Genome-wide Association Study Identifies *HLA-DPB1* as a Significant Risk Factor for Severe Aplastic Anemia

Sharon A. Savage,[1,15,*] Mathias Viard,[2,15] Colm O'hUigin,[2,3] Weiyin Zhou,[3,4] Meredith Yeager,[3,4] Shengchao Alfred Li,[3,4] Tao Wang,[5,6] Veron Ramsuran,[7] Nicolas Vince,[8] Aurelie Vogt,[3,4] Belynda Hicks,[2,4] Laurie Burdett,[3,4] Charles Chung,[2,4] Michael Dean,[1] Kelvin C. de Andrade,[1] Neal D. Freedman,[1] Sonja I. Berndt,[1] Nathaniel Rothman,[1] Qing Lan,[1] James R. Cerhan,[9] Susan L. Slager,[9] Yawei Zhang,[10] Lauren R. Teras,[11] Michael Haagenson,[6] Stephen J. Chanock,[1] Stephen R. Spellman,[5,6] Youjin Wang,[1] Amanda Willis,[12] Medhat Askar,[12] Stephanie J. Lee,[5,13] Mary Carrington,[2,14] and Shahinaz M. Gadalla[1]

Severe aplastic anemia (SAA) is a rare disorder characterized by hypoplastic bone marrow and progressive pancytopenia. The etiology of acquired SAA is not understood but is likely related to abnormal immune responses and environmental exposures. We conducted a genome-wide association study of individuals with SAA genetically matched to healthy controls in discovery (359 cases, 1,396 controls) and validation sets (175 cases, 1,059 controls). Combined analyses identified linked SNPs in distinct blocks within the major histocompatibility complex on 6p21. The top SNP encodes p.Met76Val in the P4 binding pocket of the HLA class II gene *HLA-DPB1* (rs1042151A>G, odds ratio [OR] 1.75, 95% confidence interval [CI] 1.50–2.03, $p = 1.94 \times 10^{-13}$) and was associated with HLA-DP cell surface expression in healthy individuals ($p = 2.04 \times 10^{-6}$). Phylogenetic analyses indicate that Val76 is not monophyletic and likely occurs in conjunction with different HLA-DP binding groove conformations. Imputation of *HLA-DPB1* alleles revealed increased risk of SAA associated with Val76-encoding alleles *DPB1*03:01*, (OR 1.66, $p = 1.52 \times 10^{-7}$), *DPB1*10:01* (OR 2.12, $p = 0.0003$), and *DPB1*01:01* (OR 1.60, $p = 0.0008$). A second SNP near *HLA-B,* rs28367832G>A, reached genome-wide significance (OR 1.49, 95% CI 1.22–1.78, $p = 7.27 \times 10^{-9}$) in combined analyses; the association remained significant after excluding cases with clonal copy-neutral loss-of-heterozygosity affecting class I HLA genes (8.6% of cases and 0% of controls). SNPs in the HLA class II gene *HLA-DPB1* and possibly class I (*HLA-B*) are associated with SAA. The replacement of Met76 to Val76 in certain *HLA-DPB1* alleles might influence risk of SAA through mechanisms involving DP peptide binding specificity, expression, and/or other factors affecting DP function.

Aplastic anemia is a rare heterogeneous disorder characterized by progressive pancytopenia and bone marrow hypoplasia.[1,2] An estimated 500 to 1,000 cases of severe aplastic anemia (SAA) occur in the United States each year.3 It occurs at all ages, and its etiology is not well understood. Acquired aplastic anemia is often immune-mediated and may occur after hepatitis or with certain drug or environmental exposures.[1,2] A limited number of studies with small sample sizes have evaluated the potential role(s) of common single-nucleotide polymorphisms (SNPs) in aplastic anemia etiology.[3–8] For example, a study of 170 cases and 222 controls from Pakistan found two *FAS* and two *FASL* SNPs associated with SAA.[9] Another study evaluated four *FOXP3* SNPs in 94 aplastic anemia cases and 195 controls from Korea, and this study suggested an association with disease and response to immunosuppressive therapy (IST).[10] Aplastic anemia also occurs in individuals with rare inherited bone marrow failure syndromes caused predominantly by pathogenic germline variants in DNA-repair, ribosomal, or telomere-biology genes.[11] Somatic copy neutral loss of heterozygosity in chromosome 6 (chr6-CNLOH) encompassing the HLA-class I locus has been previously described in acquired SAA.[12,13] Hematopoietic stem cells with chr6-CNLOH are thought to escape the cytotoxic T cell immune attack by deleting *HLA* alleles involved in autoantigen presentation.[14]

The cytopenias in aplastic anemia may progress to a life-threatening severe disease, and affected individuals are at high risk of progression to myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML).[15] In acquired SAA, allogeneic hematopoietic cell transplantation (HCT)

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA; [2]Basic Science Program, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA; [3]Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA; [4]Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA; [5]Center for International Blood and Marrow Transplant Research, Medical College of Wisconsin, Milwaukee, WI 53226, USA; [6]Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI 53226, USA; [7]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban, South Africa; [8]Université de Nantes, CHU Nantes, Inserm, Centre de Recherche en Transplantation et Immunologie, UMR 1064, ITUN, F-44000 Nantes, France; [9]Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55902, USA; [10]Section of Surgical Outcomes and Epidemiology, Department of Surgery, Yale Medical School, New Haven, CT 06520, USA; [11]Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA, 30303, USA; [12]Department of Pathology and Laboratory Medicine, Baylor University Medical Center, Dallas, TX 76798, USA; [13]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [14]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard University, Cambridge, MA 02139, USA
[15]These authors contributed equally to this work
*Correspondence: savagesh@mail.nih.gov
https://doi.org/10.1016/j.ajhg.2020.01.004.

is the first line of therapy for young individuals with a matched sibling donor, while IST, followed by HCT in non-responders, is recommended for all others.[2,16,17]

We conducted a genome-wide association study (GWAS) of acquired SAA to agnostically evaluate the contribution of common germline SNPs to the etiology of this highly morbid disease. Individuals with SAA (n = 895) were identified from the Center for International Blood and Marrow Transplant Research (CIBMTR) database, and biorepository and germline DNA were collected prior to HCT between 1989 and 2015. CIBMTR is a research collaboration between the National Marrow Donor Program (NMDP) "Be The Match Registry" and the Medical College of Wisconsin; CIBMTR has more than 450 reporting HCT centers, and it represents one of the world's largest databases and research repositories for HCT research. All individuals provided informed consent, and the use of the samples was approved by the NMDP Institutional Review Board (IRB-1991-0002). We excluded 93 individuals with inherited bone marrow failure based on clinical diagnoses reported to CIBMTR. Genotyping was conducted using the Illumina Infinium OmniExpress BeadChip array at the Cancer Genomics Research Laboratory (CGR) in the Division of Cancer Epidemiology and Genetics (DCEG) at the National Cancer Institute (NCI). The controls were derived from cancer-free subjects drawn from two cohort studies (the Prostate, Lung, Colon and Ovarian Cancer Prevention Trial [PLCO][18] and the American Cancer Society Cancer Prevention Study II [CPSII][19] previously scanned on the Illumina Omni 2.5M SNP microarray) as well as four other USA-based studies (Mayo Clinic Case-Control Study of Non-Hodgkin Lymphoma and Chronic Lymphocytic Leukemia [MAYO];[20] NCI Surveillance, Epidemiology, and End Results Non-Hodgkin Lymphoma [NHL] Case-Control Study [NCI-SEER];[21,22] Women's Health Initiative [WHI];[23] and the Population-based Case-Control Study in Connecticut Women [YALE][24]) which were also scanned on the Infinium OmniExpress chip. In order to reduce the effect of population stratification, population substructure analyses using STRUCTURE and principal components analyses (PCA) were used to limit the study inclusion to individuals of European ancestry. The final analyses included 534 acquired SAA cases (359 in a discovery set and 175 in a validation set) and 2,455 controls (1,396 in the discovery set and 1,059 in the validation set). Individuals included in the validation set were independent from those in the discovery set, and all study participants were genetically proven to be unrelated. Additional details are available in Supplementary Methods, Table S1, and Figure S1A and S1B.

The median age at HCT of participants with SAA was 21 years (range = 1–77), 56% were male, and the median time between SAA diagnosis and HCT was 11 months (range = 0.1–318 months). The majority of SAA-affected individuals (87%) received an unrelated donor HCT; all individuals who received a matched sibling HCT were part of the validation cohort (Table S2).

A series of strong association signals were identified across the human leukocyte antigen (*HLA*) genes encoded in the major histocompatibility complex (MHC) on chromosome 6p21 (Figure 1). Notably, 12 genotyped SNPs achieved genome-wide significance (pooled-p $< 3.46 \times 10^{-8}$; odds ratios (ORs) ranging from 1.62–1.82) (Table 1, Figure 1, and Tables S3 and S4). SNP associations with HLA class II genes were evident in two distinct regions. The top SNP, rs1042151 (pooled-p $= 1.94 \times 10^{-13}$), was in linkage disequilibrium with 18 other SNPs ($R^2 > 0.5$, chr6:33,048,661-33,064,605) that were in or near *HLA-DPB1* and *HLA-DPB2*. A second locus in HLA class II, containing 13 linked SNPs ($R^2 > 0.9$) in or near *HLA-DQA2* and *HLA-DQB2* (chr6:32,707,295-32,727,905) approached genome-wide significance (top SNP, rs9276370 T>G, pooled-p $= 5.01 \times 10^{-7}$, OR 1.41, 95% confidence interval [CI] 1.23–1.62).

The top SNP associated with SAA, rs1042151 A>G (combined OR 1.75, 95% CI 1.50–2.03, p = 1.94×10^{-13}), encodes a nonsynonymous change of p.Met76Val (RefSeq accession number NP_002112.3) in the β-1 domain of HLA-DPB1. While dbSNP annotates rs1042151 as p.Met105Val because it includes the first 30 residues of the leader peptide, p.Met76Val is the appropriate numbering according to the WHO Nomenclature Committee for Factors of the HLA System (IPD-IMGT/HLA database, see Web Resources).[25] rs1042151 A>G has a minor allele frequency (MAF) of 17.5% in the non-Finnish European, 8.9% in the East Asian, and 37.6% in the African/African American gnomAD populations.[26] In our study participants, the MAF of rs1042151 A>G (p.Met76Val) was 18.8% in controls and 29.4% in the individuals with SAA.

DPB1 comprises the beta strand of the HLA-DP protein heterodimer located on the cell surface of antigen-presenting cells.[27] Genetic variation in *HLA-DPB1* has been associated with several autoimmune diseases, including rheumatoid arthritis, Grave's disease, and multiple sclerosis.[28–30] Specifically, the p.Met76Val amino acid substitution caused by the rs1042151 G variant has previously been associated with risk of aspirin-exacerbated respiratory disease.[31] Residue 76 is involved in peptide binding, forming part of the P4 pocket, and the p.Met76Val substitution is likely to influence the binding repertoire.[32] Two other SNPs (rs9277534 and rs9277535), located in the 3′ UTR of *HLA-DPB1*, approximately 6 kilobases (kb) away from rs1042151 but not genotyped in our study (Figure 1C), have been previously associated with expression of *HLA-DPB1*, risk and persistence of hepatitis B virus (HBV) infection, and risk of graft-versus-host disease in HCT recipients.[33–36]

To evaluate the possible contribution of HLA-DP cell surface expression related to the association between the rs1042151 A>G genotypes and risk of SAA, we stained CD19+ cells isolated from peripheral blood drawn from 175 healthy donors with the HLA-DP-specific monoclonal antibodies B7/21 and BrafB6[37,38] as described previously[33] (see Supplementary Methods in Supplemental Data), and

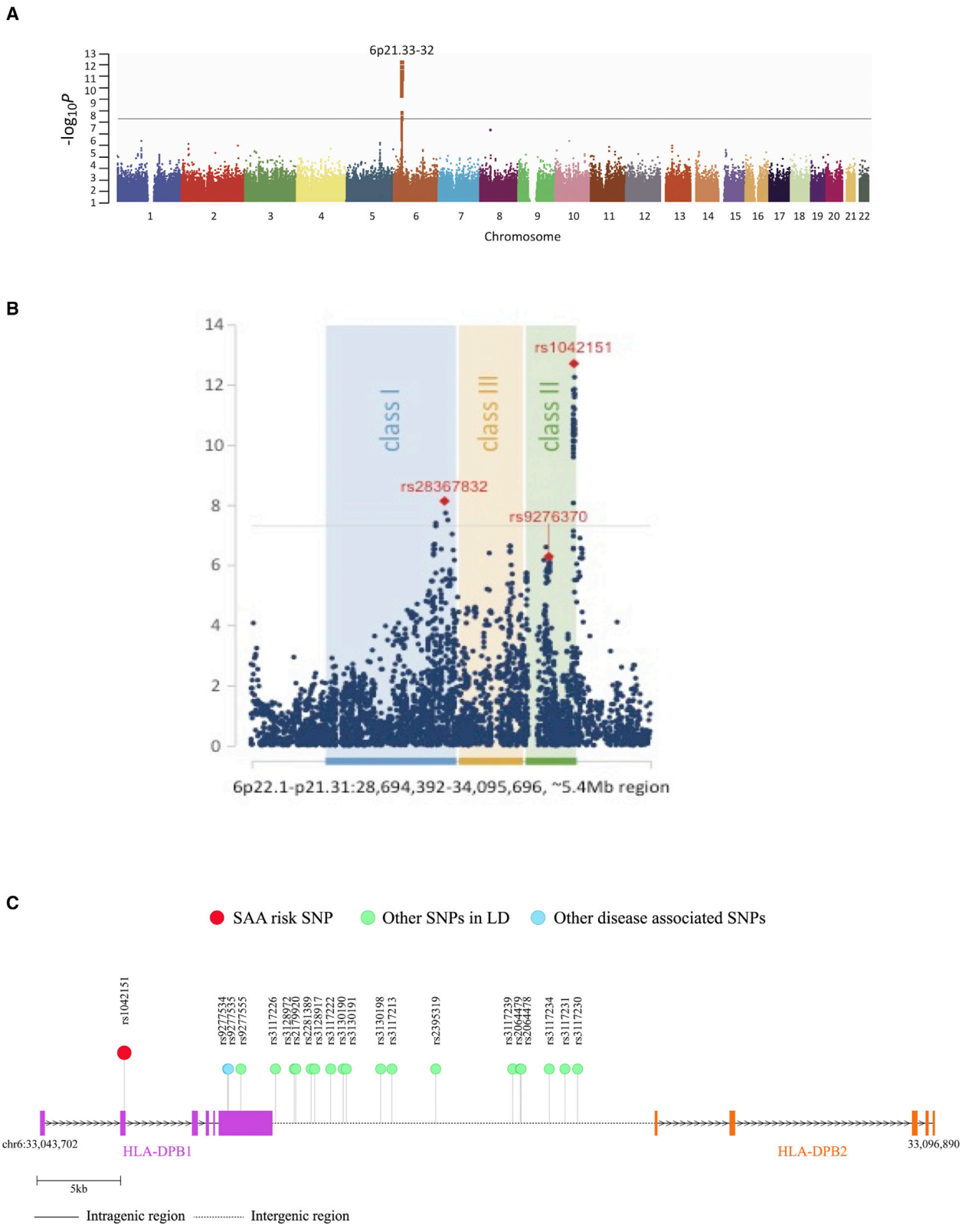


Figure 1. Genomic Locations of Single-Nucleotide Polymorphims Associated with Severe Aplastic Anemia
(A) Manhattan plot of genome-wide association result for combined discovery and validation sets of European ancestry severe aplastic anemia cases and genomically matched controls.
(B) A portion of the Manhattan plot of the major histocompatibility complex (MHC) genomic region showing locations of SNPs in human leukocyte antigen (*HLA*) genes by HLA class.
(C) Schematic of HLA-DPB1 with SNPs of interest noted.

Table 1. Summary of combined association results for severe aplastic anemia cases and controls

dbSNP	Gene	Locus	Position	Variant	Stage	Cases	Controls	*lp*	OR (95% CI)
rs1042151									
rs1042151	*HLA-DPB1*	6p21	chr6: 33048661	A > G	Discovery	356	1396	1.68E-10	1.82 (1.51-2.20)
rs1042151	*HLA-DPB1*	6p21	chr6: 33048661	A > G	Validation	175	1059	0.0015	1.52 (1.17-1.97)
rs1042151	*HLA-DPB1*	6p21	chr6: 33048661	A > G	Combined	531	2453	**1.94E-13**	**1.75 (1.50-2.03)**
rs28367832									
rs28367832	*HLA-B*	6p21	chr6: 31305731	G > A	Discovery	355	1388	3.05E-06	1.49 (1.27-1.79)
rs28367832	*HLA-B*	6p21	chr6: 31305731	rs28367832	Validation	173	1048	0.001	1.46 (1.16-1.84)
rs28367832	*HLA-B*	6p21	chr6: 31305731	rs28367832	Combined	528	2434	**7.27E-09**	**1.49 (1.22-1.78)**

SNPs shown are those with combined validation and discovery p < 10^{-8}

we compared the median fluorescence intensity across genotypes. The risk variant, G, was associated with a significant dose-response increase in HLA-DP cell surface expression on the CD19+ cell surface ($p = 2.04 \times 10^{-6}$) (Figure 2); this association suggests that higher cell surface expression of HLA-DP may contribute to SAA risk. However, because of the strong linkage disequilibrium across this region, it is likely that the identified risk SNP is not solely related to cell surface expression of DP. A variant previously associated with HLA-DP cell surface expression, rs9277534 A>G,[33] is also associated with SAA in our study (OR 1.64, $p = 8.2 \times 10^{-8}$). Additionally, rs1042151 is not the strongest HLA-DP eQTL, but based on GTExPortal (data accessed November 2019, see Web Resources),[39] it is associated with the transcript level of HLA-DPB1 in transformed fibroblasts with a p value of 1×10^{-9}.

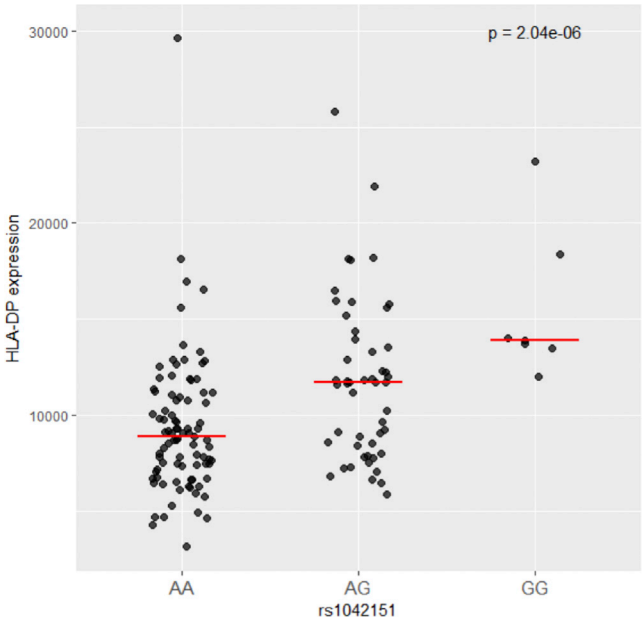


Figure 2. Cell Surface Expression of HLA-DP is Associated with rs1042151 Genotype
The cell surface expression levels of specific HLA alleles were determined in healthy individuals and plotted based on their rs1042151 genotype (see Online Methods).

To further understand the role of HLA-DPB1 in SAA etiology, we constructed a phylogenic tree of *HLA-DPB1* through the use of the neighbor-joining method based on Kimura 2-parameter distances as implemented in MEGA7 26 to visualize the distribution pattern of HLA-DP sequences as marked by Val or Met at position 76. Our results, based on both coding regions (Figure 3) and full-length gene sequences (not shown), indicate that 76Val is not monophyletic and is found in multiple distinct clades of alleles which are likely to recognize different peptide binding motifs. We then used the HIBAG software[40] to impute *HLA-DPB1* alleles for SAA cases and controls; comparisons with high-resolution clinical typing were available for 401 individuals with SAA and showed 92.7% concordance. SAA excess risk was associated with the *HLA-DPB1* alleles containing the rs1042151 Val76 variant: *DPB1*03:01* (OR 1.66, $p = 1.52 \times 10^{-7}$), *DPB1*10:01* (OR 2.12, p = 0.0003), and *DPB1*01:01* (OR 1.60, p = 0.0008). On the other hand, an inverse association was present between the Met76-encoding alleles *DPB1*04:01* (OR = 0.78, p = 0.01) and *DPB1*04:02* (OR = 0.75, p = 0.02) (Table S5). Logistic regression analysis based upon the number of Val76 residues shows an additive effect; compared to zero copies of Val76-encdig alleles, two copies are associated with higher SAA risk (OR 2.32, $p = 3.2 \times 10^{-6}$) than one copy is (OR = 1.98, $p = 1.2 \times 10^{-11}$). Peptide repertoire specificity is not determined by the P4 pocket alone, and the phylogenetic distribution of Val76 (i.e., interspersed with Met76-encoding alleles) indicates that Val76 occurs on wide variety of antigen-binding groove configurations. Taken together, the data suggest that the p.Met76Val change might influence risk of SAA through a mechanism involving DP peptide binding specificity, *HLA-DPB1* cell surface expression levels, and/or other factors affecting DP function. For example, the P4 pocket of DRB1 has been shown to facilitate immunogenic metabolite binding,[41] a scenario that may hold true for DP, as well, resulting in risk of immune-mediated diseases such as SAA.

One SNP in HLA class I, rs28367832 G>A, reached genome-wide statistical significance only in the combined analysis (pooled-$p = 7.27 \times 10^{-9}$, OR 1.49, 95% CI

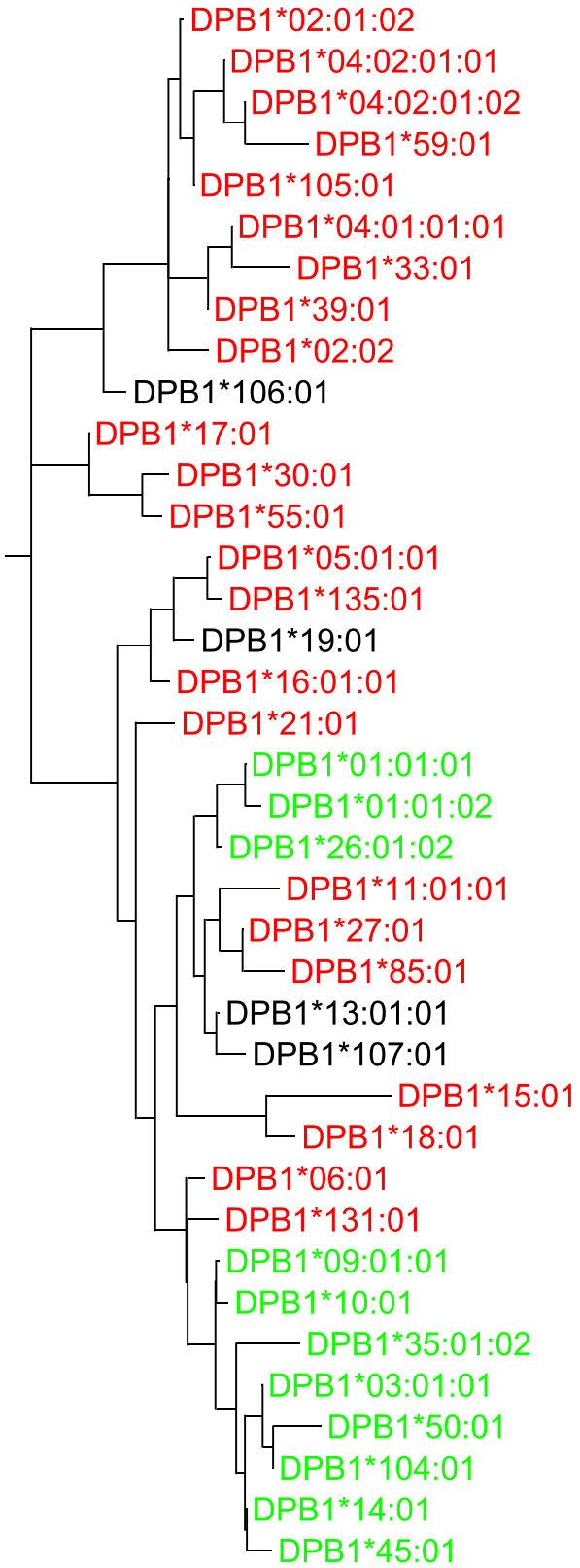


Figure 3. Phylogenic Tree of Coding Regions of HLA-DPB1
The HLA-DBP1 alleles are color coded according to the amino acid at residue 76 as follows: red, methionine, M; black, isoleucine, I; green, valine, V. The neighbor-joining tree is based on Kimura's 2-parameter distances for the entire 777 base pair coding region.

1.22–1.78) (Table 1, Table S4). rs28367832 G>A is 15.9 kb telomeric of *HLA-B* and a common noncoding SNP with a MAF of 57.7% in the SAA cases compared with 47.5% of control in the combined analyses. This SNP was not in linkage equilibrium with our study top SNP rs1042151 ($R^2 = 0.0099$). The association remained statistically significant in conditional analysis (p for rs28367832 = 1.82×10^{-8}). Because rs28367832 G>A is located in the region possibly affected by chr6-CNLOH, we calculated the Log R ratio (LRR) and B allele frequency (BAF) in order to assess somatic chr6-CNLOH in our samples (see Figure S2 and Supplementary Methods in Supplemental Data). There were 46 individuals with SAA (8.6%) with somatic chr6-CNLOH affecting class I HLA genes (Figure S2). Next-generation sequencing of the HLA genes performed with the MIA FORA NGS kit (Immucor) confirmed these SNP array results (see Figure S3 and Supplementary Methods in Supplemental Data). Somatic chr6-CNLOH appears to be specific to SAA; none of the 2,453 controls in this study nor of 73 the individuals with Fanconi anemia in our prior work had this alteration.[42] Persons with SAA and somatic chr6-CNLOH were more likely to carry the rs28367832 AA genotype, and this result suggests loss of the common rs28367832 G variant (56.1% versus 32.9%, p = 0.003). However, exclusion of SAA individuals with somatic chr6-CNLOH from the analyses did not change the germline SNP association results (data not shown).

Several additional non-HLA loci did not reach the accepted statistical genome-wide association threshold of $p < 5x10^{-8}$ (Table S6). An inverse association was present between SAA and the A allele of rs7845664 on chromosome 8p11.21, a multi-allelic SNP within the *ZMAT4* gene (OR 0.70, $p = 3.07 \times 10^{-7}$). There are little data on *ZMAT4* function, but one study previously identified copy number variants (CNVs) and gene expression changes in hematologic malignancies.[43] We did not identify CNVs in this region in our cohort of individuals with SAA (data not shown). Four additional non-HLA loci contained SNPs with borderline p values that may warrant follow-up in a larger study: rs12753487 (OR 1.75, pooled-$p = 2.85 \times 10^{-6}$), rs1731229 (OR 0.72, pooled-$p = 5.26 \times 10^{-6}$), rs4345355 (OR 1.54, pooled-$p = 5.29 \times 10^{-6}$), and rs9533317 (OR 1.4, pooled-$p = 7.87 \times 10^{-6}$). Of these, the intragenic rs9533317 SNP is located within *EPSTI1*, a gene with functions in epithelial stromal interactions.[44] These results may warrant follow-up because they may identify a set of individuals in whom SAA is not immune mediated.

This GWAS focused on individuals with SAA who required HCT. Our data suggest that idiopathic SAA may be due, in part, to germline susceptibility due to common variants in HLA class II, and possibly class I. It remains possible that some of the identified loci are associated with disease severity or lack of response to IST rather than with the etiology of SAA because all cases required HCT for their disease. However, the agreement between results from the discovery (all received unrelated donor HCT,

mostly after failing IST) and validation sets (40% received HCT from matched sibling as a first line of therapy and did not receive IST) argues against results being associated primarily with IST response. The small sample size in this study is small in the scope of GWAS but very large in the context of a rare disease. We limited our analysis to individuals of European ancestry; analyses in populations of other racial backgrounds are of interest, particularly because SAA is more common in Asians.[45]

In summary, this GWAS of acquired SAA identified a strong association at rs1042151 A>G in the *HLA-DPB1*. The rs1042151 G allele encoding HLA-DPB1 Val76 occurs on three distinct alleles, *DPB1*03:01*, *DPB1*10:01*, and *DPB1*01:01*, and this suggests that alteration of HLA-DP peptide binding specificity, as well as changes in HLA-DP cell surface expression and/or other factors affecting HLA-DP function, contribute to SAA etiology. A prospective study including the full spectrum of disease severity is warranted in order to further understand the association of HLA germline SNPs in SAA etiology.

Accession Numbers

GWAS data will be available through the dbGAP-controlled access database accession number dbGaP: phs001710.v1.p1.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2020.01.004.

Acknowledgments

This work was supported by the intramural research program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI). The Center for International Blood and Marrow Transplant Research (CIBMTR) is supported primarily by Public Health Service Grant/Cooperative Agreement 5U24CA076518 from the NCI, the National Heart, Lung and Blood Institute (NHLBI), and the National Institute of Allergy and Infectious Diseases (NIAID); Grant/Cooperative Agreement 4U10HL069294 from the NHLBI and the NCI; contract HHSH250201200018C with the Health Resources and Services Administration (HRSA/DHHS); and two grants, N00014-17-1-2388 and N0014-17-1-2850, from the Office of Naval Research. The Mayo Clinic Case-Control Study of Non-Hodgkin Lymphoma and Chronic Lymphocytic Leukemia (MAYO) study is funded through grants R01 CA92153 and P50 CA97274. The NCI Surveillance, Epidemiology, and End Results Non-Hodgkin Lymphoma Case-Control Study (NCI-SEER) study is funded by the Intramural Research Program of the NCI, National Institutes of Health (NIH), and Public Health Service (N01-PC-65064, N01-PC-67008, N01-PC-67009, N01-PC-67010, and N02-PC-71105). The Population-based Case-Control Study in Connecticut Women (YALE) is funded through the NCI (CA62006 and CA165923). This project has also been funded in part with federal funds from the Frederick National Laboratory for Cancer Research, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This research was supported in part by the Intramural Research Program of the NIH, Frederick National Lab, Center for Cancer Research. V.R. is supported by the African Academy of Sciences (AAS) and the Royal Society, which is funded by the UK Government as part of the Global Challenge Research Fund (GCRF); and V.R. is also supported by the South African Medical Research Council (SAMRC) with funds from the Department of Science and Technology (DST). V.R. is funded in part through the Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE), a Developing Excellence in Leadership Training and Science (DELTAS) Africa Initiative (grant # DEL-15-006) by the AAS.

Declaration of Interests

The authors declare no competing interests.

Received: March 23, 2019
Accepted: January 7, 2020
Published: January 30, 2020

Web Resources

GTExPortal, https://www.gtexportal.org/home/gene/DPB1

Immuno Polymorphism Database—International ImMunoGeneTics project HLA (IPD-IMGT/HLA), https://www.ebi.ac.uk/ipd/imgt/hla/

National Organization for Rare Disorders (NORD) Rare Disease Information on Acquired Aplastic Anemia, https://rarediseases.org/rare-diseases/acquired-aplastic-anemia/

References

1. Luzzatto, L., and Risitano, A.M. (2018). Advances in understanding the pathogenesis of acquired aplastic anaemia. Br. J. Haematol. *182*, 758–776.
2. Boddu, P.C., and Kadia, T.M. (2017). Updates on the pathophysiology and treatment of aplastic anemia: a comprehensive review. Expert Rev. Hematol. *10*, 433–448.
3. Savage, S.A., Calado, R.T., Xin, Z.T., Ly, H., Young, N.S., and Chanock, S.J. (2006). Genetic variation in telomeric repeat binding factors 1 and 2 in aplastic anemia. Exp. Hematol. *34*, 664–671.
4. Takaku, T., Calado, R.T., Kajigaya, S., and Young, N.S. (2009). Interleukin-23 receptor (IL-23R) gene polymorphisms in acquired aplastic anemia. Ann. Hematol. *88*, 653–657.
5. Li, B., Guo, L., Zhang, Y., Xiao, Y., Wu, M., Zhou, L., Chen, S., Yang, L., Lu, X., and Li, Y. (2016). Molecular alterations in the TCR signaling pathway in patients with aplastic anemia. J. Hematol. Oncol. *9*, 32.
6. Wu, Z., Miao, M., Qiu, Y., Qin, Z., Wang, J., Jiang, Y., Ming, Z., and Zhang, X. (2013). Association between polymorphisms in PDCD1 gene and aplastic anemia in Chinese Han population. Leuk. Lymphoma *54*, 2251–2254.
7. Ming, Z.J., Hui, H., Miao, M., Qiu, Y.H., and Zhang, X.G. (2012). Polymorphisms in PDCD1 gene are not associated with aplastic anemia in Chinese Han population. Rheumatol. Int. *32*, 3107–3112.
8. Chang, H., Zeng, F., Zhang, J.Y., Mu, X.Y., Meng, W.T., Ma, H.B., and Liu, T. (2010). Association of the interferon-gamma single nucleotide polymorphism +874(T/A) with response to

immunosuppressive therapy in patients with severe aplastic anemia. Blood Cells Mol. Dis. *45*, 313–316.
9. Rehman, S., Saba, N., Naz, M., Ahmed, P., Munir, S., Sajjad, S., Tabassum, S., and Naseem, L. (2018). Single-Nucleotide Polymorphisms of FAS and FASL Genes and Risk of Idiopathic Aplastic Anemia. Immunol. Invest. *47*, 484–491.
10. In, J.W., Lee, N., Roh, E.Y., Shin, S., Park, K.U., and Song, E.Y. (2017). Association of aplastic anemia and FoxP3 gene polymorphisms in Koreans. Hematology *22*, 149–154.
11. West, A.H., and Churpek, J.E. (2017). Old and new tools in the clinical diagnosis of inherited bone marrow failure syndromes. Hematology (Am. Soc. Hematol. Educ. Program) *2017*, 79–87.
12. Katagiri, T., Sato-Otsubo, A., Kashiwase, K., Morishima, S., Sato, Y., Mori, Y., Kato, M., Sanada, M., Morishima, Y., Hosokawa, K., et al.; Japan Marrow Donor Program (2011). Frequent loss of HLA alleles associated with copy number-neutral 6pLOH in acquired aplastic anemia. Blood *118*, 6601–6609.
13. Betensky, M., Babushok, D., Roth, J.J., Mason, P.J., Biegel, J.A., Busse, T.M., Li, Y., Lind, C., Papazoglou, A., Monos, D., et al. (2016). Clonal evolution and clinical significance of copy number neutral loss of heterozygosity of chromosome arm 6p in acquired aplastic anemia. Cancer Genet. *209*, 1–10.
14. Zaimoku, Y., Takamatsu, H., Hosomichi, K., Ozawa, T., Nakagawa, N., Imi, T., Maruyama, H., Katagiri, T., Kishi, H., Tajima, A., et al. (2017). Identification of an HLA class I allele closely involved in the autoantigen presentation in acquired aplastic anemia. Blood *129*, 2908–2916.
15. Shimamura, A. (2016). Aplastic anemia and clonal evolution: germ line and somatic genetics. Hematology (Am. Soc. Hematol. Educ. Program) *2016*, 74–82.
16. Bacigalupo, A. (2017). How I treat acquired aplastic anemia. Blood *129*, 1428–1436.
17. Scheinberg, P. (2018). Recent Advances and Long-Term Results of Medical Treatment of Acquired Aplastic Anemia: Are Patients Cured? Hematol. Oncol. Clin. North Am. *32*, 609–618.
18. Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A., et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. Nat. Genet. *40*, 310–315.
19. Amundadottir, L., Kraft, P., Stolzenberg-Solomon, R.Z., Fuchs, C.S., Petersen, G.M., Arslan, A.A., Bueno-de-Mesquita, H.B., Gross, M., Helzlsouer, K., Jacobs, E.J., et al. (2009). Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat. Genet. *41*, 986–990.
20. Cerhan, J.R., Fredericksen, Z.S., Wang, A.H., Habermann, T.M., Kay, N.E., Macon, W.R., Cunningham, J.M., Shanafelt, T.D., Ansell, S.M., Call, T.G., et al. (2011). Design and validity of a clinic-based case-control study on the molecular epidemiology of lymphoma. Int. J. Mol. Epidemiol. Genet. *2*, 95–113.
21. Chatterjee, N., Hartge, P., Cerhan, J.R., Cozen, W., Davis, S., Ishibe, N., Colt, J., Goldin, L., and Severson, R.K. (2004). Risk of non-Hodgkin's lymphoma and family history of lymphatic, hematologic, and other cancers. Cancer Epidemiol. Biomarkers Prev. *13*, 1415–1421.
22. Wang, S.S., Cerhan, J.R., Hartge, P., Davis, S., Cozen, W., Severson, R.K., Chatterjee, N., Yeager, M., Chanock, S.J., and Rothman, N. (2006). Common genetic variants in proinflammatory and other immunoregulatory genes and risk for non-Hodgkin lymphoma. Cancer Res. *66*, 9771–9780.
23. Anderson, G.L., Manson, J., Wallace, R., Lund, B., Hall, D., Davis, S., Shumaker, S., Wang, C.Y., Stein, E., and Prentice, R.L. (2003). Implementation of the Women's Health Initiative study design. Ann. Epidemiol. *13* (9, Suppl), S5–S17.
24. Zhang, Y., Hughes, K.J., Zahm, S.H., Zhang, Y., Holford, T.R., Dai, L., Bai, Y., Han, X., Qin, Q., Lan, Q., et al. (2009). Genetic variations in xenobiotic metabolic pathway genes, personal hair dye use, and risk of non-Hodgkin lymphoma. Am. J. Epidemiol. *170*, 1222–1230.
25. Robinson, J., Barker, D.J., Georgiou, X., Cooper, M.A., Flicek, P., and Marsh, S.G.E. (2020). IPD-IMGT/HLA Database. Nucleic Acids Res. *48* (D1), D948–D955.
26. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.
27. Burek Kamenaric, M., Maskalan, M., Grubic, Z., Mikulic, M., Serventi Seiwerth, R., Durakovic, N., Vrhovac, R., Stingl Jankovic, K., and Zunec, R. (2017). HLA-DPB1 matching in unrelated hematopoietic stem cell transplantation program contributes to a higher incidence of disease relapse. Hum. Immunol. *78*, 665–671.
28. Orozco, G., Barton, A., Eyre, S., Ding, B., Worthington, J., Ke, X., and Thomson, W. (2011). HLA-DPB1-COL11A2 and three additional xMHC loci are independently associated with RA in a UK cohort. Genes Immun. *12*, 169–175.
29. Field, J., Browning, S.R., Johnson, L.J., Danoy, P., Varney, M.D., Tait, B.D., Gandhi, K.S., Charlesworth, J.C., Heard, R.N., Australia and New Zealand Multiple Sclerosis Genetics Consortium, et al. (2010). A polymorphism in the HLA-DPB1 gene is associated with susceptibility to multiple sclerosis. PLoS ONE *5*, e13454.
30. Okada, Y., Momozawa, Y., Ashikawa, K., Kanai, M., Matsuda, K., Kamatani, Y., Takahashi, A., and Kubo, M. (2015). Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. Nat. Genet. *47*, 798–802.
31. Park, B.L., Kim, T.H., Kim, J.H., Bae, J.S., Pasaje, C.F., Cheong, H.S., Kim, L.H., Park, J.S., Lee, H.S., Kim, M.S., et al. (2013). Genome-wide association study of aspirin-exacerbated respiratory disease in a Korean population. Hum. Genet. *132*, 313–321.
32. Crivello, P., Zito, L., Sizzano, F., Zino, E., Maiers, M., Mulder, A., Toffalori, C., Naldini, L., Ciceri, F., Vago, L., and Fleischhauer, K. (2015). The impact of amino acid variability on alloreactivity defines a functional distance predictive of permissive HLA-DPB1 mismatches in hematopoietic stem cell transplantation. Biol. Blood Marrow Transplant. *21*, 233–241.
33. Thomas, R., Thio, C.L., Apps, R., Qi, Y., Gao, X., Marti, D., Stein, J.L., Soderberg, K.A., Moody, M.A., Goedert, J.J., et al. (2012). A novel variant marking HLA-DP expression levels predicts recovery from hepatitis B virus infection. J. Virol. *86*, 6979–6985.
34. Petersdorf, E.W., Malkki, M., O'hUigin, C., Carrington, M., Gooley, T., Haagenson, M.D., Horowitz, M.M., Spellman, S.R., Wang, T., and Stevenson, P. (2015). High HLA-DP Expression and Graft-versus-Host Disease. N. Engl. J. Med. *373*, 599–609.
35. O'Brien, T.R., Kohaar, I., Pfeiffer, R.M., Maeder, D., Yeager, M., Schadt, E.E., and Prokunina-Olsson, L. (2011). Risk alleles for chronic hepatitis B are associated with decreased mRNA expression of HLA-DPA1 and HLA-DPB1 in normal human liver. Genes Immun. *12*, 428–433.

36. Matsuda, H., Hiramatsu, K., Akazawa, Y., Nosaka, T., Saito, Y., Ozaki, Y., Hayama, R., Takahashi, K., Naito, T., Ofuji, K., et al. (2018). Genetic polymorphism and decreased expression of HLA class II DP genes are associated with HBV reactivation in patients treated with immunomodulatory agents. J. Med. Virol. *90*, 712–720.
37. Horejsí, V., Chorváth, B., Polákova, K., Duraj, J., Sedlák, J., and Karpatová, M. (1988). Characterization of a new murine monoclonal antibody against human DP antigens. Tissue Antigens *32*, 6–11.
38. Watson, A.J., DeMars, R., Trowbridge, I.S., and Bach, F.H. (1983). Detection of a novel human class II HLA antigen. Nature *304*, 358–361.
39. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al.; GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.
40. Zheng, X., Shen, J., Cox, C., Wakefield, J.C., Ehm, M.G., Nelson, M.R., and Weir, B.S. (2014). HIBAG–HLA genotype imputation with attribute bagging. Pharmacogenomics J. *14*, 192–200.
41. Misra, M.K., Damotte, V., and Hollenbach, J.A. (2019). Structure-based selection of human metabolite binding P4 pocket of DRB1*15:01 and DRB1*15:03, with implications for multiple sclerosis. Genes Immun. *20*, 46–55.
42. Wang, Y., Zhou, W., Alter, B.P., Wang, T., Spellman, S.R., Haagenson, M., Yeager, M., Lee, S.J., Chanock, S.J., Savage, S.A., and Gadalla, S.M. (2018). Chromosomal Aberrations and Survival after Unrelated Donor Hematopoietic Stem Cell Transplant in Patients with Fanconi Anemia. Biol. Blood Marrow Transplant. *24*, 2003–2008.
43. Wan, J., Gao, Y., Zhao, X., Wu, Q., Fu, X., Shao, Y., Yang, H., Guan, M., Yu, B., and Zhang, W. (2011). The association between the copy-number variations of ZMAT4 and hematological malignancy. Hematology *16*, 20–23.
44. Kim, Y.H., Lee, J.R., and Hahn, M.J. (2018). Regulation of inflammatory gene expression in macrophages by epithelial-stromal interaction 1 (Epsti1). Biochem. Biophys. Res. Commun. *496*, 778–783.
45. Kojima, S. (2017). Why is the incidence of aplastic anemia higher in Asia? Expert Rev. Hematol. *10*, 277–279.

The American Journal of Human Genetics, Volume *106*

Supplemental Data

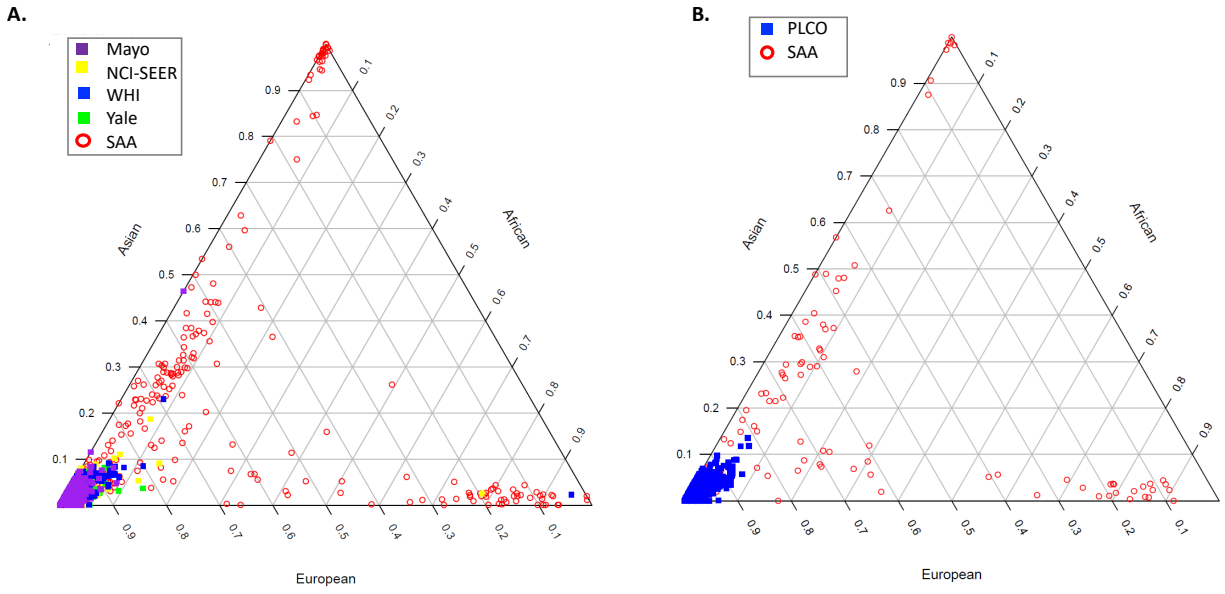
Genome-wide Association Study Identifies *HLA-DPB1* as a Significant Risk Factor for Severe Aplastic Anemia

Sharon A. Savage, Mathias Viard, Colm O'hUigin, Weiyin Zhou, Meredith Yeager, Shengchao Alfred Li, Tao Wang, Veron Ramsuran, Nicolas Vince, Aurelie Vogt, Belynda Hicks, Laurie Burdett, Charles Chung, Michael Dean, Kelvin C. de Andrade, Neal D. Freedman, Sonja I. Berndt, Nathaniel Rothman, Qing Lan, James R. Cerhan, Susan L. Slager, Yawei Zhang, Lauren R. Teras, Michael Haagenson, Stephen J. Chanock, Stephen R. Spellman, Youjin Wang, Amanda Willis, Medhat Askar, Stephanie J. Lee, Mary Carrington, and Shahinaz M. Gadalla

SUPPLEMENTARY FIGURES

Figure S1. Assessment of population admixture in severe aplastic anemia (SAA) cases and controls. Only cases and controls of >80% European ancestry were included in the association analyses **A.** Assessment of admixture in discovery set. **B.** Assessment of admixture in the validation set. **C.** Principal components analysis (PCA) was conducted using GLU struct.pca module based on the same set of population informative SNPs. The PCA was performed only on those subjects determined by GLU struct.admix to have greater than 80% of European ancestry. Abbreviations: Mayo, Mayo Clinic Case-Control Study of Non-Hodgkin Lymphoma and Chronic Lymphocytic Leukemia[1]; NCI-SEER, National Cancer Institute, Surveillance, Epidemiology, and End Results Non-Hodgkin Lymphoma Case-Control Study[2,3]; WHI, Women's Health Initiative[4]; Yale, Population-based Case-Control Study in Connecticut Women[5]; PLCO, The Prostate and Ovarian Cancer Prevention Trial

Supplementary Figure S1.



C.

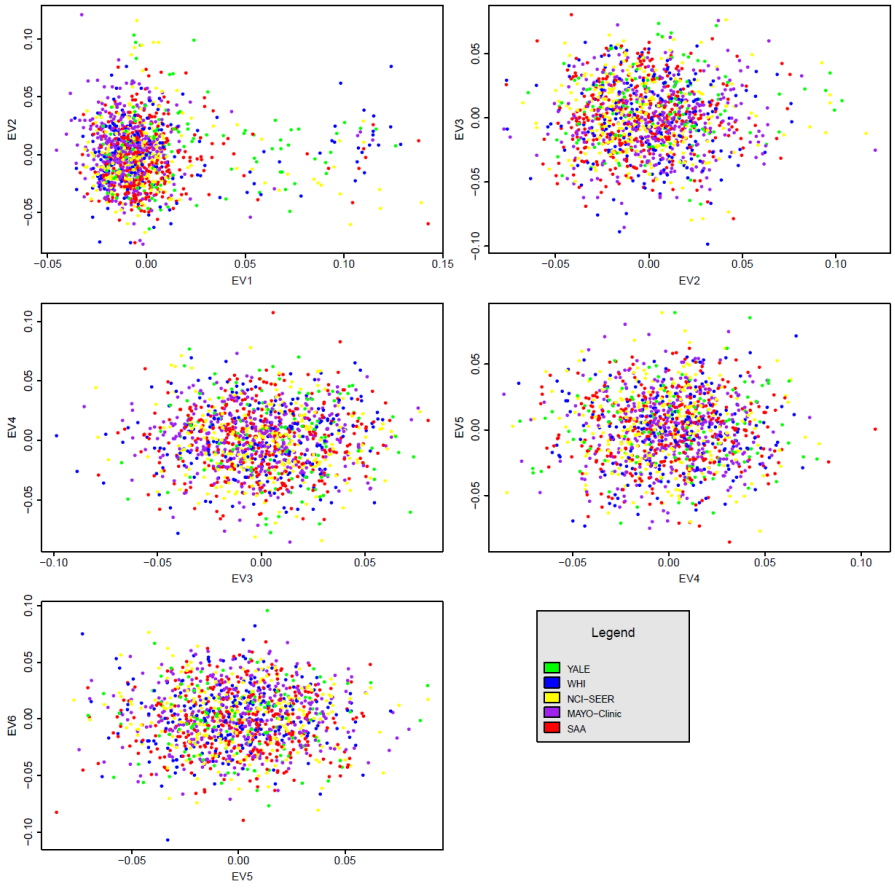

EV1
EV2
EV2
EV3
EV3
EV4
EV4
EV5
EV5
EV6
Legend
YALE
WHI
NCI−SEER
MAYO−Clinic
SAA

Figure S2. Chromosome 6 copy neutral loss of heterozygosity. **A.** location on chromosome 6p (dark blue bar indicates a second clone)**; B.** Two adjacent mosaic copy neutral loss of heterozygosity events (CNLOH) with different proportion of mosaicism for q arm of chromosome 6 covering HLA region. Each dot in the figure represents one SNP. Red dots represent B allele frequency (BAF, scale on the right side), while black dots show Log R ratio values (LRR, scale on the left side). Chr6-CNLOH event characterized by unchanged Log R ratio (mean of LRR within segment (blue line) = 0) and abnormal heterozygous BAF. The vertical gray lines indicate the breakpoint(s) of the event segment.

Figure S2.

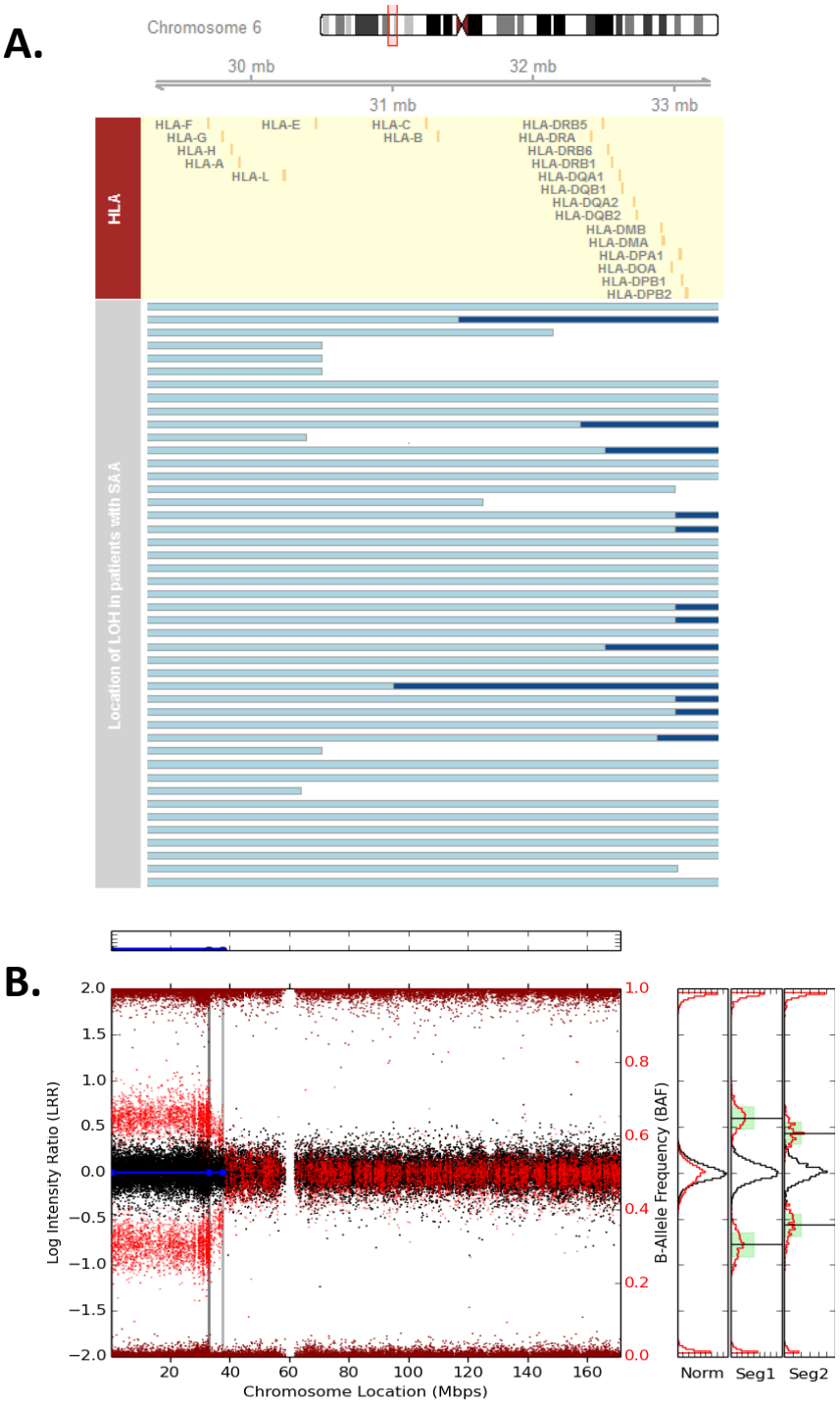
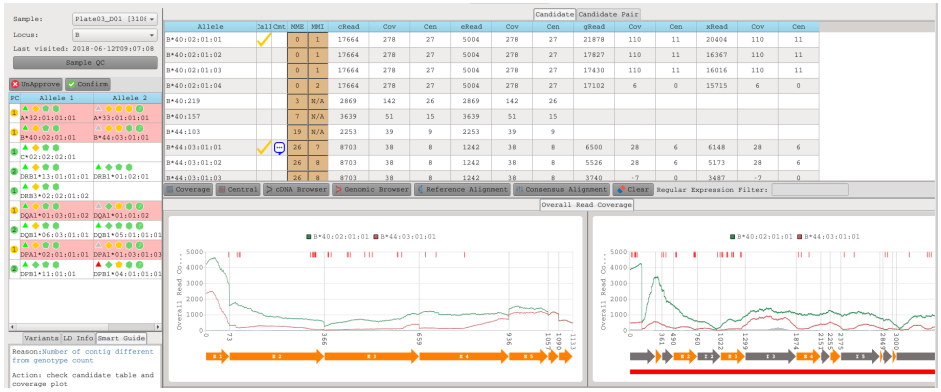

A.
Chromosome 6
30 mb
32 mb
31 mb
33 mb
HLA
HLA-F
HLA-G
HLA-H
HLA-A
HLA-E
HLA-L
HLA-C
HLA-B
HLA-DRB5
HLA-DRA
HLA-DRB6
HLA-DRB1
HLA-DQA1
HLA-DQB1
HLA-DQA2
HLA-DQB2
HLA-DMB
HLA-DMA
HLA-DPA1
HLA-DOA
HLA-DPB1
HLA-DPB2
Location of LOH in patients with SAA
B.
Log Intensity Ratio (LRR)
B-Allele Frequency (BAF)
Chromosome Location (Mbps)
Norm
Seg1
Seg2

Figure S3. Example of next generation sequencing results of HLA sequencing showing somatic loss of heterozygosity across multiple loci in the same subject. All loci show more than 5-fold difference in the number of reads among the 2 alleles at a locus that is consistent with mosaicism with a population of microdeletion in the MHC region.

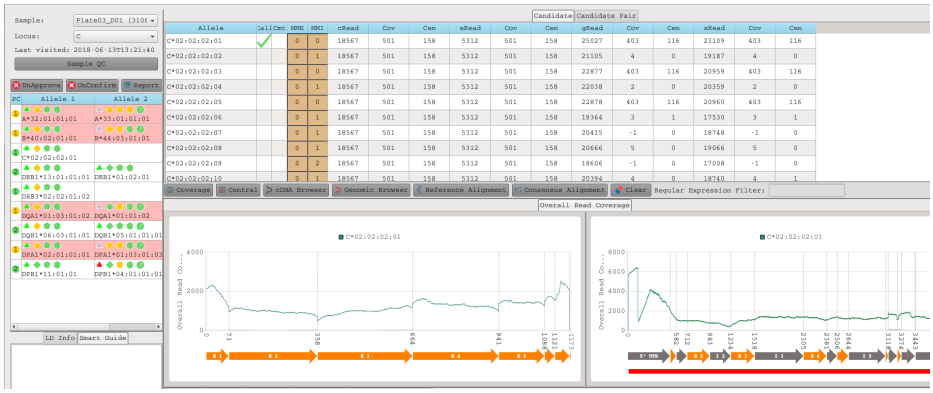
Figure S3.

A. Loss of heterozygosity

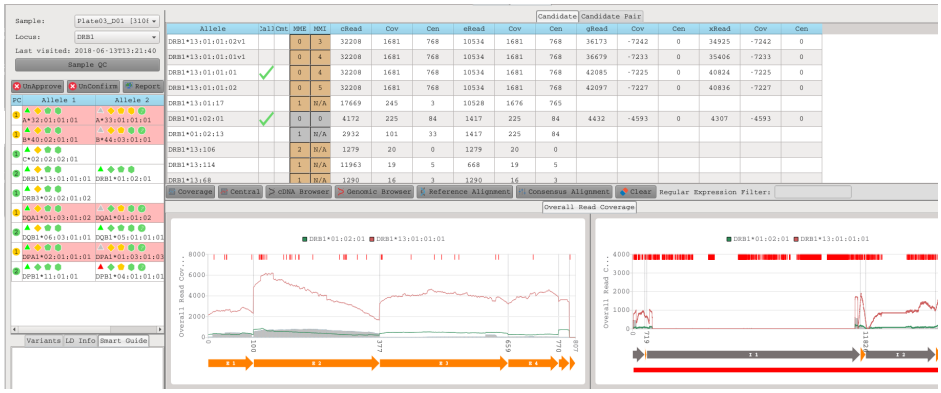
B. Loss of heterozygosity



C. Homozygous



D. Loss of heterozygosity



SUPPLEMENTARY METHODS

Genome-wide SNP Genotyping

Genotyping of SAA cases was conducted on the Illumina Infinium OmniExpress BeadChip array at the Cancer Genomics Research Laboratory (CGR) in the Division of Cancer Epidemiology and Genetics (DCEG) at the National Cancer Institute (NCI). Genotyping was done in two stages due to the timing of sample receipt. The first set, called the discovery set, consisted of 640 cases scanned on the Illumina Human OmniExpress-12v1-1_B and Illumina Human OmniExpress -24v1-0_a chip types. The second set, called the validation set, consisted of 255 cases scanned on the Illumina Infinium OmniExpress-24v1-2_A1 chip.

The controls were derived from previously scanned subjects drawn from two large cohort studies (The Prostate, Lung, Colon and Ovarian Cancer Prevention Trial (PLCO)[1] and the American Cancer Society Cancer Prevention Study II (CPSII)[2] scanned on the Illumina Omni 2.5M SNP microarray) as well as 4 other U.S.-based studies [Mayo Clinic Case-Control Study of Non-Hodgkin Lymphoma and Chronic Lymphocytic Leukemia (MAYO)[3], National Cancer Institute, Surveillance, Epidemiology, and End Results Non-Hodgkin Lymphoma (NHL) Case-Control Study (NCI-SEER)[4,5], Women's Health Initiative (WHI)[6], and the Population-based Case-Control Study in Connecticut Women (YALE)[7]] scanned on the Infinium OmniExpress chip. We selected 2,453 controls of European ancestry. The controls for the discovery set consisted of 1,396 subjects drawn from all 6 studies; the controls for the validation set consisted of 1059 subjects drawn from PLCO study.

Quality Control Assessment

We examined the distribution of the sample missing rate and sample mean heterozygosity. In the quality control analysis of discovery set, SNPs with less than a 90% completion rate were excluded from further analysis. Samples were excluded on the basis of (1) completion rates lower than 95% (n = 5 samples); (2) abnormal heterozygosity values of less than 25% or greater than 35% (n = 3) or; (3) abnormal X-chromosome heterozygosity (n = 1). After removing the low performing samples and loci described in above, the data were of high-quality at the sample level. Based on these data sets, we performed the assay concordance analysis and identified all 26 expected duplicates with the average SNP concordance rate at 99.998%. We also detected 15 inter-chip duplicates with concordance rate greater than 99.99%. There were no unexpected duplicate pairs detected. Genotypes for all subject pairs were computed for close relationships (first- and second-degree relatives) using GLU qc.ibds module (http://code.google.com/p/glu-genetics/) with an IBD0 threshold of 0.70. One pair of first degree relative was detected in the cases.

Using a set of 12,000 population informative SNPs that common to both the Illumina and Affymetrix commercial platforms and with low linkage disequilibrium (pair-wise $r^2 < 0.004$)[8] and data from HapMap build 27, we excluded 184 cases with less than 80% European ancestry[9], as determined using GLU strct.admix module. The HapMap I+II CEU, YRI, ASA (JPT+CHB) samples were used as the fixed reference populations[10]; a majority of these subjects represent 73 cases of mixed East Asian and European ancestry, 31 cases of Asian ancestries, 28 cases of African ancestry, and 26 cases of mixed European and African ancestries (Supplementary Figure S1). We also excluded 93 SAA cases reported by CIBMTR with a known inherited bone marrow failure syndrome from association analysis. The final association analysis for the discovery set included 359 cases and 1,396 controls of European ancestry. After quality

control filtering, data from 703,857 SNPs were available for the case subjects. The numbers of SNPs overlapping those of pooled controls were 688,067, respectively, and these SNPs were used in the downstream association analyses.

Similar quality control metrics were applied in the validation set; SNPs with less than a 90% completion rate were excluded from further analysis. Samples were excluded on the basis of (1) completion rates lower than 95% (n = 2 samples); (2) abnormal heterozygosity values of less than 25% or greater than 38% (n = 0); (3) abnormal X-chromosome heterozygosity (n = 0). There were nine expected duplicated cases in with the concordance rate of 99.76%. No unexpected duplicate pair detected. No first-degree relative pairs were detected. We excluded 78 case subjects with less than 80% European ancestry.

The association analysis for validation set included 175 cases and 1,059 controls of European ancestry. After quality control filtering, data from 702,117 SNPs were available for the case subjects. The numbers of SNPs overlapping controls were 663,976, respectively, and these SNPs were used in the downstream association analyses. The final association analysis for combined discovery and validation sets included 534 cases and 2,453 controls of European ancestry. After quality control filtering, data from 693,802 SNPs were available for the cases. The numbers of SNPs overlapping controls were 656,416, respectively, and these SNPs were used in the downstream association analyses. dbSNP build GRCh37/hg19 was used for annotations.

All of the controls were drawn from previous scans and passed similar quality control filtering. We excluded two control subjects that were identified as unexpected duplicates between the discovery and validation sets. We also excluded 90 control subjects with less than 80% European ancestry.

TaqMan Genotyping

Fourteen risk SNPs were validated by allele-specific TaqMan® genotyping (ThermoFisher) of a subset of 340 samples (Supplementary Table S3). TaqMan® assay validation and SNP genotyping was performed at the Cancer Genomics Research Facility (http://cgf.nci.nih.gov/). 5 ng of sample DNA, according to Quant-iT PicoGreen dsDNA quantitation (ThermoFisher), was transferred into 384-well plates (ThermoFisher) and dried down. Additionally, 5 ng of assay-specific controls, based on the validation results, were applied to pre-determined wells of the assay plates to guide analysis and overall quality and 5 ng of universal internal controls (NA07057 and NTC) were added to random locations of 384-well plates to provide a unique fingerprint for each plate for overall quality assurance.

Genotyping was performed using 5 uL reaction volumes consisting of: 2.5 uL of 2X KAPA Probe Fast MasterMix (Kapa Biosystems, Woburn, MA), 0.25 uL of 20X TaqMan® assay-specific mix of primers and probes, and 2.25 uL of MBG Water. Plates were heat-sealed using diamond optical seals (ThermoFisher) and PCR was performed using 9700 Thermal Cycler (ThermoFisher) with the following conditions: 95˚C hold for 3 min, 40 cycles of 95˚C for 3 sec and X˚C for 30 sec (where X is the optimized annealing temperature determined in validation for each assay), and 10˚C hold.

Endpoint reads were evaluated using the 7900HT Sequence Detection System (ThermoFisher) and cluster analysis was performed using SDS v2.2.2 software (ThermoFisher). Cluster analysis was performed using the Allelic Discrimination Plot, which is an X-Y scatter plot of FAM and VIC dyes, containing four distinct clusters which represent three possible genotypes: Allele 1

Homozygous (Y-Axis), Allele 2 Homozygous (X-Axis), and Allele 1/Allele 2 Heterozygous (Diagonal Axis) while the fourth cluster was at the origin and contains no amplification (NTCs). Analyzed data was imported into a LIMS where concordance of assay-specific genotyping controls and internal controls are confirmed.

The concordance rates of the 14 genotyped SNPs ranged from 99.7 to 100% (Supplementary Table S3).

GWAS Statistical Analyses

Associations between SNPs and risk of SAA were calculated using the multivariable logistic regression model from GLU assoc.logit1 module (http://code.google.com/p/glu-genetics/) assuming an additive genetic effect on the number of rare alleles presented in each genotype. For the discovery set, when included in the null model (baseline model), principal component analysis (PCA) identified three significant ($P < 0.05$) eigenvectors associated with the case/control status. The main effect model was adjusted for sex and these three eigenvectors to account for the imbalance between cases and controls. For the validation set, none of the 10 eigenvectors was significant ($P<0.05$); the main effect model was not adjusted for any covariates. For the combined discovery and validation sets, the main effect model was adjusted by sex and two eigenvectors, identified on the basis of significance ($P < 0.05$) observed in the null model of the combined sets. The estimated inflation factor λ for the test statistic from discovery, validation, and combined sets were 1.034, 1.032, and 1.047 respectively using estlambda() function with method=median from GenABEL package in R.

Identification of chromosome 6 copy neutral loss of heterozygosity (ch6CN-LOH) using SNP array data

Log R ratio (LRR) and B allele frequency (BAF) were used to assess copy number alteration derived from SNP array intensity data as previous described.[11] The Log R ratio (LRR) value is the normalized measure of total signal intensity and provides data on relative copy number and the B allele frequency derived from the ratio of allelic probe intensity is the proportion of hybridized sample that carries the B allele as designated by the Infinium Assay. The LRR and BAF values from qualified assays were re-normalized[12,13] and data analyzed using custom software pipelines that involved BAF Segmentation package (http://baseplugins.thep.lu.se/wiki/se.lu.onk.BAFsegmentation) to detect copy number change and copy neutral loss of heterozygosity. A deviation from heterozygotes band in BAF with LRR value of zero indicated a copy-neutral loss of heterozygosity (CN-LOH). To minimize false positives, the analysis was restricted to chromosomal abnormalities larger than 2 Mb. All potential events were plotted and visualized, and false positive calls were excluded from the analysis based on manual review of each plot.

Identification of chr6-LOH using HLA next generation sequencing

HLA typing was performed using MIA FORA NGS kit (Immucor, Inc., Norcross, GA). All samples were genotyped for 11 HLA loci, namely HLA-A, -B, -C, -DRB1, -DRB3, -DRB4, -DRB5, -DQA1, -DQB1, -DPA1, -DPB1. The coverage for HLA-A, -B, and -C included all exons and introns, at least 200 base pairs of the 5′ UTR and 100–1100 base pairs of the 3′ UTR; coverage for -DPA1, and -DQA1, included all exons and introns, at least 45 base pairs of the 3′ UTR and 25-190 base pairs of the 3′ UTR. Coverage for -DRB1 included all exons, introns 2–6, at least 440 base pairs of the 5′

UTR, 12 base pairs of the 3′ UTR, 275 base pairs of intron 1 adjacent to exon 1, and 210 base pairs of intron 1 adjacent to exon 2; coverage for -DRB3/4/5 included exons 2–6, introns 2–5, and 260 base pairs of intron 1 adjacent to exon 2; coverage for -DQB1 included exons 1–5 and introns 1–4; coverage for -DPB1 included exons 2–4 and introns 2–3. NGS sequencing was performed according to the manufacturer's instructions and described elsewhere [14]. Analysis was performed using MIA FORA software (Immucor, Inc.). Specimens with more than 5-fold difference in the number of reads corresponding to the 2 alleles at a locus that were consistent across all 11 tested loci and determined to be mosaic with a population of microdeletions in the MHC region. Example results from one sample are shown in Supplementary Figure S3A-D.

Imputation of HLA alleles

We performed imputation of HLA class II loci -DPB1* using the published imputation tool HIBAG.[15] The provided model trained for European ancestry was used for the imputation. The entire set of European ancestry cases and controls were imputed together. High resolution clinical typing available for 401 patients showed a concordance with the imputation of 92.7%.

Cell Surface Expression Analyses of HLA-DP

Cell surface expression levels of HLA-DP were measured using flow cytometry. A total of 175 healthy European American donors were recruited from the Frederick National Laboratory for Cancer Research Blood Donor Program. Samples were stained as follows; fresh blood was processed to generate peripheral blood mononuclear cells (PBMCs). Immediately after processing, one million PBMCs were stained with the following cocktail of

antibodies, anti-Human CD19 PE-Cyanine5 (ThermoFisher, Inc, Waltham, MA), anti-human CD3 Brilliant Violet 605™ (BioLegend, Inc., San Diego, CA) and anti-Human HLA-DP Monomorphic R-phycoerythrin (R-PE) clone B7/21 (Leinco Technologies, Inc., Fenton, MO). Cells were stained in the dark at 4 degrees ℃ for 20 minutes and washed with FACS wash followed by fixing with 300ul of BD Cytofix™ Fixation Buffer (BD Biosciences, Inc., San Jose, CA). Samples were run on the BD LSRFortessa™ and analyzed using FlowJo software (FlowJo, LLC, Ashland, OR).

Figure S3.

SUPPLEMENTARY REFERENCES

1. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet.* 2008;40(3):310-315.
2. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet.* 2009;41(9):986-990.
3. Cerhan JR, Fredericksen ZS, Wang AH, et al. Design and validity of a clinic-based case-control study on the molecular epidemiology of lymphoma. *Int J Mol Epidemiol Genet.* 2011;2(2):95-113.
4. Chatterjee N, Hartge P, Cerhan JR, et al. Risk of non-Hodgkin's lymphoma and family history of lymphatic, hematologic, and other cancers. *Cancer Epidemiol Biomarkers Prev.* 2004;13(9):1415-1421.
5. Wang SS, Cerhan JR, Hartge P, et al. Common genetic variants in proinflammatory and other immunoregulatory genes and risk for non-Hodgkin lymphoma. *Cancer Res.* 2006;66(19):9771-9780.
6. Anderson GL, Manson J, Wallace R, et al. Implementation of the Women's Health Initiative study design. *Ann Epidemiol.* 2003;13(9 Suppl):S5-17.
7. Zhang Y, Hughes KJ, Zahm SH, et al. Genetic variations in xenobiotic metabolic pathway genes, personal hair dye use, and risk of non-Hodgkin lymphoma. *Am J Epidemiol.* 2009;170(10):1222-1230.
8. Yu K, Wang Z, Li Q, et al. Population substructure and control selection in genome-wide association studies. *PLoS One.* 2008;3(7):e2551.

9. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet.* 2015;96(1):37-53.

10. International HapMap C, Frazer KA, Ballinger DG, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449(7164):851-861.

11. Peiffer DA, Le JM, Steemers FJ, et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* 2006;16(9):1136-1148.

12. Staaf J, Vallon-Christersson J, Lindgren D, et al. Normalization of Illumina Infinium whole-genome SNP data improves copy number estimates and allelic intensity ratios. *BMC bioinformatics.* 2008;9:409.

13. Diskin SJ, Li M, Hou C, et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* 2008;36(19):e126.

14. Ehrenberg PK, Geretz A, Sindhu RK, et al. High-throughput next-generation sequencing to genotype six classical HLA loci from 96 donors in a single MiSeq run. *Hla.* 2017;90(5):284-291.

15. Zheng X, Shen J, Cox C, et al. HIBAG--HLA genotype imputation with attribute bagging. *Pharmacogenomics J.* 2014;14(2):192-200.