0.5	0.0262	0.0316
SNP_A-2152188	6	53564641	rs486060		0.09	0.8782	0.7213	0.8557	0.1086	0.3241	0.124	0.0103
SNP_A-4281758	12	114362160	rs11067600		0.07	0.2257	0.6783	0.7131	0.0007	0.0090	0.0406	0.00575
SNP_A-2041620	6	131511929	rs6917207	AKAP7	0.16	0.8531	0.3997	0.9289	>0.5	>0.5	0.583	0.00204
SNP_A-2174556	1	54847650	rs2289015	ACOT11	0.05	0.2886	0.5094	0.1276	>0.5	>0.5	0.0773	0.0207
SNP_A-1727090	11	106935264	rs10502094	ALKBH8	0.04	0.6271	0.0268	0.8139	0.0070	0.0852	0.0112	0.0103
SNP_A-1793591	11	20519095	rs7128311		0.03	0.8920	0.5406	-	0.0099	0.0000	0.00574	0.0126
SNP_A-1718563	2	17888754	rs10495669		0.46	0.4493	0.1548	0.1760	>0.5	>0.5	0.0274	0.0363
SNP_A-2184177	5	27993195	rs389719		0.25	0.4633	0.3583	0.0939	>0.5	>0.5	0.0221	0.00804
SNP_A-1750447	9	23765449	rs959091	ELAVL2	0.12	0.7085	0.4046	0.9663	>0.5	>0.5	0.0055	0.189
SNP_A-4233826	7	37232877	rs4723619	ELMO1	0.07	0.0487	0.2028	0.0226	0.0057	0.0005	0.0124	0.00194
SNP_A-4285077	16	20227588	rs4078023	GP2	0.04	0.8111	0.3679	0.9158	0.0000	0.0000	0.0118	0.0129
SNP_A-1709114	10	8190719	rs10508343		0.04	0.3221	0.8737	0.5704	>0.5	>0.5	0.0648	0.00691
SNP_A-4249789	20	45642012	rs6125048	NCOA3	0.04	0.5625	0.5749	0.6604	0.0007	0.0017	0.427	0.0000858
SNP_A-1713408	2	40842565	rs1012620		0.04	0.2495	0.8475	0.0571	>0.5	>0.5	0.149	0.0247
SNP_A-1892341	10	17428149	rs359312	ST8SIA6	0.04	0.9217	0.5195	1.0000	0.0332	0.0008	0.00797	0.000297
SNP_A-1916779	11	102137403	rs17099545		0.03	0.3694	0.7429	0.2464	0.0077	0.0000	0.338	0.0184
SNP_A-2158344	4	26665455	rs6851702		0.07	0.0662	0.5900	0.9250	>0.5	>0.5	0.0364	0.0458
SNP_A-2105458	7	14412442	rs6971925	DGKB	0.02	0.5431	0.1463	1.0000	0.0731	0.0081	0.0305	0.2
SNP_A-4277336	20	45061176	rs12481102	EYA2	0.03	0.6149	0.3965	0.1146	0.0051	0.0027	0.684	0.00753
SNP_A-2242289	20	45077823	rs3827063	EYA2	0.03	0.7119	0.3472	0.1704	0.0007	0.0004	0.436	0.00637

SNP, Chr, Location, RS_ID, Gene, and MAF are defined the same as described in legend of Table 2S. MTX_CL: P values for association with methotrexate clearance; MTX_PG: P values for association with methotrexate polyglutamates in leukemic blasts; VP_CL: P values for association with etoposide clearance; Relapse_Crr: P values for association with hematologic relapse (regression test); Relapse_Cin: P values for association with hematologic relapse (Gray's test); Super_MRD_SJ: P values for association with early response in St. Jude; Super_MRD_COG: P values for association with early response in COG. SNPs in strong linkage disequilibrium (r^2>0.5) are grouped together, and each group is indicated by a distinct color.

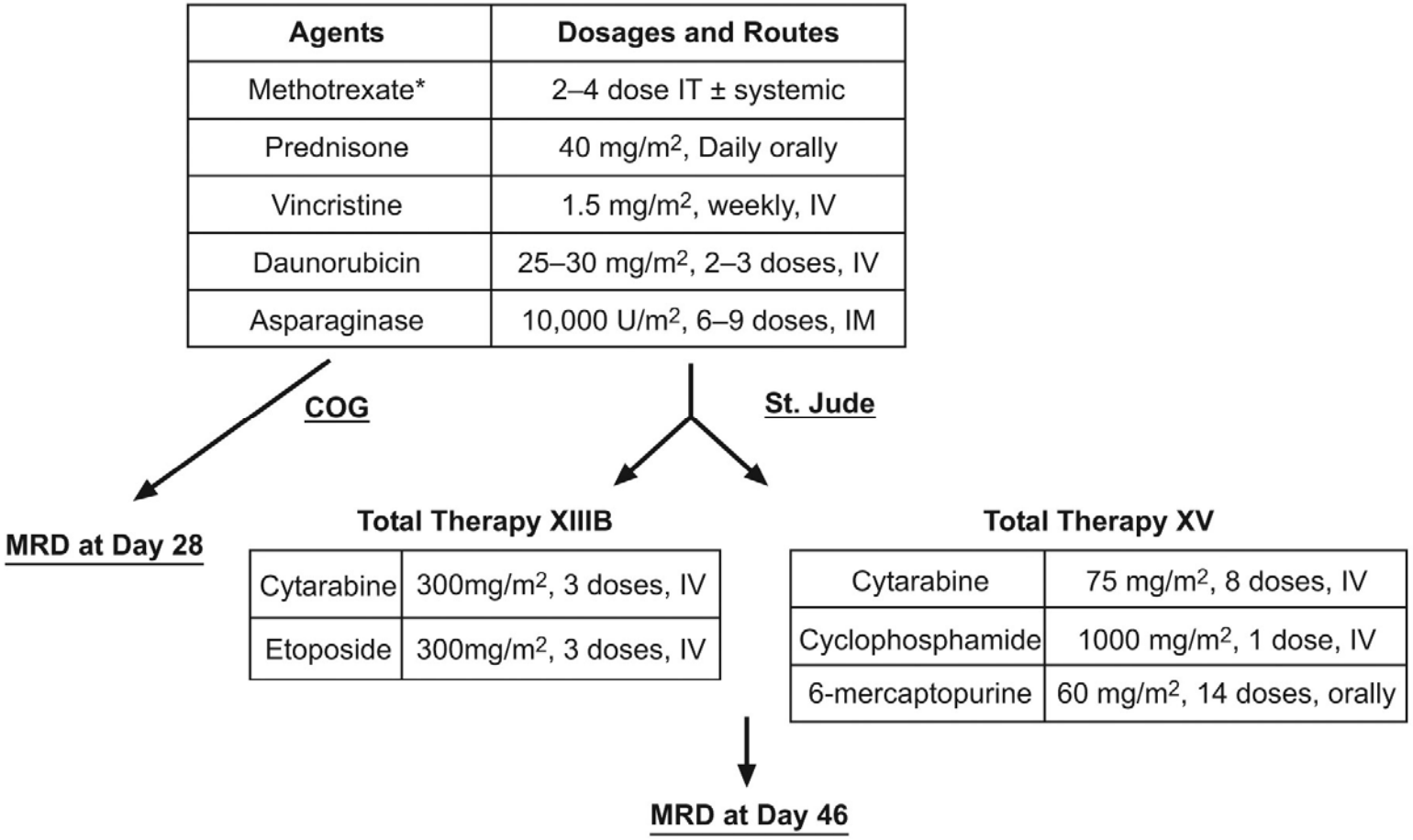

Figure 1S Remission induction regimens for St. Jude Total Therapy XIIIB, XV and COG 9906 protocols. IT: intrathecal; IV: intravenous; IM: intramuscular. *Methotrexate was given intrathecally, with or without cytarabine and hydrocortisone, as prophylaxis for central nerve system disease in both St. Jude and COG. However, some St. Jude patients also received the drug orally or intravenously for a single day, in addition to the IT methotrexate. Further details of the treatment regimens can be found in references 21, 22 and at http://www.acor.org/ped-onc/diseases/ALLtrials/9906.html.

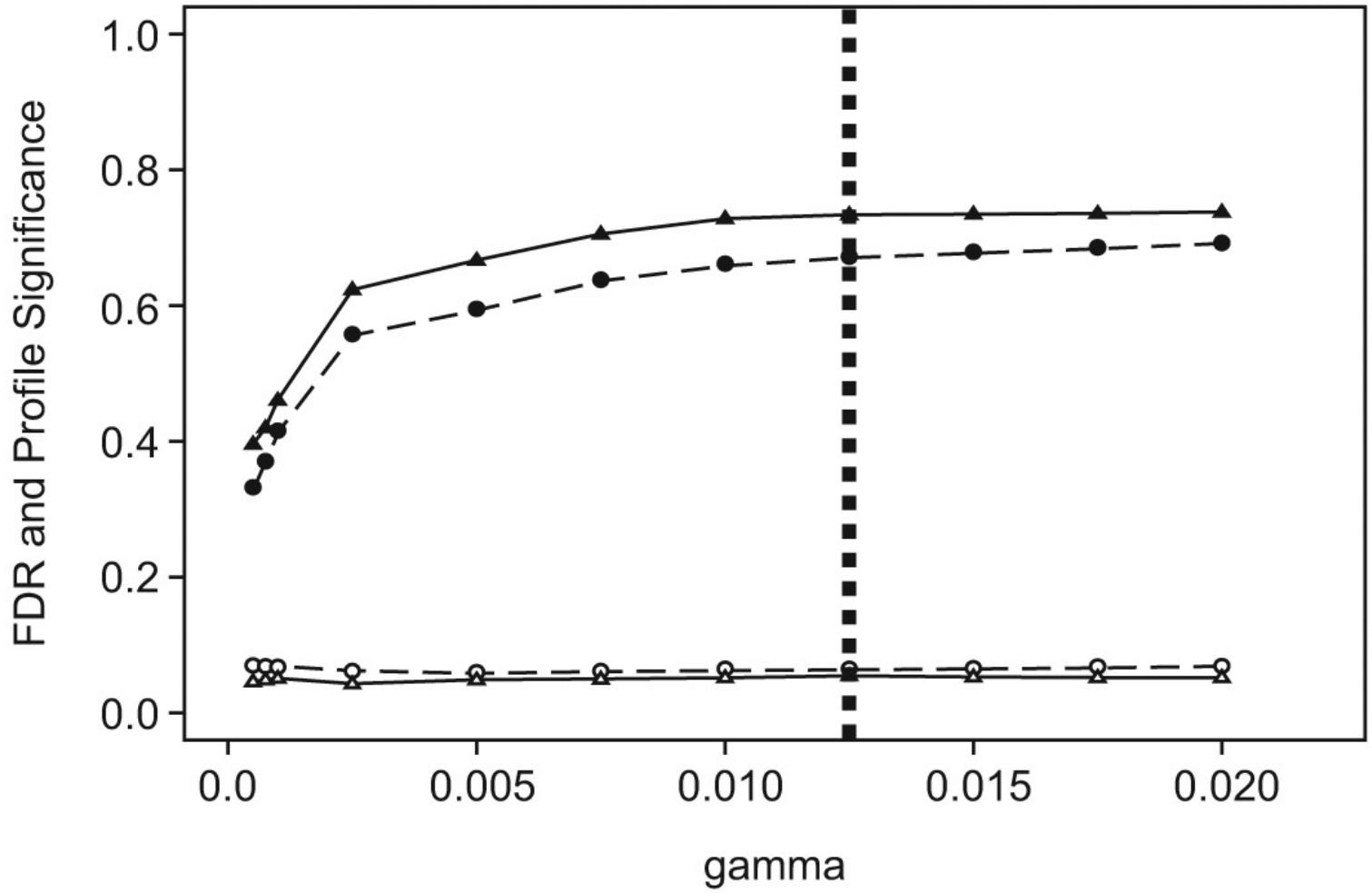

Figure 2S. False discovery rate (FDR) and profile significance in the St. Jude (n=318) and COG (n=169) MRD cohorts. Dotted lines with solid and open circles indicate FDR and profile significance, respectively, in the St. Jude cohort at pre-determined per-test significance levels (P value cutoffs/gamma) for SNP genotypes as predictors of MRD. Solid lines with solid and open triangles indicate FDR and profile significance, respectively, in the COG cohort. Dashed vertical line indicates P=0.0125.

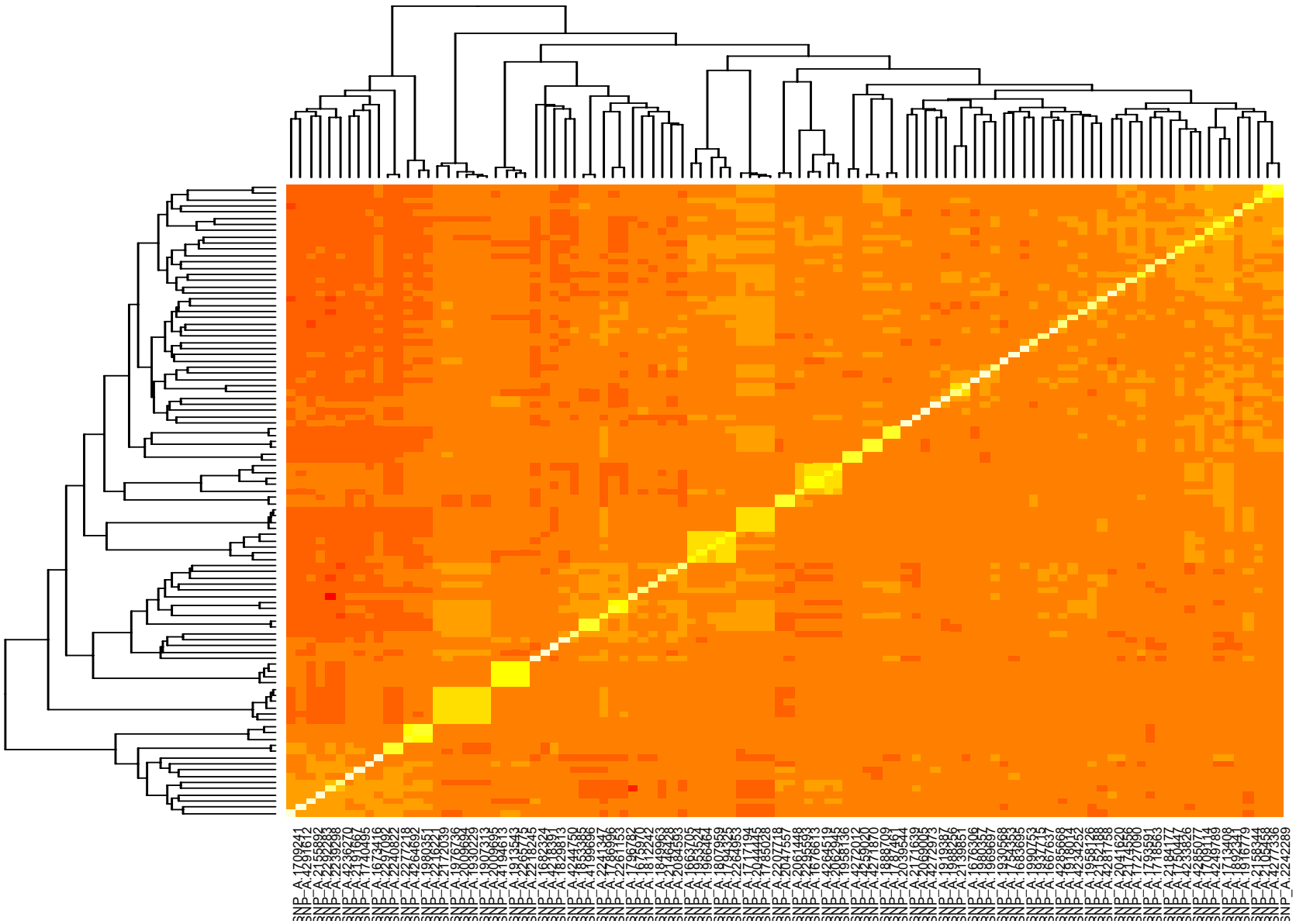

Figure 3S. Linkage disequilibrium (LD) among the 102 overlapping SNPs predicting MRD in St. Jude and COG MRD cohorts. SNPs are listed in the same order from left to right horizontally and from the bottom to the top vertically. Degree of correlation is indicated by color: yellow represents high correlation (strong LD) and red represents low correlation (weak LD).

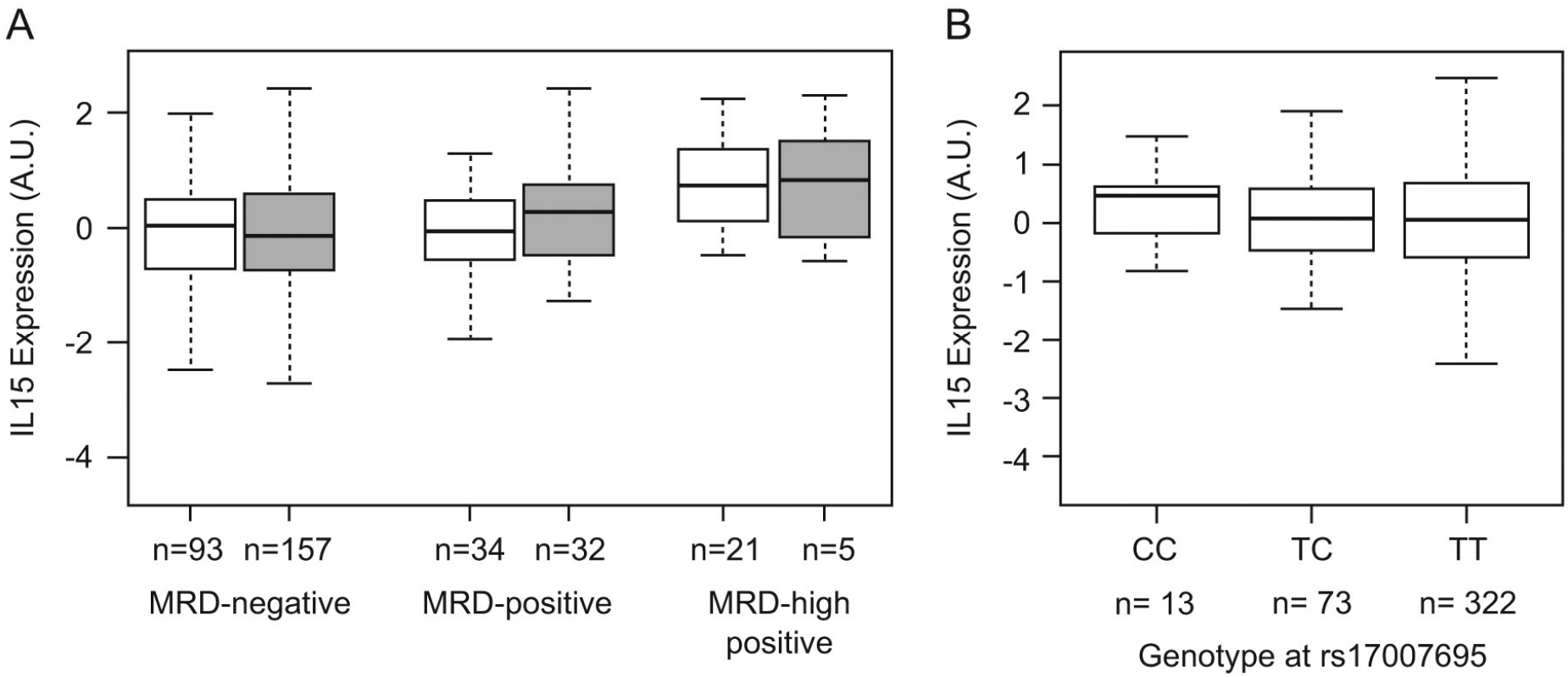

Figure 4S. Association of tumor IL15 gene expression with MRD and IL15 SNP genotype. A, Relation of IL15 gene expression in diagnostic leukemic blasts to MRD in St. Jude (gray boxes, P=0.0342) and COG (open boxes, P=0.0035). Patients were classified into MRD-negative, positive, and high-positive, as defined in Materials and Methods. **B,** IL15 gene expression in diagnostic blasts was higher in patients with the CC genotype at IL15 SNP rs17007695 (P=0.0701). IL15 expression values were normalized within the St. Jude and COG cohorts. Boxes include data between the 25th and the 75th percentiles, and whiskers represent the maximal and minimal values after excluding outliers.

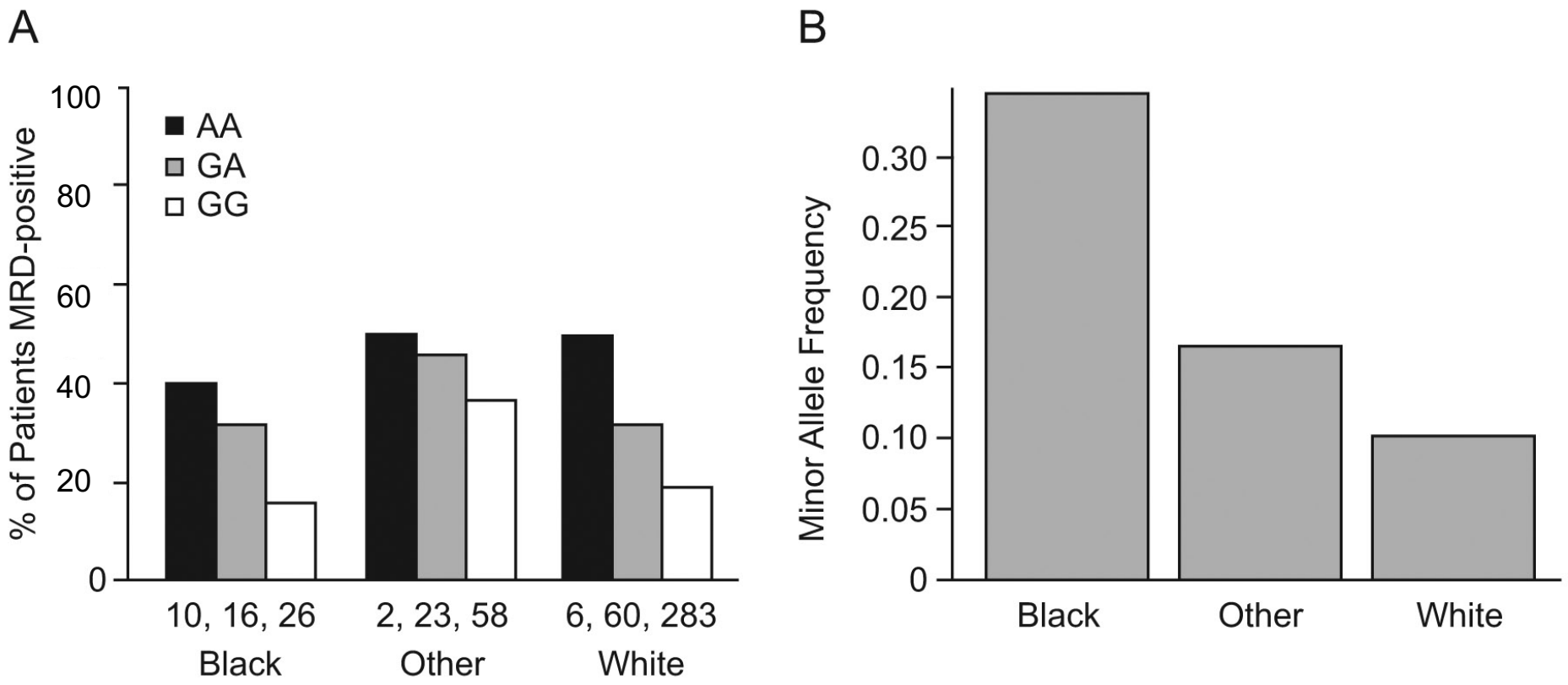

Figure 5S. MRD-positivity by genotype at rs13106616 in different racial groups. A, Comparison of association of rs13106616 genotype with MRD by race. Numbers below the X axis indicate the number of patients in that category. **B**, Minor allele frequency of rs13106616 by race. Black: African American; White: Caucasian American; Other: all other patients.