·Commentary·

The origin of genetic risk for nasopharyngeal carcinoma: a commentary on *Is nasopharyngeal cancer really a "Cantonese cancer"?*

Malcolm J. Simons

Simons Haplomics Limited, Hong Kong, P.R. China

Nasopharyngeal Carcinoma (NPC) occurs with highest incidence in Southeastern China among Cantonese-speaking Chinese, giving rise to the common name of "Cantonese cancer". Chinese in the province of Guangdong have the highest risk of NPC in China. Within Guangdong, variation in rates among the major dialect groups is threefold. The Cantonese who reside in the central region of Guangdong display the highest risk of NPC. Chinese living on boats ("Tanka") have twice the risk of those who live on land[1]. On pages 517–526 of the present issue of *Chinese Journal of Cancer*, Wee and colleagues question the appropriateness of calling NPC the "Cantonese cancer", suggesting that the term can be largely attributed to awareness promotion by the late Dr. John Ho. Less well known than the Southeastern Chinese connection, and less well promoted, is the fact that NPC occurs with similar or even higher incidence among the Naga people in Northeast India [2], the Bidayuh (Land Dyaks) in Sarawak [3], the Paiwan, Bunu and Rukai aboriginal tribes of Taiwan[4,5], and the Arctic Inuit[6].

Wee *et al*. propose a novel, unifying explanation of the NPC occurrence characteristics of high risk in the ethnic minorities / aboriginal populations of Southeast China, Northeast India and East Malaysia. They adduce archeological, linguistic and genetic evidence for the origination of risk for NPC in Bai-Yue (or proto-Bai-Yue) peoples, in the late Pleistocene era. They propose that these southern China aborigines, identified by possessing the Y chromosome marker "M119" (Haplogroup 0–1A: see Figure 1 in their article), occupied the coastal region of Sundaland. With rises in sea levels at the end of the last glaciation communities of these peoples became isolated. Wee *et al*. propose that more recent migrations such as the Austronesian expansion into insular southeast Asia and then to Oceania, and more recently (4 000–2 000 years ago) along the spice and maritime routes to East and North Africa and the Mediterranean, account for NPC occurrence patterns in countries and among peoples remote from southeast China. According to this hypothesis, the occurrence of NPC in Cantonese is a consequence of prior risk in antecedent Bai-Yue peoples. If their hypothesis is sustained then the title of their paper—*Is NPC really a "Cantonese cancer"?*—should be answered in the negative.

The advent of the hypothesis has had the effect of stimulating renewed effort to understand the genetic basis of excess NPC risk. A meeting held at the Singapore National Cancer Centre on February 20–21 reviewed the current status of NPC risk genetics, and identified populations to target in future genetic studies, both to test the Bai-Yue origination hypothesis of NPC risk, and to consider strategies to discover the genetic basis of that risk. Manuscripts resulting from presentations at the meeting will be published in a future edition of this Journal.

In this invited Commentary I will consider populations, patient investigation strategies and genetic technologies that can be expected to test the Wee *et al*. Bai-Yue NPC risk origination hypothesis. The hypothesis concerns the genetics of excess risk rather than genetic factors in the etiopathogenesis of NPC, so I do not intend to consider genetic factors in NPC development (tumorigenesis and progression). While, for completion, I will identify chromosomes that have been implicated in risk conferment, only for HLA genes of the major histocompatibility complex on chromosome 6 (6p21.3) is there reproducible evidence of genetic association with NPC risk, so HLA findings will occupy the majority of my Comment.

The origin of Wee's insight is itself the stuff of future legend. He noticed that the bamboo-pole dance (tinikling) for which the Philippines is well known is also a form of entertainment in other societies, including the Manipurians of northeast India, the Kadazans of Borneo, all over Indochina, in Polynesia and in the Zhuang and Li minority ethnic groups in southern China. While these and other socio-cultural similarities have long been recognized, the inspired insight that occurred to Wee was that these peoples also had in common a high incidence of NPC. He and his associates then set out to inquire whether historic records and more recent genetic studies of human migrations would support a unifying NPC origination hypothesis.

Whereas 40 years ago the following NPC features were only beginning to receive focused attention, in 2010 literature reference to these features is commonplace:

(1) altered immunity to Epstein-Barr herpes virus (EBV) infection, and to the presence of EBV DNA in a high proportion

Correspondence to: Malcolm J. Simons; Tel: +61-3-5278-9839;
Email: drsimons@haplomics.com, mjsimons@optusnet.com.au

This paper was edited by Wei Liu on 2010-04-10.

Received: 2010-02-13; Accepted: 2010-03-01

of NPC tumors;

(2) the candidate role of ingestants, especially N-nitroso compounds in salted fish;

(3) the unusual pattern of population occurrence, including increased risk in populations having recent or distant admixture with southern Chinese, and in populations geographically distant from, and having no known connection with, Chinese;

(4) the occurrence of multiple case families to support the strong genetic element indicated by population case patterns.

The discovery of herpes-type viral particles and Epstein Barr virus DNA in 1970 initiated the virological era of NPC research[7,8]. While descriptions of the EBV-NPC association have been emerging in the subsequent forty years a detailed understanding of how EBV infection is a causal event in NPC is still awaited[9-17]. In addition to NPC, there are several cancers in which altered immunity to EBV infection is implicated, including Burkitt's lymphoma, B-cell lymphoproliferative syndromes, Hodgkin's disease, T-cell lymphoma, salivary gland carcinoma and gastric carcinoma[11,18]. However, the role of the virus in these malignancies varies from primary etiologic agent to necessary or contributory factor[11]. It is crucially important to resolve the basis of NPC risk, and the role of EBV infection, because NPC is the prototype disease amongst not only these cancers, but also those malignancies that are initiated by exposure to other viruses[19,20]. The discovery of HLA associations with NPC allowed for the realization as early as 1975 that HLA typing of NPC predisposition could test the hypothesis that EBV is an etiological factor in NPC[21,22]. The earliest studies of HLA in the other two main cancers having a viral etiology (HPV-cervical cancer and HBV-hepatocellular carcinoma) were first reported only 6 years later, in 1981[23].

Concerning the candidate role of ingestants, advocacy of nitrosamine-contamination of salted fish as co-carcinogens was most actively asserted by Dr. John Ho in Hong Kong, supported by scientist Dr. Dolly Huang[24-26]. There are other supportive biochemical reports[27-30], and most recently polymorphisms described in genes involved in nitrosamine metabolism[31].

The hypothesis of Wee *et al*. concerns the unusual pattern of population occurrence and genetic predisposition. "Bright ideas are two a penny" the saying goes, but the ability to investigate new ideas depends on technological developments. Occurrence of NPC as multiple cases in families[32-34], and the dialect and ethnic differences of ~100 between lowest risk Caucasians and Singapore Indians and highest risk Cantonese[35], had prompted the idea of genetic factors for at least a decade before the opportunity to investigate genetics in cancer arose with the development of the scientific field of histocompatibility genetics. Studies carried out over the previous four decades on factors controlling the susceptibility to viral leukemias in mice had shown the existence of specific controlling gene loci[36]. The role of the major histocompatibility gene complex in the mouse (H-2) to susceptibility to tumor viruses, for example, mammary[37], leukemia[38], provided the rationale to type for genes of the human equivalent major histocompatibility complex (initially HL-A, later HLA) in solid tumors and hematological malignancies.

The establishment in 1970 of the first HLA genetic laboratory in a non-caucasian centre at the WHO Immunology Research & Training Centre, National University of Singapore, marked the beginning of the search for histocompatibility genetic markers of NPC risk. Singapore was an appropriate location because six years earlier, in 1964, the first international symposium on descriptive and analytic epidemiology of NPC had been held in Singapore under the Chairmanship of Professor K. Shanmugaratnam, and investigation of host factors responsible for the uniquely high risk in southern Chinese was a recommendation of the symposium[39]. The establishment of the Singapore Cancer Registry under Shanmugaratnam's stewardship provided age-standardized incidence (per 100 000 per year) evidence of likely genetic effects in NPC occurrence with differences in specific Chinese communities (dialect groups), negligible incidence in Indians, and low-medium in Malays (Cantonese 29.1 and 11.0; Teochew 18.3 and 6.2; Hokkien 14.1 and 4.7; Hainanese 14.2 and 3.3; Hakka 12.6 and 4.8; Other Chinese 12.2 and 6.0; Malays 4.7 and 0.6; Indians 0.9 and 0.0 -males and females, respectively)[35].

Since the late 1960's there had been searches for association between HLA antigens and malignant diseases. By 1974 the major findings of HLA Class I and II allele and haplotype associations with NPC began to be discovered[21,40-46]. Studies of other malignancies were not so rewarding. Only weak associations with Hodgkin's disease and acute lymphocyte leukemia were found[47]. For studies of a variety of solid tumors any differences between patients and controls either did do not attain statistical significance, or were not reproducible, except for nasopharyngeal carcinoma[48].

In the early days of blood cell marker genetics, HLA gene type was detected by serological methods using antibody reagents for detection of cell surface antigen phenotype molecules. The typing reagents were initially of Caucasian origin which was of limited help in identifying allele types in all other ethnic groups, particularly for those Chinese and African populations among whom NPC occurred with high incidence. This resulted in a sizeable frequency of less than two detectable antigens or "blanks", particularly at the HLA-B locus. Among NPC patients the frequency of less than two detectable HLA-B antigens was significantly increased, together with an increase in HLA-A2[21,40,41]. Since the phenomenon was most obvious at the HLA-B locus, the undetectable antigens were grouped as HLA-B. Blank. With time, antibody reagents were identified which detected antigens within the "blank" that were more frequent in, or detected only in, Asian populations. The first antigen detected within the HLA-B.Blank was designated Singapore-1 (Sin-1: B16-38/39), and occurred in slightly lower frequency in NPC patients relative to controls. The second HLA-B.Blank antigen, Singapore-2 (Sin-2), was found to comprise almost half the undetected HLA-B antigens in normal subjects. More importantly, the frequency of HLA-B.Sin-2 in NPC patients was almost double that of controls, being the first HLA antigen to be associated with a common human cancer[48]. Subsequently, HLA-B.Sin-2, in conjunction with independent confirmation (serum Hsieh)[49], was first conditionally approved as Bw46 by the WHO HLA Nomenclature committee, then unconditionally assigned as

HLA-B46, and most recently assigned as the genetic allele HLA-B*4601.

In the 1980's genetic typing moved from detection of the cell surface antigen phenotype to testing for DNA sequences in the nucleus, first indirectly by restriction fragment length polymorphism [50], then, with the advent of polymerase chain reaction[51], to gene exon coding sequence type[52]. For NPC and all other disease association studies, this has meant that the HLA antigen discoveries of the 1970's have progressed to HLA locus gene sequence allelotyping, with accompanying resolution of ever more alleles, and increasing confidence in definitive allele type assignment and case-control frequency comparison. For those HLA alleles that occurred in frequencies of > 5% in Chinese and have subsequently been shown to have limited allelic subtype in Chinese, such as HLA-A2, -A11, -A19/33, -B13, -B46, and -B17/58, early serology was sufficient to reveal statistically significant differences in frequency in NPC patients, increased in the cases of HLA-A2, -A19/33, -B46, -B17/58, decreased in HLA-A11, and -B13[46]. These observations of increased frequency conferring risk, and decreased frequencies denoting protection, were confirmed in a 1983 publication [53]. In this report, family studies provided insight into A-B haplotypes, such as increased A2 with B46, and decreased B13 with A11, and revealed that the risk attributable to HLA-A2/B46 accounted for only ~20% of the excess risk for NPC in Chinese[53].

In early studies of non-Chinese patients, absence of Chinese risk alleles and haplotypes in non-Chinese populations, such as Tunisian NPC patients, was not so much noteworthy as expected in view of ethnic differences in allelotype and haplotype [54]. However, in the context of the Wee *et al* hypothesis, the absence of B46 from medium-to-high-risk populations (Taiwan Aborigines; Arctic Inuit; North African Maghreb), and the apparent increase of other HLA alleles not including A2, suggests either that any excess risk in common with that in Chinese has different HLA haplotypic associations, or that high genetic risk in these populations is unassociated with HLA.

For any marker comparison of patients with the same diagnosis it is essential to ensure comparability between the patients. Features that contribute to patient heterogeneity, and which have been shown to be important in HLA genetic studies of NPC, include age of onset and whether the patients are newly diagnosed or varying duration survivors. Most reports do not distinguish between these two main groups. The former captures all patients. The latter comprise a subpopulation. Also, most reports give patient age as a mean +/- range so it is not possible to evaluate age comparability. The importance of age is well illustrated in Maghreb populations of North Africa where the age-related incidence is distinctly bimodal, with some 20% of patients developing the disease before 25 years of age. The juvenile form of NPC has distinct biological characteristics, so distinction is imperative. For instance, loss of heterozygosity on the short arm of chromosome 3, a frequent genetic change in endemic area NPC, is reduced in young-onset patients relative to adults[55]. Also young-onset patients have a low amount of p53 and Bcl2 in the tumor tissue, a low level of anti-EBV IgG and IgA in the peripheral blood [56], and an earlier tendency to metastasize[57] relative to adults. The significance of recognition of Maghreb patient age-of-onset heterogeneity will arise again later when discussing HLA patterns in North African NPC patients.

Another NPC risk population that has a special characteristic is the Arctic Inuit (Greenland, Alaska and North-West Canada). Uniquely among medium-high NPC risk populations, the Inuit have a high incidence of salivary gland carcinoma (SGC) for which there is also evidence of EBV association[58,59], in contrast to non-Inuit SGC[60]. SGC is sufficiently common in NPC families to suggest a common aetiology[61]. In first-degree relatives of Greenland patients, the relative risk for both NPC and SGC was 8.0 [95% confidence interval (95% CI), 4.1–14.0] and 8.4 (95% CI, 2.7–19.5), respectively, and for uterine cervical cancer 2.2 (95% CI, 1.1– 3.9). These results indicate that the increased risk of cancer in nasopharyngeal carcinoma families is not restricted to nasopharyngeal carcinoma, but extends to the virally associated cancers of the salivary glands and cervical uteri[62]. By contrast, NPC tends to aggregate in Cantonese families in Guangdong Province, and the malignancies in these families appear to be site specific, with no excess of any other malignancy[34]. When the basis of NPC risk in Chinese patients is revealed, it will be interesting to see whether Inuit have the same risk type despite the broader spectrum of EBV malignancy predisposition.

It has to be acknowledged that there has been very little progress in characterizing the genomic basis of HLA-associated risk for NPC over the past 3–4 decades [63]. Nonetheless, the early findings obtained by low resolution serology have been confirmed by high resolution sequence-based typing, and the NPC-associated HLA alleles and haplotypes have been better characterized. HLA typing of Asia-Pacific populations has established that the B46 allele is a Mongoloid-specific allele, in accompaniment with varying telomeric A locus and centromeric Class II (DR,DQ,DP) locus alleles. For instance, in Chinese the common DR is DRB1*0901, the common haplotype is A2-B46-DR9, and this three-locus haplotype is increased in several autoimmune diseases, suggesting an association with altered immune responsiveness. It was first noted in 1981 that there was a "breakage" (recombination) of this Chinese-common B46-DR9 pairing in NPC, where the NPC B46 haplotype had a DR6-family allele [64]. Throughout Southeast Asian populations different members of the DR6 and other families are seen with B46 haplotypes, indicating that spread from SE China has been associated with recombination between HLA-B and DR. Of great promise in resolving the HLA-NPC association is the fact that in Japanese and northern Han Chinese, among whom NPC occurs only with low incidence, the dominant HLA-B46 haplotype is neither B46-DR9 nor B46-DR6 but B46-DR8. B46 occurs predominantly as the single B*4601 allelotype. The exception, B*4602, discovered in Japanese, occurs very rarely.

A second prominent haplotype in Chinese and Chinese-related Southeast Asian populations is A19(33)-B17(58)-DR3. Like the B46-DR9 situation, this haplotype is common in normal subjects and involved in autoimmune diseases. Again, historically there have been numerous occurrences of B58-DR

recombination involving DR types of the DR6 family in Southeast Asian populations. The A33-B58-DR haplotype(s) are also associated with NPC, albeit with different disease features (age of onset, prognosis, etc). That these co-occurrence patterns are inheritable haplotypes has been supported by family studies, but the population frequency and NPC patient mechanistic significance remains to be determined. For now we can conclude that at least these two haplotype groups are involved in immune reactivity, and therefore likely to be targets of natural selection, as well as involved in NPC risk conferment, albeit probably indirectly.

It is easy to imagine that genetic variants ("mutations"), be they coding exon inter-locus recombinations as in the case of B46 (vide infra), or intergenic recombinants that disrupt haplotypes and result in different pairings, as in the case of B46, B58, B38, B13 and other risk/resistance haplotypes can, over time, and depending on environmental exposure to inhalants and/or ingestants, generate a complex pattern of selection operating on a large number of loci associated with HLA haplotypes. Whether the NPC risk locus/i are HLA loci per se, or a linked locus/i associated with the extended HLA multigene complex, under the pressure of different population dispersal events described by Wee *et al*., it is easy to imagine that patterns of NPC locus-neighbouring linked alleles may result in the loss of Mongoloid-specific B46 and B58 markers themselves, while retaining chromosomal segments carrying the NPC risk profile.

There are two main objectives of human population genetic typing. The first is analysis of demographic histories to infer the relations between populations, including migrations that influence population associations. The second is to seek genetic markers associated with risk for, or resistance to, disease occurrence within populations. Investigation of Wee *et al.*'s NPC risk origination hypothesis requires genetic analysis of high risk populations, and of NPC patients within those populations. While Wee *et al*. utilize mitochondrial DNA and Y chromosome haplogroup genetic data as the main population migration markers to support their hypothesis, they also cite HLA genetic typing as a marker of population affinities, in three situations. The first is that of the Minnan (Hokkien) and Hakka in Taiwan[65]. These major population groups in Taiwan are descendants of inhabitants of the southeast China coastal regions that have arrived in recent centuries. The HLA patterns are similar to those first identified in overseas Chinese in Singapore, and subsequently among other overseas Chinese. The prototypic findings of A2 and B46 are also found in insular southeast populations, such as Thais[66,67].

The second involves Kinh Vietnamese population[68]. As Wee *et al*. noted, the authors identified two categories of features; one of overseas Chinese, and a second that distinguished the Kinh from other populations. Wee *et al*. note that the Kinh ethnic group have a dual origin from the Bai-Yue and from Indochina Austro-Asiatic populations. In view of this duality, of the juxtaposition of present-day North Vietnam and Southern China (Yunnan province, Guangxi province), and of the historic processes of population change, it would be informative to investigate additional populations in this inter-face region of southern Chinese with mainland Southeast Asia.

The third citation of HLA studies is of Taiwan indigenous tribes who were the earliest to occupy Taiwan following its emergence as an island around the Pleistocene/ Holocene transition (~12 000 years ago)[69]. The nine groups show the least allelic diversity and the highest multi-locus haplotype frequencies of any human population, indicating a high level of isolation homogeneity. Haplotypes involving HLA-A24 and HLA-B40 are present in widely dispersed populations [Tibetans, Buryat (Mongolia-Russia), Manchuria (Orochons), Siberia (Yakut), Alaska (Tlingit), Canada (Inuit), Japanese, Thais, Papua New Guinea Highlanders, New Zealand Maoris] but there is little genetic relationship to current southern and northern Han Chinese. It is data such as this which bears on hypotheses of peopling east of China mainland and into Oceania.

For genetic epidemiological purposes, where the goal is population genetic inference from genomic sequence variation[70], genes of the HLA complex confer a unique potential because they comprise haplotypes that incorporate a varying number of loci, each of which can be regarded as a genetic unit. At any one of the Class I and II loci there are tens to hundreds of alleles which confer hundreds of multi-locus haplotypes, thus providing a very powerful resource for population affinity evaluation. The exceptional situation conferred by HLA haplotypes has been recognized [71,72], but there are still those who are 'profoundly suspicious of the new population genetic technology of reconstructing a haplotype tree and reading history off the tree' [73]. The term haplotype was coined by Ceppellini in 1967 [74] and defined in 1989 as "a region of genomic DNA on a chromosome which is bounded by recombination sites such that genetic loci within a haplotypic region are usually inherited as a unit. However, occasionally, genetic rearrangements may occur within a haplotype. Thus the term haplotype is an operational term that refers to the occurrence on a chromosome of linked loci" [75]. The limitation with full exploitation of non-recombining HLA haplotypes in single individuals is that the composite haplotypes can only be inferred by population linkage disequilibrium. This leads to the reporting of 2,3,4 and occasionally 5 loci multi-allele inferred haplotypes in populations, as in the case of the three studies referred to above. There are only two approaches to provision of direct assignment of haplotypes; pedigree inheritance or haploid analysis. The haplotype informativeness of family inheritance, long regarded as the gold standard, is compromised in situations such as where haplotypes are shared between parents, where less than the four haplotypes are transmitted, and where homozygosity disallows unambiguous assignment. Haploid typing is possible using sperm, but that only provides information on 50% of a population. The definitive approach is to investigate single chromosomes but that technology is only now becoming available[76]. Reflecting on the Poisson distribution of population haplotypes, an approach that has occurred to me is to exploit the fact that, while the variety and number of alleles amongst the world populations increases almost daily, in any one population the allelic variety of antigens is highly restricted, often to only two

varieties. In the context of testing the Wee hypothesis for relations between Chinese and Chinese-associated Asian-Pacific populations, at the HLA-A locus there are only four types of A2, 3 types of A9/23-24, two types of A11 (1101/1102), and limited varieties of others (for example 19/33). Similarly, only one type of B46 is found everywhere except very infrequently in Japanese; one type of B58, etc. It is therefore possible to examine the 3-locus allele haplotypes as A-variable / B46 or B58 or B other as invariable / and a range of recombined DRB types. Since the objective is not to estimate the frequency of any one haplotype but to examine the relation between haplotype varieties, individuals can be selected who have only a single B locus allele (either homozygous or heterozygous B-locus Blank). This approach promises to provide a higher level of population relation discernment than current strategies, in turn providing for more informative haplotype trees from which history can be better read.

In the development of HLA typing, interest in the HLA-DPB1 locus been the most recent of the nine major loci (HLA-A,-B,-C, -DRB1,-DRB3,-DRB4,-DRB5,-DQB1,-DPB1), and typing of DPB1 alleles is still not performed in many routine tissue typing laboratories. However, DPB1 may be the locus of choice for population migration studies because it is thought to be the only HLA locus that behaves as a neutral polymorphism[77]. The ideal genetic marker for demographic analysis is one in which there is high polymorphism but where the locus is under neutral selection. All other major HLA Class I (HLA-A, B, C) and Class II (DRB1, DQA1 and DQB1) loci reveal levels of intra-population diversity that are greater than that exhibited for neutral polymorphisms indicating that they are variably affected by balancing selection. DPB1 promises to be most useful as an isolated locus or in haplotypic association with HLA-DPA1 as the partner locus that underlies the heterodimeric pair as a single unit. It is also likely to be most useful when distortion of allele frequency due to centromeric haplotype extension from HLA-DR and HLA-DQ is avoided. However, this situation may not be as simple as first proposed with the finding that, amongst four Polynesian island populations, balancing selection was observed for two at DPA1 locus and for a third at DPB1[78].

Wee *et al.*'s NPC risk origination hypothesis raises the question as to the genetic molecular origin of the predisposition. They view transmission of NPC as involving relatively small numbers of "very high NPC founder" southern Chinese women (Tanka, ship convoys to Africa) and so propose a sex-linked trait carried by females to explain the male gender predominance of NPC. They cite other malignancies (colorectal, prostate, renal, oesophageal) involving comparative genomic hybridization and mtDNA haplogroups in support of their suggestion. It is important to distinguish between sex-linked traits and genetic events involving autosomal genes that are gender associated. If a single autosomal genetic event occurred then, by founder effect, closed colony breeding, or other non-random mating process, there is an increased likelihood of a double occurrence of the risk type as required for recessive inheritance.

Another epidemiological consideration is that, among southeast China Cantonese, overseas Cantonese speakers, and elsewhere where Bai-Yue speakers and their descendants travelled, NPC incidence is progressively undergoing 'risk dilution' due to intermarriage with low or no-risk populations, and possibly also by life-style changes[79]. By contrast, among peoples that remain isolated for geographic or socio-cultural reasons, such as the Nagas in Northeast India and the Tanka boat people in southeast China, NPC occurrence rates remain high. In China the incidence in southern provinces of Guangdong and Hong Kong has diminished with migration southward of low risk Han. Overseas, for instance in Singapore and California, NPC incidence has also diminished in the last 30–40 years. Is dilution by inter-marriage a sufficient explanation to account for this falling incidence, or is a changed environment, in particular viral exposure and dietary practices, also a contributory factor?

The evidence that must be considered in the attempt to understand the advent of NPC risk is that, among all the marker loci studied in NPC for patient population risk and resistance, the most consistent genetic finding is altered frequencies of HLA-A and -B loci alleles and haplotypes among Chinese patients. HLA typing of NPC patients over the last 40 years has confirmed that the combinations that were first detected in the 1970's [21, 38-44] remain the strongest associations, including HLA-A*02-HLA-B*4601; HLA-A*33-HLA-B*58; and HLA-A*11-HLA-B*13 [80-82].

The application of haplotype sharing in multiple case family studies for disease susceptibility gene identification was first described in the 1976[83]. Subsequently a sib pair linkage study indicated that the association of candidate risk alleles was incompatible with a dominant model of inheritance[84]. Rather that study provided strong evidence for the existence of a southern Chinese specific, autosomal, major risk determinant for which a recessive model provided the best fit. Further support for a recessive model was later provided by a case-control study[85]. This situation is reminiscent of the situation in Type 1 Diabetes (T1D) in which the highest risk in Caucasians is conferred by heterozygosity of DR3 and DR4 haplotypes[86]. In T1D research it is understood that resolution of HLA allele risk in different populations is best investigated by scoring the genotype, then assigning the two haplotypes, either by pedigree inheritance, or by maximum likelihood algorithms based on population allele co-occurrence frequency patterns. Genotype analysis and reporting has not been prominent in NPC reports, where the practice has been to score single alleles, and putative multi-allele combinations. Under a genetic model of recessivity, both chromosome sixes are involved in NPC so the genotype must be considered.

As early as 1975 it was realized that HLA-associated NPC risk may not be attributable to the HLA types of HLA-A2 and HLA-B.Sin2, whether as single antigens or as a haplotype, but represent the existence of Disease Susceptibility (DS) genes in linkage disequilibrium with alleles of the HLA loci[21]. The risk attributable to A2-B46 accounts for only 20% of the excess NPC risk[53]. It was proposed that NPC-DS genes may determine differences in immune responsiveness to environmental agents, and thereby determine differences in NPC incidence. If the known altered immune responsiveness of NPC patients to EBV reflected the function of DS genes linked to the high NPC risk

HL-A type, then the hypothesis that Epstein-Barr virus has an etiological role in NPC could be tested by several types of prospective studies[21].

In family studies, evaluating the joint effect of HLA-A*0207 and HLA-B*4601, individuals who were carriers of HLA-A*0207 with or without the presence of HLA-B*4601 had a 1.9-fold (95% CI = 1.0–3.4) and 2.1-fold (95% CI = 0.83–5.3) risk of NPC, respectively[81]. It is often asserted that relative risks of the order of two-fold are incompatible with a strong genetic component, and instead imply a complex multigenic mechanism. On the contrary, relative risks of this order of two-fold are the expected degree of familial risk if the risk was mediated by a single gene [87]. In a linkage study based on affected sib pairs the maximum likelihood NPC DS gene allele frequency was ~0.29, corresponding to a risk homozygote frequency of 8.4%, and the relative risk maximum likelihood estimate was 21 (upper confidence 90% bound-infinity; lower 90% level bound-9.8) [84]. These results show the existence of an NPC DS gene allele conferring an increased risk for NPC some 10-fold larger than that associated with B46 or B17/58.

In a case-control study in 2005, the number of alleles for which a statistically significant role, (either predisposing or protective) was detected increased from those reported in 1975[45] and 1983[52] (A2, A33, A11, B13, B46, B58) to include HLA-A31, -B27, -B38, -B39, -B55, and -B61[85]. The HLA allele Odds Ratio showed a 10-fold range from highest predisposing (A2–1.81) to lowest protective (B39–0.18), again suggesting that HLA alleles themselves were not the risk factors. The necessity to resolve serologically detected antigens into DNA alleles is illustrated by the B16 antigen group. Earlier I recorded that the first antigen within the HLA-B.Blank to be detected in Singapore was HLA-B. Sin-1 (B16–38/39). With sequencing resolution of B16 into the two component DNA alleles of HLA-B*3801 and HLA-B*3901, B16/38 was found to be predisposing (OR 1.73) while B16/39 was protective (OR 0.18), a 9.6 fold difference in alternative risk which is of the same order as the 10-fold span across all alleles. However, inability to resolve the serologically detectable antigen into the DNA types resulted in failure to detect this log-fold difference between one allele conferring resistance and the other being associated with susceptibility.

For HLA-A31 the difference was even more stark, and even more revealing: 34 (4.3%) positives in 778 normal subjects versus 0/734 NPC patients. As the authors point out, the fact that a single HLA-A31 allele is associated with a very low risk indicates that high risk for NPC requires an individual to be homozygous for the NPC DS allele; that is, the DS allele has to be recessive, with only DS allele homozygosity conferring high risk[86].

The populations involved in the Wee *et al*. hypothesis can be stratified into four groups:

(1) Southeastern Chinese resident in China;

(2) Chinese resident overseas;

(3) local dispersed populations-Taiwan Aborigines; Nagalanders in Northeast India; Bidayuh (Land Dyaks) in Sarawak, East Malaysia; Arctic Inuit;

(4) distant dispersed populations—North Africa Maghreb; East Africa.

Apart from China residents and overseas Chinese, HLA data is most extensive for North African Maghreb. In 154 Moroccan NPC patients the HLA-A locus allele A9/23 was significantly reduced ($pc < 0.023$, $RR = 0.45$) [88]. More dramatically, and comparable to the situation for HLA-A31 in Chinese, HLA-A9/23 occurred in 18/100 normal Tunisian subjects but was undetected in 45 Tunisian NPC patients [89]. HLA-A9/23 is a common HLA-A locus antigen in Tunisians and all other Arab populations [90]. Confirming and extending a finding by low resolution serological methods of a decreased frequency of HLA-B14 in Tunisian NPC patients [91], Marincola and colleagues used sequence-based typing to demonstrate that the HLA-B14/C locus haplotype (HLA-B14/Cw08) frequency was 0.007 in Tunisian NPC patients compared to 0.057 (OR=0.12) in both Tunisian and Moroccan controls[92]. This study considered the bimodal age of occurrence by showing subjects in 5-year age groups but it is not clear whether a distinction was made between age of testing and age at diagnosis which is the important basis of patient stratification. A third study also found a decrease in HLA-B14 (1.3% vs. 16%) [93]. These findings in two Maghrebian populations of protective effects associated with HLA Class I alleles support the possibility that excess Maghreb NPC risk might also be due to HLA-linked NPC DS allele homozygosity. Demonstration of a common NPC genetic risk basis in Chinese and North Africans would support an hypothesis of maritime migration to the Mediterranean from Asia, and thus an unitarian hypothesis of global NPC excess genetic risk.

If recessivity is the most likely genetic model, involving loci on both chromosomes, then the question arises as to how such a process began. It is difficult to imagine that a double event arose at a single time point in a single individual. However the minimum essential change arose, it must have been transmitted to subsequent generations, and to a sufficient number of descendant individuals, for chromosome NPC DS locus homozygosity to occur. Although NPC occurs in some teenage children, the numerical contribution of the "early age of onset patients", typically defined as $\leqslant$ 25 years, must have been insufficient to stem the transmission of a trait conferring risk for a cancer that mostly has an onset in the post-reproductive age, particularly in females. The Bai-Yue origination hypothesis requires a trait expansion in this progenitor population sufficient to provide for dissemination in multiple directions from its geographic region of origin.

A further question is how an HLA-associated risk can arise from a presumed single source on a single HLA haplotype as background yet manifest as a variety of HLA types in different NPC patient populations. The environment, particularly EB viral exposure, and of virotype variations in different locations[94-96], can be expected to have a significant effect on a somatic genotype, strongly influencing the strength of any selective advantage or disadvantage. In conjunction with variable break-up of multi-locus haplotypic blocks by recombination, "hitchhiker effect" of linkage to a selectively favored allele at a different locus within or across haplotype blocks, and other population genetic phenomena, it is easy to imagine how, over millennia, alleles at a locus of interest

could become differently associated with allele types at adjacent and distant loci.

It is certain that, at a single point in time, a genetic event did occur in a single individual, in a gene that is highly associated with NPC. That event was the genesis of the HLA-B locus allele, B*4601, first discovered in January 1975 [43] and independently confirmed one month later[49]. Unlike virtually all other HLA Class I alleles which arise by conversions between existing alleles of the same locus, HLA-B*4601 arose by inter-locus recombination between HLA-C*0102 and HLA- B*1501, converting B*4601 to an HLA-C-like molecule [97]. The allele is confined to Mongoloid populations, but has had sufficient time to undergo upstream and downstream recombinations in mainland and overseas Chinese, Japanese, and Korean populations. The B*4601 molecule has novel peptide-binding properties which appear to confer both selective advantage as well as negative effects. Other immunofunctional characteristics are that alloreactive cytotoxic T lymphocytes, natural killer cells and cellular glycosidases all recognize B*4601 as though it were an HLA-C allotype. Thus individuals who are homozygous for B*4601 have, in effect, four HLA-C functioning molecules and no HLA-B allelotypes.

How might molecules such as HLA-B*4601 and other HLA allele combinations confer risk to, or protection from, NPC? It was understood in the 1970's that if the aetiopathogenic mechanism in NPC involved EBV infection and immunity then HLA typing could test the hypothesis that EBV was an aetiological factor in NPC [21,22]. With convergence of EBV molecular biology and HLA immunogenetics through allele variability in viral peptide presentation, these possibilities can be investigated by examining the observed and computer-simulated EBV peptide binding avidity and affinity with candidate predisposing and protective alleles. First and even more distant relatives[98] of NPC patients have increased risk so these subjects can serve as a conservative comparison group in that they will include "false negatives"—subjects who are currently unaffected, but some of whom will become so in time. Many groups are expecting that the knowledge arising from peptide presentation research, with preferential binding or failure to bind of EBV peptide subsets, may be applied to protective manipulation of the EBV immune response by vaccination similar to HPV vaccination that is being applied for protection against development of cervical carcinoma.

In the 1970's it was not known whether the risk associated with HLA-A2 and HLA-B.Sin-2 (*4601) indicated disease susceptibility in a centromeric or a telomeric direction because only the Class I loci had been discovered. It was just emerging that the genetics of mixed lymphocyte reaction was revealing a genetic system or systems separate from, and centromeric to, Class I [99]. There were early indications from family studies of Singapore patients that what became known as HLA Class II genes were involved in NPC [42]. When HLA Class II genetics became resolved into separate loci defining the three sub-regions (DR,DQ,DP), it became possible to assign combinations of Class I and Class II loci alleles according to their population co-occurrence frequencies, or as haplotypes when family analyses revealed combinational segregation. However, in 2010, any significance of the Class I HLA-B locus—Class II HLA-DRB loci "breakage" of allele co-occurrence in NPC remains to be revealed.

Particularly during this decade, genome-wide association studies (GWAS) have come into vogue for disease gene region delineation. The first description in 1990 of GWAS as a strategy for disease gene risk identification in single individuals in the absence of family members referred to the rationale of linkage disequilibrium fine mapping as detection of haplotype heterogeneity restriction [75]. The first application of this strategy for disease-gene discovery was in the detection of the haemochromatosis locus 4Mb telomeric of HLA-A[100]. Discovery of the haemochromatosis-causing locus was delayed from 1976[101] for 20 years because of failure to imagine the extent of non-recombining haplotypes telomeric of HLA-A. Disequilibrium values were found to be high over an unusually large region beginning 150 kb centromeric of HLA-A and extending nearly 5 Mb telomeric of it[102]. Also, recombination was as low as 28% of expected. Since this is the region suspected of harbouring the NPC DS gene risk, the challenge for DS identification will be first to define and align haplotypes that extend over megabases. Use of diploid DNA will require algorithims that maximize fine locus mapping of assembled haplotypes, such as FineMap that was tested against hemochromatosis-HLA haplotypes[103]. Ideally, use of haploid DNA will merely require alignment of the sequence pairs in each individual against suitable reference haplotypes. The population-based linkage disequilibrium map of chromosome 6p derived from Singapore Chinese will be an essential resource for analysis of haploid sequence variation[104].

After more than three decades of case-control reports of chromosome 6p21.3-located HLA-associated risk, some NPC genetic researchers found it surprising that the first GWAS report of linkage analysis of familial NPC failed to identify risk associated with chromosome 6, instead suggesting the existence in a subset of NPC families of a susceptibility locus on chromosome 4 [105]. Confirmation of 6p21.3 as a region already established by case-control studies as having high relative risk by later GWA studies can be attributed to use of much larger patient and control populations and to higher SNP density sufficient to provide statistical power sufficient to overcome the effects of a recessive mode of inheritance, a DS allele frequency in the vicinity of 30%, and (given a cumulative NPC rate to age 75 [lifetime risk] of ~2%), penetrance of the homozygous genotype of ~20% . A subsequent GWAS study focused on an 18-megabase region spanning the previously implicated chromosome 4 region but the significance of candidate risk microsatellite alleles failed to withstand statistical correction [106]. Another microsatellite marker study also failed to support a risk locus in this chromosome 4 subregion[107]. More recently, a panel of polymorphic markers for the chromosome 4 subregion, and for the short arms of previously implicated risk chromosomes 3 and 9, again failed to support 4p15.1–q12 as well as a risk region on chromosome 9, but did find support for a 13.6-cM region on 3p21.31–21.2 [108]. Support for an NPC susceptibility locus in this megabase region was provided by a family-based association analysis of haplotype transmission disequilibrium [109]. A recent

microsatellite survey of autosomal chromosomes, again at average marker distance of 5 cM, did not incriminate either the HLA region at 6p21.3, or the 4p15.1–q12 region, chromosome 9, or the 3p21.31–21.2 region but instead proposed an increased LOD score for 5p13[110].

There are indications that any single HLA region of NPC risk is more likely to be around the telomeric HLA-A loci region. While the HLA-A2 association dates to 1974[41,42], it was not until 1995 that analysis of A2 allelic microvariants revealed that NPC patients had different HLA patterns, both as A2 subtypes, and as A2-B46 combinations[111]. A recent detailed allele analysis found that HLA-B*4601 lacked a positive association with NPC in the absence of HLA-A*0207, favouring a primary role for risk in association with the HLA-A locus [81]. An earlier linkage study based on the affected sib pair method that was originally described in 1976 [83] had indicated that risk is conferred within inherited HLA chromosomal haplotype segments by a gene closely linked to the HLA locus but not necessarily by the HLA loci per se for which the maximum likelihood relative risk estimate was approximately 21 [84]. Microsatellite studies supported this approach of trying to delimit the HLA complex region [112], preparatory to ever further gene fine mapping. An HLA-A candidate risk region[113] was further reduced to a 132kb segment in the vicinity of the HLA-A locus using microsatellite markers distributed over a 1 megabase region[114].

In 2009 a GWAS approach utilizing nearly one-half million single nucleotide polymorphic markers identified a similar region in around the HLA-A locus, spanning to include HLA-F and the gamma amino butyric acid b receptor 1 locus[115]. The limitations of widely separated marker GWAS should be apparent from the inconsistencies of published reports. As the inventor of this technology[75], I remain surprised that so much effort is invested in seeking statistically significant associations when employing markers that are millions of bases apart. Ultimately, there is no substitute for continuous chromosome sequencing because allelotyping separated loci can be very misleading. For instance, the Caucasian HLA haplotype B8-DR3 that is equivalent to the Chinese A2-B46 combination is increased in frequency in Type 1 Diabetes (T1D). Northern Indians also have an increased frequency of HLA-B8-DR3 in T1D. With the same major HLA locus markers in both populations it might be expected that the intervening sequence would also be the same. However, the Indian chromosome differs at all the loci over the intervening chromosomal length [116]. Clearly, the selective advantage associated with the classical HLA loci differs from that of the intervening loci.

Most recently, an extensive GWAS involving 1 615 individual patients from Guangzhou and a total of 2 033 controls from Guangzhou (1025) and Singapore (1008) revealed three strong associations, all within the HLA region, the strongest locating to the Class I HLA-A region (personal communication with Bei JX in 2010). While there were hints of associations with chromosomes 1,3,7 and 13, there were no associations with the previously reported 3p21, 4p15.1–q12 or 5p13. Thus the two most recent GWAS, as reported by Tse *et al*. [115] and stated by Bei JX (personal communication in 2010), both support the long-standing population association findings of HLA involvement in NPC genetic risk.

Considering the above studies, particularly the two extensive GWAS, the priority chromosome segment to sequence is the few megabases on either side of HLA-A. Ideally, continuous sequencing should be extended to include both ends of the 4 Mb core HLA complex, say 4 Mb telomeric of HLA-A to the Hemochromatosis locus, and ~5 Mb centromeric of HLA-DP to Glyoxylase1. The recent advent of high throughput multiplex single molecule sequencing (SMS) provides the opportunity to readily examine such 13 Mb-length targeted genome segments. However, current sequence analysis is complicated by the "phase problem" associated with using diploid DNA for SMS haplotype assembly[117]. The "phase problem" is stated as: "Whole genome sequencing is required to know which variations or mutations are on which one of the two chromosomes. Variations on the same chromosome are likely to have different functional meaning than if they are on opposite chromosomes. Sequencing a selective part of the genome does not provide full information about this 'phase' issue"[117].

Sequencing of each of the two haploid chromosomes, as targeted regions such as each of the two 13 Mb lengths, eventually extending to coverage of the entire genome chromosome pairs, will enable identification of chromosome segments as contiguous haplotype blocks, overcoming the phase problem, and ensuring that genomic analysis definitively addresses the di-haploid state of the inherited elements. This definitive strategy of "reductio ad inheritum" dispenses with all inferences and current complications conferred by the "phase problem". Reductio ad inheritum haploid sequencing that eliminates the chromosome phase problem is now a realizable goal[76].

Young NPC patients have the highest likelihood of having parents who can also be sequenced to provide for comparison of pedigree haplotype assignment with single chromosome haploid sequence. If an NPC genetic risk lesion can be identified in high risk Chinese, then a direct test of the Wee *et al*. common NPC genetic risk origination hypothesis is to seek the predicted similar risk genetic lesion in NPC patients from other high risk populations.

Acknowledgements

Over 40 years I have been inspired and supported by many colleagues, too numerous to list, but whose contribution to my thinking has been invaluable in the development of my understanding, summarized within this Commentary. Two of those colleagues of longest standing, and most significant in their contribution, are Professor Nicholas Day and Dr. Brian Tait. Professor Day's contribution has been to interpretation of HLA patterns in NPC. Unravelling HLA-NPC association typing results by a strategy focusing on protective alleles that unifies the likely NPC genetic excess risk mechanism in seemingly disparate Chinese and Maghreb populations might come to be regarded as the equivalent of deciphering the Rosetta Stone of HLA-cancer association. I thank Dr. Brian Tait for discussions concerning all aspects of HLA-disease associations and most recently on HLA allele differential presentation of EBV peptides. I also wish to

thank Dr. Jeppe Friborg for clarification of published reports on Arctic Inuit, and Dr. David Turnbull for bringing to my attention publications on inference of population demographic histories from molecular genetic data, the subject matter of which is summarized by the article entitled 'Can we read history from haplotypes?' [73].

References

[1] Li CC, Yu MC , Henderson BE. Some epidemiologic observations of nasopharyngeal carcinoma in Guangdong, People´s Republic of China [J]. Natl Cancer Inst Monogr, 1985,69:49–52.

[2] Kumar S, Mahanta J. Aetiology of nasopharyngeal carcinoma. A review [J]. Indian J Cancer, 1998,35(2):47–56.

[3] Devi BC, Pisani P, Tang TS, Parkin DM. High incidence of nasopharyngeal carcinoma in native people of Sarawak, Borneo Island [J]. Cancer Epidemiol Biomarkers Prev, 2004,13(3):482–486.

[4] Wang CJ, Lin FJ, Leung W, et al. Occurrence of nasopharyngeal carcinoma in aboriginals of Taiwan: report of 14 cases [J]. J Formos Med Assoc, 1992,91(11):1044–1048.

[5] Ko YC, Liu BH, Hsieh SF, et al. Cancer mortality analysis among aborigines in Taiwan [J]. Gaoxiong Yi Xue Ke Xue Za Zhi, 1994,10(7): 379–391. [in Chinese]

[6] Nielsen NH, Mikkelsen F, Hansen JP. Nasopharyngeal cancer in Greenland. The incidence in an Arctic Eskimo population [J]. Acta Pathol Microbiol Scand A, 1977,85(6):850–858.

[7] de-Thé G, Ambrosioni JC, Ho HC, et al. Lymphoblastoid transformation and presence of herpes-type viral particles in a Chinese nasopharyngeal tumour cultured in vitro [J]. Nature, 1969,221(5182):770–771.

[8] Zur Hausen H, Schulte-Holthausen H, Klein G, et al. EBV DNA in biopsies of Burkitt tumours and anaplastic carcinomas of the nasopharynx [J]. Nature, 1970,228(5276):1056–1058.

[9] Liebowitz D. Nasopharyngeal carcinoma: the Epstein-Barr virus association [J]. Semin Oncol, 1994,21(3):376–381.

[10] Takada K, Shimizu N, Tanabe-Tochikura A, et al. Pathogenic role of Epstein-Barr virus in human cancer [J]. Intervirology, 1995,38(3–4):214–220.

[11] Pagano JS. Epstein-Barr virus: the first human tumor virus and its role in cancer [J]. Proc Assoc Am Physicians, 1999,111(6):573–580.

[12] Hsu JL, Glaser SL. Epstein-barr virus-associated malignancies: epidemiologic patterns and etiologic implications [J]. Crit Rev Oncol Hematol, 2000,34(1):27–53.

[13] Dolcetti R, Menezes J. Epstein-Barr virus and undifferentiated nasopharyngeal carcinoma: new immunobiological and molecular insights on a long-standing etiopathogenic association [J]. Adv Cancer Res, 2003,87:127–157.

[14] Young LS, Murray PG. Epstein-Barr virus and oncogenesis: from latent genes to tumours [J]. Oncogene, 2003,22(33):5108–5121.

[15] Chou J, Lin YC, Kim J, et al. Nasopharyngeal carcinoma—review of the molecular mechanisms of tumorigenesis [J]. Head Neck, 2008,30(7): 946–963.

[16] Niller HH, Wolf H, Minarovits J. Epigenetic dysregulation of the host cell genome in Epstein-Barr virus-associated neoplasia [J]. Semin Cancer Biol, 2009,19(3):158–164.

[17] Shah KM, Young LS. Epstein-Barr virus and carcinogenesis: beyond Burkitt´s lymphoma [J]. Clin Microbiol Infect, 2009,15(11):982–988.

[18] Maeda E, Akahane M, Kiryu S, et al. Spectrum of Epstein-Barr virus-related diseases: a pictorial review [J]. Jpn J Radiol, 2009,27(1):4–19.

[19] Pagano JS, Blaser M, Buendia MA, et al. Infectious agents and cancer: criteria for a causal relation [J]. Semin Cancer Biol, 2004,14(6):453–471.

[20] de Martel C, Franceschi S. Infections and cancer: established associations and new hypotheses [J]. Crit Rev Oncol Hematol, 2009,70 (3):183–194.

[21] Simons MJ, Wee GB, Chan SH, et al. Immunogenetic aspects of nasopharyngeal carcinoma (NPC) III. HL-A type as a genetic marker of NPC predisposition to test the hypothesis that Epstein-Barr virus is an etiological factor in NPC [J]. IARC Sci Publ, 1975,(11 Pt 2):249–258.

[22] Simons MJ. Immunogenetic-immunovirologic strategy in human cancer research: application to carcinomas of the nasopharynx and liver [D], Tokyo, Japan: University of Tokyo, 1975.

[23] Simons MJ, Chan SH, Ou BX. Nasopharyngeal Carcinoma (NPC), including analysis of HLA gene patterns in Chinese patients with Cervical and Hepatocellular Carcinoma [M]. Simons MJ, Tait BD, eds. Proceeding of the Second Asia and Oceania Histocompatibility Workshop Conference: Immunopublishing, Publishing Division of Immunosearch Unit Trust, 445 Toorak Road, Toorak, 3142, Victoria, Australia (ISBN 0 9591744 0 0), 1983:369–378.

[24] Ho JH, Huang DP, Fong YY. Salted fish and nasopharyngeal carcinoma in southern Chinese [J]. Lancet, 1978,2(8090):626.

[25] Fong LY, Ho JH, Huang DP. Preserved foods as possible cancer hazards: WA rats fed salted fish have mutagenic urine [J]. Int J Cancer, 1979,23(4):542–546.

[26] Yu MC, Ho JH, Lai SH, et al. Cantonese-style salted fish as a cause of nasopharyngeal carcinoma: report of a case-control study in Hong Kong [J]. Cancer Res, 1986,46(2):956–961.

[27] Fong YY, Walsh EO. Carcinogenic nitrosamines in Cantonese salt-dried fish [J]. Lancet, 1971,2(7732):1032.

[28] Mirvish SS. N-nitroso compounds: their chemical and in vivo formation and possible importance as environmental carcinogens [J]. J Toxicol Environ Health, 1977,2(6):1267–1277.

[29] Zou XN, Lu SH, Liu B. Volatile N-nitrosamines and their precursors in Chinese salted fish—a possible etological factor for NPC in china [J]. Int J Cancer, 1994,59(2):155–158.

[30] Mirvish SS. Role of N-nitroso compounds (NOC) and N-nitrosation in etiology of gastric, esophageal, nasopharyngeal and bladder cancer and contribution to cancer of known exposures to NOC [J]. Cancer Lett, 1995,93(1):17–48.

[31] Dodd LE, Sengupta S, Chen IH, et al. Genes involved in DNA repair and nitrosamine metabolism and those located on chromosome 14q32 are dysregulated in nasopharyngeal carcinoma [J]. Cancer Epidemiol Biomarkers Prev, 2006,15(11):2216–2225.

[32] Chen CJ, Liang KY, Chang YS, et al. Multiple risk factors of nasopharyngeal carcinoma: Epstein-Barr virus, malarial infection, cigarette smoking and familial tendency [J]. Anticancer Res, 1990,10 (2B):547–553.

[33] Zhang F, Zhang J. Clinical hereditary characteristics in nasopharyngeal carcinoma through Ye-Liang´s family cluster [J]. Chin Med J (Engl), 1999,112(2):185–187.

[34] Jia WH, Feng BJ, Xu ZL, et al. Familial risk and clustering of nasopharyngeal carcinoma in Guangdong, China [J]. Cancer, 2004,101 (2):363–369.

[35] Shanmugaratnam K. Variations in nasopharyngeal cancer incidence among specific Chinese communities (dialect groups) in Singapore [J]. IARC Sci Publ, 1978,20:191–198.

[36] Meredith RF, Okunewick JP. Genetic influence in murine viral leukemogenesis [J]. Biomedicine, 1976,24(6):374–380.

[37] Mühlbock O, Dux A. Histocompatibility genes (the H-2 complex) and susceptibility to mammary tumor virus in mice [J]. J Natl Cancer Inst, 1974,53(4):993–996.

[38] Tennant JR, Snell GD. The H-2 locus and viral leukemogenesis as studied in congenic strains of mice [J]. J Natl Cancer Inst, 1968,41(2): 597–604.

[39] Hirayama T. Descriptive and analytical epidemiology of nasopharyngeal cancer [J]. IARC Sci Publ, 1978,20:167–189.

[40] Simons MJ, Wee GB, Day NE, et al. Immunogenetic aspects of nasopharyngeal carcinoma: I. Differences in HL-A antigen profiles between patients and control groups [J]. Int J Cancer, 1974,13(1):122–134.

[41] Simons MJ, Day NE, Wee GB, et al. Nasopharyngeal carcinoma V: Immunogenetic studies of Southeast Asian ethnic groups with high and low risk for the tumor [J]. Cancer Res, 1974,34(5):1192–1195.

[42] Simons MJ, Chan SH, Ho JHC, et al. A Singapore 2—Associated LD

Antigen in Chinese Patients with Nasopharyngeal Carcinoma [M]. Kissmeyer-Nielsen F, ed. Histocompatibility Testing 1975: Munksgaard, Copenhagen, 1975: 809–812.
[43] Simons MJ, Wee GB, Chan SH, et al. Probable identification of an HL-A second-locus antigen associated with a high risk of nasopharyngeal carcinoma [J]. Lancet, 1975,1(7899):142–143.
[44] Simons MJ, Day NE. Histocompatibility leukocyte antigen patterns and nasopharyngeal carcinoma [J]. Natl Cancer Inst Monogr, 1977,47:143–146.
[45] Chan SH, Simons MJ. Immunogenetics of nasopharyngeal carcinoma [J]. Annals Acad Med, 1977,6(4):342–346.
[46] Simons MJ, Chan SH, Wee GB, et al. Nasopharyngeal carcinoma and histocompatibility antigens [J]. IARC Sci Publ, 1978,20:271–282.
[47] Feingold N. HLA system and malignant diseases (author´s transl) [J]. Bull Cancer, 1978,65(1):79–82.
[48] Simons MJ, Amiel JL. HLA and Malignant Diseases [M]. Dausset J, Svejgaard A, eds. HLA and Disease: Munksgaard, Copenhagen, 1977: 212–232.
[49] Payne R, Radvany R, Grumet C. A new second locus HL-A antigen in linkage disequilibrium with HL-A2 in cantonese chinese [J]. Tissue Antigens, 1975,5(1):69–71.
[50] Simons MJ, Wheeler R, Cohen D, et al. Restriction fragment length polymorphism of HLA genes. Summary of the Tenth International Workshop Southern blot analysis [M]. Dupont B, ed. Immunobiology of HLA, Vol. 1: Histcompatability testing. New York:Springer-Verlag, 1989: 959–1023.
[51] Mullis KB, Faloona FA. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction [J]. Methods Enzymol, 1987,155: 335–350.
[52] Saiki RK, Bugawan TL, Horn GT, et al. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes [J]. Nature, 1986,324(6093):163–166.
[53] Chan SH, Day NE, Kunaratnam N, et al. HLA and nasopharyngeal carcinoma in Chinese—a further study [J]. Int J Cancer, 1983,32:171–176.
[54] Betuel H, Camoun M, Colombani J, et al. The relationship between nasopharyngeal carcinoma and the HL-A system among Tunisians [J]. Int J Cancer, 1975,16(2):249–254.
[55] Trimeche M, Braham H, Ziadi S, et al. Investigation of allelic imbalances on chromosome 3p in nasopharyngeal carcinoma in Tunisia: high frequency of microsatellite instability in patients with early-onset of the disease [J]. Oral Oncol, 2008,44(8):775–783.
[56] Khabir A, Karray H, Rodriguez S, et al. EBV latent membrane protein 1 abundance correlates with patient age but not with metastatic behavior in north African nasopharyngeal carcinomas [J]. Virol J, 2005,2:39.
[57] Frikha M, Toumi N, Ghorbel L, et al. Pediatric nasopharyngeal carcinoma: Anatomoclinic aspects, therapeutic results and evolutive particularities [J]. Cancer Radiother, 2010 Mar 26. [in French]
[58] Saemundsen AK, Albeck H, Hansen JP, et al. Epstein-Barr virus in nasopharyngeal and salivary gland carcinomas of Greenland Eskimoes [J]. Br J Cancer, 1982,46(5):721–728.
[59] Raab-Traub N, Rajadurai P, Flynn K, et al. Epstein-Barr virus infection in carcinoma of the salivary gland [J]. J Virol, 1991,65(12):7032–7036.
[60] Hamilton-Dutoit SJ, Therkildsen MH, Neilsen NH, et al. Undifferentiated carcinoma of the salivary gland in Greenlandic Eskimos: demonstration of Epstein-Barr virus DNA by in situ nucleic acid hybridization [J]. Hum Pathol, 1991,22(8):811–815.
[61] Albeck H, Nielsen NH, Hansen HE, et al. Epidemiology of nasopharyngeal and salivary gland carcinoma in Greenland [J]. Arctic Med Res, 1992,51(4):189–195.
[62] Friborg J, Wohlfahrt J, Koch A, et al. Cancer susceptibility in nasopharyngeal carcinoma families—a population-based cohort study [J]. Cancer Res, 2005,65(18):8567–8572.
[63] Simons MJ. HLA and Nasopharyngeal Carcinoma: 30 years on [J]. American Society for Histocompatibility and Immunogenetics (ASHI), 2003,27(2):52–55.
[64] Chan SH, Wee GB. HLA and Nasopharyngeal Cancer in Singapore Chinese [M]. Simons MJ, Tait BD, eds. Proceeding of the Second Asia and Oceania Histocompatibility Workshop Conference: Immunopublishing, Publishing Division of Immunosearch Unit Trust, 445 Toorak Road, Toorak, 3142, Victoria, Australia (ISBN 0 9591744 0 0), 1983:495–496.
[65] Lin M, Chu CC, Chang SL, et al. The origin of Minnan and Hakka, the so-called "Taiwanese", inferred by HLA study [J]. Tissue Antigens, 2001,57(3):192–199.
[66] Pimtanothai N, Kangwanshiratada O, Charoenwongse P. Serological analysis of human leukocyte antigens-A and -B antigens in Thai patients with nasopharyngeal carcinoma [J]. J Med Assoc Thai, 2003,86 Suppl 2: S237–S241.
[67] Pimtanothai N, Charoenwongse P, Mutirangura A, et al. Distribution of HLA-B alleles in nasopharyngeal carcinoma patients and normal controls in Thailand [J]. Tissue Antigens, 2002,59(3):223–225.
[68] Hoa BK, Hang NT, Kashiwase K, et al. HLA-A, -B, -C, -DRB1 and -DQB1 alleles and haplotypes in the Kinh population in Vietnam [J]. Tissue Antigens, 2008,71(2):127–134.
[69] Lin M, Chu CC, Lee HL, et al. Heterogeneity of Taiwan's indigenous population: possible relation to prehistoric Mongoloid dispersals [J]. Tissue Antigens, 2000,55(1):1–9.
[70] Pool JE, Hellmann I, Jensen JD, et al. Population genetic inference from genomic sequence variation [J]. Genome Res, 2010,20(3):291–300.
[71] Cavalli-Sforza L. Demic diffusion as the basic process of human expansions [M]. Bellwood P, Renfrew C, eds. Examining the farming / language dispersal hypothesis. McDonald Institute for Archeological Research, Cambridge, UK (ISBN 1-902937-20-1), 2002:79–88.
[72] Oppenheimer S. Following Populations or Molecules? Two contrasting approaches and descriptive outcomes of island colonization arising from a similar knowledge-base [M]. Matsumura S, Forster P, Renfrew C, eds. Simulations, genetics and human prehistory. McDonald Institute for Archeological Research, Cambridge, UK (ISBN 978-1-902937-45-8), 2008:27–36.
[73] Harpending H. Can we read history from haplotypes? [M] Matsumura S, Forster P, Renfrew C, eds. Simulations, genetics and human prehistory. McDonald Institute for Archeological Research, Cambridge, UK (ISBN 978-1-902937-45-8), 2008:193–197.
[74] Ceppellini R, Curtoni ES, Mattiuz PL, et al. Genetics of Leukocyte Antigens: A family study of segregation and linkage [M]. Curtoni ES, Mattiuz PL, Tosi RM, eds. Histocompatibility Testing 1967: The Williams & Wilkins Company, 1967:149–187.
[75] Simons MJ, Inventor; Genetic Technologies Pty. Ltd, Assignee. Genomic mapping method by direct haplotyping using intron sequence analysis [P]. US patent 5851762. Priority date July 11, 1990. Issue date December 22, 1998.
[76] www.haplomictechnologies.com.
[77] Sanchez-Mazas A. An apportionment of human HLA diversity [J]. Tissue Antigens, 2007,69 Suppl 1:198–202.
[78] Varney MD, Perko L, Abbott W, et al. Unusual DPA1 and DPB1 heterogeneity contributes to DP polymorphism in Polynesians [M]. Hansen J, ed. Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatability Workshop and Conference: International Histocompatibility Working Group Press Fred Hutchinson Cancer Research Center, Seattle, Washington, USA (ISBN 0-945278-03-9), 2006:157–160.
[79] Yu WM, Hussain SS. Incidence of nasopharyngeal carcinoma in Chinese immigrants, compared with Chinese in China and South East Asia: review [J]. J Laryngol Otol, 2009,123(10):1067–1074.
[80] Hildesheim A, Apple RJ, Chen CJ, et al. Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan [J]. J Natl Cancer Inst, 2002,94(23):1780–1789.
[81] Yu KJ, Gao X, Chen CJ, et al. Association of human leukocyte antigens with nasopharyngeal carcinoma in high-risk multiplex families in Taiwan [J]. Hum Immunol, 2009,70(11):910–914.
[82] Tang M, Zeng Y, Poisson A, et al. Haplotype-dependent HLA susceptibility to nasopharyngeal carcinoma in a Southern Chinese

population [J]. Genes Immun, 2010,14:1-9.
[83] Day NE, Simons MJ. Disease susceptibility genes—their identification by multiple case family studies [J]. Tissue Antigens, 1976,8(2):109-119.
[84] Lu SJ, Day NE, Degos L, et al. Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region [J]. Nature, 1990,346(6283):470-471.
[85] Hu SP, Day NE, Li DR, et al. Further evidence for an HLA-related recessive mutation in nasopharyngeal carcinoma among the Chinese [J]. Brit J Cancer, 2005,92:967-970.
[86] Bertrams J, Sodomann P, Gries FA, et al. HLA association of insulin-dependent diabetes mellitus type I (author's transl) [J]. Dtsch Med Wochenschr, 1981,106(29-30):927-932.
[87] Tulinius H, Day NE, Bjarnason O, et al. Familial Breast Cancer in Iceland [J]. Int J Cancer, 1982,29(4):365-371.
[88] Dardari R, Khyatti M, Jouhadi H, et al. Study of human leukocyte antigen class I phenotypes in Moroccan patients with nasopharyngeal carcinoma [J]. Int J Cancer, 2001,92(2):294-297.
[89] Mokni-Baizig N, Ayed K, Ayed FB, et al. Association between HLA-A/-B antigens and -DRB1 alleles and nasopharyngeal carcinoma in Tunisia [J]. Oncology, 2001,61(1):55-58.
[90] Ayed K, Bardi R, Gebuhrer L, et al. HLA-A,B,C and DR antigens in a sample of the Tunisian population [J]. Tissue Antigens, 1987,29(5):225-231.
[91] Makni H, Daoud J, Ben Salah H, et al. HLA association with nasopharyngeal carcinoma in southern Tunisia [J]. Mol Biol Rep, 2009 Aug 28.
[92] Li X, Ghandri N, Piancatelli D, et al. Associations between HLA class I alleles and the prevalence of nasopharyngeal carcinoma (NPC) among Tunisians [J]. J Transl Med, 2007,5:22.
[93] Herait P, Tursz T, Guillard MY, et al. HLA-A, -B, and -DR antigens in North African patients with nasopharyngeal carcinoma [J]. Tissue Antigens, 1983,22(5):335-341.
[94] Dardari R, Khyatti M, Cordeiro P, et al. High frequency of latent membrane protein-1 30-bp deletion variant with specific single mutations in Epstein-Barr virus-associated nasopharyngeal carcinoma in Moroccan patients [J]. Int J Cancer, 2006,118(8):1977-1983.
[95] Li DJ, Bei JX, Mai SJ, et al. The dominance of China 1 in the spectrum of Epstein-Barr virus strains from Cantonese patients with nasopharyngeal carcinoma [J]. J Med Virol, 2009,81(7):1253-1260.
[96] Chang CM, Yu KJ, Mbulaiteye SM, et al. The extent of genetic diversity of Epstein-Barr virus and its geographic and disease patterns: a need for reappraisal [J]. Virus Res, 2009,143(2):209-221.
[97] Barber LD, Percival L, Valiante NM, et al. The inter-locus recombinant HLA-B*4601 has high selectivity in peptide binding and functions characteristic of HLA-C [J]. J Exp Med, 1996,184(2):735-740.
[98] Guo X, Johnson RC, Deng H, et al. Evaluation of nonviral risk factors for nasopharyngeal carcinoma in a high-risk population of Southern China [J]. Int J Cancer, 2009,124(12):2942-2947.
[99] Bach FH, Amos DB. Hu-1: Major histocompatibility locus in man [J]. Science, 1967,156(781):1506-1508.
[100] Feder JN, Gnirke A, Thomas W, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis [J]. Nat Genet, 1996,13(4),399-408.
[101] Simon M, Bourel M, Fauchet R, et al. Association of HLA-A3 and HLA-B14 antigens with idiopathic haemochromatosis [J]. Gut, 1976,17(5):332-334.
[102] Ajioka RS, Jorde LB, Gruen JR, et al. Haplotype analysis of hemochromatosis: evaluation of different linkage-disequilibrium approaches and evolution of disease chromosomes [J]. Am J Hum Genet, 1997,60(6),1439-1447.
[103] Lam JC, Roeder K, Devlin B. Haplotype fine mapping by evolutionary trees [J]. Am J Hum Genet, 2000,66(2):659-673.
[104] Yu HX, Chia JM, Bourque G, et al. A population-based LD map of the human chromosome 6p [J]. Immunogenetics, 2005,57(8),559-565.
[105] Feng BJ, Huang W, Shugart YY, et al. Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4 [J]. Nat Genet, 2002,31(4):395-399.
[106] Guo XC, Scott K, Liu Y, et al. Genetic factors leading to chronic Epstein-Barr virus infection and nasopharyngeal carcinoma in South East China: study design, methods and feasibility [J]. Hum Genomics, 2006,2(6):365-375.
[107] Guo XC, O´Brien SJ, Winkler C, et al. Association study of chromosome 4 STRs polymorphisms with nasopharyngeal carcinoma [J]. Yi Chuan, 2006,28(7):783-790.
[108] Xiong W, Zeng ZY, Xia JH, et al. A susceptibility locus at chromosome 3p21 linked to familial nasopharyngeal carcinoma [J]. Cancer Res, 2004,64(6):1972-1974.
[109] Zeng Z, Zhou Y, Zhang W, et al. Family-based association analysis validates chromosome 3p21 as a putative nasopharyngeal carcinoma susceptibility locus [J]. Genet Med, 2006,8(3):156-160.
[110] Hu LF, Qiu QH, Fu SM, et al. A genome-wide scan suggests a susceptibility locus on 5p 13 for nasopharyngeal carcinoma [J]. Eur J Hum Genet, 2008,16(3):343-349.
[111] Ren EC, Law GC, Chan SH. HLA-A2 allelic microvariants in nasopharyngeal carcinoma [J]. Int J Cancer, 1995,63(2):213-215.
[112] Ooi EE, Ren EC, Chan SH. Association between microsatellites within the human MHC and nasopharyngeal carcinoma [J]. Int J Cancer, 1997,74(2):229-232.
[113] Lu CC, Chen JC, Jin YT, et al. Genetic susceptibility to nasopharyngeal carcinoma within the HLA-A locus in Taiwanese [J]. Int J Cancer, 2003,103(6):745-751.
[114] Lu CC, Chen JC, Tsai ST, et al. Nasopharyngeal carcinoma-susceptibility locus is localized to a 132 kb segment containing HLA-A using high-resolution microsatellite mapping [J]. Int J Cancer, 2005,115(5):742-746.
[115] Tse KP, Su WH, Chang KP, et al. Genome-wide association study reveals multiple nasopharyngeal carcinoma-associated loci within the HLA region at chromosome 6p21.3 [J]. Am J Hum Genet, 2009,85 (2):194-203.
[116] Witt CS, Price P, Kaur G, et al. Common HLA-B8-DR3 haplotype in Northern India is different from that found in Europe [J]. Tissue Antigens, 2002,60(6):474-480.
[117] http://genomics.xprize.org/archon-x-prize-for-genomics/why-whole-genome-sequencing.