Genes Involved in DNA Repair and Nitrosamine Metabolism and Those Located on Chromosome 14q32 Are Dysregulated in Nasopharyngeal Carcinoma

Lori E. Dodd,[1] Srikumar Sengupta,[4] I-How Chen,[7] Johan A. den Boon,[4] Yu-Juen Cheng,[8] William Westra,[9] Michael A. Newton,[5] Beth F. Mittl,[3] Lisa McShane,[1] Chien-Jen Chen,[8] Paul Ahlquist,[4,6] and Allan Hildesheim[2]

[1]Biometric Research Branch, Division of Cancer Treatment and Diagnosis and [2]Hormonal and Reproductive Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute; [3]Westat, Inc., Rockville, Maryland; [4]Institute for Molecular Virology, [5]Department of Statistics and of Biostatistics and Medical Informatics, and [6]McArdle Laboratory for Cancer Research and Howard Hughes Medical Institute, University of Wisconsin, Madison, Wisconsin; [7]Department of Otolaryngology, McKay Memorial Hospital; [8]Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan; and [9]Department of Pathology, Johns Hopkins University, School of Medicine, Baltimore, Maryland

Abstract

Polymorphisms in nitrosamine metabolism, DNA repair, and immune response genes have been associated with nasopharyngeal carcinoma (NPC). Studies have suggested chromosomal regions involved in NPC. To shed light on NPC etiology, we evaluated host gene expression patterns in 31 NPC and 10 normal nasopharyngeal tissue specimens using the Affymetrix Human Genome U133 Plus 2.0 Array. We focused on genes in five a priori biological pathways and chromosomal locations. Rates of differential expression within these prespecified lists and overall were tested using a bootstrap method. Differential expression was observed for 7.6% of probe sets overall. Elevations in rate of differential expression were observed within the DNA repair (13.7%; $P = 0.01$) and nitrosamine metabolism (17.5%; $P = 0.04$) pathways. Differentially expressed probe sets within the DNA repair pathway were consistently overexpressed (93%), with strong effects observed for PRKDC, PCNA, and CHEK1. Differentially expressed probe sets within the nitrosamine metabolism pathway were consistently underexpressed (100%), with strong effects observed for NQO1, CYP2B6, and CYP2E1. No significant evidence of increases in rate of differential expression was seen within the immune/inflammatory pathway. A significant elevation in rate of differential expression was noted for chromosome 4p15.1-4q12 (13.0%; $P = 0.04$); both overexpression and underexpression were evident (38% and 62%, respectively). An elevation in the rate of differential expression on chromosome 14q32 was observed (11.3%; $P = 0.06$) with a consistent pattern of gene underexpression (100%; $P < 0.0001$). These effects were similar when excluding late-stage tumors. Our results suggest that nitrosamine activation and DNA repair are important in NPC. The consistent down-regulation of expression on chromosome 14q32 suggests loss of heterozygosity in this region. (Cancer Epidemiol Biomarkers Prev 2006;15(11):2216–25)

Introduction

EBV is a ubiquitous DNA virus with worldwide distribution known to be associated with the development of ∼100% of nasopharyngeal carcinomas (NPC), particularly in southeastern Asia and among those of Chinese descent (1, 2). In addition to EBV infection, numerous environmental and host factors have been suggested to be associated with NPC in epidemiologic studies.

Exogenous exposures associated with NPC previously include sources of nitrosamines (dietary and other), nitrosamine precursors, and other agents involved in damage to DNA. Studies have shown an association between dietary consumption of salted fish and other preserved foods rich in nitrosamines and their precursors and NPC risk (1). Such exposures seem to be particularly important when occurring at an early age. In a study of the association between dietary nitrosamine levels and NPC risk in Taiwan, individuals in the upper quartile of consumption of nitrosamines during childhood had a 2.6-fold increased risk of NPC (3). Smoking of cigarettes, which contain nitrosamines and other agents capable of DNA damage, has also been shown to be associated with NPC, particularly when smoking occurs for a prolonged period (4).

Host factors believed to be associated with NPC tend to fall into two biological categories: factors linked to DNA damage and repair and those linked to immunologic control of EBV. Data available to date suggest that polymorphisms in genes involved in the activation of nitrosamines into reactive intermediates capable of DNA damage or in the repair of damage are associated with NPC (5-7). In particular, CYP2E1 variants have been associated with NPC (5, 6), and specific polymorphisms in the DNA repair genes *XRCC1* and *hOGG1* have been suggested to affect NPC risk (7).

Immunologic factors have been hypothesized to be associated with NPC development because interindividual differences in immunologic control of the ubiquitous EBV infection are likely important determinants of NPC risk. In this respect, genes that encode for human leukocyte antigens (HLA) responsible for the presentation of viruses, such as EBV, to the immune system have been studied. Studies to date have consistently shown an association between specific HLA alleles and haplotypes and risk of NPC (8, 9). Among others, HLA-A*0207, HLA-A*1101, HLA-B*4601, and HLA-B*5801 are associated with NPC risk. In addition to HLA gene

polymorphisms that might affect the adaptive immune response, recent studies have begun to evaluate the role of innate immune responses in NPC pathogenesis through the evaluation of KIR genes and their HLA ligands (10).

Most recently, studies of multiplex families at high risk of NPC have been conducted. Initial reports from these family-based studies have suggested linkage of several chromosomal regions (including 4p15.1-4q12 and 14q32) with NPC in these families[10] (11).

As summarized above, data point to the involvement of multiple factors in NPC pathogenesis. However, other than the evaluation and localization of EBV in tumor cells of the nasopharynx, little is known about events occurring locally at the nasopharynx that might support any of the hypotheses mentioned above.

Herein, we describe results of a study in which expression levels of a priori–defined genes were evaluated using micro-array technology (Affymetrix array) and compared between tissues obtained from NPC lesions and normal nasopharyngeal tissue. Our results provide important evidence linking the expression of genes involved in nitrosamine metabolism and DNA repair to NPC. The results also provide evidence for the involvement of a gene(s) on the telomeric end of the long arm of chromosome 14 in the etiology of NPC. A separate report from this study has evaluated expression levels in genes included on the Affymetrix array and related them to levels of expression of EBV genes at the nasopharynx (12).

Materials and Methods

Study Participants, Specimen Collection, and Handling. Participants were selected from among individuals who participated in a case control study of NPC conducted in Taipei, Taiwan, as described previously (6). As part of the case-control study, individuals presenting to one of two tertiary care centers in Taipei, Taiwan, with signs and symptoms compatible with NPC and who were to undergo a clinical evaluation and diagnostic biopsy were recruited after informed consent was obtained. Of the 680 potential cases who participated (participation rate, >95%), 375 had histologically confirmed NPC; the remaining 305 had benign conditions of the nasopharynx. At one of the two participating hospitals, biopsy specimens collected from the nasopharyngeal lesion were hemidissected, and one half of the collected tissue was flash frozen in liquid nitrogen within 1 minute of collection. An additional biopsy was collected from an adjacent area of the nasopharynx without macroscopic evidence of disease and similarly frozen. Lesional tissue was obtained from 147 NPC cases and 145 biopsy-negative controls, and adjacent normal tissue was collected from 244 individuals. For the present study, paired biopsies from 55 NPC cases and 6 biopsy-negative controls were available (final analysis included 41 specimens as detailed below). Financial constraints precluded a larger number of specimens from being included.

All participants in the study provided informed consent. The study was conducted with approvals by ethical review committees from the National Cancer Institute (Rockville, MD), the National Taiwan University (Taipei, Taiwan), and the University of Wisconsin-Madison Health Sciences (Madison, WI).

Tissue Histology. The frozen biopsy specimens were embedded in Tissue-Tek OCT (Sakura Finetek, Inc., Torrance, CA), and 6-μm serial cryosections were prepared. Step sections were stained with H&E, and intervening unstained sections were evaluated by wide-spectrum cytokeratin immunohisto-chemistry using monoclonal mouse anti-human cytokeratin primary antibody (AE1/AE3; 1:20 dilution) from DAKO and DAKO EnVision+ System, Peroxidase (3,3′-Diaminobenzidine) Mouse (DAKO Corp., Carpinteria, CA) as per manufacturer's protocol, and *in situ* hybridization for EBV EBERs was done using EBV Probe ISH kit (NCL-EBV-K) from Novocastra Laboratories Ltd. (Newcastle upon Tyne, United Kingdom) following manufacturer's protocol.

Tumor tissue was defined as section areas that stained positive for both cytokeratins and EBV EBERs.

Laser Capture Microdissection. For tumor enrichment from NPC sections containing <70% tumor tissue as well as to isolate epithelial compartment from normal healthy nasopharyngeal tissue specimens, we used laser capture microdissection (LCM). Most of the adjacent matched nontumor biopsies were found to harbor tumor cells and were not included for further study. To prepare and orient the sections for LCM, serial adjacent sections were briefly hematoxylin stained, dehydrated through graded ethanol, and air dried. Tumor identification and mapping were aided by referring to the adjacent slides evaluated by cytokeratin immunohistochemistry and EBV *in situ* hybridization. LCM was done using an Arcturus PixCell II laser dissecting microscope and CapSure Macro LCM caps (Arcturus Bioscience, Mountain View, CA).

RNA Extraction and Amplification. RNA was extracted from LCM-enriched or whole tissue sections using Trizol (Invitrogen, Carlsbad, CA), DNaseI treated, and amplified twice using Affymetrix, Santa Clara, CA, Small Sample Labeling Protocol VII. One twentieth of the second-round cDNA was used to evaluate success of the amplification procedure using quantitative real-time PCR for β-actin cDNA using Qiagen QuantiTect SYBR Green PCR kit (Qiagen, Valencia, CA).

RNA Labeling and Microarray Analysis. Based on histologic integrity and RNA yield, 31 NPC tumor specimens and 10 normal tissue specimens (6 obtained from the nasopharynx of NPC cases and 4 from the nasopharynx of biopsy-negative controls) were used for comprehensive human gene expression analysis. Half of the second-round cDNA was used as templates for bacteriophage T7 RNA polymerase to synthesize biotinylated antisense RNA for hybridization to Affymetrix Human Genome U133 Plus 2.0 oligonucleotide microarrays as per manufacturer's protocols.[11]

A Priori Identification of Genes and Affymetrix Probe Sets within Pathways/Chromosomal Regions of Interest. False-positive finding resultant from multiple comparisons is a problem in tissue microarray studies that evaluate genome-wide expression. To minimize this problem, we focused our evaluation on biological pathways and chromosomal regions selected for this study before initiation of the analysis. The following biological pathways were investigated based on prior evidence from published work linking them to NPC development: DNA repair, nitrosamine metabolism, immune/inflammatory response, and a group of ''other miscellaneous'' genes identified from a review of the NPC literature as being possibly involved in NPC pathogenesis. This latter group, although not representing a single biological pathway, will be referred to as such for simplicity of presentation. The chromosomal regions selected for study based on evidence from previous linkage studies of multiplex NPC families in China and Taiwan are as follows: 4p15.1-4q12 and 14q32.1-14q32.33.

---

[10] A. Hildesheim. Genome-wide linkage and disequilibrium analyses of nasopharyngeal carcinoma show complex etiology and identify multiple candidate genes, personal communication.

[11] http://www.affymetrix.com/support/technical/datasheets/human_datasheet.pdf; http://www.affymetrix.com/support/technical/technotes/hgu133_p2_technote.pdf.

The Affymetrix annotation file from October 12, 2004 provided the link between probe sets on the Affymetrix array and genes of interest. This annotation file provides gene names, chromosomal banding locations, and Gene Ontology descriptions.[12]

Genes within the DNA repair pathway were identified using the Web-based supplement[13] to the review by Wood et al. (13, 14). One hundred twenty-two genes within the base excision repair (BER), mismatch excision repair (MMR), nucleotide excision repair (NER), and other DNA repair pathways were identified (listing of genes available online).[14] These genes mapped to 211 probe sets on the Affymetrix array after filtering probe sets according to the criteria described in the statistical analysis section.

Genes within the nitrosamine metabolism pathway were identified by consultation with the National Cancer Institute Core Genotyping Facility and a review of the literature. Forty-six genes involved in nitrosamine metabolism were identified (listing of genes available online).[15] These genes mapped to 80 probe sets on the filtered Affymetrix array.

Genes within the immune and inflammatory response pathways were identified via review of Gene Ontology biological process and molecular function descriptions available through the Affymetrix annotations. Eight hundred twenty-four genes involved in immune and/or inflammatory responses were identified (listing of genes available online).[16] These genes mapped to 1,244 probe sets on the filtered Affymetrix array.

Other genes that were not included in the three pathways listed above but that were identified via a review of the literature as possibly linked to NPC pathogenesis comprised the final set of genes targeted a priori for evaluation. Twenty-three such genes were included in the list of ''other miscellaneous'' genes of interest (listing of genes available online).[17] These genes mapped to 47 probe sets on the filtered Affymetrix array.

Finally, probe sets included in the Affymetrix array representing genes located in either chromosomal regions 4p15.1-4q12 or 14q32.1-14q32.33 were targeted. The chromosomal banding locations of probe sets included on the Affymetrix array were defined using annotations available on the Affymetrix annotation file, which were based on the Unigene database.[18] For probe sets identified in this file as being in the regions of interest, more precise alignment of the probe sets was done using positional base number information obtained from the University of California at Santa Cruz database[19] (15). The vast majority of probe sets identified on the filtered Affymetrix array as being within the chromosomal bands of interest on chromosomes 4 and 14 were identified in the University of California at Santa Cruz database as being located within these two regions (159 of 162 for chromosome 4 and 250 of 256 for chromosome 14). The few discrepant probe sets were included in the differential expression analysis but excluded from the plots. The final number of probe sets from the regions of interest on chromosomes 4 and 14 used in the analysis was 162 and 256, respectively.

Statistical Analysis. Log (base 2) expression measures for each probe set were computed using robust multiarray average (16). Box plots of robust multiarray average–processed expression values showed acceptable normalization of arrays. Because our main interest was to find expression differences arising from biological processes unique to tumor specimens and not those simply due to differences between individuals, probe sets were filtered based on agreement in the direction of expression between the paired and complete set of samples. Specifically, any probe set for which the direction of gene regulation, as measured by the ratio of mean expression of tumor to normal, was not consistent between the complete set of samples and the four paired samples was excluded from the analysis. This left 44,416 filtered probe sets (out of 54,675) remaining for analysis.

Expression levels for tumor and normal specimens were compared using the two-sample pooled variance t test. The two-sample t test that treats all observations as independent might be slightly conservative. However, it should be more robust than a statistic incorporating the correlation inherent in the four paired observations that would rely more heavily on normality of the data.

The false discovery proportion was controlled at 1% (with 90% probability) using the permutation method developed by Korn et al. (17) as implemented in BRB ArrayTools.[20]

From this list of differentially expressed probe sets, a rate of differential expression was computed for the entire array as (no. of differentially expressed probe sets / no. of probe sets passing filtering criteria). For each of the prespecified lists, a rate of differential expression was computed as (no. of differentially expressed probe sets in the list / no. of probe sets on filtered array contained in the list). When a pathway is unimportant, differential expression occurs irrespective of pathway, and the differential expression rates for the complete array and the pathway are the same. We tested this null hypothesis against the alternative that the rates were higher in the list of interest by comparing the differential expression rates for probe sets included and excluded in the list. Probe sets within a list may be correlated (e.g., multiple probe sets often correspond to the same gene). Because of this potential correlation, Ps were estimated using the bootstrap with 2,500 bootstrap samples (18). For each bootstrap sample, arrays were sampled with replacement within group (i.e., tumor and normal). The differential expression list was generated as described above, and a difference in proportions was estimated. The bootstrap distribution of these differences was used to compute one-sided Ps.

Plots of expression values by chromosome location for probe sets located on regions of interest on chromosomes 4 and 14 were created using starting base numbers provided by the University of California at Santa Cruz database.[19]

To explore whether there was a consistent/preferential loss or gain of gene expression in the chromosomal regions of interest, the average expression in the specified regions was computed for each subject. To test whether the means of cancers were significantly different than that of normals, a two-sample t test was done.

To better understand relationships between genes in a pathway, Spearman rank correlations were computed for all possible pairs of differentially expressed genes. From these correlations, image plots were created in S-Plus 6.2 (Insightful Corp., Seattle, WA) with red, green, and black indicating strong positive, strong negative, and no correlation, respectively.

To explore whether results were due to late-stage cancers, expression levels of differentially expressed probe sets were examined for early-stage and normal specimens only. Twenty-one specimens were from early-stage patients, defined as patients with stage I or II tumors (including T1 and T2N2 tumors by the tumor-node-metastasis classification system). Expression ratio estimates for the differentially expressed probe sets were computed restricting to these 21 tumor and 10

[12] www.affymetrix.com.
[13] http://www.cgal.icnet.uk/DNA_Repair_Genes.html.
[14] http://linus.nci.nih.gov/Data/doddl/npc/; Appendix A.
[15] http://linus.nci.nih.gov/Data/doddl/npc/; Appendix B.
[16] http://linus.nci.nih.gov/Data/doddl/npc/; Appendix C.
[17] http://linus.nci.nih.gov/Data/doddl/npc/; Appendix D.
[18] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene.
[19] http://www.genome.ucsc.edu/; release date, May 2004.

[20] http://linus.nci.nih.gov/BRB-ArrayTools.html.

Table 1. Frequency of differentially expressed probe sets across the entire Affymetrix chip and for probe sets representing genes within a priori–defined biological pathways and chromosomal locations

A priori pathway/location	% Differential expression	(No. differentially expressed probe sets/no. probe sets)	$P*$	% Underexpression/ overexpression
Complete array	7.6	(3,367/44,416)		60/40
DNA repair	13.7	(29/211)	0.01	7/93
Immune/inflammatory response	9.9	(123/1,244)	0.06	59/41
Nitrosamine metabolism	17.5	(14/80)	0.04	100/0
Other miscellaneous	17.0	(8/47)	0.03	50/50
Chromosome 4p15.1-4q12 region	13.0	(21/162)	0.04	62/38
Chromosome 14q32.1-14q32.33 region	11.3	(29/256)	0.06	100/0

*P computed using a bootstrap method as described in Materials and Methods.

normal samples. These ratios were close to those based on the complete data set. To quantify this, a ratio of expression ratios for early-stage samples and all-stage samples was computed. The majority of ratios were ∼1, with 95% of ratios falling between 0.91 and 1.10 (listing of genes available online).[21] This suggests that the results are not simply due to an effect of late-stage tumors, and so, results based on all tumor and normal specimens are presented here.

Results

Overall Rate of Differential Expression. Gene expression in tumor and normal epithelial cells from the nasopharynx was compared. After filtering the array and controlling for the false discovery proportion as described in Materials and Methods, 7.6% of the probe sets evaluated were significantly differentially expressed in tumor versus normal specimens (Table 1; a complete listing of differentially expressed genes on filtered array is provided online).[22] Among these differentially expressed probe sets, 40% were significantly overexpressed in tumor specimens; the remaining 60% were significantly underexpressed in tumor relative to normal specimens.

Rate of Differential Expression within A Priori Pathways of Biological Interest. To evaluate whether differentially expressed probe sets were more likely than expected by chance to be contained within the DNA repair, immune/inflammatory response, nitrosamine metabolism, or "other miscellaneous" pathways, we compared the rate of differential expression observed within each of these pathways against the overall differential expression rate of 7.6% observed for the entire array. Significant elevations in the percentage differentially expressed probe sets were observed for the DNA repair (13.7%; $P = 0.01$), nitrosamine metabolism (17.5%; $P = 0.04$), and "other miscellaneous" (17.0%; $P = 0.03$) pathways (Table 1). A marginally significant elevation was observed for the immune/inflammatory response pathway (9.9%; $P = 0.06$). Within the DNA repair pathway, 93% (27 of 29) of differentially expressed probe sets were found to be overexpressed in tumor compared with normal specimens. For the nitrosamine metabolism pathway, 100% of differentially expressed probe sets were found to be underexpressed in tumor compared with normal specimens. This compares with a more balanced distribution of overexpressed and underexpressed probe sets when differentially expressed probe sets on the entire array were examined. A balanced distribution of overexpression and underexpression was also observed among differentially expressed probe sets within the "other miscellaneous" category (50%/50%), perhaps not surprising given the diversity of biological functions of these genes.

For the three pathways listed in Table 1 for which significant evidence of increased rate of differentially expressed probe sets relative to the entire filtered array was observed, a listing of genes found to be differentially expressed is presented (Table 2A-C). For each gene listed, the number of differentially expressed probe sets and total number of probe sets included on the filtered array are indicated. For each gene listed, the most extreme differential expression ratio observed for the differentially expressed probe sets within those genes is also listed as an indication of the magnitude of underexpression or overexpression observed in tumor versus normal specimens. Genes found to be underexpressed and those found to be overexpressed in tumors are listed separately in each table. A listing of differentially expressed genes within the immune/inflammatory response pathway is provided online.[23]

To further evaluate the interrelationship of genes differentially expressed within the DNA repair and nitrosamine metabolism pathways, we examined the correlation in expression levels between these genes. For this evaluation, tumor and normal specimens were combined. Results are summarized in Fig. 1A and B and indicate a high degree of interrelatedness in expression levels of genes identified as being differentially expressed within these pathways. For comparison, we evaluated the degree of correlation in expression levels for differentially expressed genes within our "other miscellaneous" group. As the genes included in this category have diverse biological function, there is no a priori expectation that their expression levels should be correlated. In fact, as evidenced in Fig. 1C and in contrast to Fig. 1A and B, there seems to be no strong pattern of correlation for differentially expressed genes within the "other miscellaneous" group.

Next, we evaluated whether differential expression observed in the DNA repair pathway could be explained by genes within a specific subset of DNA repair genes. We specifically evaluated whether probe sets for DNA repair genes within the following subpathways were preferentially involved: BER, chromatin structure (CS), DNA polymerases (catalytic subunits) (DNAP), direct reversal of damage (DRD), editing and processing nucleases (EPN), homologous repair (HR), mismatch excision repair (MER), modulation of nucleotide pools (MNP), nucleotide excision repair (NER), nonhomologous end joining (NHEJ), Rad6 pathway, repair of DNA-protein cross-links (RDPC), Fanconi anemia genes/other genes defective in diseases with sensitivity to DNA damaging agents (FANC/SDNAD), and other genes. Results are presented as supplementary information online.[24] There was no evidence to suggest that differential expression seen within the DNA repair pathway was explained by a specific repair pathway (Table 2A).

---

[21] http://linus.nci.nih.gov/Data/doddl/npc/; Appendix E.
[22] http://linus.nci.nih.gov/Data/doddl/npc/; Appendix F.

[23] http://linus.nci.nih.gov/Data/doddl/npc/; Appendix G.
[24] http://linus.nci.nih.gov/Data/doddl/npc/; Appendix Fig. 1.

Table 2. List of genes demonstrating significant differential expression

Gene symbol	Differentially expressed probe sets, n (%)	Total no. probe sets on filtered array	Most extreme ratio*	Pathway[†]
A. List of genes showing significant differential expression within the DNA repair pathway				
Down-regulated genes (expression lower in cancers than normals)				
CETN2	1 (100)	1	0.38	NER
NEIL1	1 (50)	2	0.65	BER
Up-regulated genes (expression higher in cancers than normals)				
NEIL3 (FLJ10858)	1 (100)	1	1.68	BER
TDG	1 (50)	2	1.80	BER
UNG	1 (100)	1	2.03	BER
MSH2	1 (100)	1	2.14	MMR
MSH6	1 (50)	2	2.12	MMR
PMS1	1 (25)	4	1.82	MMR
BRCA1	1 (50)	2	2.42	HR
BRCA2	1 (50)	2	1.32	HR
RAD51C	1 (100)	1	2.36	HR
RAD54B	1 (100)	1	2.34	HR
RAD54L	1 (100)	1	1.37	HR
PRKDC	2 (67)	3	2.91	NHEJ
DUT	3 (100)	3	2.17	MNP
NUDT1	1 (100)	1	2.12	MNP
PCNA	1 (100)	1	2.68	DNAP
POLQ	2 (100)	2	1.88	DNAP
RAD1	1 (20)	5	1.35	DNAP
EXO1	1 (100)	1	2.07	EPN
FEN1	2 (100)	2	2.10	EPN
FANCG	1 (100)	1	1.61	FANC
CHEK1	1 (50)	2	2.56	Other
CHEK2	1 (100)	1	1.85	Other
TOTAL	29 (67)	43		
B. List of genes showing significant differential expression within the nitrosamine metabolism pathway				
Down-regulated genes (expression lower in cancers than normals)				
CYP2A6	1 (25)	4	0.83	
CYP2B6	3 (75)	4	0.39	
CYP2C8	1 (100)	1	0.57	
CYP2E1	3 (100)	3	0.40	
NQO1	3 (100)	3	0.17	
UGT1A1/UGT1A10/UGT1A4/UGT	2 (100)	2	0.25	
UGT1A1/UGT1A4/UGT1A6	1 (100)	1	0.46	
TOTAL	14 (78)	18		
C. List of genes showing significant differential expression within the other miscellaneous pathway				
Down-regulated genes (expression lower in cancers than normals)				
MYCBP	2 (67)	3	0.53	
NESG1	1 (100)	1	0.18	
THRA	1 (25)	4	0.8	
Up-regulated genes (expression higher in cancers than normals)				
MKI67	3 (75)	4	1.92	
NME1	1 (100)	1	2.69	
TOTAL	8 (62)	13		
D. List of genes showing significant differential expression within chromosome 4 region of interest				
Down-regulated genes (expression lower in cancers than normals)				
LOC132671	2 (100)	2	0.13	
HOP	1 (50)	2	0.20	
FLJ13352	2 (67)	3	0.34	
FLJ14001	1 (20)	5	0.35	
FLJ11017	1 (100)	1	0.38	
LNX	1 (50)	2	0.39	
FLJ21511	2 (100)	2	0.48	
LOC201895	1 (50)	2	0.49	
KIAA1458	1 (17)	6	0.67	
SPINK2	1 (100)	1	0.80	
Up-regulated genes (expression higher in cancers than normals)				
POLR2B	2 (67)	3	1.43	
RFC1	1 (33)	3	1.51	
SRP72	2 (67)	6	1.85	
KLHL5	1 (100)	1	2.02	
KIAA0635	1 (50)	2	2.30	
PPAT	1 (50)	2	2.44	
TOTAL	21 (49)	43		

(*Continued on the following page*)

Table 2. List of genes demonstrating significant differential expression (Cont'd)

Gene symbol	Differentially expressed probe sets, *n* (%)	Total no. probe sets on filtered array	Most extreme ratio*	Pathway[†]
E. List of genes showing significant differential expression within chromosome 14 region of interest[‡]				
Down-regulated genes (expression lower in cancers than normals)				
AK7	1 (33)	3	0.19	
C14orf78	1 (50)	2	0.19	
CRIP1	1 (100)	1	0.23	
IgHM	2 (40)	5	0.24	
ADSSL1	1 (100)	1	0.27	
KIAA0500	1 (100)	1	0.29	
CLMN	3 (100)	3	0.30	
CKB	1 (100)	1	0.33	
IGHG1	1 (14)	8	0.33	
TCL1A	1 (50)	2	0.35	
C14orf142	1 (100)	1	0.35	
RAGE	1 (100)	1	0.42	
C14orf129	1 (100)	1	0.45	
ZFYVE21	1 (50)	2	0.49	
C14orf66	1 (100)	1	0.58	
SLC25A29	1 (25)	4	0.67	
MGC4645	1 (100)	1	0.71	
BTBD7	1 (20)	5	0.73	
C14orf79	1 (100)	1	0.73	
KIAA0284	1 (100)	1	0.74	
BAG5	1 (100)	1	0.74	
KIAA0329	1 (50)	2	0.76	
CDC42BPB	1 (50)	2	0.77	
C14orf131	1 (33)	3	0.79	
TOTAL	27 (51)	53		

Abbreviations: NER, nucleotide excision repair; BER, base excision repair; MMR, mismatch excision repair; HR, homologous recombination; NHEJ, nonhomologous end joining; MNP, modulation of nucleotide pools; DNAP, DNA polymerase (catalytic subunits); EPN, editing and processing nucleases; FANC, genes associated with Fanconi anemia.
*Defined as the maximum ratio if >1 or the minimum ratio if <1.
[†]Categories based on Web supplement to Wood et al. as described in Materials and Methods.
[‡]Two differentially expressed probe sets located in this region did not have a corresponding gene symbol and are not included in this table.

Rate of Differential Expression within A Priori Chromosomal Locations of Interest. Next, we evaluated differential expression within the two chromosomal regions previously shown to be linked to NPC in family-based studies conducted in China and Taiwan: 4p15.1-4q12 and 14q32-14q32.33. As summarized in Table 1, significant evidence for an excess rate of differential expression was evident in the region of interest within chromosome 4 when differential expression rate was compared against the rate observed across the entire array (13.0% versus 7.6%; $P = 0.04$). A listing of the specific genes in the 4p15.1-4q12 region found to be differentially expressed is presented in Table 2D. An increase in the rate of differential expression was also observed for the region of interest within chromosome 14, a difference that was marginally statistically significant (11.3% versus 7.6%; $P = 0.06$). A listing of specific genes in the 4p15.1-4q12 region differentially expressed is presented in Table 2E. Interestingly, although the proportion of overexpressed and underexpressed probe sets were relatively balanced for chromosome 4 (62% of differentially expressed probe sets were underexpressed and 38% were overexpressed in tumors compared with normal specimens), a strong tendency for differentially expressed probe sets to be underexpressed was noted for chromosome 14 (100% of differentially expressed probe sets were underexpressed in tumors compared with normal specimens; $P < 0.0001$). To further evaluate this, all filtered probe sets located within the regions of interest on chromosome 4 and 14 were plotted by location (X axis) and the ratio (tumor/normal) of mean expression (Fig. 2A and B). The tendency for underexpression at the telomeric end of chromosome 14 is clearly evident. A significant reduction was evident for chromosome 14, where the mean expression level for tumor specimens was 6.75

compared with 7.06 for normal specimens ($P < 0.0001$). No such difference was evident for chromosome 4 (mean among tumor specimens, 6.65; mean among normal specimens, 6.68; $P = 0.52$). In addition, when we evaluated the tendency for underexpression at the telomeric end of chromosome 14 by individual tumor specimens (Fig. 3), the pattern of underexpression was broadly evident across most tumor specimens evaluated. The most strong and consistent evidence for underexpression occurred in the region located between base number positions 94,500,000 and 105,900,000.

Discussion

In this study, we had the opportunity to evaluate patterns of gene expression in cells from NPC lesions and from normal nasopharynx. Because our objective was to better understand the etiology of NPC, we focused on specific biological pathways of interest and on chromosomal regions that previous studies suggested were involved in NPC pathogenesis. This strategy had the advantage of reducing the number of statistical comparisons made and of focusing our analysis on biologically relevant genes when evaluating the tens of thousands of probe sets/genes represented on the Affymetrix array.

Results from our study provide independent support for a role in NPC etiology of nitrosamines and DNA repair. A more than doubling of the expected proportion of differentially expressed genes involved in nitrosamine metabolism was observed. Similarly, rates of differentially expressed genes involved in DNA repair were observed at nearly double the expectation. In addition, within these two pathways, a
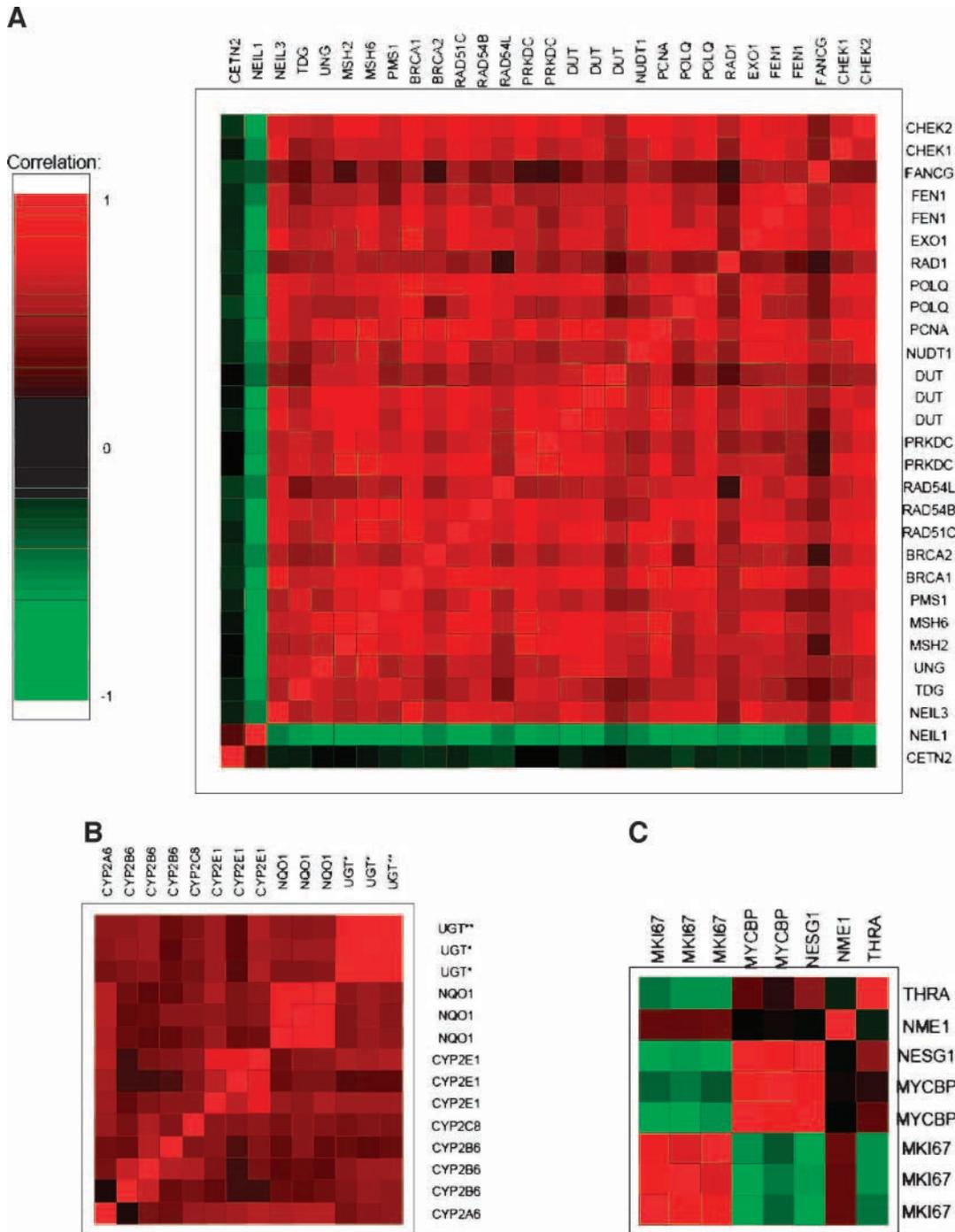
Figure 1. Evidence for correlations in expression levels seen for differentially expressed genes within biological pathways of interest. The correlation in expression levels between differentially expressed genes in the DNA repair (**A**), nitrosamine (**B**), and "other miscellaneous" (**C**) pathways is presented. *Red,* positive correlation (with *bright red* indicating complete positive correlation); *black,* lack of correlation; *green,* negative correlation (with *bright green* indicating complete negative correlation). *Vertical* and *horizontal axes,* gene labels.

consistent positive correlation was observed between the levels of expression of the various genes, supporting the notion that these genes are involved in related biological activities and that they work in concert. It is interesting to note that a strong tendency for underexpression of genes in the nitrosamine metabolism pathway and for overexpression of genes in the DNA repair pathway was observed. Although the explanation for this finding is not known, one might speculate that down-regulation of nitrosamine metabolism genes might reduce the host's ability to efficiently metabolize these potentially harmful compounds, leading to increased disease risk. Conversely,

the observed up-regulation in expression levels of genes in the DNA repair pathway might suggest an unsuccessful attempt by the host to recover from DNA damage. Lack of effective DNA repair despite high levels of expression of DNA repair genes might be expected among individuals who carry DNA repair genotypes with inherently reduced capacity for efficient repair.

Specific genes involved in nitrosamine metabolism found to be most strongly differentially expressed in NPC include *NQ01, CYP2B6,* and *CYP2E1*. Interestingly, common polymorphisms in *CYP2E1* are associated with NPC in at least two
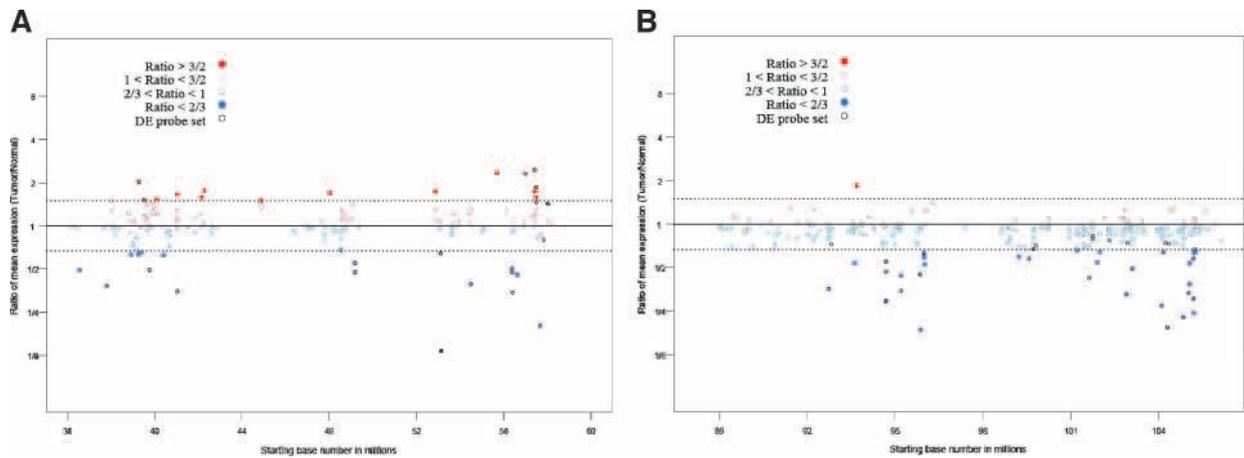
Figure 2. Evaluation of overexpression and underexpression tendencies for tumor relative to normal cells within chromosomal regions of interest. Plots of chromosomal location based on the starting base number in millions (*X axis*) and the tumor to normal ratio of mean expression levels observed (*Y axis*) are presented. A ratio of 1.0 indicates that the mean expression observed for tumor cells and normal cells is identical. Ratios >1.0 indicate that, on average, expression levels are higher in tumor cells compared with normal cells. Ratios <1.0 indicate that, on average, expression levels are lower in tumor cells compared with normal cells. All probe sets identified within the chromosomal areas of interest are included as specified in Materials and Methods. *Solid red circles,* ratios >1.5; *open red circles,* ratios between 1.0 and <1.5; *solid blue circles,* ratios <0.67; *open blue circles,* ratios between 0.67 and 1.0. *External black lines around the circles,* probe sets found to be significantly differentially expressed. **A.** Plot for centromeric region of chromosome 4 (4p15.1-4q12). **B.** Plot for the telomeric end of chromosome 14 (14q32-end).

studies (5, 6). Within the DNA repair pathway, the genes most strongly differentially expressed in NPC were *PRKDC, PCNA,* and *CHEK1*. Of note, *PRKDC* is a subunit of a nuclear DNA-dependent serine/threonine protein kinase that is involved not only in cell cycle control, response to DNA damage, and DNA repair but also in V(D)J recombination in developing B and T cells. Cells defective in DNA-dependent serine/threonine protein kinase are sensitive to killing by ionizing radiation
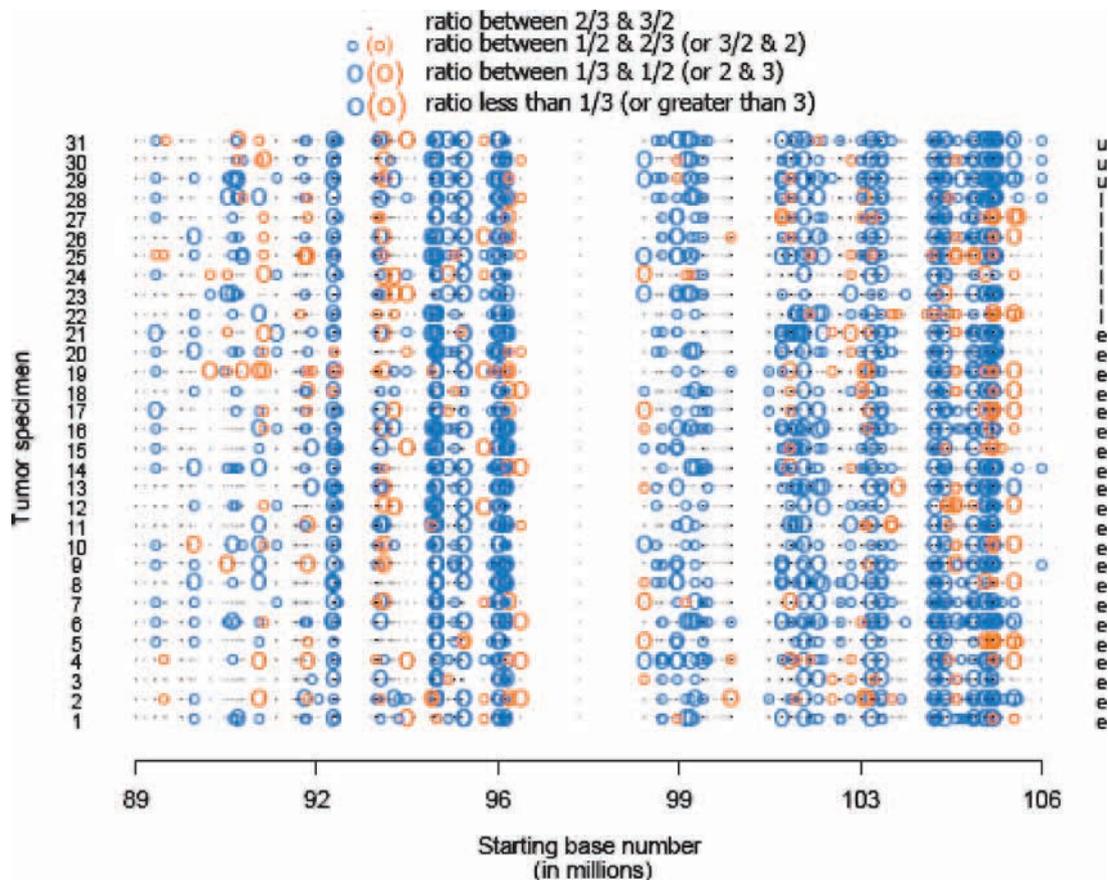
Figure 3. Evaluation of expression differences between tumor and normal cells within the telomeric region of chromosome 14 (14q-32-end). Each row (*Y axis*) plots expression ratios for individual tumor specimens relative to mean normal expression. *X axis,* chromosomal location. *Circle size* and *color,* magnitude of the expression ratio as indicated on the plot.

(because of an inability to repair dsDNA breaks) and to also be unable to do V(D)J recombination necessary to create variable regions in immunoglobulins and T-cell receptor genes (19).

Our results also provide support for the involvement of a gene(s) at the telomeric end of chromosome 14 in the etiology of NPC. A 49% increase in the rate of differentially expressed probe sets was noted in this chromosomal region, and the tendency was for gene expression to be down-regulated for all differentially expressed probe sets in this region. In addition, this systematic tendency for reduced gene expression was observed for a large fraction of individuals investigated. This phenomenon of down-regulation of genes that are biologically distinct but share in common a physical chromosomal proximity suggests that loss of the telomeric end of chromosome 14 is a frequent occurrence in NPC pathogenesis and that a gene within this region of loss is important in promoting disease formation. Interestingly, two of the genes found to be down-regulated in NPC in our study are the *IgHM* (immunoglobulin heavy constant μ chain) and *IgHG1* (immunoglobulin heavy constant γ 1 chain) genes. Studies of Burkitt's lymphoma, another neoplasia associated with EBV, have reported frequent chromosome 8q24;14q32 translocations involving the IgH genes on chromosome 14 (20). Future studies are needed to elucidate the biological significance of this finding.

Our data also suggested an excess in differentially expressed genes located in the region of chromosome 4 linked to NPC in a family study conducted in China (71% increase in the rate of differentially expressed probe sets observed). Unlike findings for chromosome 14, however, differentially expressed probe sets on chromosome 4 were both up-regulated and down-regulated. Because evidence from positional studies (i.e., genome-wide anonymous microsatellite marker scans of individuals within multiplex NPC families) points to a single or, at most, few genes in this region being linked to NPC, one would not have expected the overall rate of differentially expressed genes to be elevated in this region unless there was either consistent chromosomal gains or losses involved. If this were the case, then the differentially expressed genes should have been differentially expressed in a consistent direction (as was seen for chromosome 14). The interpretation of a significant elevation in the rate of differentially expressed genes with equal representation of overexpression and underexpression among differentially expressed genes is unclear at this time and requires further study.

In contrast to our finding for nitrosamine metabolism and DNA repair genes, no statistically significant differences were observed in the degree of expression of genes involved in immune and inflammatory responses in cells from NPC compared with normal nasopharynx. These results should be interpreted in light of the extensive data linking EBV to the development of NPC and consistent findings suggesting that immune response is an important determinant of NPC risk (e.g., HLA). Although it is unclear why statistically significant differences in the rate of differentially expressed genes within this pathway were not evident, one possible explanation is that the a priori pathway was defined too broadly (824 genes were identified in the pathway and evaluated on the Affymetrix array) and that the differential expression of genes within specific functionally distinct subcategories of immune and inflammatory responses within our broad group might exist. In fact, in a parallel evaluation by our group of joint host and EBV expression patterns in NPC tissue, we observed that increases in EBV expression levels were associated with concomitant decreases in the expression of host genes involved in antigen processing and presentation (12). Another possible explanation for the lack of a statistically significant effect seen for the immune/inflammatory pathway might be that the effect of variability in immune response to EBV is expressed phenotypically in cells other than the nasopharynx (e.g.,

B cells known to harbor latent, lifelong EBV infections) and that we evaluated expression of immune response genes in an irrelevant compartment.

Limitations of our study should be discussed. First, most of our tumor and normal specimens were derived from different individuals and so there is the possibility that differences observed between tumor and normal specimens represent differences between individuals rather than differences between cell types within individuals. To guard against this possibility, we filtered the probe sets on the Affymetrix array up front and included only those probe sets where the direction of expression differences between tumor and normal specimens matched that observed for the four paired tumor-normal pairs included in our study. Second, the sample size of our study is modest. To guard against the possibility of false-positive results resultant from large numbers of comparisons made using a relatively small number of specimens, we limited our evaluation to predefined genes within biological pathways or chromosomal locations of interest based on previous work in this area. Furthermore, our statistical approach controlled for the false discovery proportion (1% with 90% probability). These efforts, although designed to minimize the possibility that the findings presented are false signals, likely resulted in a higher rate of false negatives in our study. Third, tissue evaluated in the present study was collected at the time of diagnosis for NPC cases. Whether gene expression patterns observed at the time of diagnosis reflect patterns that led to disease development in the first place can be questioned. Finally, because individual genes/proteins are often involved in multiple distinct biological processes, we cannot rule out the possibility that our findings reflect effects in biological pathways other than those discussed herein.

In summary, our findings lend support for the involvement of nitrosamines/nitrosamine metabolism and DNA repair in the etiology of NPC. Our results also provide evidence that a gene(s) located on the telomeric end of chromosome 14 is involved in NPC pathogenesis. Future efforts will be needed to define the specific gene(s) on chromosome 14 involved in NPC etiology. Further work is also required to refine our understanding of the interplay between exogenous exposure to nitrosamines, the ability to metabolize nitrosamines efficiently, and the ability to repair DNA damage induced by reactive intermediates resultant from the nitrosamine metabolism process in the etiology of NPC.

References

1. Yu MC, Henderson BE. Nasopharyngeal cancer. In: Schottenfeld D, Fraumeni JF, editors. Cancer epidemiology and prevention. 2nd ed. New York: Oxford University Press; 1996. p. 603–18.
2. Yu MC, Yuan JM. Epidemiology of nasopharyngeal carcinoma. Semin Cancer Biol 2002;12:421–9.

3. Ward MH, Pan WH, Cheng YJ, et al. Dietary exposure to nitrite and nitrosamines and risk of nasopharyngeal carcinoma in Taiwan. Int J Cancer 2000;86:603–9.

4. Cheng YJ, Hildesheim A, Hsu MM, et al. Cigarette smoking, alcohol consumption, and risk of nasopharyngeal carcinoma in Taiwan. Cancer Causes Control 1999;10:201–7.

5. Kongruttanachok N, Sukdikul S, Setavarin S, et al. Cytochrome P450 2E1 polymorphism and nasopharyngeal carcinoma development in Thailand: a correlative study. BMC Cancer 2001;1:4.

6. Hildesheim A, Anderson LM, Chen CJ, et al. CYP2E1 genetic polymorphisms and risk of nasopharyngeal carcinoma in Taiwan. J Natl Cancer Inst 1997;89:1207–12.

7. Cho E, Hildesheim A, Chen CJ, et al. Nasopharyngeal carcinoma and genetic polymorphisms of DNA repair enzymes XRCC1 and hOGG1. Cancer Epidemiol Biomarkers Prev 2003;12:1100–4.

8. 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. Int J Cancer 2005;115:742–6.

9. 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 Natl Cancer Inst 2002;94:1780–9.

10. Butsch Kovacic M, Martin M, Gao X, et al. Variation of the killer cell immunoglobulin-like receptors and HLA-C genes in nasopharyngeal carcinoma. Cancer Epidemiol Biomarkers Prev 2005;14:2673–7.

11. Feng BJ, Huang W, Shugart YY, et al. Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. Nat Genet 2002;31:395–9.

12. Sengupta S, den Boon JA, Chen IH, et al. Genome-wide expression profiling reveals Epstein-Barr virus-associated inhibition of MHC class I expression in nasopharyngeal carcinoma. Cancer Res 2006;66:7999–8006.

13. Wood RD, Mitchell M, Lindahl T. Human DNA repair genes, 2005. Mutat Res 2005;577:275–83.

14. Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. Science 2001;291:1284–9.

15. Karolchik D, Baertsch R, Diekhans M, et al. The UCSC Genome Browser Database. Nucleic Acids Res 2003;31:e15.

16. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 2003;31:15.

17. Korn EL, Troendle JF, McShane LM, Simon R. Controlling the number of false discoveries: application to high-dimensional genomic data. J Stat Plan Infer 2004;124:379–98.

18. Efron B, Tibshirani RJ. An introduction to the bootstrap. Boca Raton: CRC Press; 1998.

19. Hartley KO, Gell D, Smith GCM, et al. DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. Cell 1995;82:849–56.

20. Hecht JL, Aster JC. Molecular biology of Burkitt's lymphoma. J Clin Oncol 2000;18:3707–21.