A germ-line deletion of APOBEC3B does not contribute to subtype-specific childhood acute lymphoblastic leukemia etiology

Amelia D. Wallace,[1] Stephen S. Francis,[2,3] Xiaorong Shao,[1] Adam J. de Smith,[2] Kyle M. Walsh,[2] Roberta Mckean-Cowdin,[4] Xiaomei Ma,[5] Gary Dahl,[6] Lisa F. Barcellos,[1] Joseph L. Wiemels[5] and Catherine Metayer[1]

[1]School of Public Health, University of California, Berkeley, CA; [2]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA; [3]Division of Health Sciences, University of Nevada, Reno, NV; [4]Preventative Medicine, University of Southern California, Los Angeles, CA; [5]Epidemiology, Yale, New Haven, CN and [6]Pediatrics Hematology/Oncology, Stanford University, Palo Alto, CA, USA

Correspondence: adw222@berkeley.edu
doi:10.3324/haematol.2017.179317

Table S1: Odds ratios for the association between the APOBEC3B deletion polymorphism and risk of childhood ALL stratified by self-reported Hispanic ethnicity

Hispanic	Cases (n)	Controls (n)	Model	OR*	95% CI	*P*-Value
Total	270	288	Additive	1.18	0.77-1.60	0.28
wt/wt	164	195	-	ref		-
wt/del	94	81	Dominant	1.28	0.90-1.83	0.18
del/del	12	12	Recessive	0.94	0.40-2.17	0.88
Non-Hispanic White	Cases (n)	Controls (n)	Model	OR*	95% CI	*P*-Value
Total	236	320	Additive	0.67	0.38-1.16	0.16
wt/wt	187	246	-	ref		-
wt/del	46	63	Dominant	0.64	0.35-1.15	0.14
del/del	3	11	Recessive	0.81	0.37-8.70	0.86

OR odds ratio; CI confidence interval; *wt* wildtype; *del APOBEC3B* deletion; ref reference
*Adjusted for genetic ancestry and sex

Table S2: Odds ratios for the association between the APOBEC3B deletion polymorphism adjusted for continental ancestral proportions in the APOBEC3 gene region

Regional ancestral proportion adjusted for:	Cases (n)	Controls (n)	Model	OR*	95% CI	*P*-Value
African	518	608	Additive	0.99	0.79-1.23	0.89
wt/wt	360	441	-	ref	-	-
wt/del	143	144	Dominant	1.05	0.80-1.37	0.73
del/del	15	23	Recessive	0.66	0.34-1.27	0.23
Native American	518	608	Additive	0.95	0.76-1.20	0.68
wt/wt	360	441	-	ref	-	-
wt/del	143	144	Dominant	1.01	0.77-1.33	0.92
del/del	15	23	Recessive	0.62	0.31-1.19	0.16
European	518	608	Additive	0.99	0.79-1.24	0.92
wt/wt	360	441	-	ref	-	-
wt/del	143	144	Dominant	1.05	0.80-1.38	0.7
del/del	15	23	Recessive	0.66	0.34-1.27	0.22
East Asian	518	608	Additive	1.02	0.82-1.28	0.86
wt/wt	360	441	-	ref	-	-
wt/del	143	144	Dominant	1.09	0.84-1.42	0.52
del/del	15	23	Recessive	0.7	0.36-1.34	0.28

OR odds ratio; CI confidence interval; *wt* wildtype; *del APOBEC3B* deletion; ref reference

SUPPLEMENTARY FIGURE LEGENDS

Figure S1 APOBEC3B PCR-based genotyping assay. Individuals 1, 2 homozygous non-carries of the deletion, individual 3 is heterozygous, carrying one copy of the ~30Kb deletion polymorphism.

Figure S2 MDS plot comparing genetic ancestry of *A3B* genotyped CCLS subjects to the Human Genome Diversity Project reference

Figure S3 CCLS *APOBEC3* Region GWAS. Blue dashed line significance threshold p=0.05; red dashed line Bonferroni adjusted significance threshold p=0.00004; rainbow box *APOBEC3B* deletion position

Figure S4 Association of immunogenetic ancestry at the APOBEC3 megalocus and ALL risk

Figure S1

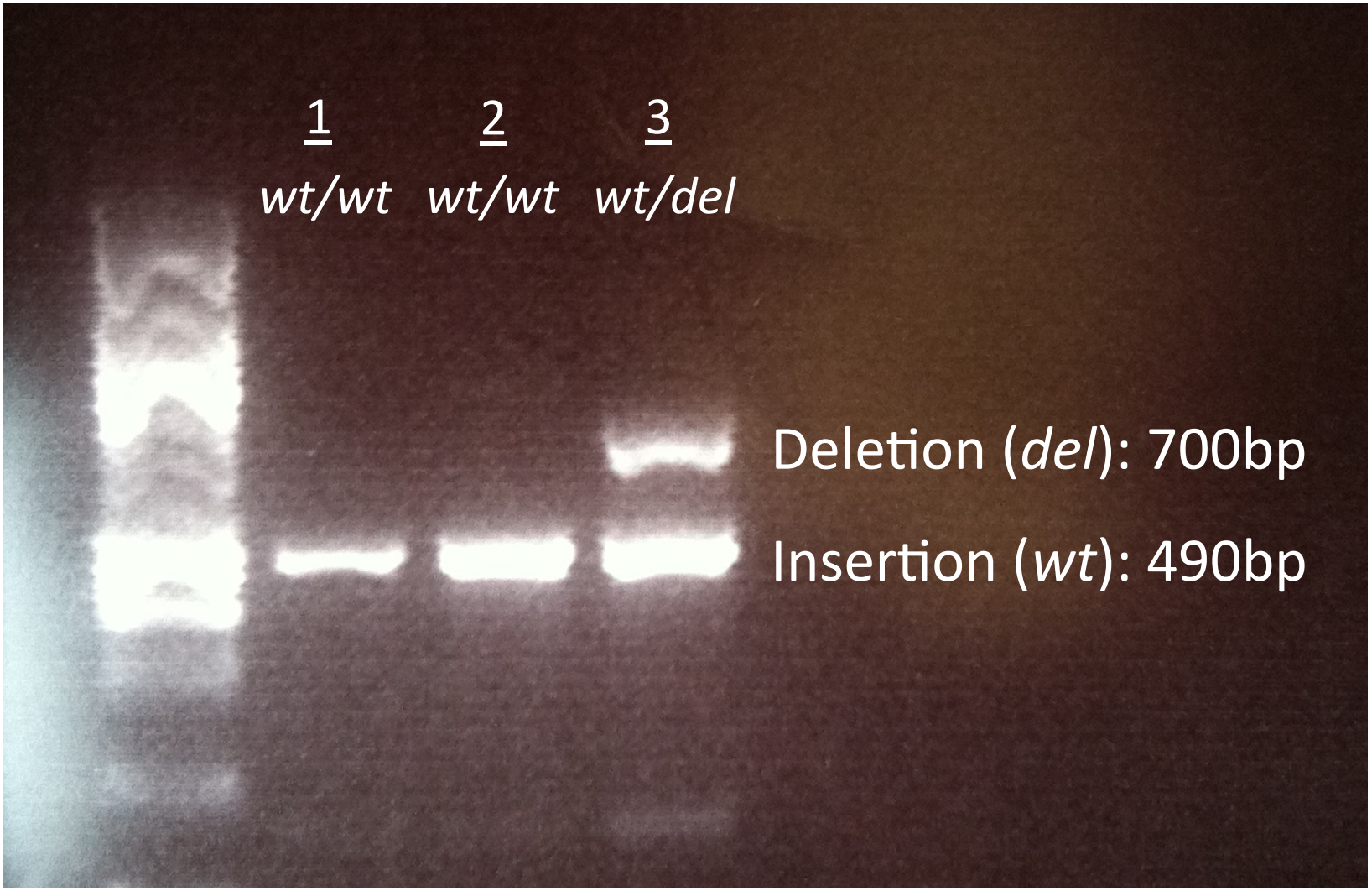


Figure S2

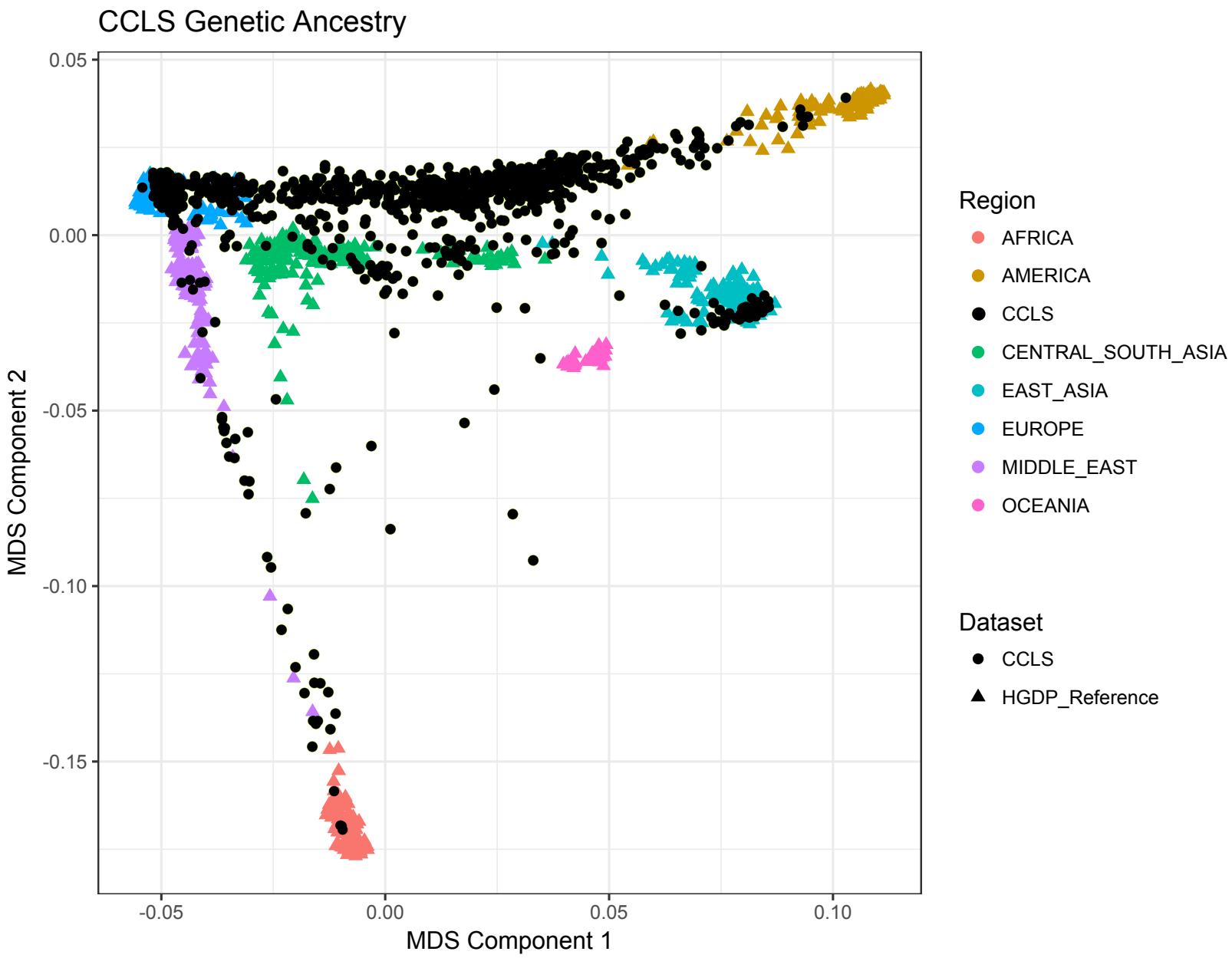

CCLS Genetic Ancestry
MDS Component 2
MDS Component 1
0.05
0.00
-0.05
-0.10
-0.15
-0.05
0.00
0.05
0.10
Region
AFRICA
AMERICA
CCLS
CENTRAL_SOUTH_ASIA
EAST_ASIA
EUROPE
MIDDLE_EAST
OCEANIA
Dataset
CCLS
HGDP_Reference

Figure S3

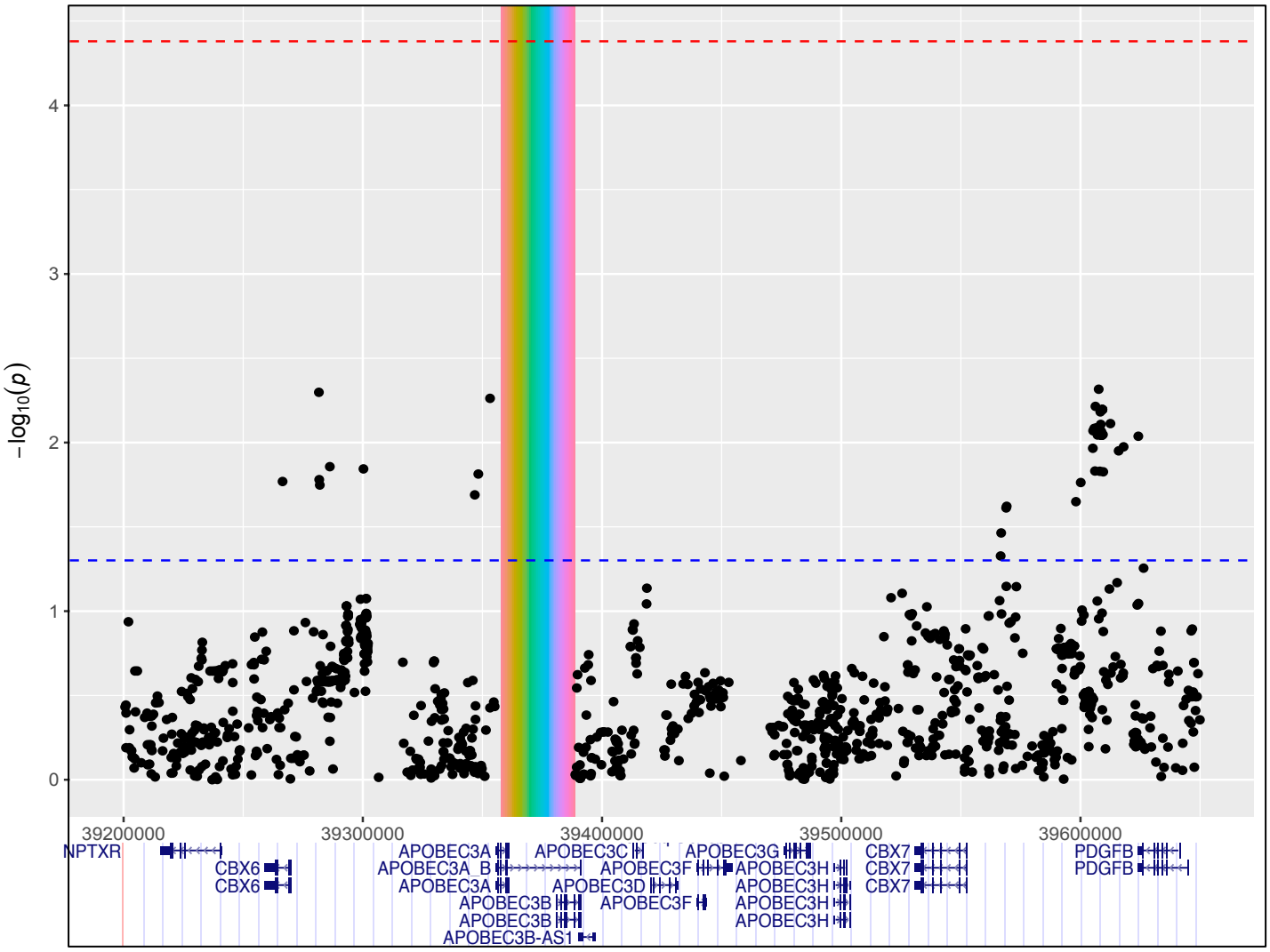
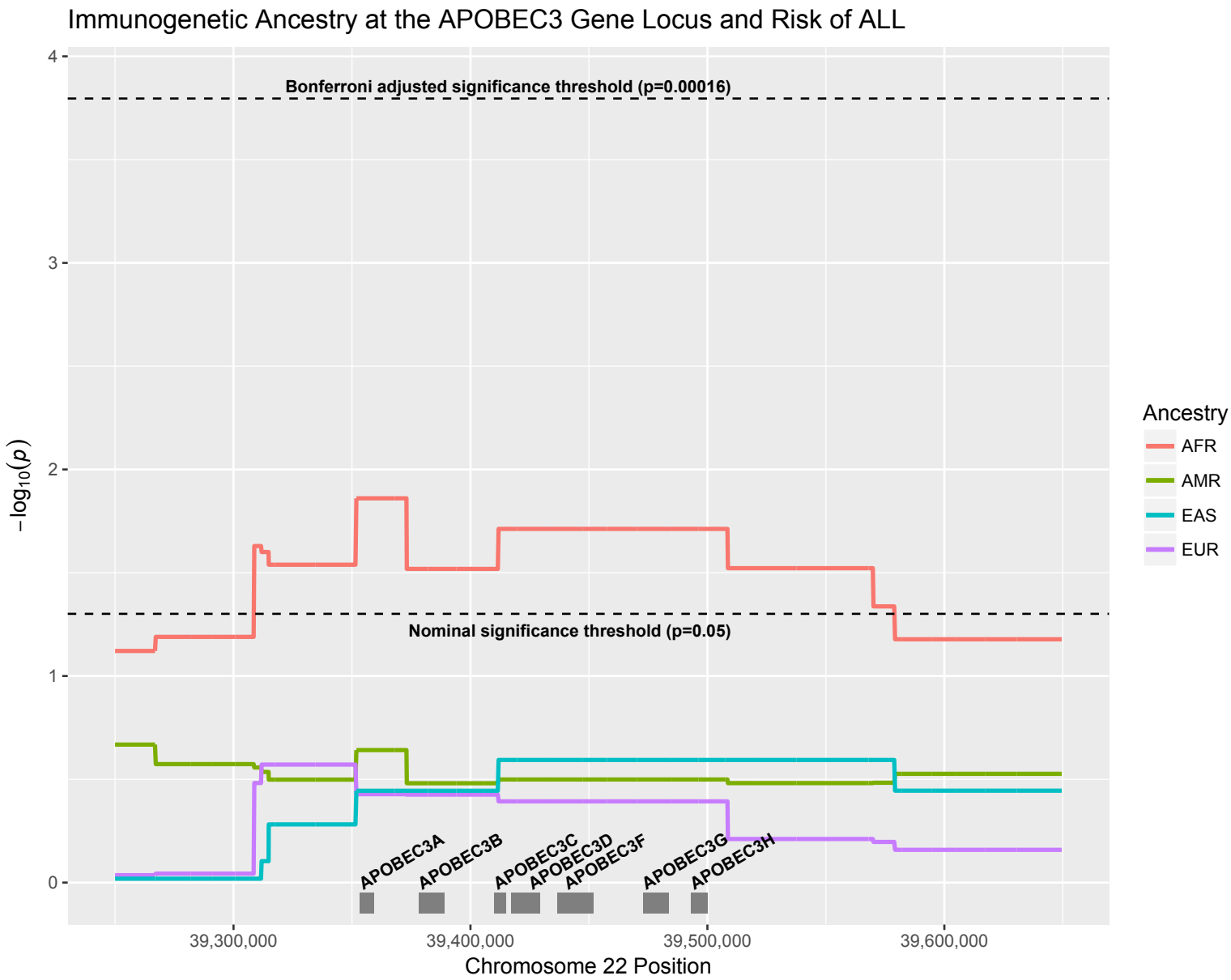

4
3
2
1
0
$-\log_{10}(p)$
39200000
39300000
39400000
39500000
39600000
NPTXR
CBX6
CBX6
APOBEC3A
APOBEC3A_B
APOBEC3A
APOBEC3B
APOBEC3B
APOBEC3B-AS1
APOBEC3C
APOBEC3F
APOBEC3D
APOBEC3F
APOBEC3G
APOBEC3H
APOBEC3H
APOBEC3H
APOBEC3H
CBX7
CBX7
CBX7
PDGFB
PDGFB
Position on Chromosome 22

Figure S4


Immunogenetic Ancestry at the APOBEC3 Gene Locus and Risk of ALL
Bonferroni adjusted significance threshold (p=0.00016)
Nominal significance threshold (p=0.05)
$-\log_{10}(p)$
4
3
2
1
0
APOBEC3A
APOBEC3B
APOBEC3C
APOBEC3D
APOBEC3F
APOBEC3G
APOBEC3H
39,300,000
39,400,000
39,500,000
39,600,000
Chromosome 22 Position
Ancestry
AFR
AMR
EAS
EUR

SUPPLEMENTARY METHODS

Individuals were enrolled in the CCLS between 1995 and 2015 with rapid, comprehensive case ascertainment from 80% of California hospitals, allowing capture of ~76% of all cases, usually within 72 hours of diagnosis. Controls were individually matched to cases on age, sex, child's Hispanic ethnicity, and maternal race (1). For consented individuals, saliva or buccal swab specimens were collected at the time of interview. Among cases, cytogenetic features of primary tumors were abstracted from medical records. The CCLS was approved by the University of California Berkeley Institutional Review Board and by all participating institutions. Informed consent was obtained from all participating subjects. A subset of subjects with biospecimens and genetic data available were selected for the present genetic study and were largely representative of the CCLS study population. Subjects with Down syndrome (n=52) were excluded.

PCR-Based APOBEC3B Deletion Genotyping: DNA was extracted from buccal and saliva samples and resuspended in Tris EDTA buffer. Copy-number of the deletion was assessed using a validated polymerase chain reaction (PCR) method described in Kidd *et al* (2). In brief, PCR amplification was carried out using the AmpliTaq Gold™ DNA Polymerase with Buffer II and $MgCl_2$ (ThermoFisher Scientific) following manufacturer's protocol. Two PCR reactions were run for each subject, one using a primer pair spanning the deletion breakpoints and resulting in a 700bp product (undeleted product is too large to be amplified), and another internal to the deletion breakpoints and producing a 490bp product. PCR products were combined and run on a 2% agarose gel by electrophoresis and visualized to determine genotype: subjects displaying no product for the "deletion" primer pair but showing the 490bp product for the "internal" primers were classed as homozygous undeleted; subjects with the 700bp deletion product plus the 490bp internal product were classed as heterozygous deleted; and subjects with the 700bp deletion product but no internal product were homozygous deleted (Figure S1). Each individual was genotyped once, and those individuals that appeared homozygous for the deletion were genotyped again using an independent set of primers (2). A positive result from the second reaction implied a heterozygous genotype (primer pair sequences available on request).

SNP Genotyping. In the CCLS, genome-wide genotype data were previously produced from DNA extracted from dried blood spots, saliva, or buccal cells genotyped using the Illumina HumanOmniExpress 12v1-1, HumanOmniExpressExome 8v1-2, and InfiniumOmniExpress 8v1-4 platforms, containing >700,000 SNP markers. SNPs with a call-rate less than 90% were excluded, as were individuals genotyped at less than 90% of markers.

SNP Imputation. Probes lying within the APOBEC3B deletion region (chr22: 39,357,694-39,388,574) were removed. Imputation was then carried out with IMPUTE2 (3) using the 1000 Genomes Phase 3 reference haplotypes.

Genetic Ancestry Estimation. Multidimensional scaling (MDS) components were derived using PLINK1.9 (5). Imputed SNPs were pruned for independence and the singular value decomposition-based algorithm was performed on an inter-sample distance matrix. The first 3 components were included in adjusted models to account for global genetic ancestry.

Local Genetic Ancestry Inference. The *APOBEC3* megalocus on chromosome 22 contains seven ABOBEC3 gene-family members (APOBEC3A-D, F-H). SNP-wise local genetic ancestry was estimated for 2,899 SNPs at the APOBEC3 megalocus (chr22:39,200,000-39,650,000) after removal of multi-allelic SNPs, SNPs with missing reference or alternate alleles, and SNPs within the *APOBEC3B* deletion region. First, the 2,899 selected SNPs were phased using BEAGLE 4.0 (6) independently in the CCLS and in five reference ancestral populations included in Phase 3 of the 1000 Genomes. Following phasing, local ancestry was inferred in the admixed CCLS data based on the five reference populations using RFMix, wherein ancestry was estimated for 65 0.1cM windows over three EM-iterations. Ancestral proportions across the APOBEC3 megalocus were calculated for each individual as the proportion of haplotypes assigned to each of the 5 reference ancestries. Dummy variables for the additive and dominant contribution of each reference ancestry were also assigned SNP-wise for each individual after pruning SNPs for local ancestry-based independence.

Statistical Analyses. Logistic regression was used to determine the association of the ~30kb deletion of *APOBEC3B* (assessed by PCR and tagging SNP, respectively) with ALL under three different models of inheritance: additive, autosomal dominant, and

recessive, adjusted for genetic ancestry and sex. Stratified analyses by Hispanic status and cytogenetic subtype were also conducted. All PCR-based analyses were carried out using R(7) and all tagging SNP-based analyses were carried out using SNPTest v2.5.2 to account for imputation uncertainty (8). Association analysis of *APOBEC3* gene region SNPs was carried out in PLINK 1.9 using logistic regression tests under the additive model, adjusting for genetic ancestry (5).

REFERENCES

1. Ma X, Buffler PA, Layefsky M, et al. Control selection strategies in case-control studies of childhood diseases. *American journal of epidemiology* 2004;159(10):915-21.
2. Kidd JNTT, E.; Kaul, R.; Eichler, E.; . Population Stratification of a Common APOBEC Gene Deletion Polymorphism. *PLoS Genetics* 2007;3(4).
3. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5(6):e1000529.
4. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015;31(21):3555-7.
5. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;4:7.
6. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American journal of human genetics* 2007;81(5):1084-97.
7. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2012.
8. Marchini J, Howie B, Myers S, et al. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics* 2007;39(7):906-13.