

# Patterns of Human Leukocyte Antigen Class I and Class II Associations and Cancer



Zhiwei Liu<sup>1</sup>, Andriy Derkach<sup>1,2</sup>, Kelly J. Yu<sup>1</sup>, Meredith Yeager<sup>3,4</sup>, Yu-Sun Chang<sup>5,6</sup>, Chien-Jen Chen<sup>7</sup>, Ulf Gyllenstein<sup>8</sup>, Qing Lan<sup>1</sup>, Mei-Hsuan Lee<sup>9</sup>, James D. McKay<sup>10</sup>, Nathaniel Rothman<sup>1</sup>, Hwai-I Yang<sup>7,9,11</sup>, Allan Hildesheim<sup>1</sup>, and Ruth M. Pfeiffer<sup>1</sup>

## ABSTRACT

Human leukocyte antigen (HLA) gene variation is associated with risk of cancers, particularly those with infectious etiology or hematopoietic origin, given its role in immune presentation. Previous studies focused primarily on HLA allele/haplotype-specific associations. To answer whether associations are driven by HLA class I (essential for T-cell cytotoxicity) or class II (important for T-cell helper responses) genes, we analyzed GWAS from 24 case-control studies and consortia comprising 27 cancers (totaling >71,000 individuals). Associations for most cancers with infectious etiology or of hematopoietic origin were driven by multiple HLA regions, suggesting that both cytotoxic and helper T-cell responses are important. Notable exceptions were observed for nasopharyn-

geal carcinoma, an EBV-associated cancer, and CLL/SLL forms of non-Hodgkin lymphomas; these cancers were associated with HLA class I region only and HLA class II region only, implying the importance of cytotoxic T-cell responses for the former and CD4<sup>+</sup> T-cell helper responses for the latter. Our findings suggest that increased understanding of the pattern of HLA associations for individual cancers could lead to better insights into specific mechanisms involved in cancer pathogenesis.

**Significance:** GWAS of >71,000 individuals across 27 cancer types suggest that patterns of HLA Class I and Class II associations may provide etiologic insights for cancer.

## Introduction

The human leukocyte antigen (HLA) complex, located within the MHC on chromosome 6p21.3, plays a central role in presenting antigens to the immune system (1). The MHC region is characterized by an enrichment of key immune genes and can be divided into three subregions: the HLA class I region, which includes genes essential for the induction of direct CD8<sup>+</sup> T-cell cytotoxicity; the HLA class II region, including genes important for the generation of CD4<sup>+</sup> helper T-cell responses; and the class III region, which contains non-HLA genes implicated in inflammatory responses, leukocyte maturation

and the complement cascade among others. Risk of numerous cancers particularly those caused by infections and/or of hematopoietic origin, has been associated with classical HLA class I and class II genes, whereas evidence for associations for non-HLA class I and class II genes within the MHC is more limited (2).

Studies of HLA and cancer have largely focused on SNP- or allele/haplotype-specific associations (3–7). Recent studies that examined HLA and cancer associations more broadly, have shown that individuals with more diverse HLA profiles (i.e., higher heterozygosity) are at reduced risk of cancer, particularly those caused by infections or of hematopoietic origin, presumably due to increased immune surveillance (8, 9). What is not known is whether associations with cancer risk are driven by HLA class I genes that are essential in the direct induction of CD8<sup>+</sup> T-cell cytotoxicity and/or HLA class II genes, involved in the generation of CD4<sup>+</sup> T-cell helper responses. A difficulty in answering this question is the strong linkage disequilibrium (LD) patterns within the MHC region (10, 11).

In this report, we used individual-level data from 24 case-control studies/consortia comprising 27 cancers and cancer subtypes (totaling 37,083 cases and 34,607 unique controls) to formally evaluate the patterns of association between HLA and cancer, focusing on the comparison of the association between HLA class I genes and class II genes. We used single SNP analysis and a novel hierarchical statistical approach (Subset Testing and Analysis of Multiple Phenotypes, STAMP) to study heterogeneity of associations for a genomic region with different cancers and to identify the subset of cancers associated with that region.

## Materials and Methods

### Study populations

Detailed information on inclusion and exclusion criteria for studies included is provided in Supplementary Fig. S1. Briefly, we obtained individual-level GWAS data from case/control GWAS genotyped at or reported to the National Cancer Institute's Cancer Genomics Research Laboratory (in total 24,414 cancer cases and 15,054 controls) or with

<sup>1</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland. <sup>2</sup>Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York. <sup>3</sup>Division of Cancer Epidemiology and Genetics, Cancer Genomics Research Laboratory, National Cancer Institute, Bethesda, Maryland. <sup>4</sup>Frederick National Laboratory for Cancer Research, Frederick, Maryland. <sup>5</sup>Department of Microbiology and Immunology and Molecular Medicine Research Center, Chang Gung University, Taoyuan, Taiwan. <sup>6</sup>Department of Otolaryngology-Head and Neck Surgery, Chang Gung Memorial Hospital, Lin-Kou, Taiwan. <sup>7</sup>Genomics Research Center, Academia Sinica, Taipei, Taiwan. <sup>8</sup>Department of Immunology, Genetics and Pathology, Science for Life Laboratory Uppsala, Uppsala University, Uppsala, Sweden. <sup>9</sup>Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan. <sup>10</sup>International Agency for Research on Cancer, World Health Organization, Lyon, France. <sup>11</sup>Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan.

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Z. Liu and A. Derkach contributed equally as co-authors of this article.

**Corresponding Author:** Zhiwei Liu, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850. Phone: 240-276-6726; E-mail: zhiwei.liu@nih.gov

Cancer Res 2021;81:1148–52

doi: 10.1158/0008-5472.CAN-20-2292

©2020 American Association for Cancer Research.

approval from the Genotypes and Phenotypes database (dbGaP; in total 10,528 cancer cases and 11,461 controls). We also included data from 4 studies of infection-associated cancers [nasopharynx, liver (caused by hepatitis B or hepatitis C virus), and cervix] not available in the public domain but of interest for this analysis due to strong prior evidence for disease associations within the MHC region for these cancers (in total 2,141 cancer cases and 8,092 controls). Finally, we obtained comprehensive summary level results (not available via NHGRI catalog) from one study (Hodgkin lymphoma, HL) due to strong prior evidence for disease associations within the MHC region but for which sharing of individual-level data was not permitted (1,200 cancer cases and 6,417 controls; ref. 7). In total, we obtained individual-level data on 37,083 cases and 34,607 unique controls from 23 studies and summary level data from one study (1,200 cancer cases and 6,417 controls), totaling 38,283 cases and 41,024 controls. Original sample size received according to each study is provided in Supplementary Table S1. Details of the case and control ascertainment and study design can be found in each individual study, as briefly summarized in Supplementary Materials, Pages S1–S7. Studies were approved by the local institutional review boards and all participants provided written informed consent.

To understand whether our results are consistent with patterns observed in other studies, we downloaded publicly available data from the NHGRI GWAS catalog (in August 2019, <https://www.ebi.ac.uk/gwas/>), including SNPs spanning a region that was 20KB wider than the coordinates defining the MHC region of chromosome 6. We identified cancers associated with the MHC regions by using key words: “cancer,” “carcinoma,” “lymphoma,” “leukemia,” “neoplasm,” and “sarcoma.”

### Statistical analysis

The data analysis pipeline is summarized in Supplementary Fig. S2.

#### Single SNP analysis

After data pretreatment and quality control that were detailed in Supplementary Materials, Pages S7–S8, each of the 23 case–control studies with individual-level data was analyzed separately and then for each cancer, results for each single SNP across studies were combined using standard fixed-effects meta-analysis. For each study, log odds ratios (log-OR) and standard errors for case–control associations were calculated using logistic regression models with SNP genotype coded as 0, 1, and 2 and fitting with a trend, assuming an additive effect. Study-specific  $P$  values for the effect size estimates were obtained by Wald tests. Because we do not have permission to access SNPs outside HLA regions for most studies, we used the same adjustments in the logistic models that produced our  $P$  values as the original studies to control type I error. Covariates used for adjustment are age (in 5 categories; unless otherwise specified), sex (if available), and principal components (PC; the top 2 PCs for each cancer type and each study were shown in Supplementary Fig. S3). Details about the covariates can be found in the Supplementary Materials.

For SNP-based analysis, results were considered significant if  $P$  values were  $<2.5 \times 10^{-5}$  (multiple testing correction for an average of 2,000 SNPs in each analysis).

#### Region-specific analysis

For all studies, except HL, we conducted analyses based on individual-level data using the SKAT test with weights set to 1 (12). For HL (EBV-positive HL and EBV-negative HL), individual-level data were unavailable, thereby only results based on summary statistics were

computed. For this analysis, we used SNP-specific log-ORs and standard errors from logistic regression model, and study-specific MAFs and pairwise correlation between SNPs estimated from cases and controls combined. Asymptotically both these region-based analyses are equivalent provided if there is low correlation between SNPs and covariates used in the logistic regression model. Details on these two approaches were shown in Supplementary Materials, Pages S8–S9.

To assess whether signal was restricted to a specific region or due to LD across regions, we conducted conditional region-specific analysis as follows. From each group of SNPs studied in the region-based analysis, we selected the 10 SNPs with the lowest  $P$  values from single SNP analysis or all SNPs with the  $P$  values below 0.05/100 (i.e., larger set of SNPs) to represent association evidence of a region. Conditional analysis was conducted exactly like the analysis described in Supplementary Section S2A and S2B in Supplementary Materials, Page S8, but logistic models were additionally adjusted for the top SNPs from the conditioned region (see Supplementary Section S2 in Supplementary Materials, Pages S8–S9). For the conditional analysis with summary statistics, used only for the HL studies, we estimated joint effects (i.e., equivalent to fitting a logistic model with additional SNP adjustment) of the LD-pruned SNPs from the region of interest and the top SNPs from the conditioning region, as described in Supplementary Section S3 in Supplementary Materials, Page S9.

#### Posterior probabilities

To accommodate LD and different study sizes across cancers, we calculated posterior probabilities for a cancer to be associated with the genomic region based on a mixture model—subset testing and analysis of multiple phenotypes (STAMP)—that allows heterogeneity of associations between all cancers included (13), although modest sample sizes for some studies still are a notable limitation of our work. As all studies had more than 50 SNPs measured in each HLA region the normal approximation assumption required by STAMP is likely to hold. We applied this model to the results from both, the analysis with summary statistics and individual-level data. We fit STAMP to the cancer-specific values of  $T_{\text{meta},c}$  and  $T_{\text{meta},c}^C$  calculated from the summary statistic (see Supplementary Sections S2 and S3 in Supplementary Materials, Pages S8–S9).

We also fit STAMP to  $T_{\text{meta}}$  and  $T_{\text{meta}}^C$  calculated from individual-level data if available and HL using summary statistics. For each study with individual-level data, we estimated joint effects of  $p_s$  SNPs in the region by

$$\hat{\xi}_s = (\hat{\xi}_{s1}, \dots, \hat{\xi}_{sp_s})' = \text{Cov}(S_s)^{-1} S_s \text{ and } \text{Cov}(\hat{\xi}_s) = \text{Cov}(S_s)^{-1},$$

where  $S_s$  is the score vector calculated from the basic logistic model described in Supplementary Section S2 in Supplementary Materials, Page S8 (14, 15). Cancer-specific test statistics are recalculated as  $T_s^S = (\text{Cov}(\hat{\xi}_s)^{-1} \hat{\xi}_s)' (\text{Cov}(\hat{\xi}_s)^{-1} \hat{\xi}_s)$ , which is asymptotically equivalent to  $T_s$ . Then, STAMP was applied to the corresponding test statistics  $T_{\text{meta},c}^S = \sum_{s \in \text{Cancer}} T_s^S$ . For the two HL cancers  $T_{\text{meta},c}$  was calculated

on the basis of summary statistics. For conditional analysis,  $T_s^S$  was estimated by using score vectors calculated from logistic model that incorporates top SNPs from another region.

For region-based analysis, results were considered significant if the STAMP posterior probability of association was  $pp_{\text{STAMP}} > 0.50$ . A  $pp_{\text{STAMP}}$  between 0.5 and 0.75 was considered a moderate association, and  $pp_{\text{STAMP}} > 0.75$  a strong association.

### Availability of data and materials

Genotyping data have been deposited at the database of Genotypes and Phenotypes (dbGaP), under accession numbers phs000863.v1.p1, phs000351.v1.p1, phs000838.v1.p1, phs000346.v2.p2, phs000801.v2.p1, phs000093.v2.p2, phs000336.v1.p1, phs000361.v1.p1, phs001286.v1.p1, phs001303.v1.p1, phs000147.v3.p1, phs000206.v5.p3, phs000207.v1.p1, phs001210.v1.p1, phs000882.v1.p1, phs001736.v1.p1, phs000147.v3.p1, phs000383.v1.p1, phs001315.v1.p1, phs001202.v1.p1, and phs000753.v1.p1. Other datasets used and/or analyzed are available from the corresponding author on reasonable request.

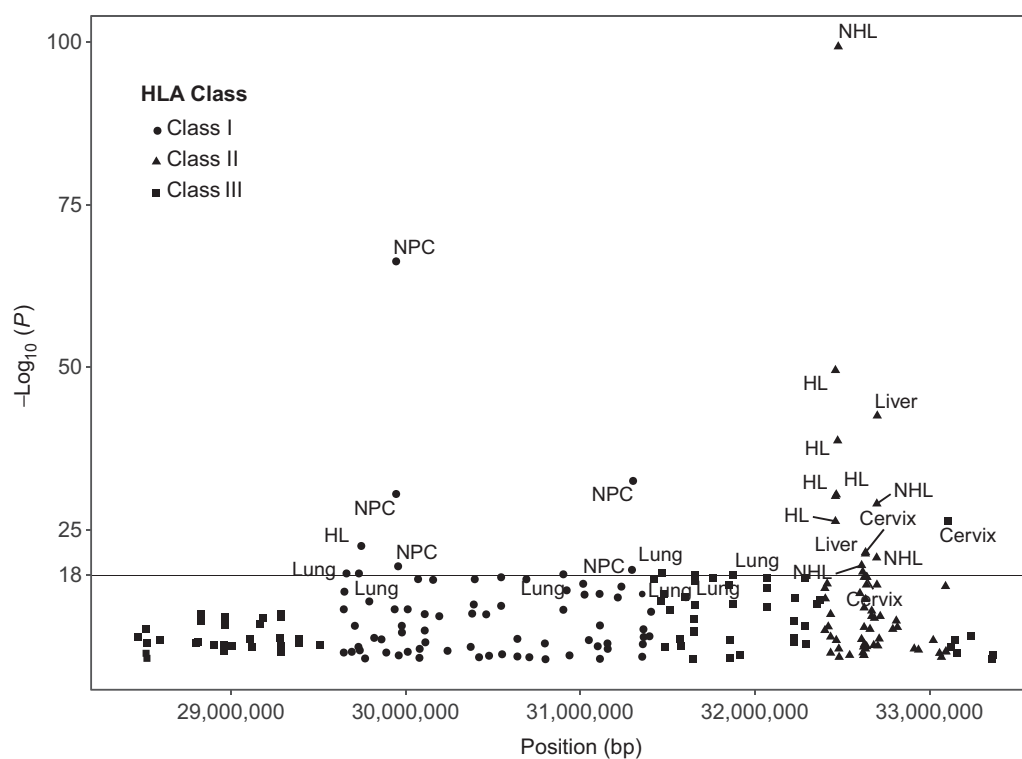
## Results

After quality control, we included 37,200 cases and 40,552 controls in the analysis (Supplementary Table S1). We first evaluated associations for SNPs within the MHC region. The most significant associations are summarized in Supplementary Table S2. Quantile–quantile (QQ) plots and 95% confidence bands of observed versus expected  $P$  values using SNPs after LD pruning within the HLA regions were shown in Supplementary Fig. S4. The QQ plots did not diverge significantly from the 95% confidence bands around the 45-degree line for cancers for which we found no associations with SNPs within HLA region (cancers of bladder, brain, colon, endometrium, esophagus, kidney, ovary, pancreas, prostate, stomach, and testis). For cancers associated with the HLA region, the QQ plots deviate, as expected, from the 45-degree line.

For cancers without evidence for heterogeneity of HLA associations by different subtypes as noted by previous studies, we did not conduct detailed stratified analysis by subtype. For breast cancer, our initial

analysis stratified by ER status did not indicate heterogeneity by subtype (Supplementary Fig. S5), which confirms previous findings (16, 17). For non-Hodgkin lymphoma (NHL) and lung cancer, we present results overall and, given previous evidence for heterogeneity of HLA associations, by histological subtype (18–20). Associations that were significant using our pre-defined statistical threshold were observed for all infection-related cancers, most hematopoietic cancers, breast cancer, and squamous cell carcinomas of the lung. Reassuringly, our results are largely consistent with patterns observed using publicly available summary data from the NHGRI GWAS catalog (Fig. 1, downloaded August 2019, <https://www.ebi.ac.uk/gwas/>), both confirming that HLA associations are most evident for cancers caused by infections and/or those of hematopoietic origin. Specifically, Class I associations were observed for NPC and Hodgkin lymphoma, both cancers linked to EBV infection. Class II associations were seen for cancers of liver and cervix, NHL, and Hodgkin lymphoma, all cancers that are known or suspected to have infectious etiologies. Strong associations were also found for lung cancer, but several SNPs involved were outside the defined HLA Class I and HLA Class II regions, suggesting a possible involvement of HLA class III genes in the MHC region. These findings suggest that our curated studies with individual-level data available for analyses are broadly representative.

We next performed a region-based analysis to better understand the independent effect of our pre-defined regions of the MHC (i.e., HLA class I region, HLA class II region, and HLA class III region), by using SNP-set (Sequence) Kernel Association Test (SKAT, results are shown in Supplementary Table S3) and estimating the posterior probability  $pp_{STAMP}$  (range 0 to 1) of association for each cancer and each MHC



**Figure 1.**

Associations between SNPs within the MHC region and cancer reported in the NHGRI GWAS catalog. Data were downloaded in August 2019 (<https://www.ebi.ac.uk/gwas/>).

**Table 1.** Posterior probability for associations with HLA Class I, HLA Class II, and HLA Class III regions of the MHC, by cancer.

Cancer categories	Cancer site	Analysis without conditioning			Analysis with conditioning		
		Class I	Class II	Class III	Class I <sup>a</sup>	Class II <sup>b</sup>	Class III <sup>c</sup>
Infection-related cancers	Nasopharynx	1.00	0.02	1.00	1.00	0.007	0.01
	EBV-positive HL	1.00	1.00	1.00	1.00	1.00	1.00
	Cervix	1.00	1.00	1.00	1.00	1.00	0.001
	Head and Neck	1.00	1.00	1.00	0.03	0.004	0.0005
	HBV-HCC	1.00	1.00	0.77	1.00	1.00	0.002
	HCV-HCC	0.13	0.84	1.00	0.07	0.02	0.008
Hematopoietic cancers	EBV-negative HL	1.00	1.00	1.00	0.62	1.00	1.00
	NHL	0.41	0.004	0.03	—	—	—
	CLL/SLL	0.11	0.67	0.31	—	—	—
	Other	1.00	0.47	0.77	0.51	0.008	0.008
Solid cancers	Breast	0.16	0.008	0.01	—	—	—
	Lung	0.015	0.002	0.01	—	—	—
	Squamous cell carcinoma	0.02	0.02	0.17	—	—	—

Abbreviations: CLL/SLL, chronic lymphocytic leukemia and small lymphocytic lymphoma; EBV, Epstein-Barr virus; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; HL, Hodgkin lymphoma; HLA, human leukocyte antigen; MHC, major histocompatibility complex; NHL, non-Hodgkin lymphoma.

<sup>a</sup>Conditional on top SNPs in HLA class II region.

<sup>b</sup>Conditional on top SNPs in HLA class I region.

<sup>c</sup>Conditional on top SNPs in HLA class I and class II regions.

region detailed in Supplementary Materials, Pages S8–S9. A high  $pp_{STAMP}$  indicates a likely association with the specific MHC region (Table 1, analysis without conditioning columns). In this analysis, despite evidence for the presence of an overall HLA association, we did not observe evidence for a significant association within any of the three regions of the HLA evaluated and cancers of the breast or lung ( $pp_{STAMP} < 0.50$  for each of the three region). Therefore, we do not report further results for these cancers. We observed a suggestive association between HLA class I region and NHL overall; however, due to heterogeneity of HLA associations in previous reports, we focused on NHL subtypes and observed evidence for modest association with HLA class II region (but not HLA class I or III regions) for chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL,  $pp_{STAMP} = 0.67$  for HLA class II). We also observed association within multiple regions of the MHC for rare subtypes of NHL ( $pp_{STAMP} > 0.50$ ). For all other cancers with significantly associated SNPs in the MHC region, there was also evidence for associations within multiple regions of the MHC ( $pp_{STAMP} > 0.50$ ). We conducted additional analyses for each MHC region conditional on the other regions to determine whether the observed signals were independent or explained by associations in other MHC regions (Table 1, analysis with conditioning columns). In most of cancers evaluated, we either observed evidence for an independent association within multiple regions [i.e., SNPs in >1 region significant in conditional analysis, including Epstein-Barr virus (EBV)-positive Hodgkin lymphoma, cervical cancer, hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC), and EBV-negative HL] or could not determine whether the observed associations were driven by a specific or multiple MHC regions [head and neck cancers and hepatitis C virus (HCV)-related HCC]. The exceptions were nasopharyngeal carcinoma (NPC,  $pp_{STAMP} = 1.00$ ) and rare subtypes of NHL ( $pp_{STAMP} = 0.51$ ), which showed evidence for independent associations with the HLA class I region (but not HLA class II or III regions).

## Discussion

In this large-scale study, we found that both NPC and rare subtypes of NHL are associated with HLA class I region (with strongest evidence

for NPC) in the absence of a statistically independent effect elsewhere in the MHC. It is hard to generate biological insights for rare subtypes of NHL due to the mix of different disease entities in this group. For NPC, however, our findings suggest that CD8<sup>+</sup> T-cell cytotoxicity regulated by HLA class I is important for its etiology. CLL/SLL was associated with HLA class II but no other MHC region, suggesting CD4<sup>+</sup> T-cell helper responses regulated by HLA class II is important for this tumor type. In contrast, for most other cancers evaluated, where evidence for both HLA class I and II associations were noted, both T-cell cytotoxicity and T-cell helper responses appear to be important.

Our study strengths include the large sample size available from multiple cohorts. We comprehensively evaluated associations with several cancers using new statistical approach largely accounting for study heterogeneity and high LD within the MHC region. Compared with other approaches (21–23), STAMP has several advantages. It borrows strength across studies of differing sample sizes, accommodates between studies heterogeneity, and allows use of summary statistics, which is computationally less burdensome. Nonetheless, our study has a few limitations. First, due to a modest sample size for some individual cancer types, we have a limited power to identify modest disease association. Second, infection status is not clear for some types of cancers; therefore, the association observed within the MHC region might be affected by the proportion of cancers caused by infections in the individual GWAS study. Third, despite the new statistical approach used in our analysis, we still encountered challenges in the selection of appropriate methods for conditional analysis that minimized false-positive results while accounting for strong LD in the MHC region.

Despite the limitations noted above, our findings are an important step toward dissecting the region of the MHC that drives associations with cancer. Biologically, HLA class I molecules present peptides to CD8<sup>+</sup> cytotoxic T-lymphocytes, activating directly killing function of CD8<sup>+</sup> T cells. On the other hand, MHC class II molecules are involved in adaptive immune responses more indirectly by affecting responses of CD4<sup>+</sup> T-helper cells. Specifically, our findings suggest that for most HLA-associated cancers variation in both cytotoxic and helper T-cell responses are

etiologically relevant whereas for a few cancers such as EBV-associated NPC, variation in one particular class of T-cell responses predominates.

### Authors' Disclosures

No disclosures were reported.

### Disclaimer

The funders had no direct role in the design of the study, the collection, analysis, or interpretation of the data, the writing of the article, or the decision to submit the manuscript for publication.

### Authors' Contributions

**Z. Liu:** Data curation, formal analysis, investigation, visualization, methodology, writing-original draft, writing-review and editing. **A. Derkach:** Data curation, formal analysis, investigation, methodology, writing-original draft, writing-review and editing. **K.J. Yu:** Data curation, project administration, writing-review and editing. **M. Yeager:** Resources, data curation, project administration, writing-review and editing. **Y.-S. Chang:** Resources, data curation. **C.-J. Chen:** Resources, data curation.

### References

- Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* 1999;401:921–3.
- Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;42:D1001–6.
- Hildesheim A, Apple RJ, Chen CJ, Wang SS, Cheng YJ, Klitz W, et al. Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. *J Natl Cancer Inst* 2002;94:1780–9.
- Johnson PC, McAulay KA, Montgomery D, Lake A, Shield L, Gallagher A, et al. Modeling HLA associations with EBV-positive and -negative Hodgkin lymphoma suggests distinct mechanisms in disease pathogenesis. *Int J Cancer* 2015; 137:1066–75.
- Magnusson PKE, Enroth H, Eriksson I, Held M, Nyren O, Engstrand L, et al. Gastric cancer and human leukocyte antigen: distinct DQ and DR alleles are associated with development of gastric cancer and infection by *Helicobacter pylori*. *Cancer Res* 2001;61:2684–9.
- Safaeian M, Johnson LG, Yu K, Wang SS, Gravitt PE, Hansen JA, et al. Human leukocyte antigen class I and II alleles and cervical adenocarcinoma. *Front Oncol* 2014;4:119.
- Urayama KY, Jarrett RF, Hjalgrim H, Diepstra A, Kamatani Y, Chabrier A, et al. Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. *J Natl Cancer Inst* 2012;104:240–53.
- Wang SS, Abdou AM, Morton LM, Thomas R, Cerhan JR, Gao X, et al. Human leukocyte antigen class I and II alleles in non-Hodgkin lymphoma etiology. *Blood* 2010;115:4820–3.
- Wang SS, Carrington M, Berndt SI, Slager SL, Bracci PM, Voutsinas J, et al. HLA class I and II diversity contributes to the etiologic heterogeneity of non-Hodgkin lymphoma subtypes. *Cancer Res* 2018;78:4086–96.
- Robinson J, Soormally AR, Hayhurst JD, Marsh SGE. The IPD-IMGT/HLA database—new developments in reporting HLA variation. *Hum Immunol* 2016; 77:233–7.
- Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. *Nat Rev Immunol* 2018;18:325–39.
- Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011;89:82–93.
- Derkach A, Pfeiffer RM. Subset testing and analysis of multiple phenotypes. *Genet Epidemiol* 2019;43:492–505.
- Lee S, Teslovich TM, Boehnke M, Lin X. General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet* 2013;93:42–53.
- Hu YJ, Berndt SI, Gustafsson S, Ganna A, Genetic Investigation of ATC, Hirschhorn J, et al. Meta-analysis of gene-level associations for rare variants based on single-variant statistics. *Am J Hum Genet* 2013;93: 236–48.
- Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 2015;47: 373–80.
- Michailidou K, Lindstrom S, Dennis J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017;551:92–4.
- Shiraishi K, Okada Y, Takahashi A, Kamatani Y, Momozawa Y, Ashikawa K, et al. Association of variations in HLA class II and other loci with susceptibility to EGFR-mutated lung adenocarcinoma. *Nat Commun* 2016;7:12451.
- McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet* 2017;49:1126–32.
- Ferreiro-Iglesias A, Lesseur C, McKay J, Hung RJ, Han Y, Zong X, et al. Fine mapping of MHC region in lung cancer highlights independent susceptibility loci by ethnicity. *Nat Commun* 2018;9:3927.
- Bhattacharjee S, Rajaraman P, Jacobs KB, Wheeler WA, Melin BS, Hartge P, et al. A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. *Am J Hum Genet* 2012;90:821–35.
- Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am J Hum Genet* 2015;96:21–36.
- Majumdar A, Haldar T, Bhattacharya S, Witte JS. An efficient Bayesian meta-analysis approach for studying cross-phenotype genetic associations. *PLoS Genet* 2018;14:e1007139.

**U. Gyllensten:** Resources, data curation. **Q. Lan:** Resources, data curation, writing-review and editing. **M.-H. Lee:** Resources, data curation, writing-review and editing. **J.D. McKay:** Resources, data curation. **N. Rothman:** Resources, data curation, writing-review and editing. **H.-I. Yang:** Resources, data curation, writing-review and editing. **A. Hildesheim:** Conceptualization, supervision, funding acquisition, investigation, methodology, project administration, writing-review and editing. **R.M. Pfeiffer:** Conceptualization, resources, formal analysis, supervision, funding acquisition, investigation, methodology, writing-original draft, project administration, writing-review and editing.

### Acknowledgments

This work was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, DHHS, Bethesda, MD.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 9, 2020; revised October 26, 2020; accepted November 30, 2020; published first December 3, 2020.