SUPPLEMENTARY MATERIALS

Supplementary Tables

Supplementary Table 1. Demographic and clinical characteristics of T-ALL cases included in this study*

Feature	Discovery GWAS (n=1191)	Replication (n=117)
Genetically-defined race/ethnicity†, No. (%)		
African	157 (13.2)	26 (22.2)
European	699 (58.7)	75 (64.1)
Hispanic	126 (10.6)	9 (7.7)
Other	209 (17.5)	7 (6)
Gender, No. (%)		
Female	330 (27.7)	27 (23.1)
Male	861 (72.3)	90 (76.9)
Median age at diagnosis, y (5th to 95th percentiles in years)	9.1 (2.3 - 18.3)	8.55 (2.45 - 16.9)
Central nervous system disease status‡, No. (%)		
CNS1	496 (41.6)	54 (46.2)
Other	695 (59.4)	63 (53.8)

*The discovery GWAS cases were from the COG AALL0434 and St. Jude Total XVI cohorts; the replication cases included the St. Jude Total XIIIA, XIIIB and XV cohorts**.** T-ALL, T-cell acute lymphoblastic leukemia; GWAS, genome-wide association study; CNS, central nervous system.
†Race/ethnicity was defined on the basis of genetic ancestry. See details in **Methods**
‡The level of leukemia in the central nervous system at diagnosis (CNS status) is classified as CNS1 for no leukemia or other.

Supplementary Table 2. Association results at the *CDKN2A* and *USP7* loci in the discovery GWAS

SNP	CHR	Genomic Position (hg19)	REF	ALT	Gene(s)	European American†		African American†		All ethnicities†		
						AAF		AAF		*P*‡	OR (95% CI)‡	*P*$_{PC}$§
						Case	Control	Case	Control			
rs2027939	9	21956492	G*	A	*MTAP,CDKN2A-AS1*	0.10	0.14	0.04	0.04	3.2×10^{-7}	0.67 (0.57-0.78)	8.2×10^{-7}
rs2188127	9	21965232	C*	G	*MTAP,CDKN2A-AS1*	0.10	0.13	0.03	0.04	5.0×10^{-8}	0.65 (0.55-0.76)	1.3×10^{-7}
rs2518719	9	21970427	A*	G	*CDKN2A*	0.10	0.14	0.13	0.13	7.9×10^{-7}	0.70 (0.61-0.81)	1.7×10^{-6}
rs3731222	9	21983914	T*	C	*CDKN2A*	0.10	0.14	0.07	0.06	7.3×10^{-7}	0.69 (0.59-0.80)	1.8×10^{-6}
rs3731203	9	21987584	T*	C	*CDKN2A*	0.10	0.14	0.08	0.07	9.8×10^{-7}	0.69 (0.60-0.80)	2.3×10^{-6}
rs3731198	9	21989477	T*	C	*CDKN2A*	0.10	0.14	0.03	0.04	6.9×10^{-8}	0.65 (0.56-0.76)	1.8×10^{-7}
rs2811711	9	21993964	T*	C	*CDKN2A*	0.10	0.14	0.17	0.16	6.4×10^{-7}	0.70 (0.61-0.81)	1.3×10^{-6}
rs2811713	9	21999328	G*	A	*CDKN2B-AS1*	0.33	0.39	0.05	0.05	7.5×10^{-7}	0.78 (0.71-0.86)	6.7×10^{-6}
rs113177770	16	8999356	G	A*	*USP7*	0.10	0.06	0.09	0.05	2.5×10^{-8}	1.52 (1.31-1.76)	2.1×10^{-8}
rs76426500	16	9000729	A	C*	*USP7*	0.10	0.06	0.03	0.01	7.1×10^{-7}	1.49 (1.27-1.74)	5.2×10^{-7}
rs75172776	16	9001814	A	G*	*USP7*	0.10	0.06	0.11	0.06	1.4×10^{-8}	1.52 (1.32-1.76)	1.2×10^{-8}
rs79430602	16	9005924	T	C*	*USP7*	0.10	0.06	0.03	0.01	4.8×10^{-7}	1.50 (1.28-1.75)	3.5×10^{-7}
rs77119485	16	9006721	A	G*	*USP7*	0.10	0.06	0.03	0.01	4.8×10^{-7}	1.50 (1.28-1.75)	3.5×10^{-7}
rs75219349	16	9007510	T	C*	*USP7*	0.10	0.06	0.07	0.05	3.6×10^{-7}	1.47 (1.27-1.70)	3.0×10^{-7}
rs45513897	16	9009286	A	G*	*USP7*	0.10	0.06	0.03	0.01	6.3×10^{-7}	1.49 (1.27-1.74)	4.6×10^{-7}
rs2164506	16	9013035	A	G*	*USP7*	0.10	0.06	0.06	0.04	3.6×10^{-7}	1.48 (1.27-1.72)	2.7×10^{-7}
rs2126999	16	9013464	T	C*	*USP7*	0.10	0.06	0.03	0.01	6.7×10^{-7}	1.49 (1.27-1.74)	4.9×10^{-7}
rs2164507	16	9013696	G	A*	*USP7*	0.10	0.06	0.03	0.01	2.2×10^{-7}	1.52 (1.29-1.77)	1.6×10^{-7}
rs118109545	16	9014660	C	T*	*USP7*	0.10	0.06	0.03	0.01	5.8×10^{-7}	1.49 (1.28-1.75)	4.3×10^{-7}
rs11075029	16	9015721	T	C*	*USP7*	0.10	0.06	0.03	0.01	5.1×10^{-7}	1.49 (1.28-1.75)	3.7×10^{-7}
rs11075031	16	9016532	C	A*	*USP7*	0.10	0.06	0.03	0.01	4.4×10^{-7}	1.50 (1.28-1.75)	3.2×10^{-7}
rs7205293	16	9019416	C	T*	*USP7*	0.10	0.06	0.03	0.01	4.2×10^{-7}	1.50 (1.28-1.76)	2.9×10^{-7}
rs7202585	16	9022767	C	T*	*USP7*	0.10	0.06	0.11	0.06	1.5×10^{-8}	1.52 (1.32-1.76)	1.2×10^{-8}

rs16964691	16	9024282	G	C*	*USP7*	0.10	0.06	0.10	0.04	1.0×10^{-8}	1.54 (1.33-1.78)	7.8×10^{-9}
rs4539593	16	9025252	T	C*	*USP7*	0.10	0.06	0.03	0.01	4.0×10^{-7}	1.50 (1.28-1.76)	2.7×10^{-7}
rs77188284	16	9026099	C	T*	*USP7*	0.10	0.06	0.03	0.01	3.5×10^{-7}	1.50 (1.29-1.76)	2.4×10^{-7}
rs117169990	16	9026656	A	C*	*USP7*	0.10	0.06	0.03	0.01	3.5×10^{-7}	1.50 (1.29-1.76)	2.4×10^{-7}
rs143625911	16	9029260	G	T*	*USP7*	0.10	0.06	0.03	0.01	3.1×10^{-7}	1.51 (1.29-1.77)	2.1×10^{-7}
rs74916209	16	9029799	A	T*	*USP7*	0.10	0.06	0.03	0.01	3.5×10^{-7}	1.51 (1.29-1.76)	2.4×10^{-7}
rs117697595	16	9032229	T	C*	*USP7*	0.10	0.06	0.03	0.01	3.7×10^{-7}	1.50 (1.29-1.76)	2.6×10^{-7}
rs139012432	16	9033796	A	T*	*USP7*	0.10	0.06	0.03	0.01	4.9×10^{-7}	1.50 (1.28-1.75)	3.4×10^{-7}
rs141361914	16	9035467	C	T*	*USP7*	0.10	0.06	0.03	0.01	2.1×10^{-7}	1.52 (1.30-1.77)	1.4×10^{-7}
rs118105617	16	9037407	C	T*	*USP7*	0.10	0.06	0.03	0.01	4.3×10^{-7}	1.50 (1.28-1.76)	3.0×10^{-7}
rs78932417	16	9037472	C	T*	*USP7*	0.10	0.06	0.03	0.01	3.5×10^{-7}	1.51 (1.29-1.76)	2.5×10^{-7}
rs77863729	16	9043185	A	C*	*USP7*	0.10	0.06	0.03	0.01	8.9×10^{-7}	1.48 (1.27-1.74)	6.8×10^{-7}
rs75593824	16	9044676	T	C*	*USP7*	0.10	0.06	0.03	0.01	8.6×10^{-7}	1.48 (1.27-1.74)	6.5×10^{-7}
rs960734	16	9046322	T	C*	*USP7*	0.10	0.06	0.03	0.01	3.9×10^{-7}	1.50 (1.28-1.75)	3.1×10^{-7}
rs112493057	16	9047370	G	C*	*USP7*	0.10	0.06	0.03	0.01	4.2×10^{-7}	1.50 (1.28-1.75)	3.3×10^{-7}
rs76682701	16	9048561	G	A*	*USP7*	0.10	0.06	0.04	0.02	4.4×10^{-7}	1.49 (1.27-1.73)	3.8×10^{-7}
rs76161145	16	9050947	T	C*	*USP7*	0.10	0.06	0.03	0.01	3.4×10^{-7}	1.50 (1.28-1.75)	2.7×10^{-7}
rs77809369	16	9052448	C	T*	*USP7*	0.10	0.06	0.03	0.01	2.4×10^{-7}	1.51 (1.29-1.76)	1.9×10^{-7}
rs61426394	16	9053519	G	C*	*USP7*	0.10	0.07	0.25	0.19	2.8×10^{-8}	1.45 (1.27-1.66)	3.6×10^{-8}
rs74010349	16	9054247	A	G*	*USP7*	0.10	0.07	0.25	0.19	2.5×10^{-8}	1.46 (1.28-1.66)	3.2×10^{-8}
rs59591814	16	9055079	A	G*	*USP7*	0.10	0.07	0.25	0.19	2.6×10^{-8}	1.46 (1.28-1.66)	3.3×10^{-8}
rs74010351	16	9056439	A	G*	*USP7*	0.10	0.07	0.26	0.19	4.5×10^{-8}	1.44 (1.27-1.64)	7.4×10^{-8}
rs111792818	16	9059241	T	A*	*USP7,C16orf72*	0.10	0.07	0.25	0.19	7.2×10^{-8}	1.44 (1.26-1.64)	1.2×10^{-7}

*asterisk indicates risk allele for T-ALL. Abbreviations: T-ALL, T-cell acute lymphoblastic leukemia; GWAS, genome-wide association study; CHR, chromosome; REF, reference allele; ALT, alternative allele; AAF, alternative allele frequency; OR, odds ratio associated with alternative allele; CI, confidence interval.

†Ethnicity was defined by single nucleotide polymorphism genotype-based European, African, East Asian, and Native American genetic ancestry (see Methods);

‡*P* values and odds ratios were estimated by the additive logistic regression test adjusting genetic ancestry, and only SNPs with $P < 10^{-6}$ were included. All tests were two-sided.

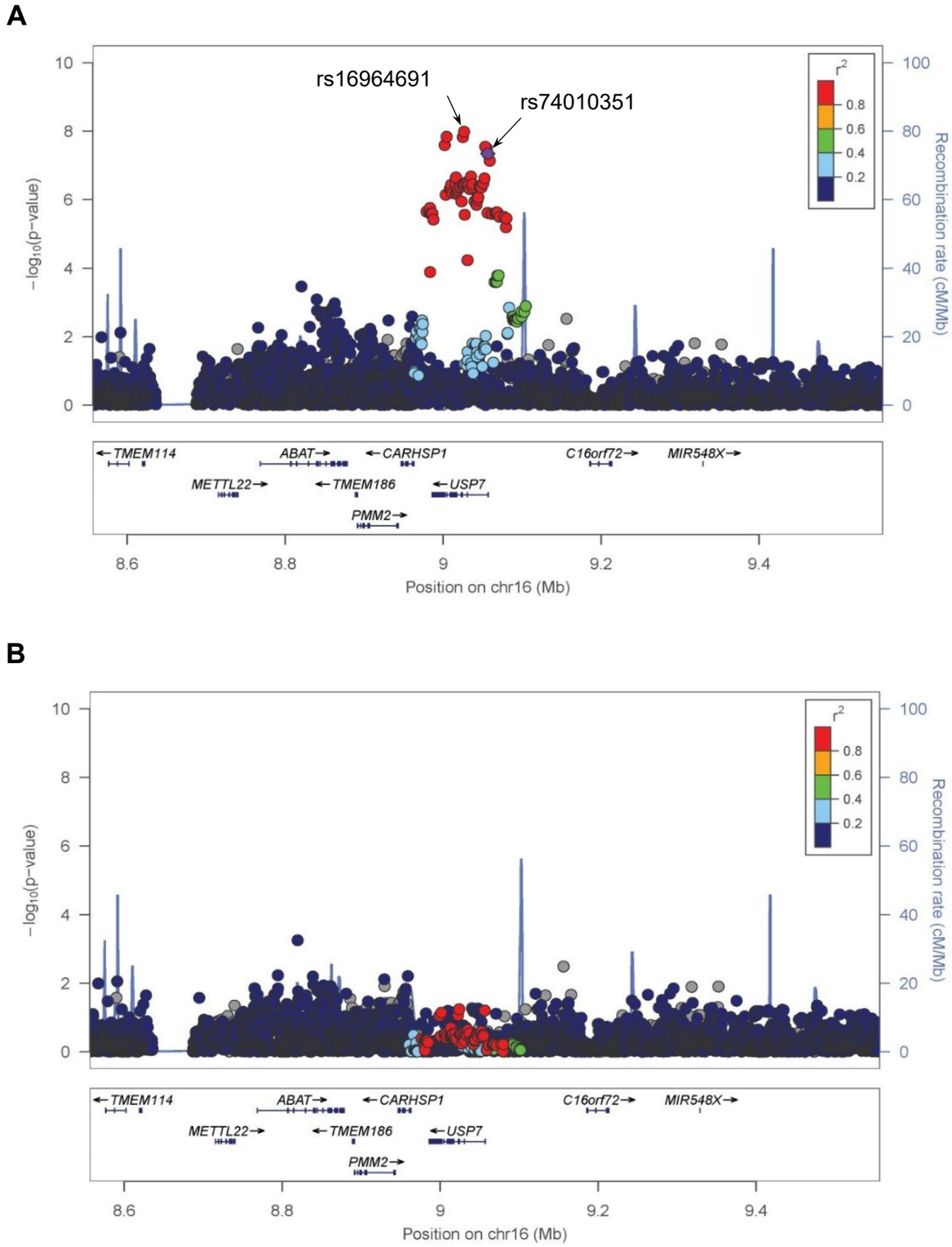
§P.pc indicates association P values estimated with principal components included as covariables. All tests were two-sided.

Supplementary Table 3. Primer sequences for cloning and site-directed mutagenesis of *USP7* variants

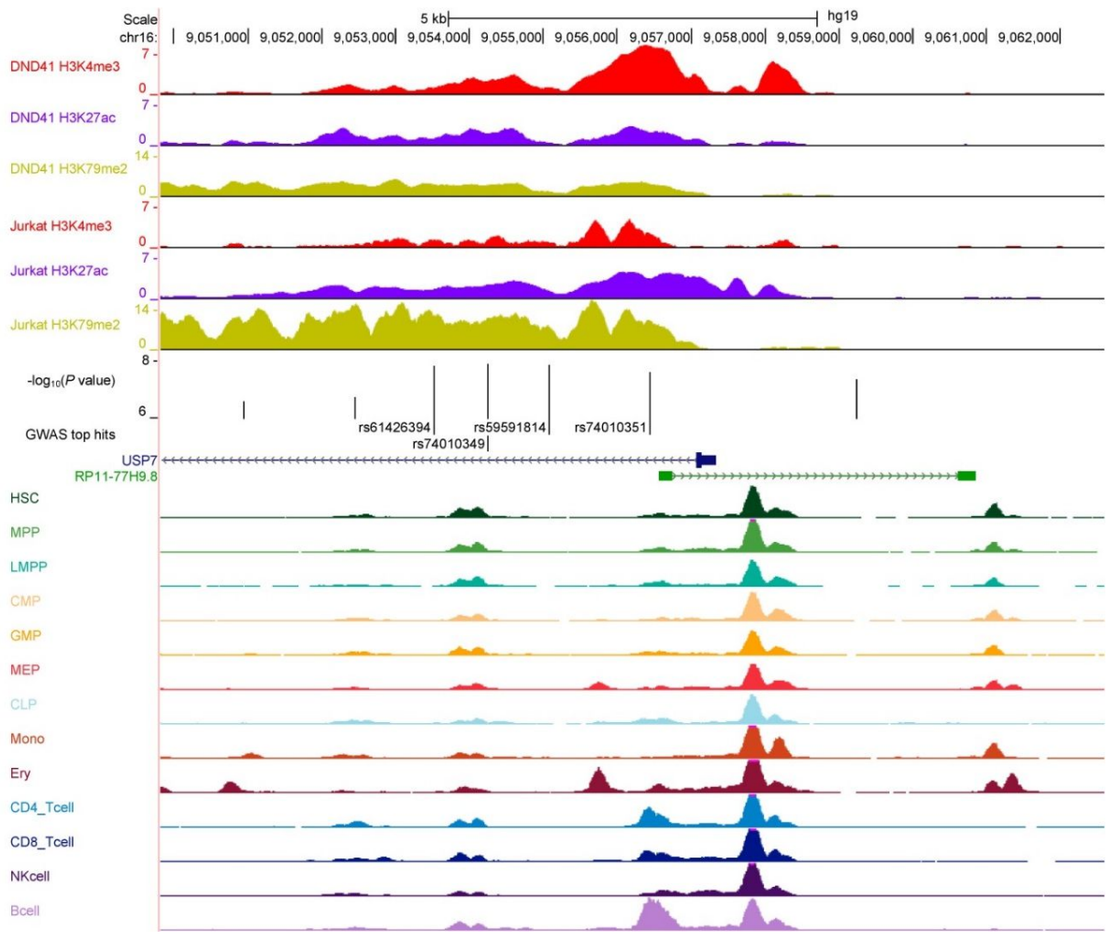
Name	Sequence (5' to 3')	Target region (hg19)
pGL4.23 miniP-rs113177770-G-F	cctaactggccggtacTTGGACAGAAAAGGATATTGGAGATTCATGG	chr16:8998957-8999748
pGL4.23 miniP-rs113177770-G-R	ccattatataccctctagtgtctaAGTGAAGAGTGCGTGTAACAGC	
pGL4.23 miniP-rs113177770-A-F	GACGGTGTGCTAGGCAAGCAGCGCCT	
pGL4.23 miniP-rs113177770-A-R	AGGCGCTGCTTGCCTAGCACACCGTC	
pGL4.23 miniP-rs75172776-A-F	cctaactggccggtacAACCCTGTCTCTACTAAAAATACAAAAATTAGCCAG	chr16:9001522-9002104
pGL4.23 miniP-rs75172776-A-R	ccattatataccctctagtgtctaCCCACATCTGATTTATTTAGATTTTCCTTCTACCT	
pGL4.23 miniP-rs75172776-G-F	AGACAACAAAATATTGGAAAATCCTGTTTAAACGAATACACTAAAAGTCCC	
pGL4.23 miniP-rs75172776-G-R	GGGACTTTTAGTGTATTCGTTTAAACAGGATTTTCCAATATTTTGTTGTCT	
pGL4.23 miniP-rs7202585-C-F	cctaactggccggtacTGCTTCTGATATGACAACCCTGAGACA	chr16:9022273-9023491
pGL4.23 miniP-rs7202585-C-R	ccattatataccctctagtgtctaATGCAGCCCCTGGAGT	
pGL4.23 miniP-rs7202585-T-F	GAGACATGTCATGACTGCAAGTGGAGGGCAG	
pGL4.23 miniP-rs7202585-T-R	CTGCCCTCCACTTGCAGTCATGACATGTCTC	
pGL4.23 miniP-rs16964691-G-F	cctaactggccggtacGTACCCAAGAAGTACCAAGCATGGA	chr16:9024064-9024513
pGL4.23 miniP-rs16964691-G-R	ccattatataccctctagtgtctaTGCAGAATATCATTGCAAATAAAGCAAATTATAGTGT	
pGL4.23 miniP-rs16964691-C-F	TAAAGAAGAAAAAGAAATTCCACGCGTTTTACATTCATGGCTTTAATAG	
pGL4.23 miniP-rs16964691-C-R	CTATTAAAGCCATGAATGTAAAACGCGTGGAATTTCTTTTTCTTCTTTA	

pGL4.23 miniP-rs61426394/rs74010349-G/A-F	cctaactggccggtacGCTCACATCAGAGTGAGATGCTAGG	chr16:9053365-9054420
pGL4.23 miniP-rs61426394/rs74010349-G/A-R	ccattatataccctctagtgtctaAACCTGGAGCATCCTGCC	
pGL4.23 miniP-rs61426394-C-F	TCGAATTATTTGACTTTTGTCTCCTTACAGTGCTATTTGAAAGAATC	
pGL4.23 miniP-rs61426394-C-R	GATTCTTTCAAATAGCACTGTAAGGAGACAAAAGTCAAATAATTCGA	
pGL4.23 miniP-rs74010349-G-F	TCCAGAGCACTGCTGTTTTCCAGTGAATATACAAGGAC	
pGL4.23 miniP-rs74010349-G-R	GTCCTTGTATATTCACTGGAAAACAGCAGTGCTCTGG	
pGL4.23 miniP-rs59591814-A/G-F	cctaactggccggtacGATTAACTGTCCTAGGGGTTACCTTTCTAATACAA	chr16:9054914-9055513
pGL4.23 miniP-rs59591814-A/G-R	ccattatataccctctagtgtctaTCCCATGACATTGTCCTACCGC	
pGL4.23 miniP-rs74010351-A-F	cctaactggccggtacTTCATTCAGGTGGCCTCCATTTGAC	chr16:9056073-9056970
pGL4.23 miniP-rs74010351-A-R	ccattatataccctctagtgtctaTAACCAGGCCCCGGGC	
pGL4.23 miniP-rs74010351-G-F	CACCACACACTGCCAGGTACAAACTCCCCGA	
pGL4.23 miniP-rs74010351-G-R	TCGGGGAGTTTGTACCTGGCAGTGTGTGGTG	

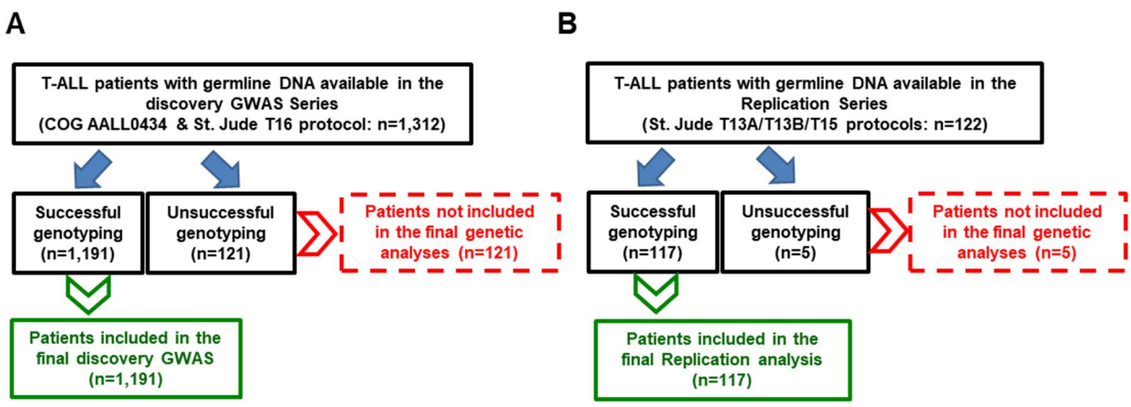
SUPPLEMENTARY FIGURES



Supplementary Figure 1. Univariate and conditional association results at the *USP7* locus. Both directly-genotyped and imputed SNPs were included in the analyses. rs74010351 was the top directly-genotyped variant in the discovery GWAS and shown as a purple diamond (OR = 1.44, 95% CI 1.27-1.65, P = 4.51 x 10^{-8}). All other SNPs are highlighted by color showing their extent of linkage disequilibrium with rs74010351. Recombination rate, genetic position, and the locations (hg19) of nearby genes (RefSeq) are indicated. The top associated SNP from imputation is rs16964691 (OR = 1.54, 95% CI 1.33-1.78, P = 1.01 x 10^{-8}). P values were calculated by additive logistic regression test in univariate analysis (**A**), or after adjusting for genotype at rs74010351 (conditional analysis, **B**). All statistical tests were two-sided.

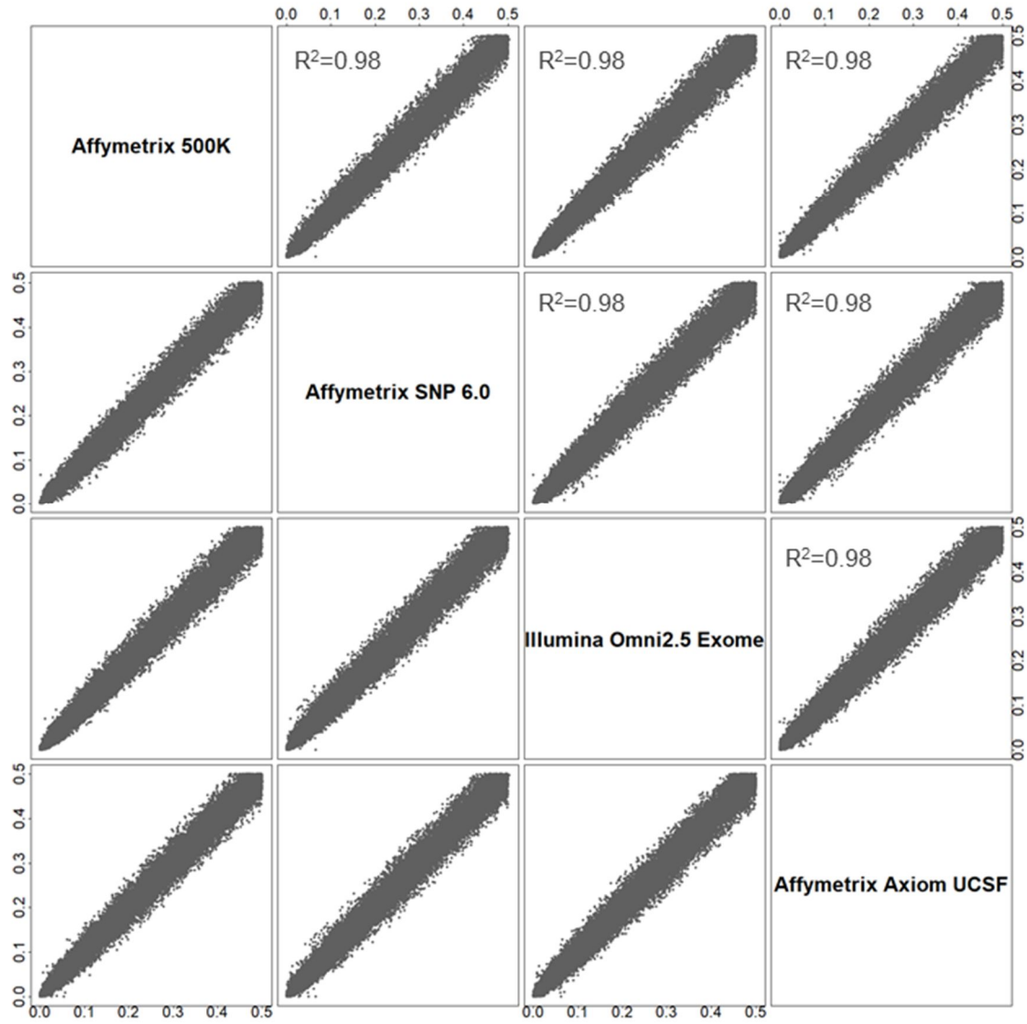


Supplementary Figure 2. Annotation and visualization of genomic region close to *USP7* promoter. Genomic positions and scale for the human genome assembly February 2009 (GRCh37/h19) are shown on the top. Genome-wide significant T-ALL risk variants in *USP7* ($P < 5 \times 10^{-8}$) were marked in the middle panel and the log-transformed *P* values (for association with T-ALL) were shown in bed graph. The gene structure, ChIP-seq signals for histone modifications (i.e., H3K4me3, H3K27ac and H3K79me2) in T-ALL cell lines (i.e., DND41 and Jurkat) (*Science*. 2014;346(6215):1373-1377), and ATAC-seq signals of hematopoietic cells (*Nat Genet.* 2016;48(10):1193-1203) were also included. T-ALL, T-cell acute lymphoblastic leukemia.

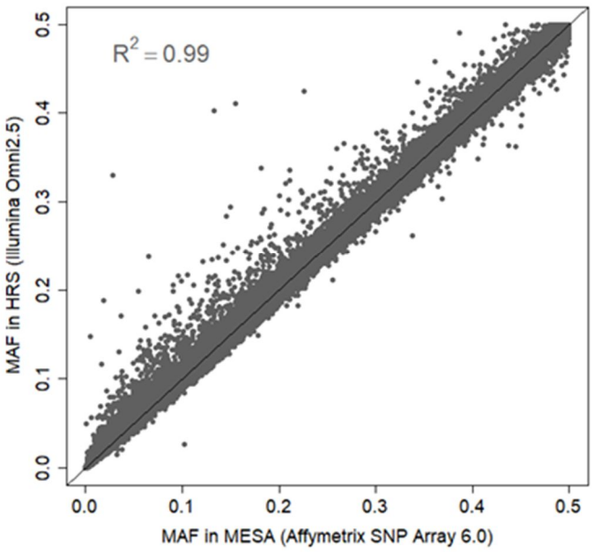


Supplementary Figure 3. Flow chart of patient inclusion and exclusion for the genomic analyses. Individuals with one of the following features were excluded: discordant sex (between clinical data and genotype-inferred sex); genotype failure rate >5%; heterozygosity rate >=5 s.d. from the mean; heterozygosity rate >=3 s.d. from the mean if genotype failure rate >= .03; identity-by-descent score > .185 and lower call rate to other individuals.

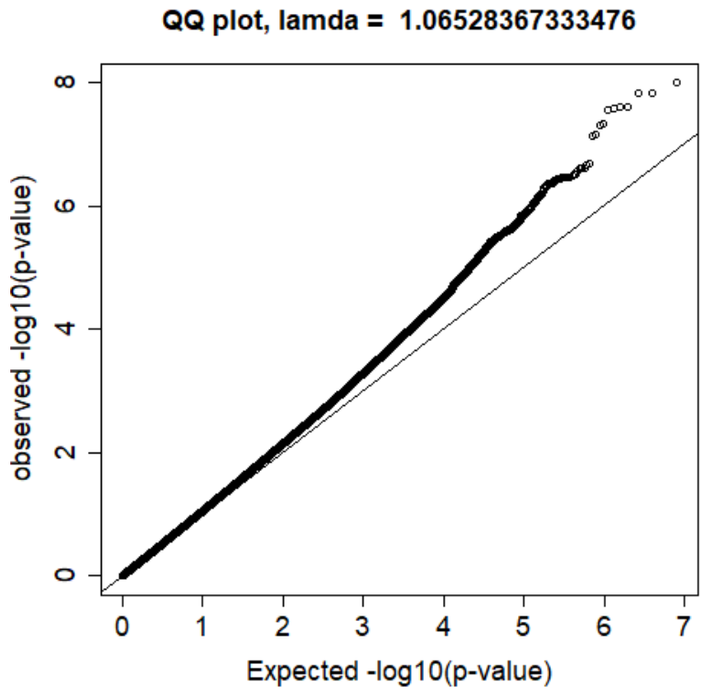
A



B



Supplementary Figure 4. Consistency in genotype imputation across cohorts with different genome-wide SNP arrays. For each cohort genotyped on a different array, we first randomly selected 30,000 imputed SNPs; the allele frequency for each SNP was calculated in each cohort with a different array platform, and then the correlation coefficient in allele frequency of these 30,000 SNPs between 2 arrays was evaluated. This analysis was performed for all possible array comparisons for cases (**A**) and for controls (**B**), respectively. The R squares of correlation across genotyping platforms were consistently more than 0.98.



Supplementary Figure 5. Quantile-quantile (Q-Q) plot of logistic regression test for GWAS. The negative logarithm of the observed (y axis) and the expected (x axis) *P* value is plotted for each SNP (dot), and the black line indicates the null hypothesis of no true association. Deviation from the expected P value distribution is evident only in the tail area (λ=1.07), suggesting that population stratification was adequately controlled by adjusting for genetic ancestry.