Supplemental Methods

Study Population

All DISCOVeRY-BMT ("Determining the Influence of Susceptibility COnveying Variants Related to One-Year mortality after BMT") patients and their respective donors were at least 8/8 high-resolution HLA (human leukocyte antigen) matched at 4 loci: HLA-A, -B, -C, and -DRB1. Specifically, Cohort 1 was 10/10 HLA matched (HLA-A, -B, -C, -DRB1, DQB1), treated for unrelated donor blood and marrow transplantation (URD-BMT) at one of the 151 United States transplant centers from 2000-2008, reported to the Center for International Blood and Marrow Transplant Research (CIBMTR) and had a banked bio-repository sample from both recipient and donor. The controls in the DISCOVeRY-BMT study have passed a rigorous medical exam and screening for potential undiagnosed conditions (eg other cancers, autoimmune diseases, heart disease and diabetes) that would exclude them from donation eligibility. Cohort 2 patients were treated from 2009-2011 and otherwise met the same criteria as the first cohort, except a small proportion of 8/8 HLA-matched patients (HLA-DQB1 mismatched). As our control population included all ages we estimated the predicted number of controls who may be diagnosed with ALL as follows. Per Surveillance Epidemiology and End Results (SEER) the incidence of ALL per 100,000 people by age groups 18-20, 21-30, 31-40 and >40 ALL is 1.0, 0.8, 0.7 and 0.7 , respectively. Applying these incidence rates by age group to the DISCOVeRY-BMT controls we estimate that less than one control in our study population will be diagnosed as a case of ALL, thus importantly it is highly unlikely a control could be a hidden case.

Genotyping Quality Control

DISCOVeRY-BMT had standard quality control (QC) measures performed on the genotypes and samples. Samples consisted of recipients treated for URD-BMT as therapy for acute myeloid leukemia (AML), myelodysplastic syndrome (MDS), or acute lymphoblastic leukemia

(ALL), and their unrelated donors. The QC process for samples included checking for recipient-donor duplicates (the recipient-donor duplicates were identified by pooling recipients and their matched donors in each cohort for checking relatedness), missingness (> 2%), gender mismatch, abnormal inbreeding coefficients (samples with inbreeding coefficients five standard deviations beyond the mean were removed), cryptic relatedness, and population outliers (identified by using the EIGENSTRAT[1] software; removing population outliers was skipped in donor sample QC as the matched donors were not necessarily from the same population as the recipients). Final sample size in cohort 1 was 2111 AML/MDS/ALL cases and 2219 controls and in cohort 2 there were 779 AML/MDS/ALL cases and 808 controls. This final clean dataset was further processed to only include B-cell ALL cases and specific B-cell ALL subtypes (hyperdiploid negative B-ALL, Ph-negative B-ALL, abnormal B-ALL, and normal cytogenetics B-ALL) (**Supplemental Figure 1**).

Histological and Cytogenetic subtyping

Histological and cytogenetic subtyping was performed at one of >200 US transplant centers where the B-ALL patient/case was diagnosed and reported to the CIBMTR. A subset of the abnormal cytogenetic reports was reviewed at Roswell Park Cancer Institute in the Clinical Cytogenetics Laboratory by two cytogeneticists (authors: AWB & SNJS) and a clinical epidemiologist (author: TH). Following quality control (60 cases were removed), 605 European American ALL cases from cohort 1 and cohort 2 were available for analyses. These cases consisted of the following histological subtypes: precursor B-cell (N=438), precursor T-cell (N=77), mature B-cell (N=8) and null and not otherwise specified (N=82). After removal of T-cell ALL cases (N=77) and collapsing B-lineage cases (82), a total of 446 European American B-cell ALL cases (364=Cohort 1 / 82=cohort 2) were available for use in all subsequent analyses.

The B-cell ALL cases were further stratified into cytogenetic subgroups for association testing: hyperdiploid negative (N: Cohort 1=333, Cohort 2=77), Ph-negative (N: Cohort 1=271, Cohort

2=62), abnormal (N: Cohort 1=205, Cohort 2=48) and normal cytogenetics (N: Cohort 1=159, Cohort 2=34) (**Supplemental Table 1**). Cytogenetic subgroups are not mutually exclusive and a case could be included in more than one subgroup.

The B-ALL cases in our study were considered high-risk because they received an URD-BMT for one or more of these high-risk features: B-cell ALL subtype that is not hyperdiploid (hyperdiploid-negative), high white blood cell count (WBC) at diagnosis, high risk chromosomal abnormalities, not achieving complete remission (CR) at the end of induction or relapsing after CR, Ph+ B-ALL, infant (<1 year of age) ALL, pediatric cases > 10 years of age, and AYA and adult cases > 30 years of age.[3;8]

GWAS Analysis Overview

Descriptive statistics were performed to understand the association between outcome and independent variables: age and sex. Standard GWAS were performed in B-ALL overall and subtype specific B-ALL in cohort 1 and cohort 2, adjusted for age and sex. These primary GWAS were aimed at replicating known ALL GWAS loci and identifying novel GWAS loci in B-ALL and B-ALL subtypes. Next, exploratory GWAS were performed in age and sex specific strata using the same control population as the overall and subtype analyses. Meta-analyses were performed with cohort 1 and cohort 2 in primary GWAS and exploratory GWAS. Due to the small sample size of cases <15 sex specific findings within this age group, we tested in two independent cohorts described in Replication dataset section (below).

Replication dataset

Hungate et al. 2016[2] previously performed a Meta-GWAS of pediatric B-ALL, consisting of cases from a German GWAS[3, 4] and from COG9904[5, 6] with Genetic Association Information Network (GAIN) controls were available for use as replication of the male and female *IKZF1* pediatric associations.[2, 3, 7-10] Briefly, all cases from COG9904 were of European ancestry diagnosed with B-ALL and treated on The Children's Oncology Group (COG) P9904 protocol

(https://members.childrensoncologygroup.org/Mtg/bookreports/Denver/reports/9904_Fall07SPR_FinalReport.pdf).[5, 6] Controls were obtained separately from the GAIN Consortium schizophrenia study cohort.[8] The German GWAS included pediatric B-ALL cases, all with the ETV6-RUNX1 rearrangement, and cancer-free controls, all of European ancestry. Cases were obtained through the Austrian-German-Italian-Swiss multicentre clinical trial AIEOP-BFM ALL 2000 and controls were obtained through the German popgen biobank.[3, 4] Cases and controls were genotyped separately by COG and the Broad Institute, respectively, using the Affymetrix Genome-Wide Human SNP Array 6.0 (Santa Clara, CA) and called using the Birdseed-v2 algorithm. The 9904-GAIN dataset included 437 cases and 958 controls and the German dataset included 427 cases and 475 controls following quality control (described in detail elsewhere).[2]

Heritability and Polygenic Risk Score Analyses

Two approaches were used to better understand heritability and the aggregate contribution of genetic variation to B-ALL risk: 1) Genome-wide Complex Trait Analysis (GCTA) [11-15], described in the main text and 2) polygenic risk scores (PRS) utilizing PRSice Software. [16]

Polygenic Risk Scores (PRS)

Polygenic risk scores (PRS) were calculated by combining significant loci (genotyped SNP/variants included with associations $P<x10^{-5}$) weighted by effect sizes estimated from the logistic regression GWAS using PRSice software.[16] The continuous PRS variable was calculated for overall, age specific, and sex specific B-ALL. These PRS variables were further categorized into low, medium and high-risk groups. These groups are based on a range of 0-0.35 and are as follows: low=0-0.54, medium=0.055-0.126, high=0.127-0.36. Control population is the same for each category within age-specific strata. Odds ratios were calculated by logistic regression assuming an additive model.

References

1. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904-9. doi: 10.1038/ng1847. PubMed PMID: 16862161.
2. Hungate EA, Vora SR, Gamazon ER, Moriyama T, Best T, Hulur I, Lee Y, Evans TJ, Ellinghaus E, Stanulla M, Rudant J, Orsi L, Clavel J, Milne E, Scott RJ, Pui CH, Cox NJ, Loh ML, Yang JJ, Skol AD, Onel K. A variant at 9p21.3 functionally implicates CDKN2B in paediatric B-cell precursor acute lymphoblastic leukaemia aetiology. *Nat Commun.* 2016;7:10635. Epub 2016/02/13. doi: 10.1038/ncomms10635. PubMed PMID: 26868379; PMCID: 4754340.
3. Ellinghaus E, Stanulla M, Richter G, Ellinghaus D, te Kronnie G, Cario G, Cazzaniga G, Horstmann M, Panzer Grumayer R, Cave H, Trka J, Cinek O, Teigler-Schlegel A, ElSharawy A, Hasler R, Nebel A, Meissner B, Bartram T, Lescai F, Franceschi C, Giordan M, Nurnberg P, Heinzow B, Zimmermann M, Schreiber S, Schrappe M, Franke A. Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. *Leukemia.* 2012;26(5):902-9. Epub 2011/11/15. doi: 10.1038/leu.2011.302. PubMed PMID: 22076464; PMCID: 3356560.
4. Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, Schreiber S. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet*. 2006;9(1):55-61. Epub 2006/02/24. doi: 10.1159/000090694. PubMed PMID: 16490960.
5. Borowitz MJ, Devidas M, Hunger SP, Bowman WP, Carroll AJ, Carroll WL, Linda S, Martin PL, Pullen DJ, Viswanatha D, Willman CL, Winick N, Camitta BM. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. *Blood.* 2008;111(12):5477-85. Epub 2008/04/05. doi: 10.1182/blood-2008-01-132837. PubMed PMID: 18388178; PMCID: 2424148.
6. Winick N. Delayed intensification (DI) enhances event-free survival (EFS) of children with B-precursor acute lymphoblastic leukemia (ALL) who received intensification therapy with six courses of intravenous methotrexate (MTX):POG 9904/9905. A Children's Oncology Group Study (COG). ASH Annual Meeting Abstract. 2007;110:583.
7. Evans TJ, Milne E, Anderson D, de Klerk NH, Jamieson SE, Talseth-Palmer BA, Bowden NA, Holliday EG, Rudant J, Orsi L, Richardson E, Lavis L, Catchpoole D, Attia JR, Armstrong BK, Clavel J, Scott RJ. Confirmation of childhood acute lymphoblastic leukemia variants, ARID5B and IKZF1, and interaction with parental environmental exposures. *PLoS One*. 2014;9(10):e110255. Epub 2014/10/14. doi: 10.1371/journal.pone.0110255. PubMed PMID: 25310577; PMCID: 4195717.
8. Manolio TA, Rodriguez LL, Brooks L, Abecasis G, Ballinger D, Daly M, Donnelly P, Faraone SV, Frazer K, Gabriel S, Gejman P, Guttmacher A, Harris EL, Insel T, Kelsoe JR, Lander E, McCowin N, Mailman MD, Nabel E, Ostell J, Pugh E, Sherry S, Sullivan PF, Thompson JF, Warram J, Wholley D, Milos PM, Collins FS. New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *NatGenet.* 2007;39(9):1045-51. Epub 2007/08/31. doi: 10.1038/ng2127. PubMed PMID: 17728769.
9. Orsi L, Rudant J, Bonaventure A, Goujon-Bellec S, Corda E, Evans TJ, Petit A, Bertrand Y, Nelken B, Robert A, Michel G, Sirvent N, Chastagner P, Ducassou S, Rialland X, Hemon D, Milne E, Scott RJ, Baruchel A, Clavel J. Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE). *Leukemia.* 2012;26(12):2561-4. Epub 2012/06/05. doi: 10.1038/leu.2012.148. PubMed PMID: 22660188.
10. Rudant J, Orsi L, Menegaux F, Petit A, Baruchel A, Bertrand Y, Lambilliotte A, Robert A, Michel G, Margueritte G, Tandonnet J, Mechinaud F, Bordigoni P, Hemon D, Clavel J. Childhood acute leukemia, early common infections, and allergy: The ESCALE Study. *AmJ*

Epidemiol. 2010;172(9):1015-27. Epub 2010/09/03. doi: 10.1093/aje/kwq233. PubMed PMID: 20807738.
11. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76-82. Epub 2010/12/21. doi: 10.1016/j.ajhg.2010.11.011. PubMed PMID: 21167468; PMCID: 3014363.
12. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010;42(7):565-9. Epub 2010/06/22. doi: 10.1038/ng.608. PubMed PMID: 20562875; PMCID: 3232052.
13. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet.* 2011;88(3):294-305. Epub 2011/03/08. doi: 10.1016/j.ajhg.2011.02.002. PubMed PMID: 21376301; PMCID: 3059431.
14. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*. 2012;28(19):2540-2. Epub 2012/07/31. doi: 10.1093/bioinformatics/bts474. PubMed PMID: 22843982; PMCID: 3463125.
15. American Cancer Society: Cancer Facts and Figures 2016 Atlanta, Ga2016 [cited 2016 Dec. 8]. http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-047079.pdf
16. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics*. 2015;31(9):1466-8. Epub 2015/01/01. doi: 10.1093/bioinformatics/btu848. PubMed PMID: 25550326; PMCID: 4410663.

Supplemental Table 1. Characteristics of European American B-ALL cases and controls

Characteristics	Cases n (%) (Cohort 1=364 / Cohort 2=82)	Controls n (%) (Cohort 1=2219 / Cohort 2=808)
Age in years		
<20	113 (31) / 21 (24)	50 (2) / 35 (4)
20-40	130 (36) / 30 (37)	1572 (71) / 586 (73)
>40	121 (33) / 31 (38)	597 (27) / 187 (23)
Sex		
Male	219 (60) / 53 (65)	1507 (68) / 591 (73)
Female	145 (40) / 29 (35)	712 (32) / 217 (27)
Cytogenetics		
Hyperdiploid Negative (<51 Chromosomes)	333 (91) / 77 (94)	--
Age in years		
<20	98 (29) / 19 (25)	50 (2) / 35 (4)
20-40	122 (37) / 28 (36)	1572 (71) / 586 (73)
>40	113 (34) / 30 (39)	597 (27) / 187 (23)
Ph Negative	271 (74) / 62 (76)	--
Age in years		
<20	101 (37) / 18 (29)	50 (2) / 35 (4)
20-40	99 (37) / 23 (37)	1572 (71) / 586 (73)
>40	71 (26) / 21 (34)	597 (27) / 187 (23)
Abnormal cytogenetics	205 (56) / 48 (59)	--
Age in years		
<20	60 (29) / 12 (25)	50 (2) / 35 (4)
20-40	67 (33) / 16 (33)	1572 (71) / 586 (73)
>40	78 (38) / 20 (42)	597 (27) / 187 (23)
Normal cytogenetics	159 (44) / 34 (41)	--
Age in years		
<20	53 (33) / 9 (27)	50 (2) / 35 (4)
20-40	63 (40) / 14 (41)	1572 (71) / 586 (73)
>40	43 (27) / 11 (32)	597 (27) / 187 (23)

ALL: Acute lymphoblastic leukemia, Ph Negative=Philadelphia chromosome negative, Abnormal Cytogenetics= Conventional or FISH (fluorescence *in situ* hybridization) cytogenetic analysis revealed >1 clonal chromosomal abnormality, Normal Cytogenetics=no chromosomal abnormalities reported

Supplemental Table 2. Most significant association signals in DISCOVeRY-BMT within a 500kb window surrounding previously published loci associated with pediatric and adolescent and young adult B-ALL

B-ALL overall (age-adjusted) (C1 / C2: N=364 / 82)						***P*$_{meta}$ for each B-ALL subtype (age-adjusted)**			
SNP / GENE	**CHR:BP**	**Alleles / EAF**	**OR, (95% CI) (C1 / C2)**	**P-value (C1 / C2)**	***P*$_{meta}$**	**Hyperdiploid Negative** (C1/C2:N=333 / 77)	**Ph Negative** (C1/C2:N=271/ 62)	**Abnormal Cytogenetics** (C1/C2:N=205 / 48)	**Normal Cytogenetics** (C1/C2:N=159 / 34)
rs3824662* *GATA3*	10:8104207	**A**/C 0.20	1.9 (1.5,2.3) / 1.9 (1.2,2.6)	$6.7x10^{-11}$ / 0.001	$3.3x10^{-13}$	$2.9x10^{-13}$	$1.2x10^{-12}$	$1.8x10^{-7}$	$1.1x10^{-8}$
rs11980379* *IKZF1*	7:50469980	**C**/T 0.30	1.6 (1.3,1.9) / 1.4 (0.9,1.9)	$4.4x10^{-7}$ / 0.04	$6.9x10^{-8}$	$2.7x10^{-7}$	$3.6x10^{-9}$	0.006	$4.6x10^{-8}$
rs1333035 *CDKN2B*	9:22044058	**A**/G 0.11	0.6 (0.5,0.8) / 0.6 (0.3,0.9)	0.0001 / 0.06	$2.1x10^{-5}$	$1.2x10^{-4}$	0.0002	0.0003	0.01
rs3731249 *CDKN2A*	9: 21970917	T/**C** 0.03	0.8 (0.4,1.3) / 0.3 (-0.5,1.1)	0.4 / 0.003	0.02	0.04	0.002	0.3	0.08
rs10994995 *ARID5B*	10:63748977	**G**/C 0.08	1.6 (1.1,2.2) / 1.7 (0.6,2.9)	0.002 / 0.1	0.0005	$7.4x10^{-4}$	0.001	0.01	0.01
rs76661504 *BMI1-PIP4K2A*	10:22843620	**C**/T 0.05	0.5 (0.3,0.7) / 1.3 (0.3,2.4)	0.0001 / 0.4	0.004	$5.3x10^{-3}$	0.06	0.004	0.2
rs2239635 *CEBPE*	14:23588731	G/**C** 0.3	0.9 (0.7,1.1) / 0.8 (0.5,1.07)	0.6 / 0.2	0.3	0.5	0.3	0.1	0.6
rs35837782 *LHPP*	10:124604740	A/**G** 0.4	0.9 (0.8,1.2) / 0.9 (0.6,1.3)	0.9 / 0.7	0.9	0.9	0.9	0.8	0.9
rs4762284 *ELK3*	12: 96218984	A/**T** 0.3	0.9 (0.7,1.1) / 0.8 (0.4,1.2)	0.3 / 0.4	0.2	0.4	0.7	0.2	0.6

*Typed SNPs, all others were imputed with info and certainty score>=0.9. DISCOVeRY-BMT: Determining the Influence of Susceptibility COnveying Variants Related to 1-Year mortality after Blood or Marrow Transplant study includes 364 cases in Cohort 1 (C1) and 82 cases in Cohort 2 (C2). Controls included (Cohort 1=2219 / Cohort 2=808). SNP: single nucleotide polymorphism, CHR: chromosome, BP: base-pair position, EAF: Effect allele frequency, bold allele is the effect allele, OR=odds ratio, 95% CI=95% confidence interval.

Supplemental Table 3. Genome-wide significant associations of *IKZF1* variant rs11980379 by sex and age (<15, >15 years) in B-cell Acute Lymphoblastic Leukemia in European Americans from DISCOVeRY-BMT and a pediatric replication dataset

	B-cell ALL DISCOVeRY-BMT*						Replication (Precursor B-cell ALL GWAS[€])			
	Males			Females			Males		Females	
Age (years)	N cases	OR (95% CI)	*P*-value	Cases	OR (95% CI)	*P*-value	Cases	OR_{meta} (95% CI) P_{meta}	Cases	OR_{meta} (95% CI) P_{meta}
<15	37	2.3 (1.4,3.8)	0.0007	35	1.9 (1.2,3.2)	0.007	422	1.4 (1.1,1.9) 3.9×10^{-5}	442	1.6 (1.2,2.3) 1.4×10^{-8}
>15	182	1.7 (1.4,2.2)	6.4×10^{-6}	110	1.1 (0.8,1.5)	0.6	NA	NA	NA	NA

OR: odds ratio of logistic additive model risk is presented in terms of the C (minor) allele (see Table 2), 95% CI: 95% confidence interval, OR_{meta}: odds ratio for fixed effects model meta-analysis, P_{meta}: P-value for fixed effects model meta-analysis, genome-wide significance $<5.0 \times 10^{-8}$

*Due to a small sample size in Cohort 2, analyses were restricted to Cohort 1 only.

[€] Cases (N): males=191, females=246 from COG9904/GAIN[1-3] and males=231, females=196 from German GWAS[4, 5]

[€] Controls (N): males=519, females=439 from Genetic Association Information Network (GAIN)[1, 6] and males=254 and females=221 from German GWAS[4, 5]

NA = data were not available for analyses

Supplemental Table 4. Summary results of Genome-wide Complex Trait Analysis (GCTA-Bivariate REML Analysis) for Males and Females associated with B-ALL

Source	Variance	SE
Genetic variance (females)	0.13	0.05
Genetic variance (males)	0.07	0.02
proportion of variance explained by all SNPs (females)	0.16	0.007
proportion of variance explained by all SNPs (males)	0.13	0.004
Genetic variance / proportion of variance explained by all SNPs for females	0.82	0.34
Genetic variance / proportion of variance explained by all SNPs for males	0.54	0.18
Genetic correlation (rG) between males and females	0.11	0.29

For a bivariate analysis of two disease traits, you can specify the prevalence rates of the two diseases in the general population so that GCTA will transform the estimate of variance explained by the SNPs from the observed 0-1 scale to that on the underlying scale for both diseases. SE=standard error. GCTA=Genome-wide complex trait analysis.

Bivariate REML analysis method: Lee et al. (2012) Estimation of pleiotropy between complex diseases using SNP-derived genomic relationships and restricted maximum likelihood. Bioinformatics. 2012 Oct 28(19): 2540-2542. [PubMed ID: 22843982]

REML analysis and GCTA software: Yang J, Lee SH, Goddard ME and Visscher PM. GCTA: a tool for Genome-wide Complex Trait Analysis. Am J Hum Genet. 2011 Jan 88(1): 76-82. [PubMed ID: 21167468]

Supplemental Table 5. Polygenic Risk Scores (PRS) by allele risk category associated with B-ALL

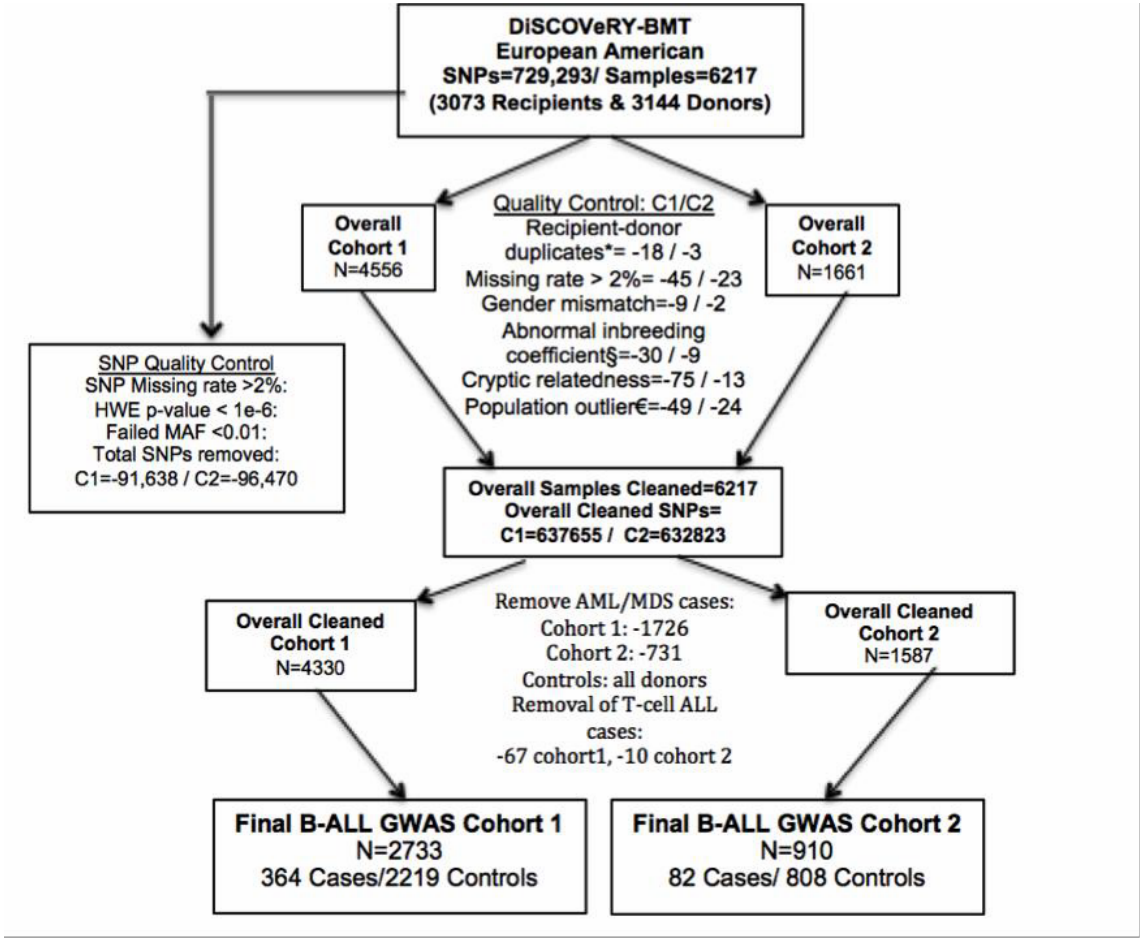
	Polygenic Risk Score Category (PRS)					
	Low Risk (reference)		**Medium Risk**		**High Risk**	
	OR (95% CI)	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
Overall (adjusted for age) Total N=2593 (364 / 2219)	1.0 (reference)	-	1.7 (1.3,2.3)	0.0006	3.1 (2.3,4.2)	**$2.4x10^{-14}$**
Total N (cases/controls)	988 (82 / 906)		837 (110 / 727)		758 (172 / 586)	
Pediatric (<20 years) Total N=2342 (113 / 2219)	1.0 (reference)	-	1.4 (0.8,2.3)	0.3	3.5 (2.2,5.8)	$1.54x10^{-7}$
Total N (cases/controls)	932 (26 / 906)		755 (28 / 727)		645 (59 / 586)	
Young Adults (20-40 years) Total N=2359 (130 / 2219)	1.0 (reference)	-	1.8 (1.1,2.9)	0.02	3.2 (2.1,5.2)	$4.3x10^{-7}$
Total N (cases/controls)	935 (29 / 906)		768 (41 / 727)		646 (60 / 586)	
Older Adults (>40 years) Total N=2350 (121 / 2219)	1.0 (reference)	-	1.9 (1.2,3.2)	0.01	3.1 (1.9,5.0)	$3.7x10^{-6}$
Total N (cases/controls)	933 (27 / 906)		768 (41 / 727)		639 (53 / 586)	
Males Total N= 1726 (219 / 1507)	1.0 (reference)	-	1.7 (1.1,2.5)	0.01	3.5 (2.4,5.1)	$4.1x10^{-11}$
Total N (cases/controls)	669 (50 / 619)		559 (61 / 498)		494 (108 / 386)	
Females Total N=857 (145 / 712)	1.0 (reference)	-	1.8 (1.1,2.9)	0.02	2.5 (1.6,4.1)	0.0001
Total N (cases/controls)	309 (32 / 277)		278 (49 / 229)		264 (64 / 200)	

(SNP/variants included with associations $P<x10^{-5}$). Control population is the same for each category. OR=odds ratio of logistic regression assuming an additive model.

95% CI=95% confidence interval (lower, upper).

Polygenic risk score ranges: 0-0.36 (0-0.54=LOW risk, 0.055-0.126=MEDIUM risk, 0.127-0.36=HIGH risk)

Supplemental Figure 1. Patient and Sample Quality Control in European Americans



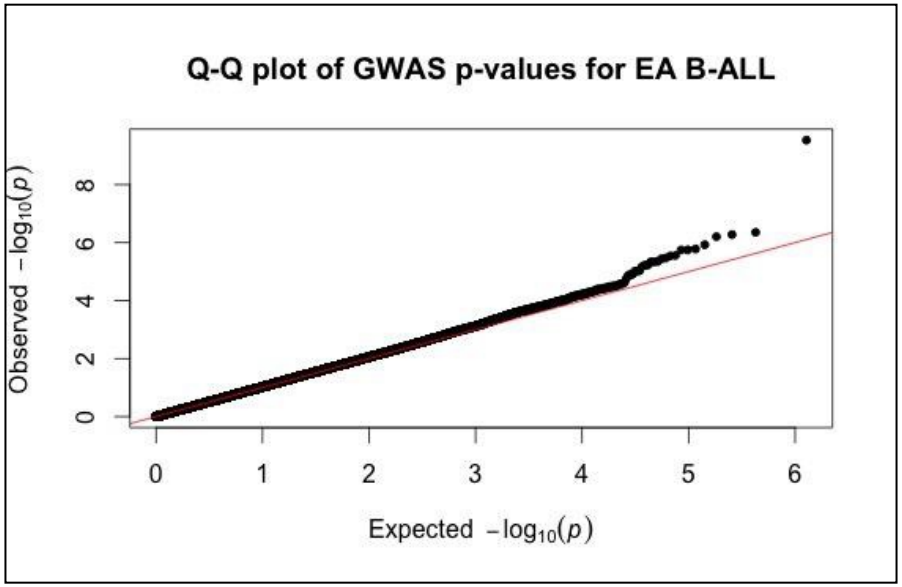
Flow chart shows removal/exclusions that occurred during quality control procedures. Minus (-) sign indicates samples/SNPs removed.

*: the recipient-donor duplicates were identified by pooling recipients and their matched donors in each cohort tor check for relatedness.

§: samples with inbreeding coefficients >5 standard deviations beyond the mean were removed.

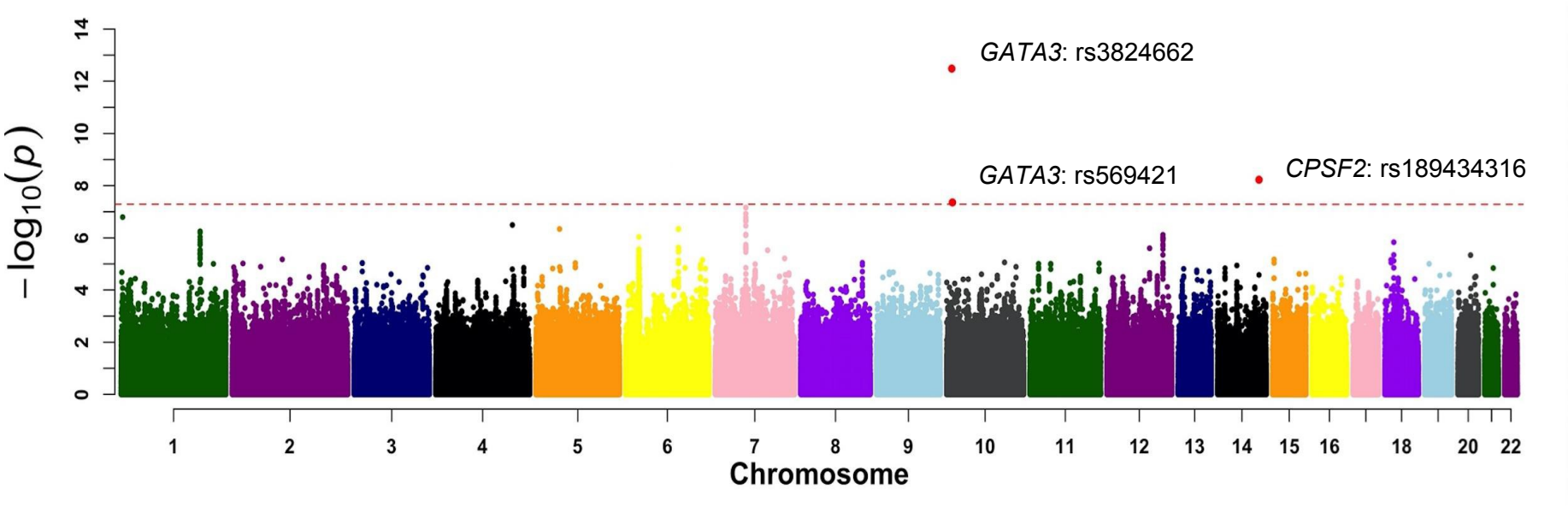
€: population outliers were identified using EIGENSTRAT software.

Supplemental Figure 2. Quantile-Quantile (QQ) plot showing genome-wide association p-values ($-\log_{10}$).



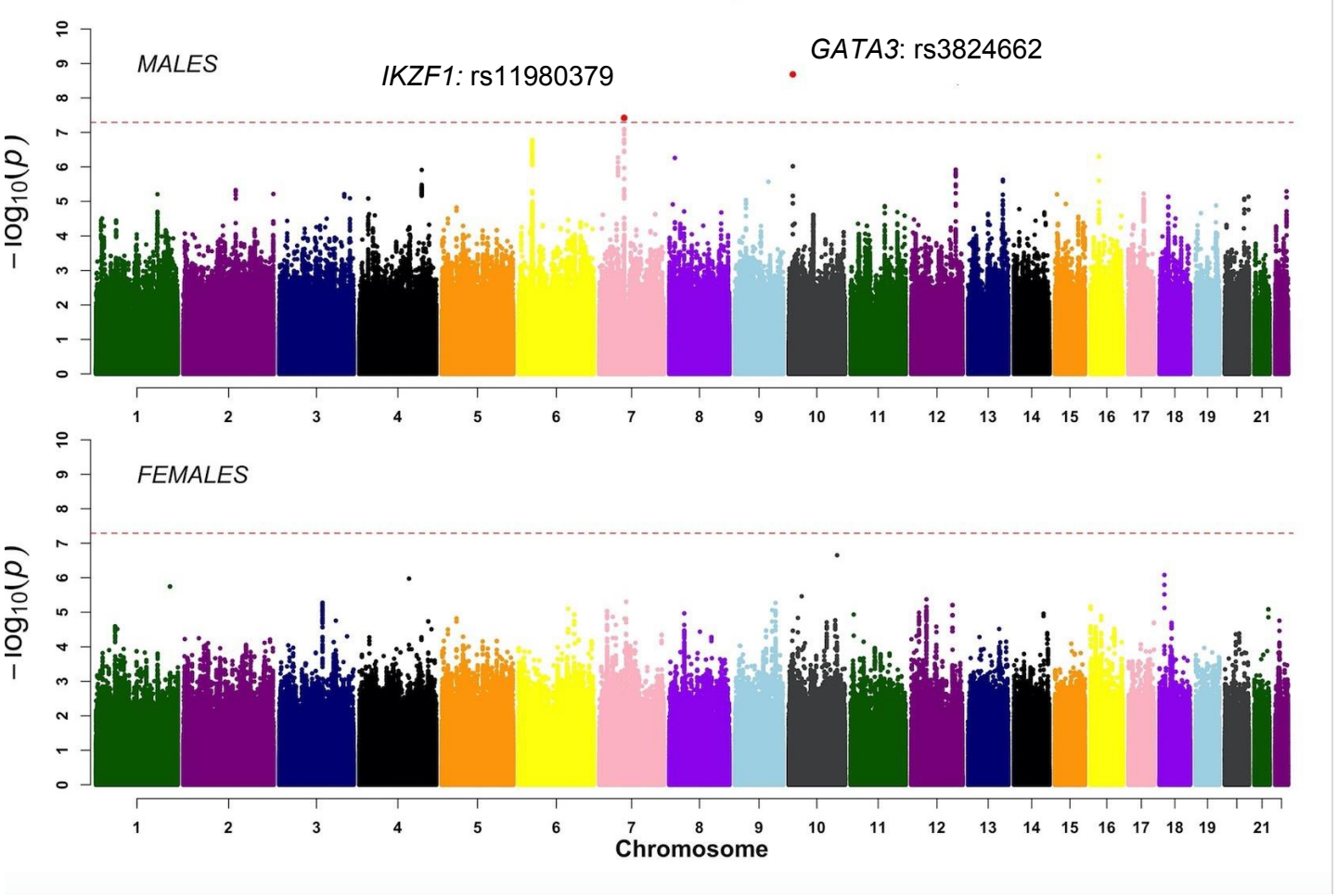
Q-Q plot of SNP association with B-ALL p-values from meta-analyses. There was no evidence of genomic inflation in Cohorts 1 or 2

Supplemental Figure 3. Manhattan plot of germline genetic variants associated with B-ALL with Normal Cytogenetics in European Americans



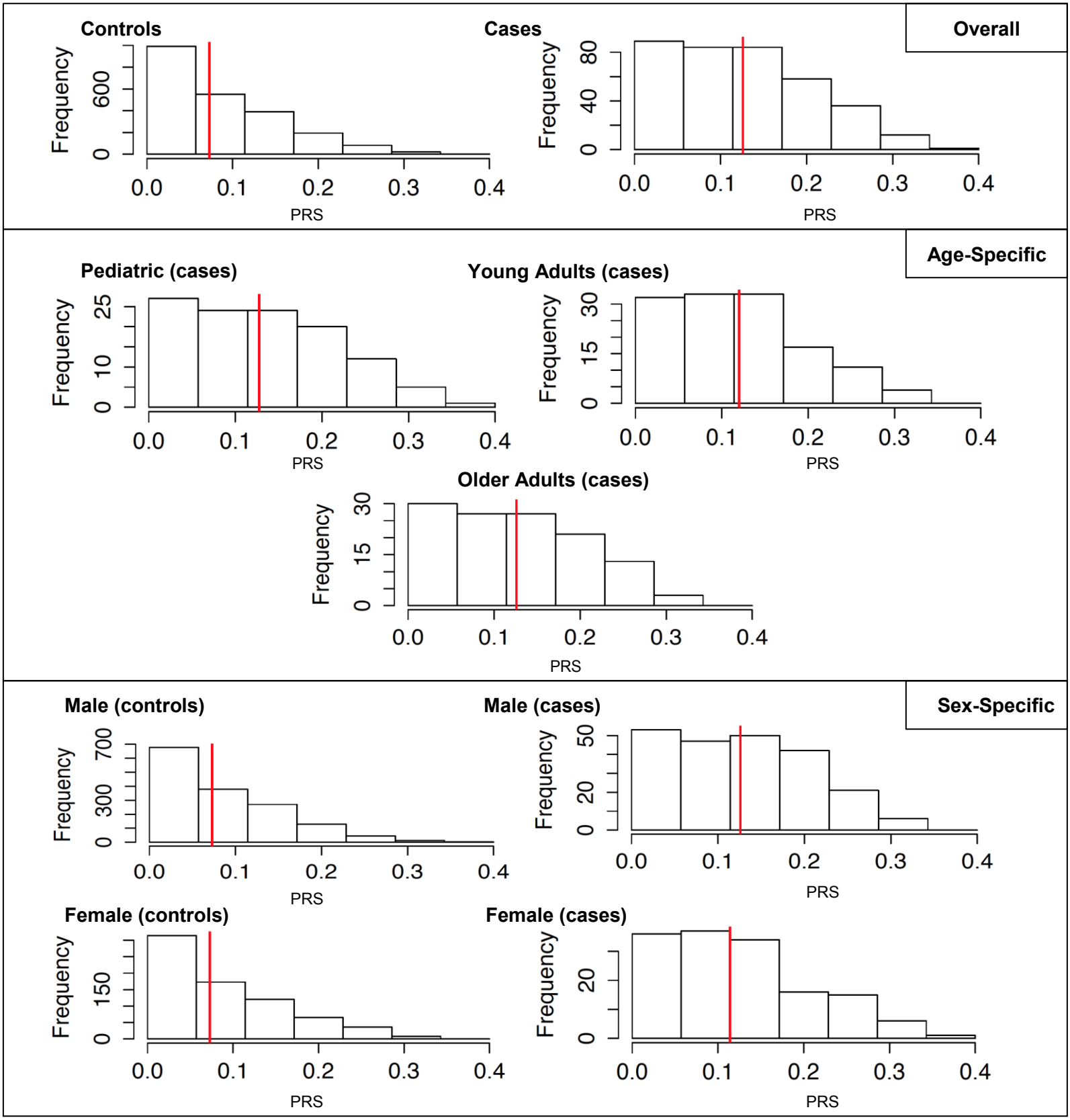
The x-axis indicates the chromosome number and the y-axis is the $-\log_{10}$ p-value showing SNP associations with meta-analyses of B-ALL with Normal Cytogenetics adjusted for age. The red horizontal line indicates a genome-wide significant P-value 5.0×10^{-8}. Genome-wide significant associations are on chromosomes 10 (*GATA3*) and 14 (*CPSF2*) indicated by red dots.

Supplemental Figure 4. Manhattan plot of germline genetic variants associated with B-ALL in European American Males and Females



Sex-specific GWAS of variants associated with B-ALL demonstrates a genome-wide significant association in males on chromosome 10: G*ATA3* and a marginal significant association on chromosome 7: *IKZF1*. GWAS of variants associated with B-ALL in females does not yield any variants that reach genome wide significance P_{meta} (5.0×10^{-8}) indicated by the red horizontal line. The x-axis indicates the chromosome and the y-axis is the $-\log_{10}$ p-value.

Supplemental Figure 5. Distribution of polygenic risk scores among cases and controls (overall, age-/ sex-specific)



Supplementary Figure 5 Legend: The x-axis is the polygenic risk score (PRS). The y-axis is the frequency of each score. The plots are shown for controls and cases overall, within case groups: Pediatric, Young Adults, and Older Adults, and sex-specific cases and controls. The red line indicates the median PRS for each group. Proportion of individuals at the higher end of the spectrum (range 0-0.35) are indicative of a higher PRS. The figures demonstrates cases, in general, have a higher proportion of those with larger PRSs and higher median PRS than the controls. This is what would be expected given the cases carry more of the high-risk susceptibility loci.

1. Hungate EA, Vora SR, Gamazon ER, Moriyama T, Best T, Hulur I, et al. A variant at 9p21.3 functionally implicates CDKN2B in paediatric B-cell precursor acute lymphoblastic leukaemia aetiology. *Nat Commun*. 2016;7:10635. PMCID: 4754340.

2. Winick N. Delayed intensification (DI) enhances event-free survival (EFS) of children with B-precursor acute lymphoblastic leukemia (ALL) who received intensification therapy with six courses of intravenous methotrexate (MTX):POG 9904/9905. A Children's Oncology Group Study (COG). ASH Annual Meeting Abstract. 2007;110:583.

3. Borowitz MJ, Devidas M, Hunger SP, Bowman WP, Carroll AJ, Carroll WL, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. Blood. 2008;111(12):5477-85. PMCID: 2424148.

4. Ellinghaus E, Stanulla M, Richter G, Ellinghaus D, te Kronnie G, Cario G, et al. Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. *Leukemia*. 2012;26(5):902-9. PMCID: 3356560.

5. Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, Schreiber S. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.* 2006;9(1):55-61.

6. Manolio TA, Rodriguez LL, Brooks L, Abecasis G, Ballinger D, Daly M, et al. New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat Genet.* 2007;39(9):1045-51.