Human leukocyte antigens and Epstein–Barr virus-associated nasopharyngeal carcinoma: old associations offer new clues into the role of immunity in infection-associated cancers

Wen-Hui Su[1,2]*, Allan Hildesheim[3] and Yu-Sun Chang[2]

[1] Department of Biomedical Sciences, Graduate Institute of Biomedical Sciences, College of Medicine, Chang Gung University, Taoyuan, Taiwan
[2] Chang Gung Molecular Medicine Research Center, Chang Gung University, Taoyuan, Taiwan
[3] Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

Nasopharyngeal carcinoma (NPC) is an Epstein–Barr virus (EBV) associated tumor. In addition to EBV, host genetic factors are believed to be important determinants of NPC risk. Of all genes studies to date, human leukocyte antigen (HLA) genes have shown the most consistent evidence for association with NPC, both from candidate-gene studies and genome-wide association studies (GWAS). In this report we summarize results from recent studies that evaluated the association between HLA and NPC, and discuss whether findings reflect direct causal associations for HLA genes and/or indirect associations that mark causal associations with other genes in the gene-dense major histocompatibility (MHC) region where HLA resides. We also compare GWAS results across cancer sites for which strong hits in the MHC region were observed to generate new hypotheses regarding the role of HLA genes in the development of EBV-associated cancers such as NPC. Of note, we report that MHC associations for EBV-associated cancers (NPC, EBV+ Hodgkin lymphoma) are driven by HLA class I genes. In contrast, MHC associations for other viral-associated cancers (cervical cancer, hepatocellular carcinoma) or other hematopoetic cancers (EBV− Hodgkin lymphoma, leukemia, non-Hodgkin lymphomas) are driven by HLA class II genes, and those for other solid tumors with less clear links to infections (lung, testicular, prostate cancers) are driven by non-HLA genes in the MHC region. Future studies should aim to better understand these patterns.

Keywords: genome-wide association study, nasopharyngeal carcinoma, HLA antigens, EBV, infection associated cancers

INTRODUCTION

Nasopharyngeal carcinoma (NPC) is an epithelial malignancy that is common in regions of Southeast Asia and the Mediterranean Basin. The age-adjusted incidence rate of this tumor in Southern China, for example, is 25–30 cases per 100,000 person years, which is approximately 50 times higher than what is observed in the Western world (1–3). Infection with Epstein–Barr virus (EBV) is believed to be a near necessary factor for the development of NPC (3, 4). EBV is a ubiquitous infection that typically occurs in early life, establishes lifelong latent infection in B-lymphocytes, and periodically reactivates in the epithelial compartment of the pharynx (5). Since EBV infection is common and NPC is rare, it is widely agreed that other environmental and genetic factors are important determinants of NPC risk. With respect to host genetic factors associated with NPC, human leukocyte antigens (HLA) have been proposed to be important, given their central role in presentation of viral antigens to the immune system (6).

The HLA genes comprise a family of highly polymorphic genes located within the major histocompatibility complex (MHC)

on chromosome 6p21.3. An association between HLA genes and NPC was first proposed by Simons and colleagues (7). Since that initial report, the association between HLA genes and NPC has been confirmed in over 100 candidate-gene-based association studies (3, 8, 9). More recently, three independent genome-wide association studies (GWAS) of NPC consistently identified SNPs within the MHC region (where HLA genes are located) as having the strongest evidence for association with NPC (10–12).
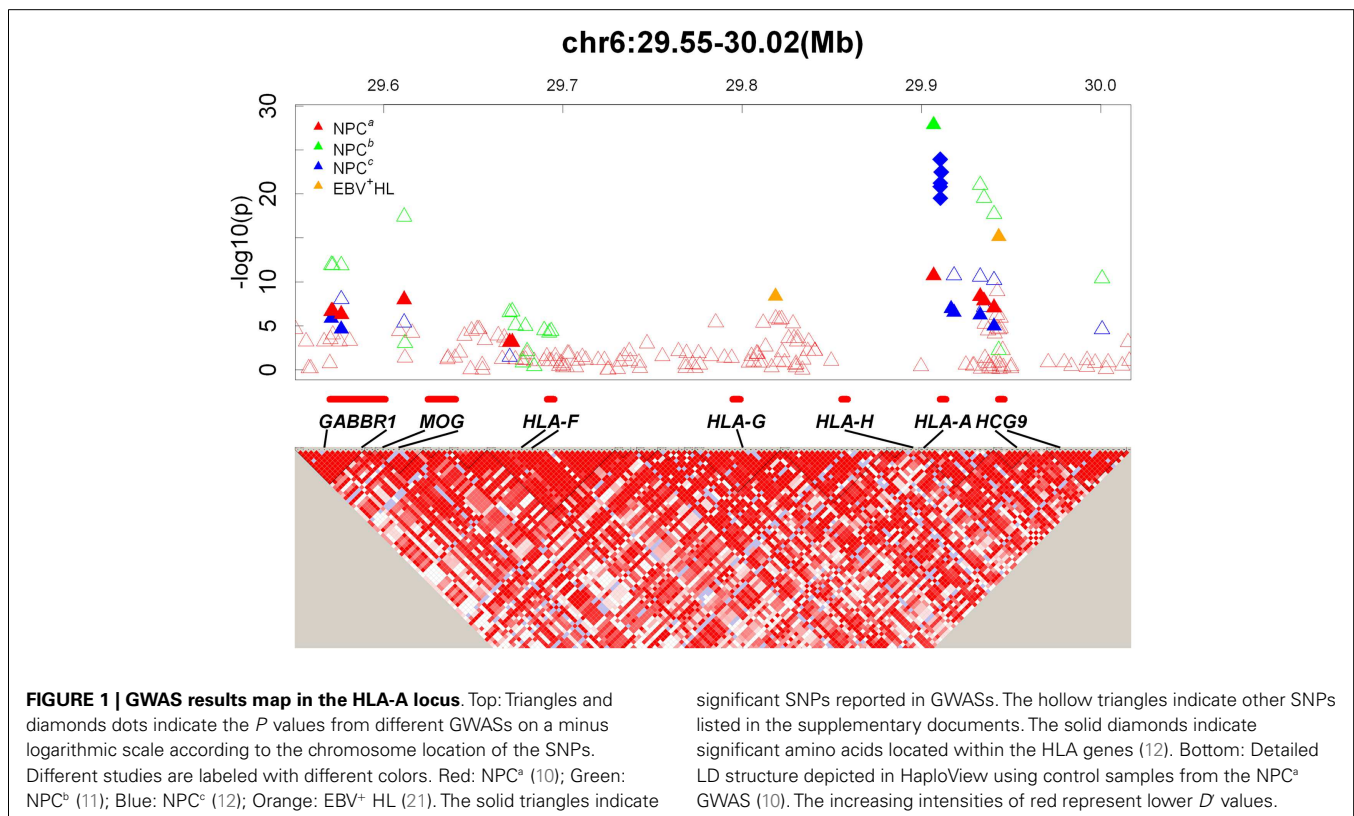
In this review, we summarize recent findings regarding the association between HLA genes and NPC susceptibility. We then discuss whether the associations observed in the gene-rich MHC region, where strong linkage disequilibrium (LD) patterns are observed, are driven only by HLA genes or whether other non-HLA genes in the region might also be involved. Finally, we compare GWAS results across cancer sites for which strong hits in the MHC region were observed, to generate new hypotheses regarding the role of HLA genes in the development of EBV-associated cancers such as NPC.

HUMAN LEUKOCYTE ANTIGEN ASSOCIATIONS WITH NASOPHARYNGEAL CARCINOMA

Human leukocyte antigen genes are located within the MHC region on chromosome 6p21. The MHC region is a gene-dense region (>150 genes) that also exhibits some of the strongest LD patterns within the human genome (13). These features of the MHC region make studies of HLA-cancer association particularly challenging because it is often difficult to determine whether reported associations are causal and/or reflect LD with other genes in this region. Nonetheless, there is a strong biological *a priori* for a causal association between HLA genes and NPC, given that HLA molecules are central to the presentation of viral peptides to cytotoxic and helper immune cells, and that infection with EBV is ubiquitously associated with the development of NPC. Of relevance to this review, there are three classical HLA class I genes, namely, *HLA-A*, *HLA-B*, and *HLA-C*, and three clusters of classical HLA class II genes, namely, *HLA-DR*, *HLA-DP*, and *HLA-DQ* that have been evaluated for their association with NPC. Both HLA class I and class II molecules can bind peptides through their peptide recognition groove and present peptides to T cells. However, HLA class I and II genes differ with respect to the types of cells in which they are expressed and the types of immune cells they regulate. HLA class I molecules are expressed on most nucleated cells and typically present foreign peptides to cytotoxic T cells. HLA class II molecules have a more restricted expression pattern being normally expressed on B-lymphocytes and antigen-presenting cells that typically present foreign peptides to helper T cells (13).

As alluded to in the Introduction, the association between HLA and NPC was first proposed in 1974 and since that time over 100 candidate-gene association studies have consistently reported associations between HLA alleles/haplotypes and NPC (3, 8, 9). More specifically, studies have reproducibly reported associations with NPC for the following HLA class I alleles: *HLA-A*0207* (risk allele in LD with *HLA-B*4601*), *HLA-A*1101* (protective allele in LD with *HLA-B*13*), and *HLA-B*5801* (risk allele in LD with *HLA-A*3303*) (9, 14). Linkage analyses also suggested that HLA-A and HLA-B genes are associated with the development of NPC (15–17). Other HLA genes, including *HLA-C* and HLA class II genes, showed less consistent findings across studies (9).

With the advent of technologies to interrogate the entire genome to better understand the genetic architecture of complex diseases (18), it has become possible to evaluate HLA-NPC associations in the context of genetic associations in other regions of the genome. To date, four NPC GWAS have been reported (10–12, 19). Of these four studies, three reported the strongest hits in the MHC region, where HLA genes reside (10–12) and only one GWAS (the smallest with a total of 110 NPC cases and 260 controls in the discovery phase) did not report genome-wide significant hits in the MHC region (19). The largest NPC GWAS to date from Guangdong, China (3,477 and 6,570 individuals in discovery and replication phases, respectively) reported their strongest finding to be located in HLA-A (rs2860580; $P_{\text{GWAS}} = 1.34 \times 10^{-28}$, $P_{\text{combined}} = 4.88 \times 10^{-67}$, OR = 0.58; **Figure 1**, NPC[b]) (11). The strongest reported associations for the Taiwan (562 and 2,275 individuals in discovery and replication phases, respectively) (10) and Guangxi,

FIGURE 1 | GWAS results map in the HLA-A locus. Top: Triangles and diamonds dots indicate the *P* values from different GWASs on a minus logarithmic scale according to the chromosome location of the SNPs. Different studies are labeled with different colors. Red: NPC[a] (10); Green: NPC[b] (11); Blue: NPC[c] (12); Orange: EBV[+] HL (21). The solid triangles indicate significant SNPs reported in GWASs. The hollow triangles indicate other SNPs listed in the supplementary documents. The solid diamonds indicate significant amino acids located within the HLA genes (12). Bottom: Detailed LD structure depicted in HaploView using control samples from the NPC[a] GWAS (10). The increasing intensities of red represent lower *D'* values.

China (1,043 and 985 individuals in discovery and replication phases, respectively) [12]. NPC GWAS were also located in HLA-A (Taiwan GWAS: rs2517713, $P_{GWAS} = 3.56 \times 10^{-8}$, $P_{combined} = 3.90 \times 10^{-20}$, OR = 1.88, **Figure 1**, NPC[a]; Guangxi, China GWAS: rs417162, $P_{GWAS} = 1.13 \times 10^{-7}$, $P_{combined} = 1.05 \times 10^{-11}$, OR = 0.63, **Figure 1**, NPC[c]). It is noteworthy to point out that rs2860580 and rs2517713 are in complete LD, and that the Guangxi, China GWAS [12] also observed strong evidence for association for rs2517713 ($P_{GWAS} = 3.03 \times 10^{-7}$, $P_{combined} = 1.63 \times 10^{-11}$, OR = 0.60, **Figure 1**, NPC[c]). While significant associations were reported for SNPs in the HLA-B/C and HLA-DR/DQ regions in the Guangdong, China NPC GWAS (rs2894207; $P_{GWAS} = 1.22 \times 10^{-16}$, $P_{combined} = 3.42 \times 10^{-33}$, OR = 0.61; rs28421666; $P_{GWAS} = 3.54 \times 10^{-9}$, $P_{combined} = 2.49 \times 10^{-18}$, OR = 0.67, respectively) [11] they were not as strong as those observed for SNPs located in HLA-A. Furthermore, in the Guangxi NPC GWAS [12], where high resolution HLA genotyping was performed, multivariate analyses suggested that associations were driven by *HLA-A*1101* and that presence of glutamine at amino acid position 62 of the HLA-A gene (which marks *HLA-A*11*) was strongly associated with NPC risk providing a biological basis for the SNP-based associations reported.

Taken together, findings from both candidate-gene-based studies and GWAS have consistently demonstrated associations between HLA genes and NPC. The strongest evidence for association has been observed for alleles within HLA class I genes, including *HLA-A*0207* (risk allele in LD with *HLA-B*4601*), *HLA-A*1101* (protective allele in LD with *HLA-B*13*), and *HLA-B*5801* (risk allele in LD with *HLA-A*3303*). The associations are biologically plausible and suggest a direct causal association defined by specific amino acids involved in defining HLA binding specificities.

NON-HUMAN LEUKOCYTE ANTIGEN ASSOCIATIONS IN THE MHC REGION WITH NASOPHARYNGEAL CARCINOMA

In addition to HLA genes, other genes in the MHC region of chromosome 6p21.3 have been found to be associated with NPC. Notably, results from the Taiwan NPC GWAS [10] suggested significant associations for the HLA complex group 9 (*HCG9*) gene (rs9260734, $P_{GWAS} = 2.49 \times 10^{-7}$, $P_{combined} = 6.77 \times 10^{-18}$, OR = 1.85), and the gamma-aminobutyric acid (GABA) B receptor 1 (*GABBR1*) gene (rs29232, $P_{GWAS} = 1.67 \times 10^{-8}$ $P_{combined} = 8.97 \times 10^{-17}$, OR = 1.67). These associations were also observed in the two NPC GWAS from China [11, 12], suggesting that the associations observed are real and not driven by chance. Given the strong LD patterns observed in the MHC region, however, the issue of whether these associations implicate new genes involved in NPC or mark one or more of the HLA associations discussed above is still an open question.

HCG9 is a non-protein coding gene located within 30 kb of the *HLA-A* gene. Given its close proximity to *HLA-A* and lack of known function, it is hypothesized that SNPs within this gene reported to be associated with NPC in recently published GWAS reflect LD with causal polymorphisms within *HLA-A*. In fact, multivariate analyses of data from the Taiwan [10] and two China [11, 12] GWAS suggested that hits within the HCG9 region are not independent of the *HLA-A* findings. Furthermore, a recent pooled analysis of two NPC case-control studies in Taiwan that

jointly evaluated HLA and HCG9 confirmed that the HCG9 signal observed in the original Taiwan GWAS is likely to be driven by HLA rather than an independent signal [20].

In contrast to *HCG9*, *GABBR1* encodes a G protein-coupled receptor that forms a heterodimer with GABAB receptor 2, thereby triggering downstream signaling events in the proliferation, differentiation, and migration of cancer cells. This known biological function provides some *a priori* support for a possible functional role of *GABBR1* in cancer development. Following the initial Taiwan GWAS finding, sequencing of full-length *GABBR1* genes, including the rs29232 region, was performed in 37 NPC cases with high-risk haplotypes and 48 controls. No germ line mutations were detected and most of the novel association signals either failed to be validated or did not reach the statistical significance of rs29232 [10]. However, in this same study expression of the GABBR1 protein in NPC tissues was also evaluated using immunohistochemical staining and the intensity of the GABBR1 signal in tumor cells was significantly higher than that detected in adjacent normal epithelial cells ($P < 0.001$) [10]. This finding provides additional support for a potential functional role for *GABBR1* in the etiology of NPC.

Statistical approaches have also been employed in an attempt to determine whether GWAS findings in *HLA-A* and *GABBR1* represent distinct, independent associations, or whether they represent a single association marked by SNPs in both gene regions. Results from these efforts have been mixed, however, multivariate logistic regression analysis of the Taiwan GWAS data indicated that the *GABBR1* signal (rs29232) remained significant after controlling for the effect of *HLA-A* GWAS SNPs and sequence-based *HLA-A* alleles, namely, *HLA-A*0207/0215N or HLA-A*110101/0121N* [10]. Similar efforts from the two China GWAS, however, suggested that the *GABBR1* effect was largely driven by the *HLA-A* SNPs/alleles [11, 12]. An analysis of pooled data from two NPC case-control studies conducted in Taiwan that jointly evaluated *HLA-A* and *GABBR1* reported that while the effect of *GABBR1* was attenuated by adjustment for *HLA-A* alleles, a significant effects for *GABBR1* (rs29232) remained after adjustment for *HLA-A* and that an independent effect of *GABBR1* could therefore not be ruled out [20].

In summary, while *HLA* genes, and *HLA-A* in particular, likely explain much of the signals observed from NPC GWAS in the MHC region on chromosome 6p21.3, the possibility that other genes in the region, such as *GABBR1*, might also be causally involved in NPC cannot be discarded at this time.

HUMAN LEUKOCYTE ANTIGEN ASSOCIATIONS IN OTHER CANCER GENOME-WIDE ASSOCIATION STUDIES

Close to 180 cancer GWASs have been published (NHGRI GWAS catalog[1]). We reasoned that examination of findings from the MHC region across these GWAS might provide interesting clues into the link between HLA genes and infection-associated cancers such as NPC. We therefore summarized findings within the MHC region from cancer GWAS available through the NHGRI GWAS catalog[1]. To organize our review, we classified cancers into mutually exclusive groups as follows: EBV-associated cancers [including NPC, classical Hodgkins lymphoma (HL), and gastric

---

[1] http://www.genome.gov/gwasstudies

cancer], other infection-related cancers [including cervical cancer and hepatocellular carcinoma (HCC)], other immune-related (hematopoetic) tumors [including various subtypes of non-Hodgkin lymphomas (NHL)], and other solid tumors (including lung, testicular, and prostate cancers). While these groupings are somewhat arbitrary (e.g., we grouped gastric cancers with EBV-associated cancers despite its strong link with *H. pylori* infection since approximately 8% of gastric cancers are known to be EBV-positive and the focus of this review is on NPC, another EBV-associated cancer), we believe that they provide a useful rubric for summarizing findings across a diverse set of cancers. We focused our review on the MHC region alone (defined as chr6: 29 –33 Mb) and considered as significant SNPs with a reported *P*-value for association $<5 \times 10^{-7}$. Findings are summarized in **Table 1**.

First, we considered cancers that are known to be linked to EBV infection (**Table 1**). In addition to NPC (for which nearly 100% of tumors are EBV-associated), we evaluated two cancers for which a proportion of tumors are known to be EBV-associated: HL and gastric cancer. Burkitt lymphoma (BL) was not considered despite its close association with EBV infection because no BL GWAS has been published to date. As discussed in previous sections of this review, results from NPC GWAS indicate that the strongest evidence for an association are with HLA-A, an HLA class I gene. Findings from HL GWAS suggest an interesting pattern. The single GWAS that evaluated EBV(+) and EBV(−) HL separately (21) found evidence for an HLA class I association for EBV(+) HL and for an HLA class II association for EBV(−) HL. The HLA class I association observed for EBV(+) HL is consistent with findings for NPC. The HLA class II association observed for EBV(−) HL is consistent with findings for other hematopoetic cancers and cancers caused by infections other than EBV (discussed further below). A second HL GWAS that did not stratify by EBV status provides evidence for association with HLA class II but failed to observe and association with HLA class I (22). Finally, a GWAS that focused on nodular sclerosis HL, a subtype of HL thought not to be EBV-associated, also reported associations with HLA class II genes only (23). This highlights the importance of stratifying analyses on viral status for this heterogeneous disease. Finally, four gastric cancer GWAS have been published (24–27). None reported evidence for a significant association within the MHC region. It should be noted, however, that <10% of gastric cancers are EBV(+) and so the lack of association observed within the MHC might reflect the lack of stratification on EBV status rather than a true lack of association for EBV(+) gastric cancers. In summary, GWAS of EBV(+) cancers have consistently reported associations in the MHC region, and within the MHC the strongest evidence points to an important role for HLA class I genes.

Next, we evaluated results from GWAS for other infection-associated cancers (**Table 1**). Of the cancers with proven link to viral infections other than EBV, GWAS have been reported for cervical cancer (linked to human papillomavirus infection) and HCC (linked to hepatitis B and C infections, HBV and HCV respectively). Two cervical cancer GWAS reported the strongest evidence for association within the HLA class II region of the MHC (28, 29). The four HCC GWAS that specifically evaluated HCV-related or HBV-related cancers all reported strong hits within the HLA class II region (30–33). In contrast, the one HCC GWAS that did

not stratify by virus status (34) observed evidence for association within the MHC region, but failed to report significant associations within HLA, again highlighting the importance of stratification by viral status for etiologically heterogeneous tumors. Of note, none the four GWAS of gastric cancer, a cancer linked with infection with the bacteria *Helicobacter pylori*, reported significant associations within the MHC region, suggesting differences in the HLA associations observed for bacterial- versus viral-associated cancers. In summary, all GWAS of non-EBV virus infection-associated cancers (HPV, HBV, and HCV) that specifically considered viral status reported evidence for HLA class II associations with disease. This contrasts with EBV-associated cancers, where evidence consistently points to HLA class I involvement.

Finally, we examined GWAS for other hematopoetic and solid tumors (**Table 1**). Again, clear and consistent patterns emerged. For other hematopoetic tumors, signals observed in the MHC region were consistently located within the HLA class II region or outside of the HLA class I or II regions. In contrasts, for other solid tumors for which significant signals were observed in the MHC region, the observed signals were consistently located outside of the HLA class I or II regions and appeared to involve non-HLA genes. The one exception to this was the HLA class II (rs2395185) association observed for lung cancer in never-smoking woman in Asia (35). It should be noted that, in contrast to observations for infection-associated cancers, significant MHC signals were not always observed for other hematopoetic and solid tumors. For example, amongst hematopoetic tumors, no evidence for significant associations within the MHC region were reported for GWAS of childhood acute lymphoblastic leukemia (36–39), chronic myeloid leukemia (40), and diffuse large B-cell lymphoma (41). For other solid tumors, no evidence for significant associations within the MHC region were reported for GWAS of basal cell carcinoma, bladder cancer, breast cancer, colorectal cancer, endometrial cancer, esophageal cancer, Ewing sarcoma, gallbladder cancer, glioblastoma, glioma, melanoma, multiple myeloma, neuroblastoma, ovarian cancer, ovarian reserve, pancreatic cancer, renal cell carcinoma, small-cell lung cancer, testicular cancer, thyroid cancer, urinary bladder cancer, and Wilms tumor (NHGRI GWAS catalog[1]).

Taken together, the present review of published cancer GWAS suggests that:

1. HLA class I genes are important for EBV-associated cancers.
2. HLA class II genes are important for HPV, HBV, and HCV-associated cancers.
3. HLA class II and/or non-HLA genes in the MHC region explain associations within the MHC for other hematopoetic malignancies, and
4. There is less evidence for HLA (class I or II) involvement in the development of other solid tumors.

Particularly striking was the tendency for EBV-associated cancers to be linked to HLA class I genes, while other cancers had stronger evidence for HLA class II gene associations (HPV, HBV, and HCV-associated cancers; and some hematopoetic malignancies) or for the involvement of non-HLA genes in the MHC region (other solid tumors and some hematopoetic malignancies). We

Table 1 | Cancer GWAS significant signals located in chromosome 6p21 MHC region.

Disease	Year	No. of cases/ No. of controls	HLA class I genes		HLA class II genes			Other MHC genes	Reference
			HLA-A	HLA-B/C	HLA-DR	HLA-DQ	HLA-DP		
EBV-related tumors[a]									
Nasopharyngeal carcinoma	2009	111/260	rs2517713 rs2975042						(19)
Nasopharyngeal carcinoma	2009	277/285						GABBR1, HCG9, HLA-F	(10)
Nasopharyngeal carcinoma	2010	1,583/1,894	rs2860580	rs2894207		rs28421666		GABBR1, HCG9	(11)
Nasopharyngeal carcinoma	2012	1,405/1,650	rs417162					GABBR1, HCG9	(12)
Hodgkin's lymphoma	2011	589/5,199			rs6903608				(22)
Hodgkin's lymphoma	2012	1,200/6,417			rs2395185			MICB	(21)
EBV-positive Hodgkin's lymphoma			rs2734986						
EBV-negative Hodgkin's lymphoma					rs6903608				
Nodular sclerosis Hodgkin's lymphoma	2011	393/3,315			rs204999 rs9268528				(23)
Other virus-related tumors									
Hepatocellular carcinoma	2011	180/271						C2	(34)
HCV-related hepatocellular carcinoma	2011	721/2,890				rs9275572		MICA	(30)
HBV-related hepatocellular carcinoma	2012	1,538/1,465				rs9272105			(31)
HBV-related hepatocellular carcinoma	2013	1,161/1,353				rs9275319			(32)
HBV-related hepatocellular carcinoma	2013	971/1,938			rs9269081	rs2856718, rs7453920	rs9277535, rs3077	EHMT2, TCF19	(33)
Cervical Cancer	2013	1,034/3,948			rs9272143[e]	rs9272143[e]	rs3117027	MICA	(28)
Cervical Cancer	2013	1,364/3,028					rs4282438		(29)
Other hematopoetic tumors[b]									
Follicular lymphoma	2009	189/592			rs6457327				(44)
Follicular lymphoma	2010	681/750				rs10484561		C6orf15	(45)
Follicular lymphoma	2011	379/791				rs2647012			(46)
Chronic lymphocytic leukemia	2010	407/296			rs674313	rs9272535			(47)
Chronic lymphocytic leukemia	2012	1,121/3,745			rs674313	rs9273012, rs9273363		BAK1, IRF4	(48)
Chronic lymphocytic leukemia	2013	3,100/7,667						IRF4	(49)
Lymphoma[c]	2013	1,245/2596			rs4530903[e] rs9268853	rs4530903[e] rs2647045 rs2621416			(50)

(Continued)

Table 1 | Continued

Disease	Year	No. of cases/ No. of controls	HLA class I genes		HLA class II genes			Other MHC genes	Reference
			HLA-A	HLA-B/C	HLA-DR	HLA-DQ	HLA-DP		
Other solid tumors[d]									
Lung cancer	2008	1,952/1,438						BAG6/APOM	(51)
Lung cancer	2013	5,510/4,544			rs2395185				(35)
Lung adenocarcinoma	2009	5,739/5,848						BAG6/APOM	(52)
Lung adenocarcinoma	2012	1,695/5,333						BTNL2	(53)
Testicular germ cell tumor	2009	730/1,435						BAK1	(54)
Prostate cancer	2011	6,621/6,939						CCHCR1	(55)
Prostate cancer	2013	11,085/11,463						NOTCH4	(56)
Multiple cancers[f]	2012	5,368/4,006						LRFN2	(27)

[a]GWAS study in gastric cancer (25–27) and non-cardia gastric cancer (24) with no significant findings reported in the MHC region.

[b]GWAS studies in acute lymphoblastic leukemia (childhood) (36–39), chronic myeloid leukemia (40), and large B-cell lymphoma (41) with no significant findings reported in the MHC region.

[c]This study included multiple types of lymphoma: 275 follicular non-Hodgkin lymphoma cases, 269 diffuse large B-cell non-Hodgkin lymphoma cases, 198 other non-Hodgkin lymphoma cases, 202 Hodgkin lymphoma cases, and 4,044 controls.

[d]GWAS studies in basal cell carcinoma, bladder cancer, breast cancer, colorectal cancer, endometrial cancer, esophageal cancer, Ewing sarcoma, gallbladder cancer, glioblastoma, glioma, melanoma, multiple myeloma, neuroblastoma, ovarian cancer, ovarian reserve, pancreatic cancer, renal cell carcinoma, small-cell lung cancer, testicular cancer, thyroid cancer, urinary bladder cancer, and Wilms tumor with no significant findings reported in the MHC region (NHGRI GWAS catalog http://www.genome.gov/gwastudies).

[e]rs9272143 and rs4530903 located between HLA-DRB1 and HLA-DQA1.

[f]This study included 2,331 lung cancer cases, 1,006 non-cardia gastric cancer cases, 2,031 esophageal squamous-cell carcinoma cases, and 4,006 Han Chinese controls.

therefore considered further the characteristics of EBV infection that might explain the "predilection" of EBV-associated cancers to be associated with HLA class I genes.

Little is known about the genetic basis for immunological responses to EBV infection, despite the belief that such responses are important mediators of cancer risk. The only published GWAS that evaluated genetic factors associated with antibodies against EBV reported evidence for the involvement of HLA class II genes in EBV seroreactivity, measured as anti-EBV EBNA-1 IgG levels (42). This finding appears at face value to be inconsistent with the strong HLA class I association observed for NPC, but needs to be interpreted with caution since the anti-EBV antibody evaluated (IgG against EBNA-1) is not a good marker of NPC risk. In the future, it will be important to evaluate genetic factors associated with anti-EBV EBNA-1 and VCA IgA responses because (1) EBNA-1 and VCA are the antigens for which strong and consistent associations with risk of NPC development have been observed and (2) IgA responses (rather than IgG responses) are thought to better reflect chronic EBV reactivation at mucosal surfaces believed to be required for NPC development.

One hypothesis to explain the specific association of HLA class I genes with EBV-associated cancers while HPV, HBV, and HCV-associated cancers have more clear associations with HLA class II genes invokes site of infection. Of all infectious agents known to be directly linked to cancer, EBV is the only one that establishes lifelong latency in B-lymphocytes and requires infection of the epithelial compartment during lytic reactivation for viral shedding and transmission. Perhaps HLA class I genes are important mediators of this complex biological lifecycle, although precisely how this might be remains to be established. Of interest to note in this respect is the fact that progression to AIDS among HIV infected individuals has been consistently shown to be associated with HLA class I alleles (43). Since HIV is a virus that, like EBV, establishes lifelong infection in lymphocytes, the commonality of these findings might suggest a parallel HLA-mediated immunological response pathway for these two pathogens that is distinct from that for other infectious agents linked to cancer development.

CHALLENGES AND OPPORTUNITIES

Human leukocyte antigen genes have long been suggested to be associated with NPC risk. Recent NPC GWAS have confirmed this association and have further indicated that the strongest genetic associations with NPC reside within HLA class I genes, particularly *HLA-A*. This contrasts with what has been observed for other tumors, suggesting a unique relationship between EBV and HLA class I genes and the need for studies to more specifically define how and which host genetic differences in antigen presentation of specific EBV antigens affects the EBV tumorigenesis process by promoting viral escape of host immune surveillance leading to NPC. These studies could provide clues not only about EBV and NPC, but also about the relationship between other oncogenic infectious agents and their respective cancers and on the nature of HLA associations for hematopoetic malignancies.

While challenges remain, given the strong LD patterns observed in the MHC, the large number of genes in this region and the highly polymorphic nature of HLA genes themselves, the prospect of studying diverse populations with distinct HLA patterns and

LD structure, and of applying new technologies such as high-throughput sequencing and molecular profiling techniques to elucidate the complex structure of the MHC and its association with NPC and other cancers could lead to better insights into our understanding of the specific mechanisms involved in cancer pathogenesis. This, in turn, could lead to new interventions aimed at interrupting this pathogenic process as a cancer prevention measure.

REFERENCES

1. Vokes EE, Liebowitz DN, Weichselbaum RR. Nasopharyngeal carcinoma. *Lancet* (1997) **350**(9084):1087–91. doi:10.1016/S0140-6736(97)07269-3

2. Boyle P, Levin B. *World Cancer Report 2008: International Agency for Research on Cancer.* Lyon: WHO Press (2008).

3. Yu MC, Jian-Min Y. Nasopharyngeal cancer. 3rd Ed. In: Schottenfeld D, Fraumeni JF editors. *Cancer Epidemiology and Prevention.* Oxford, NY: Oxford University Press (2006). p. 603–19.

4. zur Hausen H, Schulte-Holthausen H, Klein G, Henle W, Henle G, Clifford P, et al. EBV DNA in biopsies of Burkitt tumours and anaplastic carcinomas of the nasopharynx. *Nature* (1970) **228**(5276):1056–8. doi:10.1038/2281056a0

5. Longnecker RM, Kieff E, Cohen J. Epstein-Barr virus. 6th ed. In: Knipe DM, Howley PM editors. *Fields Virology* (Vol. 1), Philadelphia: Wolters Kluwer (2013). p. 2606–700.

6. Nikolich-Zugich J, Fremont DH, Miley MJ, Messaoudi I. The role of MHC polymorphism in anti-microbial resistance. *Microbes Infect* (2004) **6**(5):501–12. doi:10.1016/j.micinf.2004.01.006

7. Simons MJ, Wee GB, Day NE, Morris PJ, Shanmugaratnam K, De-The GB. Immunogenetic aspects of nasopharyngeal carcinoma: I. Differences in HLA antigen profiles between patients and control groups. *Int J Cancer* (1974) **13**(1):122–34. doi:10.1002/ijc.2910130114

8. Chang ET, Adami HO. The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev* (2006) **15**(10):1765–77. doi:10.1158/1055-9965.EPI-06-0353

9. Hildesheim A, Wang CP. Genetic predisposition factors and nasopharyngeal carcinoma risk: a review of epidemiological association studies, 2000-2011: Rosetta Stone for NPC: genetics, viral infection, and other environmental factors. *Semin Cancer Biol* (2012) **22**(2):107–16. doi:10.1016/j.semcancer.2012.01.007

10. Tse KP, Su WH, Chang KP, Tsang NM, Yu CJ, Tang P, et al. Genome-wide association study reveals multiple nasopharyngeal carcinoma-associated loci within the HLA region at chromosome 6p21.3. *Am J Hum Genet* (2009) **85**(2):194–203. doi:10.1016/j.ajhg.2009.07.007

11. Bei JX, Li Y, Jia WH, Feng BJ, Zhou G, Chen LZ, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat Genet* (2010) **42**(7):599–603. doi:10.1038/ng.601

12. Tang M, Lautenberger JA, Gao X, Sezgin E, Hendrickson SL, Troyer JL, et al. The principal genetic determinants for nasopharyngeal carcinoma in China involve the HLA class I antigen recognition groove. *PLoS Genet* (2012) **8**(11):e1003103. doi:10.1371/journal.pgen.1003103

13. Margulies DH, NAtarajan K, Rossjohn J, MacCluslory J. The major histocompatibility complex and its proteins. 7th Ed. In: Paul WE editor. *Fundamental Immunology.* Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins (2013). p. 487–523.

14. Bei JX, Jia WH, Zeng YX. Familial and large-scale case-control studies identify genes associated with nasopharyngeal carcinoma. *Semin Cancer Biol* (2012) **22**(2):96–106. doi:10.1016/j.semcancer.2012.01.012

15. Lu SJ, Day NE, Degos L, Lepage V, Wang PC, Chan SH, et al. Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region. *Nature* (1990) **346**(6283):470–1. doi:10.1038/346470a0

16. Lu CC, Chen JC, Jin YT, Yang HB, Chan SH, Tsai ST. Genetic susceptibility to nasopharyngeal carcinoma within the HLA-A locus in Taiwanese. *Int J Cancer* (2003) **103**(6):745–51. doi:10.1002/ijc.10861

17. Lu CC, Chen JC, Tsai ST, Jin YT, Tsai JC, Chan SH, et al. Nasopharyngeal carcinoma-susceptibility locus is localized to a 132 kb segment containing HLA-A using high-resolution microsatellite mapping. *Int J Cancer* (2005) **115**(5):742–6. doi:10.1002/ijc.20946

18. Hardy J, Singleton A. Genome wide association studies and human disease. *N Engl J Med* (2009) **360**(17):1759–68. doi:10.1056/NEJMra0808700

19. Ng CC, Yew PY, Puah SM, Krishnan G, Yap LF, Teo SH, et al. A genome-wide association study identifies ITGA9 conferring risk of nasopharyngeal carcinoma. *J Hum Genet* (2009) **54**(7):392–7. doi:10.1038/jhg.2009.49

20. Hsu WL, Tse KP, Liang S, Chien YC, Su WH, Yu KJ, et al. Evaluation of human leukocyte antigen-A (HLA-A), other non-HLA markers on chromosome 6p21 and risk of nasopharyngeal carcinoma. *PLoS One* (2012) **7**(8):e42767. doi:10.1371/journal.pone.0042767

21. 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**(3):240–53. doi:10.1093/jnci/djr516

22. Enciso-Mora V, Broderick P, Ma Y, Jarrett RF, Hjalgrim H, Hemminki K, et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). *Nat Genet* (2010) **42**(12):1126–30. doi:10.1038/ng.696

23. Cozen W, Li D, Best T, Van Den Berg DJ, Gourraud PA, Cortessis VK, et al. A genome-wide meta-analysis of nodular sclerosing Hodgkin lymphoma identifies risk loci at 6p21.32. *Blood* (2012) **119**(2):469–75. doi:10.1182/blood-2011-03-343921

24. Shi Y, Hu Z, Wu C, Dai J, Li H, Dong J, et al. A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat Genet* (2011) **43**(12):1215–8. doi:10.1038/ng.978

25. Saeki N, Ono H, Sakamoto H, Yoshida T. Genetic factors related to gastric cancer susceptibility identified using a genome-wide association study. *Cancer Sci* (2013) **104**(1):1–8. doi:10.1111/cas.12042

26. Abnet CC, Freedman ND, Hu N, Wang Z, Yu K, Shu XO, et al. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat Genet* (2010) **42**(9):764–7. doi:10.1038/ng.649

27. Jin G, Ma H, Wu C, Dai J, Zhang R, Shi Y, et al. Genetic variants at 6p21.1 and 7p15.3 are associated with risk of multiple cancers in Han Chinese. *Am J Hum Genet* (2012) **91**(5):928–34. doi:10.1016/j.ajhg.2012.09.009

28. Chen D, Juko-Pecirep I, Hammer J, Ivansson E, Enroth S, Gustavsson I, et al. Genome-wide association study of susceptibility loci for cervical cancer. *J Natl Cancer Inst* (2013) **105**(9):624–33. doi:10.1093/jnci/djt051

29. Shi Y, Li L, Hu Z, Li S, Wang S, Liu J, et al. A genome-wide association study identifies two new cervical cancer susceptibility loci at 4q12 and 17q12. *Nat Genet* (2013) **45**(8):918–22. doi:10.1038/ng.2687

30. Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, Hosono N, et al. Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat Genet* (2011) **43**(5):455–8. doi:10.1038/ng.809

31. Li S, Qian J, Yang Y, Zhao W, Dai J, Bei JX, et al. GWAS identifies novel susceptibility loci on 6p21.32 and 21q21.3 for hepatocellular carcinoma in chronic hepatitis B virus carriers. *PLoS Genet* (2012) **8**(7):e1002791. doi:10.1371/journal.pgen.1002791

32. Jiang DK, Sun J, Cao G, Liu Y, Lin D, Gao YZ, et al. Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus-related hepatocellular carcinoma. *Nat Genet* (2013) **45**(1):72–5. doi:10.1038/ng.2483

33. Kim YJ, Kim HY, Lee JH, Yu SJ, Yoon JH, Lee HS, et al. A genome-wide association study identified new variants associated with the risk of chronic hepatitis B. *Hum Mol Genet* (2013) **22**(20):4233–8. doi:10.1093/hmg/ddt266

34. Clifford RJ, Zhang J, Meerzaman DM, Lyu MS, Hu Y, Cultraro CM, et al. Genetic variations at loci involved in the immune response are risk factors for hepatocellular carcinoma. *Hepatology* (2010) **52**(6):2034–43. doi:10.1002/hep.23943

35. Lan Q, Hsiung CA, Matsuo K, Hong YC, Seow A, Wang Z, et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat Genet* (2012) **44**(12):1330–5. doi:10.1038/ng.2456

36. Trevino LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet* (2009) **41**(9):1001–5. doi:10.1038/ng.432

37. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E, et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet* (2009) **41**(9):1006–10. doi:10.1038/ng.430

38. Ellinghaus E, Stanulla M, Richter G, Ellinghaus D, te Kronnie G, Cario G, et al. Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. *Leukemia* (2012) **26**(5):902–9. doi:10.1038/leu.2011.302

39. Yang JJ, Cheng C, Devidas M, Cao X, Campana D, Yang W, et al. Genome-wide association study identifies germline polymorphisms associated with relapse of childhood acute lymphoblastic leukemia. *Blood* (2012) **120**(20):4197–204. doi:10.1182/blood-2012-07-440107

40. Kim DH, Lee ST, Won HH, Kim S, Kim MJ, Kim HJ, et al. A genome-wide association study identifies novel loci associated with susceptibility to chronic myeloid leukemia. *Blood* (2011) **117**(25):6906–11. doi:10.1182/blood-2011-01-329797

41. Kumar V, Matsuo K, Takahashi A, Hosono N, Tsunoda T, Kamatani N, et al. Common variants on 14q32 and 13q12 are associated with DLBCL susceptibility. *J Hum Genet* (2011) **56**(6):436–9. doi:10.1038/jhg.2011.35

42. Rubicz R, Yolken R, Drigalenko E, Carless MA, Dyer TD, Bauman L, et al. A genome-wide integrative genomic study localizes genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1). *PLoS Genet* (2013) **9**(1):e1003147. doi:10.1371/journal.pgen.1003147

43. International HIV Controllers Study, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* (2010) **330**(6010):1551–7. doi:10.1126/science.1195271

44. Skibola CF, Bracci PM, Halperin E, Conde L, Craig DW, Agana L, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. *Nat Genet* (2009) **41**(8):873–5. doi:10.1038/ng.419

45. Conde L, Halperin E, Akers NK, Brown KM, Smedby KE, Rothman N, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nat Genet* (2010) **42**(8):661–4. doi:10.1038/ng.626

46. Smedby KE, Foo JN, Skibola CF, Darabi H, Conde L, Hjalgrim H, et al. GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma. *PLoS Genet* (2011) **7**(4):e1001378. doi:10.1371/journal.pgen.1001378

47. Slager SL, Rabe KG, Achenbach SJ, Vachon CM, Goldin LR, Strom SS, et al. Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. *Blood* (2011) **117**(6):1911–6. doi:10.1182/blood-2010-09-308205

48. Slager SL, Skibola CF, Di Bernardo MC, Conde L, Broderick P, McDonnell SK, et al. Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia. *Blood* (2012) **120**(4):843–6. doi:10.1182/blood-2012-03-413591

49. Berndt SI, Skibola CF, Joseph V, Camp NJ, Nieters A, Wang Z, et al. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nat Genet* (2013) **45**(8):868–76. doi:10.1038/ng.2652

50. Vijai J, Kirchhoff T, Schrader KA, Brown J, Dutra-Clarke AV, Manschreck C, et al. Susceptibility loci associated with specific and shared subtypes of lymphoid malignancies. *PLoS Genet* (2013) **9**(1):e1003220. doi:10.1371/journal.pgen.1003220

51. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet* (2008) **40**(12):1407–9. doi:10.1038/ng.273

52. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet* (2009) **85**(5):679–91. doi:10.1016/j.ajhg.2009.09.012

53. Shiraishi K, Kunitoh H, Daigo Y, Takahashi A, Goto K, Sakamoto H, et al. A genome-wide association study identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population. *Nat Genet* (2012) **44**(8):900–3. doi:10.1038/ng.2353

54. Rapley EA, Turnbull C, Al Olama AA, Dermitzakis ET, Linger R, Huddart RA, et al. A genome-wide association study of testicular germ cell tumor. *Nat Genet* (2009) **41**(7):807–10. doi:10.1038/ng.394

55. Kote-Jarai Z, Olama AA, Giles GG, Severi G, Schleutker J, Weischer M, et al. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet* (2011) **43**(8):785–91. doi:10.1038/ng.882

56. Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* (2013) **45**(4):385–91. doi:10.1038/ng.2560