From NPC Therapeutic Target Identification to Potential Treatment Strategy

Ming-Ying Lan,1,3,10, Chi-Long Chen,6,7, Kuan-Ting Lin,3,8 Sheng-An Lee,11, Wu-Lung R. Yang,9 Chun-Nan Hsu,8,12, Jaw-Ching Wu,3, Ching-Yin Ho,5,10, Jin-Ching Lin,2,3, and Chi-Ying F. Huang,3,4

Abstract

Nasopharyngeal carcinoma (NPC) is relatively rare in Western countries but is a common cancer in southern Asia. Many differentially expressed genes have been linked to NPC; however, how to prioritize therapeutic targets and potential drugs from unsorted gene lists remains largely unknown. We first collected 558 upregulated and 993 downregulated NPC genes from published microarray data and the primary literatures. We then postulated that conversion of gene signatures into the protein-protein interaction network and analyzing the network topologically could provide insight into key regulators involved in tumorigenesis of NPC. Of particular interest was the presence of cliques, called fully connected subgraphs, in the inferred NPC networks. These clique-based hubs, connecting with more than three queries and ranked higher than other nodes in the NPC protein-protein interaction network, were further narrowed down by pathway analysis to retrieve 24 upregulated and 6 downregulated bottleneck genes for predicting NPC carcinogenesis. Moreover, additional oncogenes, tumor suppressor genes, genes involved in protein complexes, and genes obtained after functional profiling were merged with the bottleneck genes to form the final gene signature of 38 upregulated and 10 downregulated genes. We used the initial and final NPC gene signatures to query the Connectivity Map, respectively, and found that target reduction through our pipeline could efficiently uncover potential drugs with cytotoxicity to NPC cancer cells. An integrative Web site (http://140.109.23.188:8080/NPC) was established to facilitate future NPC research. This in silico approach, from target prioritization to potential drugs identification, might be an effective method for various cancer researches.

Introduction

Nasopharyngeal carcinoma (NPC) is a rare malignancy in most parts of the world but is one of the most common cancers among those of Chinese or Asian ancestry. The etiology of NPC is thought to be associated with a complex interaction of genetic, EBV exposure, environmental, and dietary factors. Although some oncogenes, tumor suppressor genes, and microarray expression data have been previously reported in NPC, a complete understanding of the pathogenesis of NPC in the context of global gene expression remains to be elucidated (1–9).

Protein-protein interactions (PPI) are important for virtually every biological process. In a PPI network, nodes having more than one connection with another node are defined as hubs and are more likely to be essential (10, 11). The key challenge facing a disease PPI network is the identification of a node or combination of nodes in the network whose perturbation might result in a desired therapeutic outcome. We have previously constructed an integrated PPI web service, POINeT (12), as a bioinformatics tool to construct and to analyze the NPC network in this study.

In addition to elucidating the pathogenesis of NPC, the refinement of current treatment modalities is also important. Although NPC is highly radiosensitive and chemosensitive, the treatment of patients with locoregionally advanced disease remains problematic. Many specific molecular-targeted therapies, epigenetic therapies, and EBV-based immunotherapy have been developed and are in clinical trials (Supplementary Table S1; ref. 1).

Authors’ Affiliations: Departments of 1Otolaryngology and 2Radiation Oncology, Taichung Veterans General Hospital, Taichung, Taiwan; Institutes of 3Clinical Medicine and 4Biomedical Informatics, and 5Department of Medicine, School of Medicine, National Yang-Ming University; 6Department of Pathology, Taipei Medical University; 7Department of Pathology, WanFang Hospital; 8Institute of Information Science, Academia Sinica; 9Department of Computer Science and Information Engineering, National Taiwan University; and 10Department of Otolaryngology, Taipei Veterans General Hospital, Taipei, Taiwan; 11Department of Computer Science and Information Engineering, Kainan University, Taoyuan, Taiwan; and 12Information Sciences Institute, University of Southern California, Marina del Rey, California

Note: Supplementary material for this article is available at Molecular Cancer Therapeutics Online (http://mct.aacrjournals.org/).

Corresponding Authors: Chi-Ying F. Huang, Institute of Clinical Medicine, National Yang-Ming University, Taipei 112, Taiwan. Phone: 886-228267904; Fax: 886-228224045. E-mail: cychuang56@ym.edu.tw and Jin-Ching Lin, Department of Radiation Oncology, Taichung Veterans General Hospital, Taichung 114, Taiwan. Phone: 886-423592525 ext. 5613; Fax: 886-423741316. E-mail: jclin@vghtc.gov.tw
doi: 10.1158/1535-7163.MCT-09-0966 ©2010 American Association for Cancer Research.
It is not clear how to elucidate key regulators and identify potential drugs for NPC treatment. To address these questions, we first collected NPC-associated genes (http://140.109.23.188:8080/NPC) and hypothesized that the PPI network, derived from the gene signature, could be analyzed topologically to prioritize potential targets. We further performed pathway analysis and applied gene signatures to drug-gene interaction databases and Connectivity Map (cMap, refs. 13, 14) to find potential drugs for the treatment of NPC. It is supposed that a small molecule may potentially reverse the disease signature if the molecule-induced signature is significantly negatively correlated with the disease-induced signature in cMap (15–17). In short, identifying potential drugs to treat NPC with an in silico screening approach followed by empirical validation might be easier and faster than those traditional drug discovery pipelines that require tremendous effort and time.

Materials and Methods

Computational methods

Acquiring NPC-related gene sets and constructing NPC PPI network. Two major components constituted the NPC-related gene expression signatures in this study. One component included the collection of the microarray profiles from three studies (Supplementary Table S2; refs. 4, 5, 7). All microarray data were the result of nontreated NPC tissues compared with normal nasopharyngeal tissues.

The second part of the gene collections consisted of the text mining of NPC-related PubMed abstracts. By March of 2008, we extracted 4,939 abstracts from PubMed containing the keyword “Nasopharyngeal carcinoma” but not having the keywords “SNP” or “polymorphism.” To further extract the genes mentioned in the abstracts, we first entered all these abstracts into the Adaptive Internet Intelligent Agents laboratory’s Gene Mention Tagger (18). The Gene Name Service (19) was used to translate these gene names into corresponding gene identifiers, such as the official gene symbol and the Entrez gene ID. Then, we manually read the top 10 abstracts with most genes mentioned from the above method and another 150 abstracts published from 2007 to 2008 to further annotate the genes as upregulated or downregulated genes. These genes and annotations can be accessed through our Web site (http://140.109.23.188:8080/NPC), and the Web site tools used in this study are summarized in Supplementary Table S3. We inputted the above-collected NPC-related genes as query terms into the POINeT (12) to detect the PPI in NPC.

Evaluation of cliques and complexes from the PPI network. The cliques of the PPI network were calculated from the following definition of cliques, a term borrowed from Graph Theory. A clique is a part of a graph in which all its nodes are completely connected to each other. In other words, a 3-clique is a completely connected graph of three nodes, which is a triangle. From this definition, we have developed CliquePOINT, which was embedded into POINeT, to calculate these cliques in the NPC PPI network. Expanding the definition of the 3-clique, we also counted the number of 4-cliques and 5-cliques in the NPC PPI network, and there is no clique larger than 5-cliques in the NPC PPI network.

We further collected and integrated the complex information to obtain an abundant data set from public domain databases, including the Human Protein Reference Database (20), the Protein Interacting in the Nucleus database (21), and the Comprehensive Resource of Mammalian protein complexes (22), and asked whether the cliques identified from the PPI network were involved in protein complexes. The cliques having more than three proteins involved in complexes were reported.

Ranking the hubs in the PPI network. To elucidate the relative roles of each node, we analyzed node centrality through POINeT, including degree centrality (DC), closeness centrality (CC), and eccentricity centrality (EC). DC is the number of link incident upon a node. CC represents the closeness between nodes in the biological network. EC is the longest distance required for a given node to reach the entire network. By conducting centrality calculation, nodes in global networks can be ranked and filtered using various network analysis formulas.

The enriched pathways from the ConsensusPathDB overrepresentation analysis. We used ConsensusPathDB (23) to perform overrepresentation analysis on the four sets of gene lists: (a) upregulated genes in NPC, (b) downregulated genes in NPC, (c) upregulated genes after clique analysis, and (d) downregulated genes after clique analysis. The significant pathway results were ranked by using an F score instead of the P value given by ConsensusPathDB. The F score was used to normalize two parameters: (a) the percentage of overlapping genes in the pathway and (b) the percentage of overlapping genes in the input list. To normalize these, we used the following formula:

\[
\text{Fscore} = \frac{2(A \times B)}{(A + B)}
\]

We compared the P values to evaluate whether the P values degrade after clique analysis and thereby give each pathway a score of degradation (0 for No and 1 for Yes).

The final NPC gene signature. The 98 up-clique and 51 down-clique genes were used as queries to perform functional annotation clustering on the Database for Annotation, Visualization, and Integrated Discovery (DAVID; ref. 24), respectively. The clustering was done on seven pathway resources: BBID, BIOMAP, EC_NUMBER, KEGG_COMPOUND, KEGG_PATHWAY, KEGG_REACTION, and PANTHER_PATHWAY. The classification stringency was set to “Medium.” For each cluster, we further intersected the genes of the pathways to obtain the
“bottleneck” genes. Twenty-four upregulated and six downregulated bottleneck genes were obtained.

Among cliques, those including oncogenes, tumor suppressor genes, genes involved in complex, and genes found by group functional profiling, were added into the bottleneck genes list to obtain the final gene signature of NPC. It includes 38 upregulated and 10 downregulated genes (Supplementary Table S4).

Hierarchical clustering the final gene signature in KEGG pathways. We used the final gene signature as queries to conduct the functional annotation clustering of DAVID against the KEGG pathway database. A perl script was written to convert the pathway records (P < 0.05) into a gct file, which can be uploaded onto GenePattern to perform hierarchical clustering and visualization. For upregulated and downregulated genes, the values are 1 (red) and −1 (green), respectively. The distance measure for both genes (row) and pathways (column) was set to “Pearson correlation, absolute value.”

Food and Drug Administration–approved drug targets. To collect target genes of Food and Drug Administration (FDA)–approved drugs, the chemical-protein links from STITCH (25) was downloaded. Then, Gene Name Service (19) was used to translate the protein ID to its corresponding HUGO-approved gene symbol and Entrez gene ID. The DrugCard file from Drug Bank (26) was downloaded. We selected FDA-approved drugs, mapped the drugs’ corresponding genes with the NPC upregulated genes, and finally identified known drug targets in the NPC upregulated PPI network.

Applying NPC gene signatures to cMap. Functional connections between various NPC gene signatures and gene signatures induced by small molecules were explored using the cMap database (13, 14). The upregulated genes were grouped and their probe sets formed the up tag file and so did the downregulated genes. These two files were used to query the cMap database, and the results showed the most significant similarities and dissimilarities to the database profiles. The 558 up and 993 down genes would convert to >1,000 probe sets. Because the cMap could only take up to 1,000 probe sets per input, three groups of NPC genes were used. The first group consists of 100 randomly chosen sets of 100 upregulated/downregulated probe sets from whole 558 up and 993 down NPC gene signature. The second group consists of 399 upregulated and 443 downregulated probe sets, which represent first 70% ranked queries served as hubs. The third group, the final gene signature, consists of 38 up genes and 10 down genes. Only drugs with negative scores and a P value of <0.05 were retained.

Biological methods

Immunohistochemical analysis in NPC. Formalin-fixed, paraffin-embedded biopsy specimens of 143 NPC cases were collected and analyzed for detection of the expression of p53 (mouse anti-human p53, 1:50) and BCL2 (mouse anti-BCL2; 1:80; DAKO), BAX (mouse anti-BAX, 1:400), and MYC (mouse anti-MYC, 1:50; Santa Cruz) by immunohistochemistry (IHC) with the institutional review board approval. Briefly, 5 to 6 μm of paraffin sections were deparaffinized and placed into citrate buffer for antigen retrieval once placed inside a microwave oven. After the sections were cooled down and rinsed with PBS, the sections were incubated with 5% normal goat serum followed by reaction with primary antibody for 30 minutes at room temperature, then washed with PBS thrice at 3 minutes each. The sections were reacted with biotinylated secondary antibody followed by streptavidin–biotin complex in the LSAB detection kit (DAKO) at room temperature for 10 minutes and washed with PBS again. The sections were colorized using freshly prepared diaminobenzidine solution containing H2O2 for 2 to 5 minutes. After washing with running water and counterstaining with hematoxylin, the sections were dehydrated and mounted. Positive staining showed brownish granular deposits in the nuclei of cells. Adenocarcinoma and normal mucosa gland of the colon were used as positive and negative controls, respectively, for the expression of p53 and MYC, whereas follicular lymphoma was used for the positive and negative control of the expression of BCL2 and BAX (Supplementary Fig. S1).

Cell culture and cell viability test. NPC cell lines, TW01, TW03, and TW04 provided by Dr. C.T. Lin (National Taiwan University, Taoyuan, Taiwan), were derived from primary nasopharyngeal tumors of Chinese patients with de novo NPC and had been tested and authenticated (27). NPC cell line BM1, provided by Dr. S.K. Liao (Chang Gung University, Taoyuan, Taiwan), was derived from bone metastatic lesions of an NPC patient (28). NPC cell lines were maintained in DMEM with 10% fetal bovine serum containing penicillin (100 U/mL) and streptomycin (100 μg/mL) in 5% CO2 at 37°C. Cell viability was determined using the 2,3-bis[2-methoxy-4-nitro-5-sulfophenyl]-2H-tetrazolium-5-carboxanilide inner salt (XTT) cell viability assay kit (Sigma-Aldrich), according to the manufacturer’s instructions. Twenty-four hours after seeding cells at a concentration of 2 × 10^3 cells per well in 100 μL culture medium in a 96-well microplate, cells were then treated with Trichostatin A (Sigma-Aldrich) and Trifluoperazine (Sigma-Aldrich), the selected small molecules from cMap. Cells were exposed with or without small molecules for 72 hours at different concentrations. Then, the cells were incubated with medium containing XTT in an amount equal to 20% of the culture medium volume for 2 hours. Absorbance was measured using a microplate reader (Spectral Max250) at 450 nm.

Results

NPC gene collections

To systematically analyze the gene expression signatures of NPC and identify potential drugs for NPC, we have set up in silico approaches (Fig. 1). We collected the NPC gene sets from two sources: one gene set from PubMed with 70 upregulated and 78 downregulated
genes, and the other from three major microarray studies (Supplementary Table S2; refs. 4, 5, 7) with 512 upregulated genes and 936 downregulated genes. By merging these two data sets, the gene expression signature of NPC contained 558 upregulated genes and 993 downregulated genes (http://140.109.23.188:8080/NPC).

Inferred NPC PPI network

To uncover the potential interaction networks of these seemingly unrelated NPC upregulated and downregulated genes, the Web site tool, POINeT, was used to detect the PPI in NPC, and results were summarized in Supplementary Table S5. Despite many queries without interacting proteins, based on our PPI collections in POINeT, the queries of NPC-related proteins formed a highly connected interactome. A total of 8,231 and 7,728 PPIs were identified in the upregulated and downregulated NPC PPI networks, respectively. The fundamental structural details revealed that 257 of 558 NPC upregulated genes interact with each other and form 492 query-query PPIs, constituting the interaction networks. On the other hand, 324 of 993 NPC downregulated queries form 395 query-query PPIs.

The inferred NPC network consists of highly interactive cliques and complexes

Of particular interests in the inferred NPC PPI network is the presence of cliques (29), which refer to completely connected subgraphs. Nodes within a clique have interactions with all the others. In our analysis, the NPC query-query network contains 198 and 21 subgraphs of cliques in upregulated and downregulated genes, respectively. In the upregulated PPI network, there are 170 3-cliques, 26 4-cliques, and 2 5-cliques (Supplementary Table S6; Fig. 2A). The count of cliques in NPC PPI network is much higher than the count of cliques in a random PPI network (Supplementary Fig. S2). The top 30 proteins involved in cliques are listed and ranked by the number of associated cliques (Supplementary Table S7). BRCA1, MYC, EGFR, TP53, and CDC2 are the top five proteins participating in a large number of cliques.

The analysis of node centrality characteristics may provide insights into the relative roles and features of each node. To address whether clique proteins are relatively more important hubs in the PPI network, we prioritized the nodes of the major subnetwork, which consists of 247 query proteins (or nodes), in the NPC upregulated PPI network (Supplementary Table S8). The 3,725 nodes of level one major subnetwork, which consists of query proteins with neighbor nodes, were also ranked. Different ranking methods, including DC, EC, and CC, were used. Those nodes, which are also clique proteins, are ranked higher than those that are not clique proteins (Fig. 3).

Because cliques have more interactions than the rest of the graph, and these protein interactions may be responsible for the formation of protein complexes or functional modules (30), we further integrated and searched for protein complexes from the Human Protein Reference Database (20), the Comprehensive Resource of Mammalian protein complexes (22), and the Protein Interacting in the Nucleus database (21). Of upregulated cliques, there are five 3-cliques and four 4-cliques involved in five protein complexes (Table 1; Fig. 2B). The DNA synthesome, also known as the DNA replication complex, consists of 15 subunits, including DNA polymerase,
DNA topoisomerase, and the replication factor C complex (31). The replication factor C complex is a heteropentameric protein that is essential for DNA replication and repair, and is also a clamp loader required for the loading of PCNA onto dsDNA (32–34). The BASC complex, BRCA1-associated genome surveillance that consists of ATM, BLM, MSH2, MSH6, MLH1, and RF-C, is involved in the recognition and repair of aberrant DNA structure (35). Another complex, the hNop56p-associated preribosomal ribonucleoprotein complex, is associated with ribosome biogenesis (36). Interestingly, many proteins are shared in these complexes. Finally, there is one complex involved in the tumor necrosis factor-α (TNF-α)/NF-κB pathway (37). The above finding raises the possibility that NPC pathogenesis might be related to aberrant DNA replication, DNA repair, and the TNF-α/NF-κB pathway. To the best of our knowledge, this finding will be the first report to provide the relationship between these complexes and NPC carcinogenesis. Few proteins in the above five complexes that are related to NPC include RFC1, PCNA, TOP1, ATM, MLH1, RPL21, and RPL31.

Oncogenes and tumor suppressor genes in NPC clique genes

Six oncogenes, including EGFR, ERBB2, MYC, RELB, NFKB2, and CCND1, were found in the 4-cliques and 5-cliques from the inferred upregulated NPC network. Overexpression of these oncogenes in NPC, except ERBB2, was suggested to be related to NPC carcinogenesis (Supplementary Table S7; refs. 38–42). Three tumor suppressor genes were found in the 51 downregulated

Figure 2. Highly interactive cliques and complexes are associated with NPC gene signature. A, 4-cliques and 5-cliques of NPC PPI network. The query-query interaction network of the NPC upregulated genes is a highly connected network that contains 26 4-cliques and two 5-cliques. The two 5-cliques are grouped in red circles. Yellow, oncogenes; green, tumor suppressor genes. BRCA1, TP53, MYC, EGFR, and CDC2 are the top five proteins involved in the largest number of cliques. B, five major complexes associated with NPC upregulated gene signatures. Red, upregulated genes; green, downregulated genes. Dark red, clique genes (upregulated cliques); dark green, clique genes (downregulated cliques).
clique genes, including CDKN1A, MLH1, and ATM. Both CDKN1A and ATM are downregulated in NPC (7, 43). The description of these genes can be accessed through our Web site (http://140.109.23.188:8080/NPC).

It is interesting to note that there are three tumor suppressor genes found in the 98 upregulated cliques, including BRCA1, TP53, and FAS. Briefly, BRCA1, a nuclear phosphoprotein, plays a role in maintaining genomic stability. Mutations in BRCA1 are responsible for ~40% of inherited breast cancers and >80% of inherited breast and ovarian cancers; however, its expression in NPC is still unknown. TP53 encodes the tumor protein p53, which responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, senescence, and DNA repair. In normal cells, p53 is rapidly turned over by a negative feedback loop mediated by MDM2. Mutant p53, noted in 30% to 50% of cancer, is unable to induce MDM2 transcription and escapes degradation, thereby leading to its accumulation at a very high level in cancer (44). Although p53 levels are high in NPC, the mutation of TP53 gene is relatively rare. Accumulated p53 in NPC is believed to be mediated by EBV LMP1 (9, 40, 45). Two reasons have been proposed to explain why wild-type p53 fails to induce apoptosis in NPC: low ARF levels due to promoter hypermethylation and excess mutated p63. Wild-type p53 function may be eliminated by the inactivation of the ARF gene, which encodes proteins that sequester MDM2 from antagonizing p53 (44). Mutated p63, which lacks the NH2-terminal transactivation domain required to activate apoptosis, binds to normal p63 (and p53; ref. 9). FAS protein is a member of the TNF receptor superfamily and contains a death domain. It plays a central role in the physiologic regulation of programmed cell death and has been implicated in the pathogenesis of various malignancies and diseases of the immune system. Fas ligand overexpression is an unfavorable prognostic marker in NPC (46, 47).

**Findings by gene group functional profiling**

To address how our NPC signature might turn biological process term groups (by Gene Ontology) on or off, 98 upregulated and 51 downregulated clique genes were subjected to g:Profiler, respectively (48). A large biological process term group is shared by both upregulated and downregulated clique genes. The group is mainly related

<table>
<thead>
<tr>
<th>Table 1. Five major complexes associated with NPC after analysis using three public domain databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex (total proteins numbers)</td>
</tr>
<tr>
<td>DNA synthesome (15)</td>
</tr>
<tr>
<td>Replication factor C complex (5)</td>
</tr>
<tr>
<td>BASC complex (9)</td>
</tr>
<tr>
<td>hNop56p-associated preribosomal ribonucleoprotein complexes (104)</td>
</tr>
<tr>
<td>TNF-α/NF-κB pathway (12)</td>
</tr>
</tbody>
</table>

NOTE: The proteins involved in complexes and proteins that are in NPC upregulated cliques are listed.
to the regulation of biological processes, cell cycle, cell death, and cell development. These important biological functions are altered, thereby leading to the activation of p53 to deal with the disturbed physiologic circumstances. Among the downregulated clique genes, three genes, including CDKN1A, HDAC3, and PRKCZ, are shown to be related to the “regulation of programmed cell death” and the “regulation of apoptosis” by using Traceable author. The genes with Traceable author references in the upregulated clique genes in the phosphorylation group are ERBB2, STAT1, and TYK2 (Supplementary Fig. S3). Overall, we used gene group profiling to further identify three downregulated genes and three upregulated genes that relate to the growth of tumors.

Pathway analysis of NPC gene signature

To find the enriched pathways of our NPC gene signature, we performed an overrepresentation pathway analysis on ConsensusPathDB (23). Under the threshold of a P value of <0.01, there were 484 enriched pathways for upregulated genes and 222 enriched pathways for downregulated genes in the original NPC signature; 409 and 294 enriched pathways were found for upregulated and downregulated genes, respectively, by using the clique analysis. To avoid the complication that small pathways are relatively easier to rank higher according to their P value, we used the F score to normalize the ranking. From the results of the intersection of the top 100 enriched pathways of upregulated gene signature, many pathways are directly related to cancer, such as the p53 signaling pathway, cell cycle–related pathways, bladder cancer pathways, lung cancer pathways, prostate cancer pathways, and pancreatic cancer pathways (Supplementary Table S9 and S10). Moreover, most of the enriched pathways and their P values did not degrade after clique analysis, suggesting that the clique analysis tends to remove genes not involved in the enriched pathways of our NPC gene signature.

Furthermore, we performed another pathway analysis for NPC final gene signature (Supplementary Table S4) by using DAVID. The clustering result shows that the final gene signature can be divided into three groups (Supplementary Table S11). All groups are closely related to cancers, signaling, and cell communications. This analysis provides a convenient way to biologically interpret at the “biological module” level (24). To provide a more insightful view of the relationships between the final gene signature and KEGG pathways, we downloaded the pathway records (P < 0.01) to perform hierarchical clustering (Fig. 4) using GenePattern. Most of the pathways are shown to have downregulated genes that might cause disruption, whereas there are five pathways having no downregulated blocks. They are amyotrophic lateral sclerosis, Jak-STAT signaling, adipocytokine signaling, neurodegenerative disease, and cell communication pathways. In addition, the tumor suppressor, ATM, is shown to be downregulated in only antitumor pathways such as apoptosis, p53 signaling pathway, and cell cycle. It implies that the ATM could be an important missing piece in NPC.

To investigate how the final NPC gene signature connect with each other in pathways, we manually referred the KEGG pathways to draw a possible molecular mechanism of NPC carcinogenesis (Fig. 5A). To confirm the expression of selected final upregulated genes, IHC studies of TP53, BCL2, BAX, and MYC were done. All of them are overexpressed in tumor cells (Fig. 5B).

FDA-approved drug targets

To annotate the NPC upregulated genes with FDA-approved drug targets, we integrated databases from STITCH (25) and Drug Bank (26). We thereby derived 566 and 827 drug target upregulated and downregulated genes, respectively. Two hundred eighty-nine and 203 FDA-approved drugs target up-clique and down-clique genes, respectively (Supplementary Table S12). The 191 drugs target up-bottleneck genes and oncogenes (Supplementary Table S13), whereas 100 drugs target down-bottleneck genes and tumor suppressor genes (Supplementary Fig. S4). Some well-known chemotherapeutic agents already used in several cancers are among the top 100 drug target up-clique genes. These drugs include paclitaxel, doxorubicin, etoposide, and cisplatin. Many of these drugs are being studied in NPC clinical trials (Supplementary Table S1), suggesting that our target prioritizations, particularly those not currently being used in clinical trials, might reveal potential therapeutic agents for the treatment of NPC, alone or in combination with older chemotherapeutic agents.

Finding candidate drugs for NPC from drugs being used or being studied in clinical trials in cancers whose pathways are related to NPC

From the results of the pathway analysis, NPC may be related to several cancer pathways, including prostate cancer, bladder cancer, pancreatic cancer, chronic myeloid leukemia, colorectal cancer, and small cell lung cancer. We derived 1,692 chemical names with 3,603 clinical trial records of the six types of cancers with refined search limited on drug from the ClinicalTrials database (Supplementary Table S3). By intersecting the chemical names with 289 up-clique drugs, we obtained 106 up-clique drugs under clinical trials. We then manually selected 83 drugs that are used as antitumor drugs in those clinical trials. Of the 83 drugs, 11 drugs are under NPC clinical trials (Supplementary Table S1). Moreover, 66 of the 83 drugs are targeting up-bottleneck genes and oncogenes. After excluding the drugs already in clinical trial for NPC, 57 drugs remain (Supplementary Table S14). These candidate drugs might be important potential drugs for future NPC treatment. In addition, 26 chemotherapeutic agents suggested to treat these cancers at different stages were retrieved from the National Comprehensive Cancer Network clinical practice guidelines (Supplementary Table S15). Individual or combined usage of the above FDA-approved drugs may improve current
NPC treatment with enhanced therapeutic effects and minimized side effects.

**Identifying potential small molecules for NPC treatment by applying NPC gene signatures to cMap**

Bioactive small molecules in cMap that reverse the gene signature of NPC may be the potential drugs to kill NPC cells. We used three groups of NPC gene signatures to query the cMap database. The first group are genes randomly selected from whole NPC 559 upregulated and 993 downregulated gene signature; the second group consists of first 70% ranked queries served as hubs; the third group are the final gene signature, consisting of 38 upregulated and 10 downregulated genes. By querying cMap with the first, the second, and the third group of genes, there are 6, 8,

![Figure 4. The heat map showing KEGG pathways with corresponding NPC final gene signature. In a given pathway, the upregulated genes and the downregulated genes are denoted as the red blocks and the green blocks, respectively. Amyotrophic lateral sclerosis, Jak-STAT signaling, adipocytokine signaling, neurodegenerative disease, and cell communication are the pathways without downregulated genes in the figure.](image)
and 8 drugs respectively among the 10 top-ranked small molecules with antitumor effect (either from cell viability tests or PubMed literatures; Table 4). Here, we show cell viability tests of two drugs, trichostatin A and trifluoperazine, whose gene signatures in cMap significantly negatively correlated with gene signatures of NPC (Fig. 6A and B). Trichostatin A, a member of histone deacetylase inhibitors, has been used with other antineoplastic agents in several clinical trials. Trifluoperazine, a typical antipsychotic drug of the phenothiazine group, can induce apoptosis of B16 melanoma cells (49) and leukemic cells (50). Both of them may have potential for treating NPC in the future.

**Discussion**

Here, we constructed NPC gene signature from the microarray data of NPC tissues due to poor correlation between the microarray data of the NPC cell lines and tissues (4). Although we can also find potential drugs, e.g., trichostatin A and trifluoperazine, for NPC by NPC cell line signature (Supplementary Table S16), we believe the gene signature collected from many patient profiles, but not cell line derived from one patient, is more representative of the NPC disease. In fact, several studies have identified promising drugs by querying cMap with the gene signatures of tumor
samples (15–17). Nevertheless, continuous effort to collect additional NPC signature from various NPC cell lines is required.

To ask whether genes participate individually or interact with each other to form a network, we hypothesize that converting gene signature into PPI network might help reveal the potential “hubs” in the inferred network. Hence, we use POINeT to construct the NPC PPI network. From the inferred NPC PPI network, although the downregulated network contained more query genes than the upregulated network, the number of PPIs of the downregulated network is less than that of the upregulated network (Supplementary Table S5). The data suggest that the upregulated PPI network in NPC is more compact than the downregulated PPI network.

It is difficult to prioritize targets involved in the carcinogenesis of NPC, especially for those differentially expressed signatures from multiple sources. Cliques having more interactions within themselves than with the rest of the graph may be an essential modularity in the network (30). As clique proteins are ranked higher than those that are not clique proteins (Tables 2 and 3; Fig. 3), this implies that clique proteins might play important roles in the NPC PPI network. Further validating these ranking results is needed to find the most appropriate hub-ranking method. From both the clique and complex analysis, we hypothesize that perturbation of these cliques and/or complexes, which do not act independently but act intimately in an intertwined network, might result in a defective NPC PPI network. It is likely that targeting them with new therapies may be another way to treat NPC.

Furthermore, the clique proteins were used for pathway analysis to obtain the bottleneck genes, which

![Table 2. The top 15 proteins ranked by different centrality in the major subnetwork](image)

<table>
<thead>
<tr>
<th>Rank</th>
<th>DC</th>
<th>EC</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GRB2</td>
<td>EGFR</td>
<td>EGFR</td>
</tr>
<tr>
<td>2</td>
<td>MYC</td>
<td>EPB41</td>
<td>GRB2</td>
</tr>
<tr>
<td>3</td>
<td>TP53</td>
<td>ERBB2</td>
<td>MYC</td>
</tr>
<tr>
<td>4</td>
<td>EGFR</td>
<td>ADA</td>
<td>TP53</td>
</tr>
<tr>
<td>5</td>
<td>BRCA1</td>
<td>CDH3</td>
<td>BRCA1</td>
</tr>
<tr>
<td>6</td>
<td>NFKB2</td>
<td>PDCD6IP</td>
<td>NPM1</td>
</tr>
<tr>
<td>7</td>
<td>CDC2</td>
<td>PDCD6</td>
<td>CDC2</td>
</tr>
<tr>
<td>8</td>
<td>EPB41</td>
<td>SH2B3</td>
<td>PRKDC</td>
</tr>
<tr>
<td>9</td>
<td>SMAD1</td>
<td>SH2D3A</td>
<td>MAPK1</td>
</tr>
<tr>
<td>10</td>
<td>ERBB2</td>
<td>SCAMP3</td>
<td>KPNB1</td>
</tr>
<tr>
<td>11</td>
<td>STAT3</td>
<td>LRPPRC</td>
<td>ESR1</td>
</tr>
<tr>
<td>12</td>
<td>NFKBIA</td>
<td>TOB1</td>
<td>RPS27A</td>
</tr>
<tr>
<td>13</td>
<td>HDAC2</td>
<td>CDK2</td>
<td>ABL1</td>
</tr>
<tr>
<td>14</td>
<td>STAT1</td>
<td>TNK2</td>
<td>NCL</td>
</tr>
<tr>
<td>15</td>
<td>BCL2</td>
<td>CDK4</td>
<td>STAT3</td>
</tr>
</tbody>
</table>

NOTE: The clique proteins are in bold.

![Table 3. The top 15 proteins ranked by different centrality in the level one major subnetwork](image)

<table>
<thead>
<tr>
<th>Rank</th>
<th>DC</th>
<th>EC</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MYC</td>
<td>GRB2</td>
<td>MYC</td>
</tr>
<tr>
<td>2</td>
<td>TP53</td>
<td>PTMA</td>
<td>EGFR</td>
</tr>
<tr>
<td>3</td>
<td>GRB2</td>
<td>STAT3</td>
<td>BRCA1</td>
</tr>
<tr>
<td>4</td>
<td>CDC2</td>
<td>MYC</td>
<td>TP53</td>
</tr>
<tr>
<td>5</td>
<td>EGFR</td>
<td>ADA</td>
<td>CDC2</td>
</tr>
<tr>
<td>6</td>
<td>BRCA1</td>
<td>CDC2</td>
<td>KPNB2</td>
</tr>
<tr>
<td>7</td>
<td>EPB41</td>
<td>NUP155</td>
<td>GRB2</td>
</tr>
<tr>
<td>8</td>
<td>STAT3</td>
<td>CD44</td>
<td>KPNB1</td>
</tr>
<tr>
<td>9</td>
<td>STAT1</td>
<td>GTF3C4</td>
<td>NPM1</td>
</tr>
<tr>
<td>10</td>
<td>HDAC2</td>
<td>CCT6A</td>
<td>PRKDC</td>
</tr>
<tr>
<td>11</td>
<td>NFKB2</td>
<td>NPM1</td>
<td>CSE1L</td>
</tr>
<tr>
<td>12</td>
<td>PRKDC</td>
<td>NFYA</td>
<td>STAT1</td>
</tr>
<tr>
<td>13</td>
<td>EZR</td>
<td>NFKBIA</td>
<td>STAT3</td>
</tr>
<tr>
<td>14</td>
<td>CDK4</td>
<td>NFKB2</td>
<td>HDAC2</td>
</tr>
<tr>
<td>15</td>
<td>KPNB1</td>
<td>NCL</td>
<td>EPB41</td>
</tr>
</tbody>
</table>

NOTE: The clique proteins are in bold.
were combined with those being also oncogenes, tumor suppressor genes, and genes involved in complex and genes found by group functional profiling to form the final gene signature for NPC (Supplementary Table S4). The final gene signature is effective to depict the possible molecular mechanism of NPC carcinogenesis (Fig. 5A). We further performed IHC studies of selected four upregulated genes from final gene signatures, including TP53, BCL2, BAX, and MYC. Consistent with our analysis, all of them were upregulated (Fig. 5B). High level of TP53 has been suggested to be related to EBV as previously discussed (9, 40, 45). In regard to the apoptosis pathway, although BCL2 acts as an antiapoptosis protein by inhibiting BAX, upregulation of BAX seems to lessen the effect of BCL2. High level of MYC, an oncoprotein, activates the CCND1 and DNK4/6 complex, which results in cell proliferation. Validation the expression of additional prioritized targets by IHC is needed to further elucidate NPC carcinogenesis.

The final gene signature is also used to query cMap for finding potential drugs (Table 4). In cMap analysis, there are fewer genes in the third group (the final gene signature) compared with the first group (randomization from the original gene set), but more promising drugs on the top-ranked drug list are obtained (Table 4). This indicated that the narrowed down method is effective to filter somehow noisy genome-wide data. Although there are only 38 upregulated and 10 downregulated genes in

<table>
<thead>
<tr>
<th>Rank</th>
<th>Group 1 Description</th>
<th>Group 2 Description</th>
<th>Group 3 Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Trichostatin A</td>
<td>HDAC inhibitors</td>
<td>HDAC inhibitors</td>
</tr>
<tr>
<td>2</td>
<td>LY-294002</td>
<td>PI3-kinase inhibitor</td>
<td>PI3-kinase inhibitor</td>
</tr>
<tr>
<td>3</td>
<td>Vorinostat</td>
<td>HDAC inhibitors</td>
<td>Mammalian target of rapamycin inhibitor</td>
</tr>
<tr>
<td>4</td>
<td>Medrysone</td>
<td>Corticosteroid</td>
<td>Mammalian target of rapamycin inhibitor</td>
</tr>
<tr>
<td>5</td>
<td>Sirolimus</td>
<td>Antibacterial and antifungal phytoalexin</td>
<td>HSP 90 inhibitor</td>
</tr>
<tr>
<td>6</td>
<td>Resveratrol</td>
<td>diuretic</td>
<td>Antipsychotic, phenothiazine, calmodulin inhibitor</td>
</tr>
<tr>
<td>7</td>
<td>Meticrane</td>
<td>A synthetic opioid, potent analgesic</td>
<td>Dopamine D2-receptor antagonist</td>
</tr>
<tr>
<td>8</td>
<td>Phthalysulfathiazole</td>
<td>Antifungal</td>
<td>A potent analgesic, analogue of pethidine</td>
</tr>
<tr>
<td>9</td>
<td>Sulconazole</td>
<td></td>
<td>Antibiotic, also targeting forkhead box M1</td>
</tr>
<tr>
<td>10</td>
<td>Tanespimycin</td>
<td></td>
<td>Drug used for joint and muscular pain</td>
</tr>
</tbody>
</table>

NOTE: Drugs in boldface are those with antitumor effects shown either by cell viability tests or literatures (PubMed).
the final gene signature, these genes are highly representative of NPC gene signature. However, which gene selection method (including the randomization method) is most appropriate for querying cMap needs further experimental investigation.

By finding FDA-approved drug targets in the NPC upregulated PPI network, we obtain possible candidate drugs for treating NPC (Supplementary Table S12 and S13). It is supposed that when the drug disturbs more crucial nodes in the NPC PPI, the NPC network may be destroyed. We further surveyed the ClinicalTrials database. Those drugs being used in clinical trials for treating several cancers whose pathway related to NPC pathogenesis may have a better chance to treat NPC, especially those 66 drugs targeting up-bottleneck genes and oncogenes. There are 9 drugs (13.6%) of the 66 drugs already under NPC clinical trials. One can choose the remaining 57 potential drugs (Supplementary Table S14), which fit the above criteria, for future NPC clinical trials.

Conclusions
To the best of our knowledge, this is the first integrative Web site (http://140.109.23.188:8080/NPC) to depict the expression patterns of differentially expressed NPC genes. This inventory provides a niche for NPC PPI network construction, target prioritization, and potential drug identification. The interaction between prioritized NPC targets (e.g., cliques and bottleneck genes) and drugs highlights a promising approach to address disease-related networks and to uncover potential new therapeutic agents. Validating potential drugs found by in silico approaches both in vitro and in vivo are required in the future.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Grant Support
Grants from National Service Center grant NSC98-3112-B-010-021-CC1 (C.Y. Huang) and grant NSC92-2320-B-038-057 (C.L. Chen), grants from Ministry of Education, AIM for the Top University Plan (National Yang Ming University), Taipei Veterans General Hospital (V99ER2-011) and Center of Excellence for Cancer Research at Taipei Veterans General Hospital (DOH99-TD-C-111-007; C.Y. Huang), and a grant from Taipei Medical University-Center of Excellence for Cancer Research (C.L. Chen and C.Y. Huang).

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 10/25/2009; revised 06/02/2010; accepted 06/28/2010; published OnlineFirst 08/17/2010.

References


Molecular Cancer Therapeutics

From NPC Therapeutic Target Identification to Potential Treatment Strategy

Ming-Ying Lan, Chi-Long Chen, Kuan-Ting Lin, et al.

Mol Cancer Ther Published OnlineFirst August 17, 2010.

Updated version

Access the most recent version of this article at: doi:10.1158/1535-7163.MCT-09-0966

Supplementary Material

Access the most recent supplemental material at: http://mct.aacrjournals.org/content/suppl/2010/08/17/1535-7163.MCT-09-0966.DC1

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.