Integrated Analysis of Multiple Gene Expression Profiling Datasets Revealed Novel Gene Signatures and Molecular Markers in Nasopharyngeal Carcinoma

Chen Huang1,3, Hailin Tang1, Wenling Zhang2, Xiaoling She1, Qianjin Liao1, Xiaoling Li1, Minghua Wu1,4, and Guiyuan Li1

Abstract

Purpose: To identify the novel gene signatures and molecular markers of nasopharyngeal carcinoma (NPC) by integrated bioinformatics analysis of multiple gene expression profiling datasets.

Experimental Design: Seven published gene expression profiling studies and one of our unpublished works were reanalyzed to identify the common significantly dysregulated (CSD) genes in NPC. Overrepresentation analysis of cytogenetic bands, Gene Ontology (GO) categories, pathways were used to explore CSD genes functionally associated with carcinogenesis. The protein expressions of selected CSD genes were examined by immunohistochemistry on tissue microarrays, and the correlations of their expressions with clinical outcomes were evaluated.

Results: Using the criteria (genes reported deregulated in more than one study), a total of 962 genes were identified as the CSD genes in NPC. Four upregulated (BUB1B, CCND2, CENPF, and MAD2L1) and two downregulated (LTF and SLPI) genes were markedly reported in six studies. The enrichments of chromosome aberrations were 2q23, 2q31, 7p15, 12q15, 12q22, 18q11, and 18q12 in upregulated genes and 14q32 and 16q13 in downregulated genes. The activated GO categories and pathways related to proliferation, adhesion, invasion, and downregulated immune response had been functionally associated with NPC. SLPI significantly downregulated in nasopharyngeal adenocarcinoma. Furthermore, the high expression of BUB1B or CENPF was associated with poor overall survival of patients.

Conclusion: It was first clearly identified the dysregulated expression of BUB1B and SLPI in NPC tissues.

Impact: Further studies of the CSD genes as gene signatures and molecular markers of NPC might improve the understanding of the disease and identify new therapeutic targets.

Cancer Epidemiol Biomarkers Prev; 21(1); 166–75. ©2011 AACR.

Introduction

Nasopharyngeal carcinoma (NPC) is a malignant epithelial head and neck carcinoma that is highly prevalent in southern China and Southeast Asia (1). Multiple molecular studies have reported that NPC is a highly aneuploid and genetically complex tumor that develops in a multistep process involving alterations of numerous genes. However, the mechanism of carcinogenesis and progression of NPC is not well understood. Moreover, the prognosis of NPC patients is related to its potent localized invasion and metastatic spread. A clearer understanding of the molecular signature and underlying mechanisms would be beneficial in providing effective biomarkers for therapeutic planning and intervention.

Gene expression profiling analysis using microarray or array technologies has emerged in the last decade as a powerful tool to search for reliable gene markers of specific cancers and to determine the molecular features of cancer progression and the clinical outcome in greater detail (2). Moreover, because the genome-wide set of dysregulated genes and proteins are known to function in regulated biological networks or pathways associated with cancers, the focus of data obtained from gene expression profiling in cancers has been directed toward analysis of these sets of genes or pathways. In the literature, several reports have described dysregulation of gene expression...
in NPC and have found a vast amount of genome-wide information (3–11). However, the overall molecular signature of NPC is still undefined. Focusing on dysregulated genes and pathways in a single gene expression profiling dataset is not the most efficient method for understanding the molecular signature of NPC because only a few clinical samples can be validated and studied in detail at any given time. Hence, an integrated analysis would be useful in understanding the cross-talk among the gene expression profiling datasets of NPC clinical samples from different laboratories and would maximize the biological information that can be collected from genome-wide, gene expression profiling datasets of NPC.

We hypothesized that common significant gene expression changes between different microarray studies may be identified by reclassifying and reanalyzing publicly available gene expression profiling data. To understand the underlying mechanisms of NPC carcinogenesis, and to develop new therapeutic approaches targeted to NPC-specific molecular abnormalities, gene expression data were examined by bioinformatics analysis tools that searched for chromosome aberrations, biological annotations, and pathways. Then, expression of selected dysregulated genes in NPC was examined by immunohistochemistry (IHC) on tissue microarrays (TMA), and the correlation of protein expression with clinical outcomes was evaluated.

**Material and Methods**

**Literature search and identification of the common significantly dysregulated genes in NPC**

For data collection, a literature survey of gene expression data published in NPC was carried out. Articles were selected that satisfied the predetermined criteria: sample from human tissues and the preparation were detailed, technology for gene expression studies was defined, and detailed results of the gene expression changes were available. Studies concerning single genes or arbitrarily selected genes were discarded. Results obtained from animal models were not considered. Eight NPC dysregulated gene lists from 7 published gene expression profiling studies and one of our unpublished works were included in our analysis (Table 1). All given gene details were converted into the National Center for Biotechnology Information official gene symbol by The Database for Annotation, Visualization and
Integrated Discovery (DAVID) v6.7 (12, 13). The dysregulated genes reported in 2 or more studies were selected.

Cytogenetic band analysis

Enrichment analysis for the cytogenetic band was carried out by using WebGestalt toolkit (14, 15). For identification of the enriched cytogenetic band of genes, the whole human genome was used as a reference list. The hypergeometric statistical method test was used, and only statistically enriched terms ($P < 0.05$) with at least 2 genes were selected.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was the WebGestalt toolkit. This toolkit includes information from Gene Ontology Tree Machine software (GOTM; ref. 16). WebGestalt queries were carried out with lists of Entrez Gene IDs. For identification of enriched GO and KEGG pathway terms (i.e., GO terms with a number of associated genes significantly higher than expected), the whole human genome was used as a reference list. The hypergeometric statistical method test was used, and only statistically enriched terms ($P < 0.05$) with at least 2 genes were selected.

IHC

The TMA prepared from NPC and normal tissue specimens collected at the Second Xiangya Hospital (Changsha, Hunan, China) was those reported in our previous studies (11, 17). IHC was carried out using the peroxidase antiperoxidase technique after a microwave antigen retrieval procedure. Antibodies against CENPF (clone 14C10/1D8, dilution 1:200), BUB1B (clone 8G1, dilution 1:100), and SLPI (clone 31, dilution 1:50) were obtained from Abcam. Antibodies against CENPF, BUB1B, and SLPI were overlaid on TMA and incubated overnight at 4°C. After incubation with secondary antibody (Dako Envision+ system, peroxidase conjugate) at room temperature for 30 minutes and staining with DAB (Dako Envision system). Two pathologists independently scored the results of IHC staining. Both staining intensity and positive areas of each tissue core on the TMA were recorded. The staining intensity was scored as 0 (no staining), 1 (weak staining exhibited as light yellow), 2 (moderate staining exhibited as yellow brown), or 3 (strong staining exhibited as brown). The extension of staining was scored as 0 (0%), 1 (1%–25%), 2 (26%–50%), or 3 (51%–100%), according to the proportion of positive cells of interest. Cases with a combined staining score (intensity + extension) of 1 or greater were considered positive. For the purpose of statistical evaluation, the expression level of protein was classified that the combined staining score (intensity + extension) of 2 or more was low expression and the score between 3 and 6 was high expression.

Statistical analysis

Associations between pathologic parameters and protein expressions were evaluated using the Pearson $\chi^2$ test or Fisher’s exact test. Overall survival and disease-specific survival curves were constructed by the Kaplan–Meier method, and the differences between the curves were compared by log-rank tests. Overall survival was calculated from the date of diagnosis of NPC to the date of death. Patients who died as a result of or who had the disease at the time of follow-up were included in the analysis. For disease-specific survival, data for patients who died from other causes were censored at the time of death. Univariate analysis with Cox proportional hazards model was used to determine each identified prognostic factor.

Results

Identification of common significantly dysregulated genes in NPC from gene expression profiling datasets

A total of 3,916 unique and mapped dysregulated genes were differentially expressed in the 8 independent gene expression profiling studies on NPC (Supplementary Fig. S1). From these genes, 2,135 genes were upregulated and 1,781 were downregulated in NPC samples. Then, 962 dysregulated genes (651 upregulated and 311 downregulated), which were identified in more than 1 study were collocated and referred to as the “common significantly dysregulated genes” (CSD genes) in NPC (Supplementary Table S1). Of these CSD genes, 30 upregulated and 6 downregulated genes overlapped in 5 and more studies (Table 2). Four upregulated genes ($BUB1B$, $CCND2$, $CENPF$, and $MAD2L1$) and 2 downregulated genes ($LTF$ and $SLPI$) were identified in 6 different studies.

Chromosomal distribution of CSD genes in NPC

The chromosomal distributions of CSD genes associated with NPC were analyzed in an attempt to identify the genetic hotspots and gene expression patterns of NPC. As shown in Fig. 1, CSD genes were represented as the peaks of gene expression in chromosomes 1, 2, and 7. Chromosome 1 contained 112 (11.64%) CSD genes, followed by chromosomes 2 and 7, which contained 88 (9.15%) and 67 (6.96%) CSD genes, respectively. Chromosome Y did not present any CSD genes. The upregulated CSD genes showed peaks in chromosomes 2 (71/651, 10.91%), 1 (70/651, 10.75%), and 7 (60/651, 9.22%). The downregulated CSD genes exhibited peaks of gene expression in chromosomes 1 (42/311, 13.50%), 3 (27/311, 8.68%), and 6 (21/311, 6.75%). Enrichment analysis for the cytogenetic band revealed the marked enrichments in the upregulated genes residing at 2q, 2q23, 2q31, 7p, 7p15, 12q, 12q15, 12q22, 18q, 18q11, and 18q12 and in the downregulated genes residing at 14q32 and 16q13 (Supplementary Table S2).

GO analysis of CSD genes in NPC

By WebGestalt, 120, 120, and 77 GO categories were enriched in CSD genes, upregulated genes, and
downregulated genes in NPC, respectively (Supplementary Table S3). The Directed Acyclic Graph (DAG) diagram visualized the close interaction of the enriched GO terms (Supplementary Fig. S2). As shown in Fig. 2, 5 main enriched biological processes of CSD genes were outlined: cell cycle and proliferation, DNA replication, development, immune process, and response to stimulus. The most significantly enriched biological processes of upregulated genes were related to the cell cycle and proliferation, DNA replication, and development. The top 5 enriched biological processes in the downregulated genes’ list were related to immune system process and response to stimulus, including the humoral immune response, protein maturation by peptide bond cleavage, the activation of plasma proteins involved in acute inflammatory response, the protein maturation, and the defense response. The dominant molecular function of CSD and upregulated genes was “binding,” especially protein binding and nucleotide binding, whereas that of downregulated genes was “oxidoreductase activity.”

### Table 2. CSD genes in NPC identified in 5 and more studies

<table>
<thead>
<tr>
<th>Official gene symbol</th>
<th>Gene name</th>
<th>Cytoband</th>
<th>Entrez gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upregulated genes in NPC</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BUB1B</td>
<td>Budding uninhibited by benzimidazoles 1 homolog β (yeast)</td>
<td>15q15</td>
<td>701</td>
</tr>
<tr>
<td>CCND2</td>
<td>Cyclin D2</td>
<td>12p13</td>
<td>894</td>
</tr>
<tr>
<td>CENPF</td>
<td>Centromere protein F, 350/400 ka (mitosin)</td>
<td>1q32-q41</td>
<td>1063</td>
</tr>
<tr>
<td>MAD2L1</td>
<td>MAD2 mitotic arrest deficient-like 1 (yeast)</td>
<td>4q27</td>
<td>4085</td>
</tr>
<tr>
<td>C12orf5</td>
<td>Chromosome 12 open reading frame</td>
<td>12p13.3</td>
<td>57103</td>
</tr>
<tr>
<td>CAPRIN1</td>
<td>Cell-cycle–associated protein 1</td>
<td>11p13</td>
<td>4076</td>
</tr>
<tr>
<td>CCNA2</td>
<td>Cyclin A2</td>
<td>4q25-q31</td>
<td>890</td>
</tr>
<tr>
<td>CCNB1</td>
<td>Cyclin B1</td>
<td>5q12</td>
<td>891</td>
</tr>
<tr>
<td>CCT6A</td>
<td>Chaperonin containing TCP1, subunit 6A (zeta 1)</td>
<td>7p11.2</td>
<td>908</td>
</tr>
<tr>
<td>CDK1</td>
<td>Cell division cycle 2, G1 to S and G2 to M</td>
<td>10q21.1</td>
<td>983</td>
</tr>
<tr>
<td>E2F3</td>
<td>E2F transcription factor 3</td>
<td>6p22</td>
<td>1871</td>
</tr>
<tr>
<td>EZH2</td>
<td>Enhancer of zeste homolog 2 (Drosophila)</td>
<td>7q35-q36</td>
<td>2146</td>
</tr>
<tr>
<td>FZD7</td>
<td>Frizzled homolog 7 (Drosophila)</td>
<td>2q33</td>
<td>8324</td>
</tr>
<tr>
<td>GAD1</td>
<td>Blutamate decarboxylase 1 (brain, 67 kDa)</td>
<td>2q31</td>
<td>2571</td>
</tr>
<tr>
<td>GJA1</td>
<td>Gap junction protein, α 1, 43 kDa</td>
<td>6q21-q23.2</td>
<td>2697</td>
</tr>
<tr>
<td>KIF14</td>
<td>Kinesin family member 14</td>
<td>1q32.1</td>
<td>9928</td>
</tr>
<tr>
<td>MELK</td>
<td>Maternal embryonic leucine zipper kinase</td>
<td>9p13.2</td>
<td>9833</td>
</tr>
<tr>
<td>NUSAP1</td>
<td>Nucleolar and spindle associated protein 1</td>
<td>15q15.1</td>
<td>51203</td>
</tr>
<tr>
<td>PAICS</td>
<td>Phosphoribosylaminomimidazole carboxylase,</td>
<td>4q12</td>
<td>10606</td>
</tr>
<tr>
<td></td>
<td>phosphoribosylaminomimidazole succinocarboxamide synthetase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCNA</td>
<td>Proliferating cell nuclear antigen</td>
<td>20pter-p12</td>
<td>5111</td>
</tr>
<tr>
<td>PMAP1</td>
<td>Phorbol-12-myristate-13-acetate-induced protein 1</td>
<td>18q21.32</td>
<td>5366</td>
</tr>
<tr>
<td>PRC1</td>
<td>Protein regulator of cytokinesis 1</td>
<td>15q26.1</td>
<td>9055</td>
</tr>
<tr>
<td>RBBP8</td>
<td>Retinoblastoma binding protein 8</td>
<td>18q11.2</td>
<td>5932</td>
</tr>
<tr>
<td>RFC3</td>
<td>Replication factor C (activator 1) 3, 38 kDa</td>
<td>13q12.3-q13</td>
<td>5983</td>
</tr>
<tr>
<td>SMC4</td>
<td>Structural maintenance of chromosomes 4</td>
<td>3q26.1</td>
<td>10051</td>
</tr>
<tr>
<td>TMOPOETIN</td>
<td>Thymopoietin</td>
<td>12q22</td>
<td>7112</td>
</tr>
<tr>
<td>TNFαP6</td>
<td>Tumor necrosis factor, α-induced protein 6</td>
<td>2q23.3</td>
<td>7130</td>
</tr>
<tr>
<td>TOP2A</td>
<td>Topoisomerase (DNA) II α 170 kDa</td>
<td>17q21-q22</td>
<td>7153</td>
</tr>
<tr>
<td>TYMS</td>
<td>Thymidylate synthetase</td>
<td>18p11.32</td>
<td>7298</td>
</tr>
<tr>
<td>ZW10INTER</td>
<td>ZW10 interactor</td>
<td>10q21-q22</td>
<td>11130</td>
</tr>
<tr>
<td><strong>Downregulated genes in NPC</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTF</td>
<td>Lactotransferrin</td>
<td>3p21.31</td>
<td>4057</td>
</tr>
<tr>
<td>SLP</td>
<td>Secretory leukocyte peptidase inhibitor</td>
<td>20q12</td>
<td>6590</td>
</tr>
<tr>
<td>CLU</td>
<td>Clusterin</td>
<td>8p21-p12</td>
<td>1191</td>
</tr>
<tr>
<td>DHCR24</td>
<td>24-Dehydrocholesterol reductase</td>
<td>1p33-p31.1</td>
<td>1718</td>
</tr>
<tr>
<td>PIGR</td>
<td>Polymeric immunoglobulin receptor</td>
<td>1q31-q41</td>
<td>5284</td>
</tr>
<tr>
<td>SERPINB6</td>
<td>Serpin peptidase inhibitor, clade B (ovalbumin), member 6</td>
<td>6p25</td>
<td>5269</td>
</tr>
</tbody>
</table>

*Gene was identified in 6 studies.*
KEGG pathway analysis of CSD genes in NPC

To identify the molecular events involving in the tumorigenesis of NPC, CSD genes were overlaid onto the KEGG pathway database using WebGestalt. This analysis identified 115, 78, and 56 KEGG pathways containing 2 or more CSD, upregulated, or downregulated genes in NPC, respectively (Supplementary Table S4). The functions of metabolic pathways have the highest significance with respect to enrichment in 110 CSD, 66 upregulated, and 44 downregulated genes ($P = 6.00E-26$), small cell lung cancer (25 genes, $P = 1.83E-20$), glioma (13 genes, $P = 1.07E-08$), colorectal cancer (13 genes, $P = 2.29E-07$), prostate cancer (13 genes, $P = 4.39E-07$), chronic myeloid leukemia (12 genes, $P = 4.77E-07$), melanoma (11 genes, $P = 1.79E-06$), pancreatic cancer (11 genes, $P = 1.98E-06$), bladder cancer (8 genes, $P = 1.24E-05$), non–small cell lung cancer (8 genes, $P = 7.87E-05$), renal cell carcinoma (8 genes, $P = 0.0003$), endometrial cancer (6 genes, $P = 0.0019$), basal cell carcinoma (5 genes, $P = 0.0101$), acute myeloid leukemia (5 genes, $P = 0.0019$), and thyroid cancer (4 genes, $P = 0.0057$).

Besides cancer pathways, KEGG pathways of CSD genes involved in classical cancer development were significantly enriched, such as cell cycle (41 genes, $P = 7.34E-35$), focal adhesion (27 genes, $P = 5.72E-13$), ECM-receptor interaction (20 genes, $P = 2.02E-14$), p53 (19 genes, $P = 6.65E-15$), MAPK (18 genes, $P = 9.29E-05$), apoptosis (15 genes, $P = 6.44E-09$), insulin signaling (15 genes, $P = 1.57E-06$), cell adhesion molecules (14 genes, $P = 5.92E-06$), Toll-like receptor (13 genes, $P = 1.56E-06$), Jak-STAT (12 genes, $P = 0.0003$), complement and coagulation cascades (11 genes, $P = 1.51E-06$), and the Wnt signaling pathway (10 genes, $P = 0.0032$). It was notable that the complement and coagulation cascades pathway was significantly enriched in 8 downregulated gene C3, C4A, CD55, CFB, CFH, CR2, F3, and PROS1 ($P = 1.17E-06$), and also in 3 upregulated gene C1QB, PLAU, and PLAUR.

BUB1B, CENPF, and SLPI protein expression by IHC

To validate the results of CSD genes in NPC from gene expression profiling datasets, we evaluated the expression of BUB1B, CENPF, and SLPI proteins in a wide range of NPC tissues to characterize their

Figure 1. Chromosomal distribution of CSD genes in NPC. Black bar plots represent the upregulated genes. White bar plots represent the downregulated genes.

Figure 2. The DAG diagram of the GO terms from the Biological Process ontology enriched in the CSD genes (A), upregulated genes (B), or downregulated genes (C) in NPC when compared with the human genome. Analysis was carried out by using WebGestalt toolkit. Only the enriched GO sets (hypergeometric; $P < 0.05$) containing 2 or more genes were included in the analysis. Enriched functions of genes were outlined.
expressions in situ by IHC using the NPC TMA (Fig. 3). BUB1B, CENPF, and SLPI were the dysregulated genes in NPC identified in 6 microarray studies, but the details of protein expressions in NPC were still lacking. Consistent with the mRNA transcript data, there were remarkable staining differences of BUB1B and CENPF protein between NPC tissues and chronic inflammation of nasopharyngeal mucosa and normal adjacent epithelia of NPC ($P < 0.001$). BUB1B protein was detected in 367 of 540 (67.96%) NPC tissues, 12 of 70 (17.14%) chronic inflammation of nasopharyngeal mucosa, and 15 of 100 (15.00%) normal adjacent epithelia of NPC. The positive expression rates of CENPF protein were 69.25% (374/540) in NPC tissues, 21.42% (15/70) in chronic inflammation of nasopharyngeal mucosa, and 17.00% (17/100) in normal adjacent epithelia of NPC. IHC revealed that SLPI protein was significantly expressed in glands of normal nasopharyngeal mucosa and normal adjacent epithelia of NPC (109/109, 100%), but was underexpressed in nasopharyngeal adenocarcinoma (4/21, 19.05%, $P < 0.001$), which is a rare histologic type of nasopharyngeal neoplasm. There was no significant difference between nasopharyngeal squamous cell carcinoma (13/519, 2.50%) and squamous and columnar epithelia cells of normal nasopharyngeal mucosa (E) and downexpressed in nasopharyngeal adenocarcinoma (F). Scar bar = 100 μm.

**Figure 3.** Protein expressions of BUB1B, CENPF, and SLPI in NPC TMA. BUB1B weakly expressed in chronic inflammation of nasopharyngeal mucosa (A) and strongly expressed in nonkeratinizing NPC (B). CENPF showed weak expression chronic inflammation of nasopharyngeal mucosa (C) and strong staining in nonkeratinizing NPC (D). SLPI protein significantly expressed in glands of normal nasopharyngeal mucosa (E) and downexpressed in nasopharyngeal adenocarcinoma (F). Scar bar = 100 μm.

**Prognostic value of BUB1B and CENPF in NPC**

To investigate whether BUB1B and CENPF protein expression are associated with the outcome of NPC, survival curves were calculated for 103 cases of NPC patients. The clinical characteristics of the NPC patients were described in detail in Supplementary Table S5. The mean follow-up period was 46.4 ± 23.2 months with a range from 3 to 96 months. Survival curves calculated by the Kaplan–Meier method and analyzed using the log-rank test showed that the survival rates of the patients with high expression of BUB1B ($P = 0.000$) and CENPF ($P = 0.000$) were significantly decreased (Fig. 4). Using univariate analysis (Cox’s proportional hazards model), the high expression of BUB1B ($P = 0.000$, 95% CI: 1.817–6.340) and CENPF ($P = 0.000$, 95% CI: 2.219–8.085) were significantly associated with the prognosis.
Discussion

The conceptual basis for the present study is the observation that the genomic-wide molecular signature predicts the development and progression of NPC. NPC, like other cancers, develop in a multistep process involving alterations of numerous genes (18). Recently, using different array technologies and tissue samples, several groups have reported dysregulated genes in NPC (3–11).

In this study, we have investigated and integrated the global gene expression profiling of NPC from 8 different studies for the first time. Multiplatform integrative analysis of gene expression profiling datasets provides an opportunity to study in detail the molecular mechanisms regulating the development and progression of NPC. Furthermore, the identified significantly dysregulated gene in NPC represent promising biomarkers and targets for the clinical diagnosis and therapy.

Despite the substantial progress made by different investigators during recent years, the knowledge of the molecular signature of NPC and the factors that induce tumorigenesis remains limited. Using gene expression profiling and integrative criteria, 962 CSD genes (651 upregulated and 311 downregulated genes) in NPC were identified multiple studies.

Chromosome mapping of CSD genes may be prove to be a useful tool in the detection of the genetic hotspots of dysregulated gene expression patterns in NPC. Gains of 2q31 (19), 7p15 (19), 12q (19–22), and 18q (19, 21) in NPC have been previously reported. Chromosome 12 has been reported to undergo a gain of its long arm in NPC, with the common band region indicating gain at 12q13-q15, and high-copy-number increases of chromosomal materials in 12q14-q15 (21). Interestingly, on chromosomes 2q3, 7p15, 12q15, 12q22, 18q11, and 18q12, the enriched genes all belonged to the upregulated genes in NPC. Previous studies suggested that loss of heterozygosity on chromosome 14q32 was the common genetic event in NPC (23–27). The enriched genes on 14q32 were CLMN, CRIP2, AKT1, AHNAK2, ZFYVE21, IGH@, IGHG1, IGHG3, all belonged to the downregulated genes in NPC. There are no reports for gain of 2q23 and loss of 16q13 in regard to the relationship to NPC.

Functional gene annotation of GO analysis and pathway analysis provided further insight into the expression profiling of CSD genes in NPC. The functions of CSD genes were enriched in the biological processes and pathways involving classical cancer development and progression. As expected, the most significantly enriched biological processes and pathways of upregulated genes in NPC were related to cell proliferation, adhesion, and invasion, which was common in all cancers.

Notably, 40 and 89 downregulated genes were enriched in the GO terms, the immune system process and the response to stimulus, respectively. Tumor growth, invasion, and metastasis are important aspects of the tumor immune escape response. As mentioned previously, downregulation or loss of expression of immune molecules are strategies used by tumor cells to escape the immune system. Complement components play an important role in regulation of inflammation and the immune response. In this study, we found 8 downregulated genes, C3, C4A, CD55, CFB, CFH, CR2, F3, and PROS1, and 3 upregulated genes, C1QB, PLAU, and PLAUR involved in the GO term "complement activation," or the complement and coagulation cascades pathway. These dysregulated genes suggested that the immune response was silenced in NPC. In addition, NPC is closely associated with persistent Epstein–Barr virus (EBV) infection, which has evolved a plethora of strategies to evade immune system recognition and to establish latent infection in memory B cells (28). Seven downregulated genes, CR2, C3, CLU, CD55, FOXJ1, C4A, and BCL6 were enriched in the B-cell–mediated immunity response.

It is not surprising that many of CSD genes have been reported to be associated with a variety of distinct human cancer types in the literature. Several of the dysregulated genes identified in 5 and more studies, such as MAD2L1 (29), CCND2 (30), PCNA (31), CDK1 (32), CCNB1 (33),
ular mechanisms of the upregulated gene BUB1B and the downregulated gene SLPI involved in NPC were still lacking.

BUB1B, a member of the BUB (budding uninhibited by benzimidazole) gene family, is one of the key molecules in the spindle assembly checkpoint. High expression of BUB1B has often been reported in several malignancies and correlated with chromosomal instability (43, 44). In this study, our results showed for the first time that BUB1B protein was upregulated in NPC compared with chronic inflammation of nasopharyngeal mucosa and normal adjacent epithelia of NPC, and that the high expression of BUB1B was associated with the prognosis of NPC patients. The data suggest that BUB1B might be a prognostic biomarker and potential therapeutic target for NPC.

Previous studies have shown that SLPI (secretory leukocyte peptidase inhibitor) is produced by the epithelial cells lining the respiratory, digestive, and reproductive tracts, and protect proepithelin from elastase cleavage in wound healing (45). SLPI inhibits cell growth through the apoptotic pathway in ovarian cancer (46) and suppresses cancer cell invasion but promotes blood-borne metastasis via an invasion-independent pathway (47). In our work, the results showed for the first time that SLPI protein was significantly expressed in the glands of normal nasopharyngeal mucosa and normal adjacent epithelia of NPC but underexpressed in nasopharyngeal adenocarcinoma. There was weak expression of SLPI in squamous and columnar epithelia cells of normal nasopharyngeal tissues and nasopharyngeal squamous cell carcinoma. Although the normal nasopharyngeal tissues used in the microarray analysis generally were the mixtures containing glands and squamous epithelium, the NPC tissues contained few or no glandular structures. The difference in glands in tissues may be the main reason for promoting the downregulated expression of SLPI in the microarray datasets. The data suggest that SLPI might be a biomarker and potential therapeutic target for nasopharyngeal adenocarcinoma.

Moreover, we examined the expression of CENPF protein. CENPF, a member of the human centromeric protein family, is involved in centromere formation and kinetochore organization during mitosis (48). Cao and colleagues reported that CENPF, which were identified from their cDNA microarray data, is a valuable marker of NPC progression (41). The protein expression of CENPF is associated with poor overall survival of patients. In our study, CENPF was expressed in 374 (69.25%) of 540 NPC tissues and the high expression of CENPF was associated with the survival of NPC patients, according to the previous study (41). Thus, these data suggest that CENPF might be a biomarker for the carcinogenesis and prognosis of NPC and may have therapeutic implications.

One limitation of our study is that it only directly integrates differentially expressed genes in the datasets generated from the different array platforms. Even with this limitation, we still obtain a large number of common dysregulated genes (962 genes) in NPC. With the rapid increase of available microarray studies of NPC, the amount of gene expression profiling data generated from the different platforms will continue to grow, which will make our results of the integration analysis more useful.

In conclusion, by integrated analysis of multiple gene expression profiling datasets, the common gene signatures and molecular markers of NPC have been identified. The marked enrichments of chromosome aberrations in the upregulated genes were 2q23, 2q31, 7p15, 12q15, 12q22, 18q11, and 18q12, and in the downregulated genes were 14q32 and 16q13. The activated category related to cell proliferation, adhesion, and invasion and the downregulated immune response have been functionally related to the progression of NPC. Significant overexpression of BUB1B and CENPF in NPC and downregulation of SLPI in nasopharyngeal adenocarcinoma were identified. Furthermore, survival curves and univariate analysis revealed that the expression of BUB1B or CENPF was the prognostic indicator for survival of the patients. Further studies of the CSD genes as gene signatures and molecular markers of NPC might improve our understanding of NPC and identify new drug targets of clinical treatment.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Grant Support**

This work was supported by Hunan Province Natural Sciences Foundations (10JJ7003; 11JJ1013).

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

Received June 27, 2011; revised October 7, 2011; accepted October 31, 2011; published OnlineFirst November 8, 2011.

---

**References**


Huang et al.


Integrated Analysis of Multiple Gene Expression Profiling Datasets Revealed Novel Gene Signatures and Molecular Markers in Nasopharyngeal Carcinoma

Chen Huang, Hailin Tang, Wenling Zhang, et al.

Cancer Epidemiol Biomarkers Prev 2012;21:166-175. Published OnlineFirst November 8, 2011.

Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-11-0593

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2011/12/01/1055-9965.EPI-11-0593.DC1

Cited articles
This article cites 48 articles, 7 of which you can access for free at:
http://cebp.aacrjournals.org/content/21/1/166.full#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/21/1/166.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.