# Genetic factors leading to chronic Epstein–Barr virus infection and nasopharyngeal carcinoma in South East China: Study design, methods and feasibility

*Xiu Chan Guo,[1,5] Kevin Scott,[1] Yan Liu,[2] Michael Dean,[2] Victor David,[2] George W. Nelson,[1] Randall C. Johnson,[1] Holli H. Dilks,[2] James Lautenberger,[2] Bailey Kessing,[1] Janice Martenson,[2] Li Guan,[1] Shan Sun,[2] Hong Deng,[3] Yuming Zheng,[3] Guy de The,[4] Jian Liao,[5] Yi Zeng,[6]* Stephen J. O'Brien[2]** and Cheryl A. Winkler[1]*

[1] Laboratory of Genomic Diversity, SAIC Frederick, National Cancer Institute-Frederick, Frederick, MD 21702, USA
[2] Laboratory of Genomic Diversity, National Cancer Institute-Frederick, Frederick, MD 21702, USA
[3] Cancer Institute of Wuzhou, Wuzhou 543002, Guangxi, China
[4] Institut Pasteur, 75724 Paris, France
[5] Cangwu Institute for Nasopharyngeal Carcinoma Control and Prevention, Wuzhou, Guanxi, China
[6] Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China
*Correspondence to*: *Tel/Fax: +86 6354432; E-mail: zengy@public.bta.net.cn; **Tel: +1 301 846 1296; Fax: +1 301 846 1686; E-mail: obrien@ncifcrf.gov

*Date received (in revised form): 12th April 2006*

## Abstract

Nasopharyngeal carcinoma (NPC) is a complex disease caused by a combination of Epstein–Barr virus chronic infection, the environment and host genes in a multi-step process of carcinogenesis.The identity of genetic factors involved in the development of chronic Epstein–Barr virus infection and NPC remains elusive, however. Here, we describe a two-phase, population-based, case-control study of Han Chinese from Guangxi province, where the NPC incidence rate rises to a high of 25–50 per 100,000 individuals. Phase I, powered to detect single gene associations, enrolled 984 subjects to determine feasibility, to develop infrastructure and logistics and to determine error rates in sample handling. A microsatellite screen of Phase I study participants, genotyped for 319 alleles from 34 microsatellites spanning an 18-megabase region of chromosome 4 (4p15.1-q12), previously implicated by a linkage analysis of familial NPC, found 14 alleles marginally associated with developing NPC or chronic immunoglobulin A production ($p = 0.001 - 0.03$). These associations lost significance after applying a correction for multiple tests. Although the present results await confirmation, the Phase II study population has tripled patient enrolment and has included environmental covariates, offering the potential to validate this and other genomic regions that influence the onset of NPC.

***Keywords:*** *nasopharyngeal carcinoma, chromosome 4, microsatellite, association study, Epstein–Barr virus*

## Introduction

Nasopharyngeal carcinoma (NPC) is a disease with distinct racial and geographical distributions. In southern China, Taiwan, Vietnam and the Philippines, the incidence of NPC is 15–20 per 100,000 individuals per year, and in some local Chinese regions bordering the Xijiang River drainage in Guangdong and Guangxi provinces, the incidence is as high as 25–50 per 100,000 individuals.[1,2] An intermediate incidence is observed among the Arab populations of Northern Africa,[3] including Saudi Arabia;[4] in the Caribbean; and in the Eskimo populations of Alaska and Greenland.[5] Elsewhere, NPC is rare, with an incidence of less than 1 per 100,000. In the USA, NPC comprises only 0.2 per cent of all malignancies, with an incidence is 1 per 100,000. The male:female ratio for NPC is usually 2 or 3:1, with an incidence peak between 50 and 59 years of age.[6]

A link between NPC and Epstein–Barr virus (EBV) was reported in 1966.[7] Ten years later, the presence of

immunoglobulin (Ig) A antibodies to EBV viral capsid antigens (EBV/IgA/VCA) was found to serve as a predictive marker for the development of NPC in Chinese populations.[8] More than 95 per cent of adults in all ethnic groups across the world are healthy carriers of EBV. In high NPC incidence regions, EBV infection of the nasopharyngeal epithelium induces IgA antibodies against VCA, suggesting that reactivation of EBV replication at the mucosal surface precedes the development of NPC. Consistent with this, approximately 2.5 per cent of the general population are EBV/IgA/VCA antibody positive. Of these, less than 3 per cent will develop NPC, while >95 per cent of all NPC patients are EBV/IgA/VCA antibody positive.[9–14] In addition to EBV infection, case control studies have indicated a role for environmental factors, including food preservatives (carcinogenic nitrosamines), salt-preserved fish and phorbol esters in herbs and plants that are commonly consumed among ethnic populations with the highest NPC rates.[15,16]

Evidence for genetic modulation of NPC risk has accumulated recently. Familial aggregation of NPC has been observed in China and in other countries.[17–19] Familial aggregation of NPC is uncommon in low-risk or non-Chinese populations. The proportion of NPC with affected first-degree family history is >5 per cent in south China, 7.2 per cent in Hong Kong, 6.0 per cent in Yulin and 5.9 per cent in Guangzhou.[20] Descendants of south Chinese immigrants to western countries show progressively lower risk, but their NPC incidence remains higher than that of the indigenous population,[21] suggesting both environmental and genetic components to disease susceptibility. Several studies have shown associations between *HLA* genes and NPC,[22–28] and the D6S1624 microsatellite within the *HLA* class I region has been associated with NPC.[29] Studies comparing age of NPC onset report conflicting results for familial versus sporadic NPC. In a study comparing 200 probands with and without NPC-affected first-degree relatives from Singapore, the age of onset was 48 and 49 years, respectively.[30] In another Chinese study, the average age of onset was 35.5 years in 32 Guangdong families with 4–5 relatives with NPC compared with 46.6 years for sporadic cases.[20] In a third study, however, the age of onset decreased from 44.5 years to 40.4 as the number of NPC-affected relatives increased from one to four.[31] There is, therefore, some suggestion that age of onset may be lower in families with one or more NPC-affected first-degree relatives.

A genome-wide linkage analysis of 20 NPC families from a high incidence region in Guangdong identified a susceptibility region on the short arm of chromosome 4.[32] Two chromosome 4p15.1-q12 markers, D4S405 and D4S3002, yielded high logarithm of the odds (LOD) scores (>3.5) by both parametric and multipoint non-parametric analysis in 70 per cent of the NPC families studied. A subsequent study of 18 families from Hunan province genotyped a panel of markers on the short arms of chromosomes 3, 9 and 4 that included D4S405 and D4S3002 and failed to detect an obvious susceptibility locus on 4p15.1-q12.[33] A region on chromosome 3p21.31-21.2 containing a tumour suppressor gene cluster, however, showed a modest association with NPC incidence.[33]

Here, we describe the design of a new case — control study population recruited for the discovery of genetic factors that are involved in the development of chronic EBV infection and in the development of NPC. In a preliminary test to resolve the discrepancy between the two family-based studies, we performed a population-based case — control association analysis of 34 microsatellite markers within 4p15.1-q12 (Figure 1) to determine if specific alleles within the region: 1) were associated with a propensity to develop chronic EBV replication, as evidenced by IgA antibodies against EBV viral capsid antigen (EBV/IgA/VCA); or 2) were associated with NPC susceptibility.

## Materials and methods

### Study design

Enrolment into the study occurred in two collection phases. The Phase I pilot was powered to detect single gene associations and to determine feasibility for meeting recruitment goals, accuracy of data collection and sample handling, and to develop the infrastructure for a large international collaboration. Cases and controls ($n = 984$) were recruited in 2000 from Wuzhou City and Cangwu County, bordering the Xijiang River in the Guangxi province of South East China. An effort was made to enrol triads consisting of a proband, an unaffected spouse and an adult child or parent. Family triads were enrolled for haplotype inference and for quality control assessment. Three clinically described disease categories were collected: 1) incident or prevalent NPC biopsy-confirmed ($NPC^+$) cases ($n = 350$) who were EBV/IgA/VCA antibody positive ($IgA^+$); 2) $IgA^+$ cases ($n = 288$) who were defined as EBV/IgA/VCA antibody positive and NPC free at the time of study enrolment (EBV/IgA/VCA titres were confirmed by serological testing at the time of study enrolment); 3) $IgA^-$ controls ($n = 346$). For each case, his or her spouse was tested for EBV/IgA/VCA antibodies, and the spouse and parent or adult child were invited to enrol. The $IgA^-$ group consisted of 346 spouses who were $IgA^-$ at the time of study enrolment (Table 1). A dominant model was selected for power calculations for two reasons: 1) if the true model is additive, there is little difference in power using either an additive model or a dominant model for power calculations (data not shown); 2) if, however, the true model is dominant, a dominant model is the most powerful. Assuming a dominant genetic model and at least a 10 per cent allele frequency, this number of NPC, $IgA^+$, and $IgA^-$ cases and controls provided >90 per cent power to detect associations with an odds ratio (OR) ≥3, at the $p = 0.01$ level for a two-tailed test (Table 2).

Phase II enrolment was initiated in 2004 and after the completion of Phase I collection. The Phase II design is a

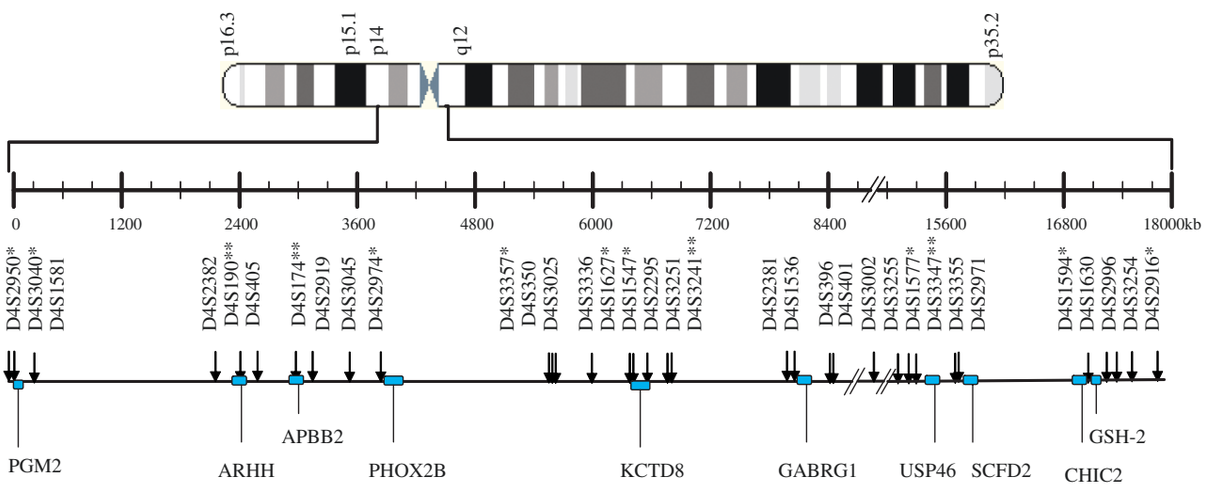


**Figure 1.** Map of markers for the nasopharyngeal carcinoma susceptibility locus on chromosome 4.[28] Thirty-four short tandem repeat markers distributed across an 18 megabase region were selected between *D4S2950* to *D4S2916*, with intervals of 10–3,500 kilobases. The positional relationship among markers and selected genes are indicated below the map. Asterisks indicate levels of statistical significance: $*p < 0.05$ and $**p < 0.01$.

cross-sectional, case control study: family members were not recruited. A questionnaire capturing environmental factors, including occupational, dietary and tobacco exposures, was administered to each study participant at enrolment (Table 3). NPC cases were recruited from the Wuzhou Red Cross Hospital in collaboration with the Cancer Institute of Wuzhou, Wuzhou City and the Cangwu Institute for Nasopharyngeal Carcinoma Control and Prevention, Cangwu County. NPC cases, $IgA^+$ subjects and $IgA^-$ participants were recruited from cities and villages bordering the Xijiang River. Power was determined for single gene and gene environment interactions for participants in each group (Table 2). For single-gene associations at a 10 per cent allele frequency, power will range from 83 per cent to $> 99$ per cent and from 35 per cent to $> 99$ per cent to detect associations with an odds ratio (OR) of 1.5–3.0 at $p < 0.05$ and $p < 0.001$, respectively, for the dominant genetic model and a two-sided significance level. For gene-environment interactions, there is power to detect gene-environment effects for genotype and exposures with frequencies $\geq 0.1$ for genotype and exposure, if the main exposure effect and genotype have an OR $\geq 1$ and an interaction effect of OR $\geq 3$.[34]

Exclusion criteria for Phases I and II were ethnicity other than Han Chinese, birth or residency for more than six months outside of the NPC endemic region or failure to provide informed consent. Internal review board approval was obtained from all participating institutions and informed consent was obtained from each study participant or their guardian for subjects between 16 and 18 years of age.

## Sample and data handling

A total of 10–20 ml of blood was collected in acid citrate dextrose (ACD) vacutainers for serology testing, direct DNA extraction and for cryopreservation of peripheral blood

**Table 1.** Characteristics of the Phase I study groups.

| | NPC positive | NPC negative | |
|---|---|---|---|
| | | EBV/IgA/VCA positive | EBV/IgA/VCA negative |
| No. study participants | 350 | 288 | 346 |
| Male/female (%male) | 233/117 (67%) | 142/146 (49%) | 129/217 (59%) |
| Age* | 48 ± 10 (16–79) | 44 ± 9 (20–77) | 47 ± 10 (18–75) |
| IgA/VCA titre | 1:10–1:640 | 1:5–1:80 | <1:5 |
| IgA/EA titre | 1:10–:160 (57.4%) | 1:5–1:20 (5%) | <1:5 |

Abbreviations: EBV, Epstein–Barr virus; IgA, immunoglobulin A; NPC, nasopharyngeal carcinoma; VCA, viral capsid antigens.
*Age at NPC diagnosis and at study enrolment for IgA serostatus.

**Table 2.** Phase I and phase II sample power.

| | | | Phase I | | Phase II | | All | |
|---|---|---|---|---|---|---|---|---|
| | | | Case/control | Power | Case/control | Power | Case/control | Power |
| NPC+ vs IgA+ | | | 350/288 | | 1024/1009 | | 1374/1297 | |
| | OR | 0.75 1.5 | | 17% | | 64% | | 79% |
| | | 0.50 2.0 | | 62% | | >99% | | >99% |
| | | 0.33 3.0 | | 98% | | >99% | | >99% |
| IgA+ vs IgA− | | | 288/346 | | 1009/1022 | | 1297/1368 | |
| | OR | 0.75 1.5 | | 18% | | 64% | | 79% |
| | | 0.50 2.0 | | 64% | | 99% | | >99% |
| | | 0.33 3.0 | | 99% | | >99% | | >99% |
| NPC+/IgA+ vs IgA− | | | 638/346 | | 2033/1022 | | 2671/1368 | |
| | OR | 0.75 1.5 | | 25% | | 77% | | 90% |
| | | 0.50 2.0 | | 78% | | >99% | | >99% |
| | | 0.33 3.0 | | >99% | | >99% | | >99% |

Note. Allele frequency = 10 per cent; Type 1 error = 1 per cent.
Power calculated for a 10% allele frequency, dominant genetic model and a $p$ value of $\leq 0.01$ for given case and control numbers using a two-tailed test. For example, with the Phase I cases/controls available, we would discover a genetic association of strength OR = 1.5, only 14% of the time, but a stronger gene (OR = 3.0) would be detected 96% of the time with available patient numbers.
Abbreviations: IgA, immunoglobulin A; NPC, nasopharyngeal carcinoma.

mononuclear cells (PBMCs). Blood samples were separated into plasma and PBMCs. Serum was tested at the Wuzhou Centre for EBV/IgA/VCA antibodies and antibodies to EBV early antigen by immunoenzymatic assay. The PBMCs were EBV-transformed to establish lymphoblastoid cell lines (LCLs) as a renewable DNA source. In addition, 3 cc of whole blood were preserved in DNA Tris, ethylene diamine tetraacetic acid and sodium dodecyl sulphate extraction buffer as a back-up DNA source. Questionnaires capturing demographic, laboratory and social history were administered at enrolment. Two individuals entered responses to the questionnaire and laboratory results into a FileMaker Pro database independently as a method of capturing data entry errors.

## Genomic DNA extraction

DNA was extracted from whole blood or lymphoblastoid cell lines using the QIAamp DNA blood maxi kit (Qiagen, Valencia, CA, USA, catalog #51194). More than 80 per cent

**Table 3.** Phase II study design and non-genetic covariates.

| Exposure | EBV/IgA/VCA status | | |
|---|---|---|---|
| | NPC/IgA+ | $IgA^+$ | $IgA^-$ |
| No. enrolled[1] | 1024 | 1009 | 1022 |
| Male:female (% male) | 735/289 (72%) | 555/454 (55%) | 684/338 (67%) |
| Consume dried meat[2] | 292 (28.5%) | 172 (17.0%) | 283 (27.7%) |
| Wood cooking fires[2] | 995 (97.2%) | 960 (95.1%) | 914 (89.4%) |
| Occupational exposure to solvents[2] | 78 (07.6%) | 31 (03.1%) | 87 (08.5%) |
| Smoking >10 years | 531 (51.9%) | 408 (40.4%) | 372 (36.4%) |

Abbreviations: EBV, Epstein–Barr virus; IgA, immunoglobulin A; NPC, nasopharyngeal carcinoma; VCA, viral capsid antigens.
[1] Greater than 99 per cent of $IgA^+$ and $IgA^-$ participants and 100 per cent of the NPC cases were born in Guangdong or Guangzhou provinces.
[2] Participants reporting any level exposure.

of the genotypes were determined from DNA directly extracted from whole blood.

### Microsatellite genotyping

Microsatellite loci ($n = 34$) containing 319 alleles were selected between D4S2950 and D4S2916 (18 megabases [Mb]) on chromosome 4 (Figure 1). The markers consist of 22 dinucleotide repeats, two trinucleotide repeats and ten tetranucleotide repeats. The genetic and physical distances between marker pairs are as follows: mean = 0.51 centimorgans (cM), 562 kilobases (kb); median = 0.34 cM, 230 kb; and range = 0.00–2.79 cM, 11–4,185 kb. The primer sequences were obtained from the University of Santa Cruz Genome Bioinformatics database.[35] All of the forward primers were 5′-tailed with the M13 sequence 5′-CACGACGTTGTAAAACGAC-3′. The M13-forward primers were used in combination with an M13 primer that had the same sequence but was labelled at its 5′ end with a fluorescent reagent from Applied Biosystems (ABI) Foster City, CA, USA such as 6-FAM, VIC or NED. The latter primer is the sole source of label and can be used with any M13-forward primer to generate a labelled amplified allele.[36] The polymerase chain reaction (PCR) amplifications of individual microsatellite loci were performed in 10 μl volumes containing 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 2.0 mM $MgCl_2$, 0.2 mM of each dinucleotide triphosphate, 1 μM labelled M13 primer and reverse primer, 0.07 mM M13-tailed primer (15:1 molar ratio of labelled M13 primer versus a M13-tailed forward primer), 25 ng genomic DNA, 0.5 U TaqGold DNA polymerase (Applied Biosystems, Foster City, CA, USA). PCR amplification was performed in a PE Applied Biosystems Model 9700 using 384 high-throughput format plates. The PCR conditions were a modified touchdown PCR procedure: 95°C, 10 minutes; two cycles of 95°C, 15 seconds; annealing temperature, 30 seconds; 72°C, 45 seconds, at annealing temperatures of 60°C, 58°C, 56°C, 54°C, 52°C; 30 cycles at an annealing temperature of 50°C; 72°C, 30 minutes. Six PCR products were pooled together for multiplex loading according to the label colour and marker size. Samples were diluted appropriately, pooled and then 3 μl of sample was mixed with 9 μl of formamide containing Liz 350 size standard (ABI). Samples were electrophoresed in a 22 cm capillary array using POP5 polymer and 3700 running buffer (ABI) on an ABI Model 3100 Automated DNA Sequencer using data collection software version 1.0.1 and Genescan Analysis software version 3.7. Genotyping was performed using Genotyper Version 2.5 and allele sizes were binned using Allelogram (Carl Manaster, available at http://s92417348.onlinehome.us/software/allelogram/index.html). For quality control between plates, DNAs from 22 per cent of subjects were duplicated across plates. Mendelian errors were tested within the triad families using the PedChek program.[37]

### Genetic association analyses

Allele frequencies were computed and compared between cases and controls using Pearson's $\chi^2$ test or Fisher's exact test. ORs, 95 per cent confidence intervals (CIs) and $p$ values were computed for dominant and recessive genetic models adjusted for age and sex. Logistic regression adjusted for age and sex was used to compute ORs using SAS PROC LOGISTIC software (SAS Institute, Cary, NC, USA). ORs were computed for a dominant model, comparing the combined homozygous and heterozygous genotypes against all other genotypes. When the allele frequency of the minor allele was ≥5 per cent, ORs were calculated for the recessive model, comparing the homozygous genotype against all other genotypes. Conformance to Hardy–Weinberg equilibrium expectations was calculated for all loci. Tests for D' as a measure of linkage disequilibrium (LD) were conducted for allele pairs using SAS Genetics software (SAS Institute, Cary, NC, USA).

## Results

The Phase I pilot study enrolled participants from the Cancer Institute in Wuzhou City and the Cangwu Institute for Nasopharyngeal Carcinoma Control and Prevention, Cangwu County in Guangxi province in the autumn of 2000. For NPC cases, 71.3 per cent of spouses and 81 per cent of adult children were enrolled. For cases with EBV/IgA/VCA titres consistent with chronic EBV infection, 72.4 per cent of spouses and 67.4 per cent of adult children or parents were enrolled. Complete triad sets were available for 366 NPC probands. As predicted for this highly endemic NPC region, 71.8 per cent of the NPC cases were male. PBMCs cryopreserved on-site were transported to the Laboratory of Genomic Diversity-National Cancer Insitute (LGD-NCI) for EBV immortalisation: 83 per cent of 633 transformation attempts resulted in LCLs.

Sample and genotyping errors were estimated by including 10 per cent duplicate sampling with one sample derived from DNA isolated directly from peripheral blood and the second from DNA isolated from LCLs. Less than 0.5 per cent mismatches within duplicate samples were observed, all of which were resolved using family trios, indicating that tubes collected from a single individual were appropriately labelled (data not shown) and that error was not introduced during cell line development or sample handling. A second test for Mendelian errors using PedChek was performed using the chromosome 4 microsatellite data (described below). Two unresolved Mendelian errors were observed within the 366 family triads. Near-complete genotyping and complete clinical data were available for 350 NPC cases, 288 IgA seropositives and 346 IgA seronegatives (Table 1).

Phase II enrolment occurred between November 2004 and July 2005 in Guangxi province. Subjects were enrolled if at least one parent was from the Guangxi or Guangdong provinces. NPC cases were identified as seroincident or seroprevalent cases presenting at Red Cross hospitals and $IgA^+$ and $IgA^-$ controls were identified from field stations in

cities and villages bordering the Xijiang River drainage. Table 3 presents summary data of environmental exposures for the Phase II $NPC^+$, $IgA^+$ and $IgA^-$ groups and the numbers of participants enrolled.

We have addressed the questions of whether a locus within the chromosome 4p15.1-q12 region leads to the development of NPC or the development of EBV/IgA/VCA in response to EBV replication using the Phase I cases and controls. Microsatellite loci ($n = 34$) were distributed over an 18 Mb region on chromosome 4p15.1-q12, with intervals of 10–3,500 kb and an average distance of 530 kb. Four Phase I genetic association comparisons were made: 1) NPC cases versus EBV/IgA/VCA seropositive controls (Table 4); 2) EBV/IgA/VCA seropositive cases without NPC versus EBV/IgA/VCA seronegative controls (Table 5); 3) NPC cases plus EBV/IgA/VCA seropositive cases versus EBV/IgA/VCA seronegative controls (Table 6); and 4) NPC cases versus EBV/IgA/VCA seronegative controls (data not shown). No distortions in Hardy-Weinberg equilibrium were observed. Alleles with at least one significant result ($p < 0.05$) for either the dominant or recessive genetic models are reported in Tables 4–6. The results are presented without correction for multiple comparisons because the interrogated 4p15.1-q12 region was previously implicated as a susceptibility locus in a family-based study and we were specifically testing the prior hypothesis that markers within the region would also be associated with NPC in a population-based study.[32] It should be noted that associations with $p > 0.0015$ would not remain significant after correction for multiple comparisons considering the 34 independent loci.

### Linkage disequilibrium among the 34 loci

The spacing of markers varied from 10–3,500 kb, with denser coverage flanking the microsatellite markers with the highest LOD scores from the family study (Figure 1). We calculated two-point D' as a measure of LD between all alleles at neighbouring short tandem repeat loci; however, a D' value of 1 (complete LD) was observed for only 60 two-point allele combinations. Using HapMap single nucleotide polymorphism (SNP) data (http://www.hapmap.org), we examined whether the microsatellites were included in reasonably strong LD blocks. The $r^2$ between any given marker pairs were set at a 0.8 cut-off threshold to determine the LD blocks. Only 11 of the 34 microsatellite markers occurred within an LD block: *D4S396* and *D4S401* occurred within the same 17 kb block. Of the two NPC-linked markers,[32] *D4S405* was not within a block and *D4S3002* occurred within an 8 kb block. The mean size of the blocks was 17.9 kb (range 8–50 kb).

### Genetic association with NPC

Table 4 presents the locus name, location, allele length, allele frequencies, ORs, *p* values and 95 per cent CIs for 350 NPC cases and 288 EBV/IgA/VCA seropositive controls (93.0–99.7 per cent of NPC cases and 91.0–97.6 per cent of $IgA^+$ subjects were genotyped successfully). The genotype frequencies among NPC cases were significantly higher than those among control subjects for five alleles (OR 1.51–5.36; $p = 0.01–0.03$): for the recessive model, *D4S3040-215* and *D4S1547-251*; and for the dominant model, *D4S3040-213*, *D4S2974-137* and *D4S2916-204*. The genotype frequencies among NPC cases was statistically lower than among control subjects for four alleles (OR 0.3–0.71; $p = 0.02–0.045$): for the recessive model, *D4S2950-141* and *D4S2974-135*; and for the dominant model, *D4S3357-271* and *D4S2381-277*.

### Genetic association with persistent $IgA^+$ status

To test the hypothesis that genetic factors may influence EBV/IgA/VCA formation in response to EBV infection, we compared genotype frequencies between 288 $IgA^+$ cases and 346 $IgA^-$ controls (Table 1). Table 5 provides the allele frequencies, *p* values, ORs and 95 per cent CIs in cases and controls for significant results. Eleven alleles were significantly associated with $IgA^+$ persistence: five risk alleles (OR 1.51–2.38; $p = 0.004–0.040$) and six protective alleles (OR 0.33–0.70; $p = 0.002–0.050$).

Because all NPC cases in our study were $IgA^+$, we then pooled NPC and $IgA^+$ cases together to increase power, with the hypothesis being that the alleles associated with $IgA^+$ serostatus would be shared among $NPC^+IgA^+$ and $NPC^-IgA^+$ individuals. Significant associations are presented in Table 6: four alleles were associated with risk for $IgA^+$ (OR 1.5–1.63; $p = 0.001–0.030$) and seven were protective (OR 0.46–0.76; $p = 0.001–0.050$). Based on the two comparisons (Tables 5 and 6), ten alleles associated with IgA were shared in both comparisons. Five were highly significant ($p < 0.01$) associations with $IgA^+$ serostatus. Alleles D4S190-170 ($p = 0.005$; OR 1.5, 95% CI 1.13–2.0), D4S3241-136 ($p = 0.004$; OR 1.91, 95% CI 1.2-3.0) and D4S3347-213 ($p = 0.001$; OR 1.58, 95% CI 1.2-2.1) significantly increased the risk of developing EBV/IgA/VCA. Alleles D4S174-202 ($p = 0.001$; OR 0.46, 95% CI 0.3-0.7) and D4S2950-137 ($p = 0.0036$; OR 0.56, 95 per cent CI 0.38-0.83) significantly decreased the risk of EBV/IgA/VCA. Within a single locus (D4S3357), one allele increased susceptibility (D4S3357-271) and the other allele was protective (D4S3357-275).

## Discussion

We have described the design and recruitment efforts for a genetic association study to investigate the role of host genetic factors in the development of chronic EBV infection leading to NPC in subjects born and living in a region with one of the world's highest incidence rates of NPC. This study was conducted in two phases. Phase I was a pilot study to explore the feasibility of conducting a cross-sectional study

**Table 4.** Significant allele frequencies between NPC versus $IgA^+$ groups.

| cM* | Locus | Individuals # | | | Allele freq (%) | | | Dominant | | | Recessive | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Allele | NPC | IgA+ | NPC | | IgA+ | OR | p value | 95% CI | OR | *p* value | 95% CI |
| 37.64 | D4S2950 | 141 | 334 | 270 | 12.9 | < | 19.6 | 0.70 | 0.06 | 0.48–1.01 | 0.30 | 0.02 | 0.11–0.84 |
| 37.74 | D4S3040 | 213 | 324 | 261 | 17.4 | > | 14.0 | 1.52 | 0.03 | 1.04–2.22 | 1.47 | 0.42 | 0.57–3.75 |
| 37.74 | D4S3040 | 215 | 324 | 261 | 15.6 | > | 12.5 | 1.21 | 0.34 | 0.82–1.79 | 5.36 | 0.03 | 1.17–24.53 |
| 41.57 | D4S2974 | 135 | 347 | 281 | 60.7 | < | 64.9 | 1.22 | 0.41 | 0.76–1.97 | 0.67 | 0.02 | 0.48–0.93 |
| 41.57 | D4S2974 | 137 | 347 | 281 | 16.9 | > | 11.4 | 1.51 | 0.03 | 1.03–2.20 | 3.43 | 0.07 | 0.92–12.85 |
| 43.27 | D4S3357 | 271 | 340 | 271 | 29.7 | < | 33.4 | 0.71 | 0.04 | 0.51–0.99 | 1.01 | 0.98 | 0.60–1.71 |
| 43.33 | D4S350 | 256 | 346 | 275 | 64.6 | > | 58.7 | 1.61 | 0.05 | 1.01–2.56 | 1.24 | 0.21 | 0.89–1.7 |
| 44.11 | D4S1547 | 251 | 349 | 276 | 56.4 | > | 52.7 | 0.91 | 0.64 | 0.62–1.35 | 1.53 | 0.02 | 1.07–2.19 |
| 44.28 | D4S2295 | 222 | 345 | 278 | 76.1 | < | 81.3 | 0.52 | 0.12 | 0.23–1.19 | 0.71 | 0.05 | 0.50–1.0 |
| 45.71 | D4S2381 | 277 | 347 | 277 | 61.8 | < | 65.9 | 0.59 | 0.03 | 0.36–0.95 | 0.92 | 0.60 | 0.66–1.28 |
| 53.64 | D4S2971 | 165 | 344 | 275 | 45.1 | > | 43.3 | 1.43 | 0.05 | 1.0–2.05 | 0.76 | 0.20 | 0.50–1.1 |
| 55.73 | D4S2916 | 204 | 336 | 265 | 13.8 | > | 10.6 | 1.55 | 0.03 | 1.03–2.33 | 0.57 | 0.39 | 0.16–2.07 |

Abbreviations: CI, confidence interval; cM, centimorgan; IgA, immunoglobulin A; NPC, nasopharyngeal carcinoma; OR, odds ratio.
*See Figure 1.

 

**Table 5.** Group 2: Markers and alleles showing significant allele frequencies among $IgA^+$ cases without NPC and $IgA^-$ subjects.

| cM | Locus | Individuals # | | | Allele freq (%) | | | Dominant | | | Recessive | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Allele | IgA+ | IgA − | IgA+ | | IgA − | OR | p value | 95% CI | OR | *p* value | 95% CI |
| 37.64 | D4S2950 | 137 | 230 | 319 | 13.3 | < | 18.5 | 0.56 | 0.004 | 0.38–0.83 | 0.75 | 0.56 | 0.28–2.00 |
| 40.14 | D4S190 | 170 | 236 | 332 | 23.9 | > | 18.7 | 1.40 | 0.06 | 0.98–1.99 | 2.38 | 0.04 | 1.03–5.49 |
| 40.75 | D4S174 | 202 | 237 | 331 | 23.4 | < | 32.6 | 0.59 | 0.003 | 0.42–0.84 | 0.33 | 0.002 | 0.16–0.67 |
| 43.27 | D4S3357 | 271 | 232 | 326 | 34.3 | > | 27.8 | 1.51 | 0.02 | 1.06–2.13 | 1.46 | 0.20 | 0.82–2.61 |
| 44.09 | D4S1627 | 218 | 239 | 335 | 33.3 | < | 38.1 | 0.81 | 0.24 | 0.58–1.15 | 0.57 | 0.039 | 0.33–0.97 |
| 44.48 | D4S3241 | 136 | 233 | 325 | 13.7 | > | 8.9 | 1.91 | 0.004 | 1.24–2.94 | 2.11 | 0.28 | 0.54–8.27 |
| 45.71 | D4S2381 | 301 | 238 | 330 | 18.9 | < | 23.0 | 0.70 | 0.05 | 0.49–1.0 | 0.90 | 0.80 | 0.42 |
| 45.84 | D4S1536 | 284 | 233 | 327 | 33.7 | < | 40.7 | 0.79 | 0.19 | 0.56–1.13 | 0.48 | 0.01 | 0.27–0.84 |
| 52.95 | D4S1577 | 143 | 238 | 332 | 34.2 | < | 41.6 | 0.70 | 0.05 | 0.49–1.0 | 0.52 | 0.01 | 0.31–0.86 |
| 53.03 | D4S3347 | 213 | 237 | 332 | 34.8 | > | 29.5 | 1.65 | 0.004 | 1.17–2.33 | 0.89 | 0.67 | 0.50–1.56 |
| 53.03 | D4S3347 | 217 | 237 | 332 | 45.4 | < | 50.0 | 0.87 | 0.48 | 0.59–1.28 | 0.66 | 0.05 | 0.43–1.0 |
| 54.86 | D4S1594 | 266 | 236 | 330 | 66.1 | > | 64.2 | 1.76 | 0.04 | 1.02–3.03 | 1.01 | 0.95 | 0.71–1.43 |

Abbreviations: CI, confidence interval; cM, centimorgan; IgA, immunoglobulin A; OR, odds ratio.

**Table 6.** Group 3: Markers and alleles showing significant allele frequencies among NPC cases plus $IgA^+$ cases and $IgA^-$ subjects.

| cM | Locus | Individuals # | | | Allele freq (%) | | | Dominant | | | Recessive | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Allele | NPC | IgA− | NPC+IgA+ | | IgA− | OR | *p* value | 95% CI | OR | *p* value | 95% CI |
| 37.64 | D4S2950 | 137 | 604 | 319 | 15.0 | < | 18.5 | 0.69 | 0.01 | 0.51–0.93 | 1.06 | 0.88 | 0.50–2.26 |
| 40.14 | D4S190 | 170 | 627 | 332 | 24.1 | > | 18.7 | 1.50 | 0.005 | 1.13–1.99 | 1.99 | 0.07 | 0.96–4.14 |
| 40.75 | D4S174 | 202 | 616 | 331 | 27.4 | < | 32.6 | 0.79 | 0.10 | 0.60–1.05 | 0.46 | 0.001 | 0.28–0.74 |
| 40.75 | D4S174 | 204 | 616 | 331 | 11.5 | > | 8.5 | 1.45 | 0.04 | 1.02–2.08 | 3.17 | 0.15 | 0.67–15.0 |
| 43.27 | D4S3357 | 275 | 611 | 326 | 21.4 | < | 24.8 | 0.72 | 0.02 | 0.55–0.96 | 0.95 | 0.87 | 0.51–1.75 |
| 44.09 | D4S1627 | 218 | 625 | 335 | 31.8 | < | 38.1 | 0.76 | 0.05 | 0.58–1.0 | 0.57 | 0.007 | 0.38–0.86 |
| 44.48 | D4S3241 | 136 | 611 | 325 | 13.4 | > | 8.9 | 1.63 | 0.007 | 1.14–2.31 | 1.85 | 0.30 | 0.58–5.89 |
| 45.84 | D4S1536 | 284 | 615 | 327 | 34.5 | < | 40.7 | 0.78 | 0.09 | 0.59–1.04 | 0.55 | 0.004 | 0.37–0.83 |
| 46.25 | D4S401 | 213 | 628 | 333 | 10.8 | < | 11.1 | 1.09 | 0.63 | 0.77–1.53 | 0.33 | 0.05 | 0.11–1.00 |
| 52.72 | D4S3255 | 189 | 614 | 328 | 8.5 | > | 6.1 | 1.50 | 0.05 | 1.0–2.25 | 3.18 | 0.30 | 0.36–28.0 |
| 52.95 | D4S1577 | 143 | 624 | 332 | 35.8 | < | 41.6 | 0.72 | 0.02 | 0.54–0.95 | 0.67 | 0.03 | 0.46–0.96 |
| 53.03 | D4S3347 | 213 | 623 | 332 | 34.3 | > | 29.5 | 1.58 | 0.001 | 1.20–2.08 | 1.02 | 0.93 | 0.66–1.58 |
| 53.03 | D4S3347 | 217 | 623 | 332 | 45.2 | < | 50.0 | 0.90 | 0.48 | 0.66–1.22 | 0.62 | 0.004 | 0.45–0.86 |
| 54.86 | D4S1594 | 266 | 620 | 330 | 66.5 | > | 64.2 | 1.57 | 0.03 | 1.04–2.38 | 1.03 | 0.86 | 0.78–1.35 |

Abbreviations: CI, confidence interval; cM, centimorgan; IgA, immunoglobulin A; OR, odds ratio.

 

in China (Table 1). The pilot provided strong support for expanding the study in several important ways: export permits for genetic material were obtained, sample handling was excellent — with few detectable errors — and recruitment goals were attainable. Upon the successful completion of Phase I, we increased the catchment area for $IgA^+$ cases to cities and villages along the Xijiang River and tributaries, expanded the study to include more subjects and added a detailed questionnaire to capture environmental exposures that may interact with host genes in the development of NPC (Table 2). Complementing previous studies, we also attempted to determine if Phase II of the study was powered for the detection of both gene–gene and gene–environment interactions.

To revisit the recent linkage analysis in NPC families implicating a susceptibility locus linked to chromosome 4p15.1-q12, we selected 34 microsatellite loci spanning the 18 Mb region at intervals of 10–3,500 kb. Unlike in previous studies, we first also attempted to determine if the chromosome 4 region was associated with EBV/IgA/VCA antibody formation and, secondly, if the chromosome 4 region was associated with NPC incidence in the setting of EBV replication as indicated by EBV/IgA/VCA. We identified several loci that showed significant associations with either EBV/IgA/VCA or NPC status. The associations tended to be marginally significant for NPC (Table 4), with somewhat stronger associations observed for EBV/IgA/VCA ($IgA^+$) (Tables 5 and 6).

Few NPC families have been identified outside of NPC endemic areas. More than 90 per cent of all NPC cases do not show familial aggregation or family history, implying either environmental causes or geographical family clustering. Two family-based NPC linkage studies implicated different chromosomes as harbouring an NPC susceptibility locus.[32,33] Although the studies differed in strategy, both used multiple families with two or more NPC cases from two separate high NPC incident provinces in China and included similar numbers of families and affected cases. Although it is possible that environmental exposures may differ between the two provinces, it is unlikely that different environmental factors account for the lack of concordance between the studies. More likely, multiple genes predispose to chronic EBV replication and the development of NPC, each of which may contribute only a small part of the total genetic influence. Family-based linkage studies are ideal for identifying single genes with large effects, but are relatively insensitive for localising genetic factors with small effects. By contrast, case-control association studies are ideal for identifying genetic factors with small or moderate effects once a candidate gene or region has been identified.[38]

We cannot exclude the possibility that there may be causal alleles in the chromosome 4 region that may be associated with chronic EBV replication or a predisposition to develop NPC. Marker associations within this region (Tables 4–6) may be tracking a susceptible locus through LD. Because included alleles predominantly occurred at very low frequencies and haplotype inferences were unreliable, we could not reliably assess associations with either EBV persistence or NPC (data not shown). A denser placement of polymorphic markers is required to survey the genetic variation content of the region more thoroughly.

Although this study did not find associations with robust *p* values for NPC, making conclusions tentative, a number of loci did show moderate to strong risk, suggesting that this region warrants further attention, particularly for chronic EBV replication. For one of the microsatelite loci *D4s3347* (Tables 5 and 6), two alleles were associated with EBV/IgA/VCA, suggesting that these alleles may be tracking a potential causative allele (see Figure 1). Of potential interest is the association of two microsatellites with IgA incidence: *D4S3347*, which shows three significant associations with $p < 0.01$ and one with $p < 0.05$ for two alleles (*213* and *217*), and the tightly linked (<20 kb) *D4S1577* locus, which also shows four significant associations ($p < 0.05$) (Tables 5 and 6, Figure 1). Microsatellite D4S190 occurs within the oncogene ARHH. D4S190 was associated with risk for EBV/IgA/VCA seropositive status but not with NPC. ARHH, a member of the ras homolog gene family, encodes a small GTP-binding protein belonging to the RAS superfamily and is transcribed by only haemopoietic cells. *ARHH* non-coding variants that may affect expression are observed in 46 per cent of diffuse large-cell lymphomas.[39] It is possible that one or more variant alleles of *ARHH* in LD with associated *D4S190-170* may modify EBV replication.

Given the similar geographical distribution of familial and non-familial NPC, it is likely that both forms share similar aetiological risk factors, particularly environmental and viral factors; however, it is likely that the genetic factors underpinning familial, early-onset and non-familial NPC susceptibility may also overlap. It is also possible that different genes contribute to familial NPC cases, analogous to the situation in breast cancer, where *BRCA1* and *BRCA2* account for only a small proportion of non-familial breast cancer cases.[40,41] The best approach to identifying NPC susceptibility factors may be the organisation of well-designed and highly powered case–control studies for whole-genome and targeted candidate gene association investigations, as we describe here.

## Acknowledgments

We gratefully acknowledge Beth Binns-Roemer and Maidar Jamba for excellent technical assistance and Dr Michael Smith for valuable discussions. This project has been funded, in whole or in part, by federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400. The content of this paper does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organisations imply endorsement by the US Government. The publisher or recipient acknowledges the right of the US Government to retain a non-exclusive, royalty-free licence in and to any copyright covering the paper.

## References


1. deThe, G. (1982), 'Epidemiology of Epstein Barr Virus and associated diseases in man', in Roizman, B. (ed.), *The Herpesviruses*, Springer, New York, NY, pp. 25–103.
2. deThe, G. (1995), 'Viruses and human cancers: Challenges for preventive strategies', *Environ. Health Perspect.* Vol. 103(Suppl. 8), pp. 269–273.
3. Jeannel, D., Hubert, A., de Vathaire, F. *et al.* (1990), 'Diet, living conditions and nasopharyngeal carcinoma in Tunisia — A case-control study', *Int. J. Cancer* Vol. 46, pp. 421–425.
4. Laramore, G.E., Clubb, B., Quick, C. *et al.* (1988), 'Nasopharyngeal carcinoma in Saudi Arabia: A retrospective study of 166 cases treated with curative intent', *Int. J. Radiat. Oncol. Biol. Phys.* Vol. 15, pp. 1119–1127.
5. Johansen, L.V., Mestre, M. and Overgaard, J. (1992), 'Carcinoma of the nasopharynx: Analysis of treatment results in 167 consecutively admitted patients', *Head Neck* Vol. 14, pp. 200–207.
6. Lee, A.W., Foo, W., Mang, O. *et al.* (2003), 'Changing epidemiology of nasopharyngeal carcinoma in Hong Kong over a 20-year period (1980–99): An encouraging reduction in both incidence and mortality', *Int. J. Cancer* Vol. 103, pp. 680–685.
7. Old, L.J., Boyse, E.A., Oettgen, H.F. *et al.* (1966), 'Precipitating antibody in human serum to an antigen present in cultured Burkitt's lymphoma cells', *Proc. Natl. Acad. Sci. USA* Vol. 56, pp. 1699–1704.
8. Henle, G. and Henle, W. (1976), 'Epstein–Barr virus-specific IgA serum antibodies as an outstanding feature of nasopharyngeal carcinoma', *Int. J. Cancer* Vol. 17, pp. 1–7.
9. Deng, H., Zeng, Y., Lei, Y. *et al.* (1995), 'Serological survey of nasopharyngeal carcinoma in 21 cities of south China', *Chin. Med. J. (Engl.)* Vol. 108, pp. 300–303.
10. Sham, J.S., Wei, W.I., Zong, Y.S. *et al.* (1990), 'Detection of subclinical nasopharyngeal carcinoma by fibreoptic endoscopy and multiple biopsy', *Lancet* Vol. 335, pp. 371–374.
11. Zeng, Y., Zhang, L.G., Li, H.Y. *et al.* (1982), 'Serological mass survey for early detection of nasopharyngeal carcinoma in Wuzhou City, China', *Int. J. Cancer* Vol. 29, pp. 139–141.
12. Zeng, Y., Zhong, J.M., Li, L.Y. *et al.* (1983), 'Follow-up studies on Epstein–Barr virus IgA/VCA antibody-positive persons in Zangwu County, China', *Intervirology* Vol. 20, pp. 190–194.
13. Zeng, Y., Zhang, L.G., Wu, Y.C. *et al.* (1985), 'Prospective studies on nasopharyngeal carcinoma in Epstein–Barr virus IgA/VCA antibody-positive persons in Wuzhou City, China', *Int. J. Cancer* Vol. 36, pp. 545–547.
14. Zong, Y.S., Sham, J.S., Ng, M.H. *et al.* (1992), 'Immunoglobulin A against viral capsid antigen of Epstein–Barr virus and indirect mirror examination of the nasopharynx in the detection of asymptomatic nasopharyngeal carcinoma', *Cancer* Vol. 69, pp. 3–7.
15. Jalbout, M., Bel Hadj Jrad, B., Bouaouina, N. *et al.* (2002), 'Autoantibodies to tubulin are specifically associated with the young age onset of the nasopharyngeal carcinoma', *Int. J. Cancer* Vol. 101, pp. 146–150.
16. Yu, M.C. and Yuan, J.M. (2002), 'Epidemiology of nasopharyngeal carcinoma', *Semin. Cancer Biol.* Vol. 12, pp. 421–429.
17. Brown, T.M., Heath, C.W., Lang, R.M. *et al.* (1976), 'Nasopharyngeal cancer in Bermuda', *Cancer* Vol. 37, pp. 1464–1468.
18. Coffin, C.M., Rich, S.S. and Dehner, L.P. (1991), 'Familial aggregation of nasopharyngeal carcinoma and other malignancies. A clinicopathologic description', *Cancer* Vol. 68, pp. 1323–1328.
19. Yu, M.C., Garabrant, D.H., Huang, T.B. *et al.* (1990), 'Occupational and other non-dietary risk factors for nasopharyngeal carcinoma in Guangzhou, China', *Int. J. Cancer* Vol. 45, pp. 1033–1039.
20. Jia, W.H., Feng, B.J., Xu, Z.L. *et al.* (2004), 'Familial risk and clustering of nasopharyngeal carcinoma Guangdong, China', *Cancer* Vol. 101, pp. 363–369.
21. Buell, P. (1974), 'The effect of migration on the risk of nasopharyngeal cancer among Chinese', *Cancer Res.* Vol. 34, pp. 1189–1191.
22. Hildesheim, A., Apple, R.J., Chen, C.J. *et al.* (2002), 'Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan', *J. Natl. Cancer Inst.* Vol. 94, pp. 1780–1789.
23. Li, P.K., Poon, A.S., Tsao, S.Y. *et al.* (1995), 'No association between HLA-DQ and -DR genotypes with nasopharyngeal carcinoma in southern Chinese', *Cancer Genet. Cytogenet.* Vol. 81, pp. 42–45.
24. Lu, C.C., Chen, J.C. and Jin, Y.T. (2003), 'Genetic susceptibility to nasopharyngeal carcinoma within the HLA-A locus in Taiwanese', *Int. J. Cancer* Vol. 103, pp. 745–751.
25. Mokni-Baizig, N., Ayed, K., Ayed, F.B. *et al.* (2001), 'Association between HLA-A/-B antigens and -DRB1 alleles and nasopharyngeal carcinoma in Tunisia', *Oncology* Vol. 61, pp. 55–58.
26. Pimtanothai, N., Charoenwongse, P., Mutirangura, A. and Hurley, C.K. (2002), 'Distribution of HLA-B alleles in nasopharyngeal carcinoma patients and normal controls in Thailand', *Tissue Antigens* Vol. 59, pp. 223–225.
27. Thomas, J.A., Iliescu, V., Crawford, D.H. *et al.* (1984), 'Expression of HLA-DR antigens in nasopharyngeal carcinoma: An immunohistological analysis of the tumour cells and infiltrating lymphocytes', *Int. J. Cancer* Vol. 33, pp. 813–819.
28. Wu, S.B., Hwang, S.J., Chang, A.S. *et al.* (1989), 'Human leukocyte antigen (HLA) frequency among patients with nasopharyngeal carcinoma in Taiwan', *Anticancer Res.* Vol. 9, pp. 1649–1653.
29. Ooi, E.E., Ren, E.C. and Chan, S.H. (1997), 'Association between microsatellites within the human MHC and nasopharyngeal carcinoma', *Int. J. Cancer* Vol. 74, pp. 229–232.
30. Loh, K.S., Goh, B.C., Lu, J. *et al.* (2006), 'Familial nasopharyngeal carcinoma in a cohort of 200 patients', *Arch. Otolaryngol. Head Neck Surgery* Vol. 132, pp. 82–85.
31. Zeng, Y.X. and Jia, W.H. (2002), 'Familial nasopharyngeal carcinoma', *Semin. Cancer Biol.* Vol. 12, pp. 443–450.
32. Feng, B.J., Huang, W., Shugart, Y.Y. *et al.* (2002), 'Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4', *Nat. Genet.* Vol. 31, pp. 395–399.
33. Xiong, W., Zeng, Z.Y., Xia, J.H. *et al.* (2004), 'A susceptibility locus at chromosome 3p21 linked to familial nasopharyngeal carcinoma', *Cancer Res.* Vol. 64, pp. 1972–1974.
34. Saunders, C.L. and Barrett, J.H. (2004), 'Flexible matching in case-control studies of gene-environment interactions', *Am. J. Epidemiol.* Vol. 159, pp. 17–22.
35. http://genome.ucsc.edu/cgi-bin/hgGateway
36. Boutin-Ganache, I., Raposo, M., Raymond, M. and Deschepper, C.F. (2001), 'M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods', *Biotechniques* Vol. 31, pp. 24–26, 28.
37. O'Connell, J.R. and Weeks, D.E. (1998), 'PedCheck: A program for identification of genotype incompatibilities in linkage analysis', *Am. J. Hum. Genet.* Vol. 63, pp. 259–266.
38. Risch, N. and Merikangas, K. (1996), 'The future of genetic studies of complex human diseases', *Science* Vol. 273, pp. 1516–1517.
39. Preudhomme, C., Roumier, C., Hildebrand, M.P. *et al.* (2000), 'Nonrandom 4p13 rearrangements of the RhoH/TTF gene, encoding a GTP-binding protein, in non-Hodgkin's lymphoma and multiple myeloma', *Oncogene* Vol. 19, pp. 2023–2032.
40. Malone, K.E., Daling, J.R., Neal, C. *et al.* (2000), 'Frequency of BRCA1/BRCA2 mutations in a population-based sample of young breast carcinoma cases', *Cancer* Vol. 88, pp. 1393–1402.
41. Peto, J., Collins, N., Barfoot, R. *et al.* (1999), 'Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer', *J. Natl. Cancer Inst.* Vol. 91, pp. 943–949.