Identification of novel biomarkers and small-molecule compounds for nasopharyngeal carcinoma with metastasis

Jing-Lin Mi, MD<sup>a,b</sup>, Meng Xu, PhD<sup>a,b</sup>, Chang Liu, PhD<sup>a,b</sup>, Ren-Sheng Wang, PhD<sup>a,b,#</sup>

Abstract
The purpose of this study was to investigate novel biomarkers and potential mechanisms in nasopharyngeal carcinoma (NPC) patients with metastasis.

Two microarray datasets (GSE103611 and GSE36682) were obtained from GEO database, differentially expressed genes (DEGs) and differentially expressed miRNA (DEMs) were identified. Gene ontology (GO) as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were conducted with DEGs and DEMs targeted genes. Protein-protein interactions (PPI) network of the DEGs and DEMs targeted genes were constructed, furthermore, Connectivity Map (CMap) database was applied to select the potential drugs with therapeutic effects.

Overall, we identified 396 upregulated and 19 downregulated DEGs. Additionally, we identified 1 upregulated DEM, miR-135b, and a downregulated DEM, miR-574-5p. Functional enrichment analysis indicated that both DEGs and DEMs targeted genes participated in biological process (BP) of regulation of transcription from RNA polymerase II promoter, DNA-templated positive regulation of transcription, and Epstein-Barr virus infection signaling pathway. Besides, upregulated EP300 gene was a hub node both in DEGs and DEMs target genes. CMap database analysis indicated that sanguinarine, verteporfin, and chrysian are potential drugs for prevention and treatment of NPC metastasis.

In summary, the common hub gene, biological process and pathway identified in the study provided a novel insight into the potential mechanism of NPC metastasis. Furthermore, we identified several possible small molecule compounds for treatment of NPC metastasis.

Abbreviations: BP = biological process, CMap = Connectivity Map CMap, DEGs = differentially expressed genes, DEM = differentially expressed miRNA, EP300 = E1A binding protein p300, GEO = Gene Expression Omnibus, GO = Gene Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, NPC = nasopharyngeal carcinoma.

Keywords: biomarkers, metastasis, microarray technology, nasopharyngeal carcinoma, small-molecule compounds

1. Introduction
Differ from other head and neck cancer, nasopharyngeal carcinoma (NPC) has a special geographical distribution, with a peak incidence approaching 50 cases per 100,000 in Southern China and Southeast Asia. Over the past few decades, microarray technology established as a promising and efficient tool to detect phenotypic characteristics of cancer. For example, Zhang et al revealed that 8 genes and immune-related
pathways associated with the development of osteosarcoma metastasis by weighted gene co-expression network analysis (WGCNA). Cai et al.\(^8\) and Lu et al.\(^9\) identified 5 hub genes as prognostic biomarkers for breast cancer metastasis by bioinformatics analysis. However, studies integrating microarray datasets for investigation of key genes and regulatory networks of metastatic NPC are lacking. Thus, the goal of the present study was to systematically explore the molecular mechanisms underlying metastatic NPC.

In our study, we identified differentially expressed genes (DEGs) and differentially expressed microRNAs (DEM}s) between metastatic NPC and non-metastatic NPC tissue based on information obtained from 2 GEO databases (GSE103611 and GSE36682). Next, we performed bioinformatic analyses of DEGs and DEMs targeted genes by utilizing Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis and construction of protein–protein interaction (PPI) network. The common hub gene, biological process, and pathway were identified between DEGs and DEMs targeted genes, Diagram of data analysis process was showed in Fig. 1. Furthermore, DEGs were uploaded to CMap database to detect small active compounds with possible therapeutic effects on NPC metastasis.

2. Materials and methods

2.1. Data source

Microarray data (GSE103611 and GSE36682) was downloaded from NCBI Gene Expression Omnibus (GEO) database. The GSE103611 dataset contained gene expression from 48 samples, including 24 NPC tumor tissues with distant metastasis and 24 NPC tumor tissues without distant metastasis. The GSE36682 dataset contained the miRNA expression profiles of 68 samples, which included 17 NPC tumor tissues with distant metastasis and 45 NPC tumor tissues without distant metastasis, and were utilized for further analysis in our study. All relevant datasets were publicly available through open access. Therefore, an ethics committee approval was not necessary.

2.2. Screening results of DEGs and DEMs

The interactive online tool GEO2R was utilized to compare between 2 datasets from GEO.\(^{10}\) In our study, GEO2R was used to analyze and screen DEGs and DEMs between the metastatic and non-metastatic NPC samples using the GSE103611 and GSE36682 data sets. Then, microarray data of DEGs and DEMs were obtained in text format. The gene probes were transformed into gene names by referencing the annotation file of probes in the platform. DEGs were screened due to the criteria of \(P < .01\) and \(|\log 2 \text{ fold–change}| > 1.5\). DEMs were screened due to the criteria of \(P < .05\) and \(|\log 2 \text{ fold–change}| > 1\).

2.3. DEMs related genes were predicted using miRTarBase

DEM target genes were screened by the miRTarBase online databases (http://mirtarbase.mbc.nctu.edu.tw/php/index.php).\(^{11}\)
This database contains extensive information of experimentally validated miRNA-target interactions. The DEMs target genes were supported by strong experimental evidence, such as Western blot, quantitative polymerase chain reaction (qPCR), or reporter assay. If the DEMs target genes were not observed in the strong experimental evidence, which would screened by weak experimental evidence, such as NGS, pSILAC, or microarray. In our study, the intersections of target genes of DEMs were screened for further analysis.

2.4. Analyses of GO and KEGG pathway enrichment

GO database was applied to conduct functional studies based on the 2 gene datasets. The gene annotation terms were represented as biological process (BP), cellular component (CC), and molecular function (MF). KEGG was applied to identify functional pathways and practical application of genes. Analyses were assessed using Database for Annotation, Visualization and Integrated Discovery (DAVID). P-value of <.05 indicated statistical significance with regards to GO terms and KEGG pathway. Venn Diagram online tool was utilized to identified the common biological process and pathways between the 2 gene datasets.

2.5. Construction of the PPI network

The STRING database is a web tool that is used for studying target protein interactions. Cytoscape, an open source tool, was applied to visualize biological pathways and molecular interaction networks by integrating networks with gene expression profiles and annotations. We downloaded the DEGs and DEMs from the GSE103611 and GSE36682 dataset, respectively, and we uploaded them onto STRING’s official website. Then, the PPI interaction network of DEGs and DEMs target genes was visualized using the Cytoscape software (v3.6.1). Molecular Complex Detection (MCODE) of Cytoscape was utilized to established modules of DEGs and DEMs according to the following criteria: the highest MCODE score and node number. In addition, we applied Cytohubba of Cytoscape to screen hub nodes that are commonly linked to proteins.

2.6. Screening of small-molecules compounds

Connectivity Map (CMap) is a unique gene-expression-based drug development platform that is focused on finding associations between genes, drugs, and diseases. In this study, we transformed DEGs into probes and compared these probe sets with those that participate in small active molecular interference using the CMap website and obtained the corresponding small-molecule compounds. Enrichment score representing similarity and P-value were calculated. Negatively related small-molecule compounds (P<.01 and enrichment <0) were suggested to be therapeutically effective for treatment of NPC metastasis.

3. Results

3.1. Identification of DEGs and DEMs

In our study, expression profiles from 2 datasets (GSE103611 and GSE36682) were obtained. Based on criteria of P<.01 and |log FC| >1.5, we identified 415 DEGs, including 396 upregulated mRNAs and 19 downregulated mRNAs, volcano plot of DEGs were identified and showed in Fig. 2. Additionally, based on P<.05 and |log FC| >1, we identified 1 upregulated microRNA (miR-135b) and 1 downregulated microRNA (miR-574-5p). Furthermore, we identified microRNA related genes using the miRTarBase online tool. One hundred eighty nine mRNAs were found to be miR-135b-3p target genes while 415 mRNAs were found to be miR-574-5p target genes. All the DEGs and DEMs were identified by comparing expression profiles between the NPC tumor tissues with distant metastasis and NPC tumor tissues without distant metastasis.

3.2. Gene ontology analysis of DEGs and DEMs

In order to further explore the potential molecular mechanism of NPC metastasis. GO analyses of the DEGs and DEMs target genes were conducted by using the DAVID web tool. With regards to BP, DEGs were mainly enriched in ATP-dependent chromatin remodeling, DNA repair, and RNA processing. DEMs target genes were mainly involved in regulation of transcription from RNA polymerase II promoter, DNA-templated positive regulation of transcription and Wnt signaling pathway. For the CC category, DEGs were significantly highly expressed in nucleoplasm, membrane, and cytosol. DEMs were mainly associated with nucleus, membrane raft and postsynaptic membrane. Additionally, for the MF category, DEGs were found to be largely involved in protein binding, poly (A) RNA binding, and nucleotide binding, DEMs target genes, on the other hand, were mainly involved in transcription corepressor activity, transmembrane receptor protein serine/threonine kinase activity and transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding. Result of Venn Diagram indicated that regulation of transcription from
RNA polymerase II promoter and DNA-templated positive regulation of transcription were the common significant biological process across both DEGs and DEMs target genes (Fig. 3A). The top 5 most significant enriched GO terms (*P* value < .05) are summarized in Table 1.

### 3.3. Pathway enrichment analyses of DEGs and DEMs

We analyzed cell signaling pathway enrichment of DEGs and DEMs target gene by using KEGG. A total of 30 significant pathways were selected. DEGs were found to be involved in phosphatidylinositol signaling system, Epstein-Barr virus infection, and choline metabolism in cancer. DEMs target genes were enriched in TGF-beta signaling pathway, signaling pathways regulating pluripotency of stem cells, and Hippo signaling pathway. We further identified that the Epstein-Barr virus infection was common pathway both DEGs and DEMs target genes (Fig. 3B). The top 5 most significant KEGG pathways (*P* value of < .05) are summarized in Table 2.

### 3.4. PPI network construction

We input all DEGs and DEMs target genes into STRING database, as shown in Figs. 4A and 5A, red colored nodes represent up-regulation, while blue colored nodes represent down-regulation. The PPI network of DEGs contained 363 nodes and 1470 edges. We visualized the network using Cytoscape software. Next, we modularized the network using the plug-in MCODE and identified the highest MCODE score and node number. The data indicated that the DEGs module was mainly correlated with the spliceosome and RNA polymerase (Fig. 4B), and the seed gene was heterogeneous nuclear ribonucleoprotein D (HNRNPD). On the other hand, the DEMs target genes module mainly correlated with the cell adhesion molecules (CAMs), and the seed gene was ring finger protein 138 (RNF138) (Fig. 5B). We also used cytohubba, a plug-in, to select the hub nodes. The top 10 hub nodes with higher degrees of interaction in DEGs were identified, including EP300, MAPK10, CCND1, CDKN1A, IGF1, TGFBR1, RUNX1, MYC, KLF4, FOXO1 (Fig. 4C). Among the above genes, EP300 had a high degree in the protein–protein interaction network and was identified as a hub gene in both DEGs and DEMs target genes (Fig. 3C). Therefore, upregulated EP300 may play an important role in metastatic NPC.

### 3.5. CMap analysis of DEGs

We screened the compounds with molecular features that have the potential to treat NPC metastasis. In order to do so, we uploaded previously selected DEGs into the CMap database, the potential small molecular compounds were ranked according to negative connectivity scores and *P* < .05. The top 10 small-molecule compounds that demonstrated high correlation with NPC metastasis are shown in Table 3. Among these molecules, sanguinarine, verteporfin, and chrysin were described to be significantly associated with NPC metastasis (Fig. 6).

### 4. Discussion

NPC is a malignant disease of the head and neck area and causes 34.1 million deaths annually.[16] The disease is characterized by unique geographic, etiologic, and biologic features. As the widely application of intensity modulated radiotherapy and chemotherapy, the survival rate and quality of NPC patients have increased largely. Nevertheless, distant metastases are still an obstacle to successful treatment. It is reported that approximately 30% to 40% of patients with locoregionally advanced NPC develop distant metastasis after receiving treatment.[17] In fact, distant metastasis accounts for cancer-specific mortality among approximately 70% of NPC patients, there is no obvious clinical evidence of metastases in the initial diagnosis and regular examinations is hard to detected subclinical micrometastases.[18] On the other hand, studies to date have indicated that some molecules, such as specific mRNAs, microRNAs, and proteins regulate metastasis through different biological process, including methylation, cell cycle, and adhesion.[19–21] Whereas, the interaction and molecular mechanisms between mRNAs and microRNAs in NPC metastasis remain unclear. Therefore, it is
urgent to screen the key genes and explore the potential mechanisms for NPC patients with metastasis.

Microarray data sets as well as other omics data sets, such as proteomics, phosphoproteomics, and single cell RNA seq data have been increasingly used in cancer research.\[22\] This study is focused on DEGs and DEMs in metastatic NPC compared with non-metastatic NPC based on 2 expression datasets downloaded from the GEO database. Up-regulated gene EP300 was a core node between DEGs and DEMs target genes in the PPI network. In the study of Liao et al.,\[26\] result of western blotting showed

### Table 1
Go function annotation for the DEGs and DEMs targeted genes (top 5).

| Expression | Category | Term | Count | P value |
|------------|----------|------|-------|---------|
| DEGs       | BP       | 00:043044–ATP-dependent chromatin remodeling | 6 | 1.41E-04 |
|            |          | 00:008281–DNA repair | 16 | 3.25E-04 |
|            |          | 00:008396–RNA processing | 10 | 3.57E-04 |
|            |          | 00:00723–telomere maintenance | 6 | 0.01238 |
|            |          | 00:008380–RNA splicing | 12 | 0.01455 |
|            | CC       | 00:005654–nucleoplasm | 130 | 5.54E-19 |
|            |          | 00:016020–membrane | 104 | 5.71E-15 |
|            |          | 00:005283–cytosol | 129 | 2.11E-12 |
|            |          | 00:005737–cytoplasm | 164 | 1.73E-08 |
|            |          | 00:005634–nucleus | 163 | 4.17E-07 |
|            | MF       | 00:000515–protein binding | 276 | 3.38E-15 |
| DEMs target genes | BP       | 00:008366–transcription from RNA polymerase II promoter | 31 | 9.42E-05 |
|            |          | 00:045890–positive regulation of transcription, DNA-templated | 30 | 2.35E-04 |
|            |          | 00:016055–Wnt signaling pathway | 15 | 7.00E-04 |
|            |          | 00:032924–activin receptor signaling pathway | 5 | 7.98E-04 |
|            |          | 00:007178–transmembrane receptor protein serine/threonine kinase signaling pathway | 5 | 0.01021 |
|            | CC       | 00:005634–nucleus | 187 | 9.80E-05 |
|            |          | 00:045121–membrane raft | 17 | 1.72E-04 |
|            |          | 00:045211–postsynaptic membrane | 17 | 2.25E-04 |
|            |          | 00:005923–bicellular tight junction | 12 | 2.93E-04 |
|            |          | 00:000986–cell surface | 29 | 9.54E-04 |
|            | MF       | 00:003714–transcriptional corepressor activity | 19 | 1.24E-05 |
| DEMs target genes | BP       | 00:004675–transmembrane receptor protein serine/threonine kinase activity | 5 | 6.26E-05 |
|            |          | 00:001077–transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding | 18 | 2.81E-04 |
|            |          | 00:003700–transcription factor activity, sequence-specific DNA binding | 46 | 3.10E-04 |
|            |          | 00:000978–RNA polymerase II core promoter proximal region sequence-specific DNA binding | 23 | 3.29E-04 |

**BP** = biological process, **CC** = cellular component, **DEGs** = differentially expressed genes, **DEM** = differentially expressed mRNA, **MF** = molecular function.

### Table 2
Kyoto encyclopedia of genes and genomes pathway analysis for the DEGs and DEMs targeted genes (top 5).

| Expression | Term | Count | P value | Genes |
|------------|------|-------|---------|-------|
| DEGs       | hsa04070:Phosphatidylinositol signaling system | 8 | 0.12857 | PLCB3, INPP5J, DGKJ, PKP5K2, OCRL, MMPI4, PKR2 |
|            | hsa05169:Epstein-Barr virus infection | 9 | 0.12097 | CD44, PSMD1, W, RGD1, JAK3, INPP5K2, STAT3, PIK3R2 |
|            | hsa05231:Choline metabolism in cancer | 8 | 0.14096 | PLD2, PLD1, HF1A, DGKJ, CHKB, PLA2G4F, RPS6KB2, PKR2 |
|            | hsa04666:Fc gamma R-mediated phagocytosis | 7 | 0.02711 | PLD2, PLD1, LMK2, LMK1, RPS6KB2, VAV2, PKR2 |
|            | hsa03008:Ribosome biogenesis in eukaryotes | 7 | 0.02417 | EF6, DROSHA, XPO1, EMG1, HEAT1, SPATA5, XRNI |
| DEMs target genes | hsa04350:Transforming growth factor-beta signaling pathway | 13 | 1.86E-06 | EP300, PPP2R1B, SMAD9, TGFBR1, TGFBR2, SMAD5, BMP2, BMP2, SMAD4, INHBA, ACVR1B, ACVR2B, MYC, BMP11A |
|            | hsa04565:Signal transduction pathways regulating pluripotency of stem cells | 16 | 3.98E-06 | SMAD9, SMAD5, BMP2, SMAD4, IGF1, FZD3, ACVR1B, INHBA, ACR2B, WNT4, SKL, MYC, KLF4, BMP11A, APC, KAT6A |
|            | hsa04390:Insulin signaling pathway | 15 | 4.41E-05 | PPP2R1B, WNK4, TGFBR1, TGFBR2, BMP2, SMAD4, BRCC3, GL2, FZD5, LATS2, CCND1, WNT4, MYC, BMPR1A, APC, KAT6A |
|            | hsa04310:Wnt signaling pathway | 14 | 7.05E-05 | EP300, NKD1, ROCK2, NLK, SMAD4, MAPK10, FZD5, PORCN, PRKCB, CCND1, WNT4, MYC, FOSL1, APC |
|            | hsa05200:Pathways in cancer | 25 | 1.05E-04 | EP300, ROCK2, FGF14, TGFBR1, TGFBR2, MITF, RUNX11, SMAD4, FOXO1, IGF1, BRC5, MAPK10, FZD5, GL2, PRKCB, WNT4, CCND1, CDKNC1A, HHP, TRAF6, TRAF5, CRK, MYC, TRAF3, APC |

**DEGs** = differentially expressed genes, **DEM** = differentially expressed mRNA.
that EP300 protein was increased expression in NPC cell lines compared with normal nasopharyngeal cells. What’s more, upregulation of EP300 promote invasion and metastasis of NPC cells by the induction of epithelial-mesenchymal transition (EMT). A previous study indicated that the expression levels of EP300 was significantly higher in NPC tissues than adjacent non-cancerous tissues, additionally, higher E300 expression is associated with poor overall survival and progression-free survival in NPC patients. EP300 has been proved to be upregulated in several malignancies, such as non-small cell lung cancer, colorectal carcinomas, hepatocellular carcinoma. A recent study demonstrated that mutation of EP300 is correlated with higher tumor mutation burden (TMB) and promotes antitumor immunity in bladder cancer patients. Dou et al. showed that phosphorylation of EP300 associated with liver metastasis. Besides, our results also suggested that EP300 is

Figure 4. PPI network of the differentially expressed genes (DEGs) (A). A significant module of the DEGs (B). Top 10 hub nodes of DEGs (C). Red colored nodes represent up-regulated genes, blue colored nodes represent down-regulated genes. PPI = protein–protein interactions.

Figure 5. PPI network of the differentially expressed miRNAs (DEMs) target genes (A). A significant module of the DEMs target genes (B). Top 10 hub nodes of DEMs target genes (C). Red colored nodes represent down-regulated DEM target genes, blue colored nodes represent up-regulated DEM target genes. PPI = protein–protein interactions.
## Table 3
The top 10 most significant small-molecule compounds that could reverse the tumoral metastasis status of NPC.

| Rank | Cmap name       | Mean  | Count | Enrichment score | P value |
|------|-----------------|-------|-------|------------------|---------|
| 1    | etomidate       | −0.709| 3     | −0.929           | .00052  |
| 2    | scopoletin      | −0.697| 2     | −0.944           | .00692  |
| 3    | sanguinarine    | −0.644| 2     | −0.898           | .02072  |
| 4    | verteporfin     | −0.638| 3     | −0.853           | .00637  |
| 5    | chrysin         | −0.627| 3     | −0.84            | .00827  |
| 6    | talampicillin   | −0.624| 4     | −0.735           | .00977  |
| 7    | piperidolate    | −0.603| 3     | −0.799           | .01661  |
| 8    | gliclazide      | −0.599| 4     | −0.699           | .01737  |
| 9    | proscillaridin  | −0.597| 3     | −0.827           | .01022  |
| 10   | tyloxapol       | −0.592| 4     | −0.791           | .0039   |

NPC = nasopharyngeal carcinoma.

**Figure 6.** The 2D structures of the 3 compounds that could reverse the changes of gene expression in NPC metastasis. Sanguinarine (A), verteporfin (B), chrysin (C). NPC = nasopharyngeal carcinoma.
involved in the TGF-beta signaling pathway, Wnt signaling pathway, and pathways in cancer. Currently, it is well recognized that TGF-beta contributes to the metastatic potential of tumor cells through promoting EMT, cell migration, and invasion.\(^{[23]}\) A reported by Liao et al.\(^{[26]}\) proved that EP300 enhance EMT pathway in metastatic NPC. Up-regulation of EP300 is enriched in TGF-beta signaling pathway and Wnt signaling pathway, and thus promoted the NPC metastasis, which is similar to the result of our study. In addition, several studies demonstrated an important role for aberrant Wnt signaling in NPC involving diverse cellular processes, including cell migration, hypermethylation, and stemness.\(^{[34]}\) Taken together, our study reveals that high expression of EP300 is associated with cell growth, differentiation, migration, and apoptosis.\(^{[35]}\)

Mi et al.\(^{[42]}\) reported that the Epstein-Barr virus-encoded latent membrane protein-1 could induce MicroRNA-10b and promote the metastasis of human nasopharyngeal carcinoma cells. Therefore, we can infer that Epstein-Barr virus signaling process potentially contributes to the mechanism of NPC metastasis and may promote the development of metastatic NPC.

MicroRNAs (miRNA) constitute a family of small non-coding RNAs and help control a wide range of biological pathways such as cell growth, differentiation, migration, and apoptosis.\(^{[35]}\) Moreover, miRNAs play a potential role in cancer development and are often associated with cancer metastasis. In the present study, an upregulated DEM (miR-135b) and a downregulated DEM (miR-574-5p) were selected. The hub gene EP300 was one of the miR-574-5p upregulated genes. MiR-135b was first identified as having a role in differentiation of somatic stem cells.\(^{[36]}\) Prior studies have indicated that miR-135b encourages cell migration and invasion by downregulation of LZTS1 in tumor tissue in the beginning stages of squamous cell carcinoma progression.\(^{[37]}\) Similarly, results of our study indicated that MiR-135b was upregulated in NPC metastatic tissue. On the other hand, another study demonstrated that miR-574-5p suppresses colorectal cancer liver metastasis by negatively regulating MACC-1 expression.\(^{[38]}\) Results of our study also demonstrated that miR-574-5p was downregulated in metastatic NPC tissue. Therefore, our results suggest that the above mentioned microRNAs, particularly miR-574-5p, may play an important role in the metastasis of NPC, and can potentially predict metastasis of NPC.

Our study showed that many pathways were detected in the GO enrichment analysis as related to NPC metastasis. Additionally, we identified 2 significant biological processes across both DEGs and DEMs target genes, such as regulation of transcription from RNA polymerase II promoter and DNA-templated positive regulation of transcription. Previous reports indicated that these processes were involved in metastasis of cancer, such as liver cancer, breast cancer, and osteosarcomas.\(^{[39-41]}\) Furthermore, a KEGG pathway, the Epstein-Barr virus infection was enriched and regulated by DEGs and DEMs target genes together. Li et al.\(^{[42]}\) reported that the Epstein-Barr virus-encoded latent membrane protein-1 could induce MicroRNA-10b and promote the metastasis of human nasopharyngeal carcinoma cells. Therefore, we can infer that Epstein-Barr virus signaling process likely contributes to the mechanism of NPC metastasis and therefore, may be a possible therapeutic target.

Chemotherapy plays an important role in the treatment of cancer metastasis.\(^{[43,44]}\) We identified a few small molecules that hold promise of therapeutic efficacy against NPC metastasis. These small molecules including sanguinarine, verteporfin, and chrysin, have been reported to display anti-metastatic activity. Sanguinarine (SN) serves as quaternary benzophenanthridine alkaloid, which stimulates cytotoxicity across various human cancers and inhibits specific pro-tumorigenic processes including invasion, angiogenesis, and metastasis.\(^{[45]}\) Vertipofen is a Yap1 inhibitor that blocks metastasis by effectively reducing the expression of Yap1 and FGFR1 in lung cancer.\(^{[46]}\) Chrysin, 5,7-dihydroxyflavone, is a naturally-occurring flavonoid that is known to block angiogenesis and metastasis.\(^{[47]}\) Lin et al.\(^{[48]}\) showed that chrysin inhibits IL-6-mediated angiogenesis by reducing the soluble IL-6 receptor/gp130/JAK1/STAT3/VEGF. Hence, our results suggest that these small molecules can function to combat NPC metastasis.

5. Conclusion
In conclusion, we analyzed GEO data for metastatic NPC, the result indicated that an upregulated gene EP300, which was potentially targeted by downregulated miR-574-5p, may associated with the NPC metastasis. The DEGs and DEMs target genes might be involved in the biological process of regulation of transcription from RNA polymerase II promoter and DNA-templated positive regulation of transcription and ultimately affect the development of NPC metastasis through Epstein-Barr virus signaling pathway. Besides, our study identified small-molecule compounds including sanguinarine, verteporfin, and chrysin which may be efficacious in the treatment of NPC metastasis. Nonetheless, our study had some limitations. First, microarray datasets and platform obtained from the GEO were different. Second, the hub gene, biological process, and pathway should be validated using in vitro or in vivo experiments. Despite the limitations in our study, the findings in the study might contribute to advance our understanding of the development of NPC metastasis and provide reference information for the treatment of NPC metastasis.

Author contributions
Conceptualization: Ren-Sheng Wang.
Data curation: Jing-Lin Mi.
Formal analysis: Meng Xu.
Funding acquisition: Ren-Sheng Wang.
Project administration: Chang Liu.
Software: Jing-Lin Mi.
Supervision: Ren-Sheng Wang.
Validation: Jing-Lin Mi, Chang Liu.
Visualization: Meng Xu.
Writing – original draft: Jing-Lin Mi.
Writing – review & editing: Rensheng Wang.

References
[1] Chua MLK, Wee JTS, Hui EP, et al. Nasopharyngeal carcinoma. Lancet 2016;387:1012–24.
[2] Xia WX, Zhang HB, Shi JL, et al. A prognostic model predicts the risk of distant metastasis and death for patients with nasopharyngeal carcinoma based on pre-treatment serum C-reactive protein and N-classification. Eur J Cancer 2013;49:2152–60.
[3] Zhang L, Huang Y, Hong S, et al. Gemcitabine plus cisplatin versus fluorouracil plus cisplatin in recurrent or metastatic nasopharyngeal carcinoma: a multicentre, randomised, open-label, phase 3 trial. Lancet 2016;388:1883–92.
[4] Lee AW, Ma BB, Ng WT, et al. Management of nasopharyngeal carcinoma: current practice and future perspective. J Clin Oncol 2015;33:3156–64.
[5] Tang XR, Li YQ, Liang SB, et al. Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study. Lancet Oncol 2018;19:382–93.
[11] Chou CH, Shrestha S, Yang CD, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res 2018;46:D296-302.

[12] Kanehisa M, Sato Y, Kawashima M, et al. The KEGG Pathway Maps. Nucleic Acids Res 2013;41:D991-5.

[13] Dwivedi P, Chutipongtanate S, Muench DE, et al. SWATH-proteomics and phospho-tyrosine analysis reveals Bruton’s Tyrosine kinase mediated signaling downstream of the mutated granulocyte colony stimulating factor receptors. Sci Data 2019;6:21.

[14] Dou C, Liu Z, Tu K, et al. P300 acetylation regulates the metastasis of hepatic stellate cells into tumor-promoting myofibroblasts. Gastroenterology 2018;154:209-21.

[15] Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene expression signatures to connect small molecules, genes, and disease. Science 2006;313:1929-35.

[16] Liao ZW, Zhou TC, Tan XJ, et al. High expression of p300 is linked to aggressive features and poor prognosis of nasopharyngeal carcinoma. J Transl Med 2012;10:110. doi: 10.1186/1479-5876-10-110.

[17] Sun Y, Li WF, Chen NY, et al. Induction chemotherapy plus concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase 3, multicentre, randomised controlled trial. Lancet Oncol 2016;17:1524-75. doi: 10.1016/j.prp.2019.152475.

[18] Barrett T, Wilhite SE, Leduc P, et al. NCBI GEO: archive for functional genomics data sets-update. Nucleic Acids Res 2013;41:D991-5.

[19] Chou CH, Shrestha S, Yang CD, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res 2018;46:D296-302.

[20] Yi B, Tan SX, Tang CE, et al. Inactivation of 14-3-3 sigma by promoter methylation correlates with metastasis in nasopharyngeal carcinoma. J Clin Pathol 2009;62:20. doi: 10.1186/s41065-019-0096-6.

[21] Chen L, Wang L, Zhu L, et al. Cell cycle-dependent expression of encoded latent membrane protein-1 promotes the metastasis of human unrestricted somatic stem cells. Stem Cells Dev 2010;19:877-85.

[22] Dwivedi P, Muench DE, Wagner M, et al. Phospho serine and threonine protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447-52.

[23] Su G, Morris JH, Demchak B, et al. MicroRNA-135b regulates mineralization in osteogenic differentiation of human unrestricted somatic stem cells. J Biol Chem 2010;285:23133-23141.

[24] Lamh J, Crawford ED, Peck D, et al. The Connectivity Map: using gene expression signatures to connect small molecules, genes, and disease. Science 2006;313:1929-35.

[25] Sun PF, Chen YH, Peng XB, et al. High-dose static and dynamic intensity-modulated radiotherapy combined with chemotherapy for patients with locally advanced nasopharyngeal carcinoma improves survival and reduces brainstem toxicity. Med Sci Monit 2018;24:8849-59.

[26] Sun Y, Li WF, Chen NY, et al. Induction chemotherapy plus concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase 3, multicentre, randomised controlled trial. Lancet Oncol 2016;17:1524-75. doi: 10.1016/j.prp.2019.152475.

[27] Liao ZW, Zhou TC, Tan XJ, et al. High expression of p300 is linked to aggressive features and poor prognosis of nasopharyngeal carcinoma. J Transl Med 2012;10:110. doi: 10.1186/1479-5876-10-110.