Exploring nasopharyngeal carcinoma genetics: Bioinformatics insights into pathways and gene associations

Che Ismail Che Lah, MSc[1], Nik Yusnoraini Yusof, PhD[2], Norhafiza Mat Lazim, MMed[1], Sarina Sulong, PhD[3], Zilfalil Alwi, PhD[3], Baharudin Abdullah, MMed[1]

[1]Department of Otorhinolaryngology-Head & Neck Surgery, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia, [2]Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Kubang Kerian, Kelantan Malaysia, [3]Human Genome Center, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia

ABSTRACT

Introduction: The pathogenesis of nasopharyngeal carcinoma (NPC) is intricate, influenced by a combination of factors including host genetics, viral infection and environmental elements, resulting in genetic and epigenetic modifications. Despite a positive prognosis for early-stage patients, most NPC cases are diagnosed at an advanced stage, highlighting the pressing need for enhanced accessibility to early diagnosis and treatment. The underlying molecular pathways driving NPC progression remain elusive. This study focuses on the use of bioinformatics techniques and databases in carrying out research to gain insights into gene relevance and potential applications in NPC.

Materials and Methods: Searches encompassed articles published in English from January 2017 to June 2024, utilising keywords such as 'nasopharyngeal carcinoma,' 'bioinformatics,' 'gene expression' and 'gene microarrays' across PubMed, MEDLINE and Scopus databases. The Gene Expression Omnibus (GEO) database was utilised to access NPC messenger RNA (mRNA) expression profiling studies.

Results: Most studies utilised the GEO database to identify differentially expressed genes (DEGs) between normal and NPC tissues, followed by functional analysis using gene ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways. Protein protein interaction (PPI) networks of DEGs were commonly constructed using STRING and visualised with Cytoscape software. The integration of GO and KEGG pathway analysis alongside PPI network construction provides valuable insights into the dysregulated pathways and molecular mechanisms underlying NPC pathogenesis. Microarray analysis, particularly datasets such as GSE12452, GSE64634 and GSE34573, has enabled the identification of DEGs associated with NPC. PPI network analysis identifies hub genes, such as DNALI1, DNAI2 and RSPH9, implicated in NPC pathogenesis. Validation of gene expression patterns through platforms like GEPIA and Oncomine validates the clinical relevance of identified biomarkers. Furthermore, studies employing RNA sequencing and bioinformatics approaches uncover novel genes involved in NPC radio resistance and prognosis, paving the way for personalised therapeutic strategies.

Conclusion: Integration of bioinformatics analysis provides insights into the complexity of tumour biology and potential molecular pathways, enabling the development of enhanced strategies for early detection, outcome prediction, recurrence detection and therapeutic approaches for NPC.

KEYWORDS:
Nasopharyngeal carcinoma, bioinformatics, differentially expressed genes, clinical application

INTRODUCTION

Nasopharyngeal carcinoma (NPC) is among the most ubiquitous head and neck malignancies in China, Southeast Asia, the Middle East and North Africa.[1] Due to the interaction of host genetics, viral infection and environmental factors, NPC pathogenesis is multifactorial and strongly correlated with genetic and epigenetic alterations.[2] Despite the promising outlook for early-stage NPC patients, the majority of newly diagnosed cases manifest with locally advanced disease. Presently, there is limited access to early NPC diagnosis and treatment, and the precise molecular pathways driving the evolution of NPC remain unclear. Therefore, it is crucial to investigate potential biomarkers to facilitate early identification and enhance the prognosis of NPC patients. Analysing biological signals, understanding them and managing data are all components of bioinformatics. This has been made easier and more feasible by developments in artificial intelligence and machine learning algorithms, which have increased the usefulness of bioinformatics in the biology of cancer.[3] There is currently a wide range of bioinformatics tools being developed to help address complicated issues ranging from predicting clinical outcomes to identifying factors responsible for altering the tumour-immune milieu. With the rapid development of gene chip and RNA sequencing technologies, bioinformatics analysis is being used in screening for potential biomarkers for various diseases, particularly cancers, in addition to its emerging role in precise screening, prompt diagnosis and molecular-targeted treatment for various types of cancers.[4] The clinical application of bioinformatics uses bioinformatics information and techniques to aid in disease diagnosis, treatment, prevention and control, as well as the development of chemical, structural and biochemical methodologies for clinical

This article was accepted: 14 July 2024
Corresponding Author: Baharudin Abdullah
Email: profbaha@gmail.com

research.[5] Bioinformatics techniques, for example, are employed in cancer research to detect biomarkers in many types of malignancies at various phases - start, progression and late. Thus, the use of bioinformatics enables personalised clinical treatment for each type of cancer. Bioinformatics techniques are frequently used in research that offer the data required to establish a link between a genetic mutation or variation and a given clinical result.

Bioinformatics gives new insights and core data for discovering reliable and functional differentially expressed genes (DEGs), microRNAs (miRNAs), circular RNAs (circRNAs) and long non-coding RNAs (lncRNAs). Various computational techniques are utilised for investigating biological databases in an effort to develop novel methods and protocols for processing genomic and proteomic data to use in a range of sectors such as the development of pharmaceuticals, medicine and other related industries.[6] The Gene Expression Omnibus (GEO) managed by National Centre for Biotechnology Information is a public archive for high-throughput experimental data from a variety of sources.[7] It is the largest and most complete public gene expression database available for free download, with the option for users to contribute their own data for sharing and validation. These data comprise proteomic research from mass spectrometry and serial analysis of gene expression (SAGE), as well as data from single and dual channel microarray-based investigations measuring messenger RNA (mRNA), genomic DNA and protein abundance.[8] The functional study of DEGs included pathway enrichment using Gene Ontology (GO) and the Kyoto Encyclopaedia of Genes and Genomes (KEGG).[9,10] The KEGG database was used to investigate the metabolic and functional pathways. With the aid of the Search Tool for the Retrieval of Interacting Genes (STRING) and Cytoscape software, a protein protein interaction (PPI) network of DEGs can be found and visualised.[1,11]

Furthermore, web-based technologies for cancer research and diagnostics are made easier by Gene Expression Profiling Interactive Analysis (GEPIA).[12] Advances in microarray technology allow for the collection of large volumes of data on DEGs from specific cancer cells. This enormous amount of data needs the use of computational tools and databases to store, decipher and extract valuable information from the obtained data, such as the identification of new biomarkers for cancer diagnosis. The present work focuses on the application of bioinformatics techniques and databases in NPC through an extensive literature search. In this review, we explored bioinformatics approaches to get additional insight into the relevance and possible uses of genes in NPC.

MATERIALS AND METHODS

Electronic literature searches were conducted in PubMed, MEDLINE and Scopus for articles published in English from January 2017 to June 2024. Search words used were 'nasopharyngeal carcinoma', 'bioinformatics', 'gene expression' and 'gene microarrays', in combination with 'AND' and 'OR'. Additional relevant articles pertinent to this review were identified by reviewing the references of articles that had been retrieved. The inclusion criteria were bioinformatics tools including genomics, transcriptomics or proteomics for the analysis of data in NPC. Studies without available data were excluded. The PRISMA guideline was followed wherever possible.[13] In addition, the GEO database (http://www.ncbi.nlm.nih.gov/gds/)[7] was accessed to retrieve related NPC mRNA expression profiling studies.

RESULTS

Selections of Studies

A total of 308 articles were extracted from electronic databases. Following the removal of duplicates, there were 298 studies left. After the titles and abstracts were screened, 249 papers were eliminated. Of the remaining 49 articles assessed for eligibility, 11 were excluded following full-text review. Finally, 38 articles were included for the review (Figure 1).[7,9-12,14-46]

Tools of Bioinformatics in NPC

The findings from the bioinformatics tools play a crucial role in the management of NPC by facilitating the analysis and interpretation of biological data, which is essential for understanding the disease, developing personalised treatment strategies, and predicting outcomes. Furthermore, bioinformatics is essential for the identification of biomarkers and clinical decision assistance for NPC. By leveraging advanced computational techniques to analyse complex molecular data, bioinformatics tools enable personalised medicine approaches and facilitate more precise and effective management of NPC patients.

Among the bioinformatics tools used were GO and KEGG, PPI, GEPIA, Weighted Gene Co-expression Network Analysis (WGCNA), Database for Annotation, Visualisation and Integrated Discovery (DAVID), STRING and Cytoscape software.[7,9-12,14,15] The majority of the investigations used the associated GEO database for selecting the DEGs between normal nasopharyngeal tissues and NPC tissues from the microarray expression profiles. Targeted DEG functional analysis was reviewed using GO and KEGG pathway enrichment analysis.[9,10] In most of the studies, a PPI network of DEGs was built by STRING and visualised using Cytoscape software.[11]

MyCancerGenome is a resource for understanding the genetic underpinnings and clinical implications of various cancers, including NPC.[16] This platform provides detailed insights into genetic mutations associated with NPC, helps in identifying targeted therapies, and connects clinicians and patients with relevant clinical trials. Understanding the mutation status of MSH6 and CDKN2A can help oncologists decide whether certain targeted treatments or immunotherapies might be effective. This is particularly important for NPC, which has a distinct genetic and aetiological profile often associated with EBV infection. MSH6 is a DNA mismatch repair gene.[17] Alterations in this gene can lead to microsatellite instability, which has implications for cancer progression and treatment response. MSH6 status is used as an inclusion criterion in several clinical trials for NPC, indicating its significance in the development of targeted therapies. CDKN2A is involved in cell cycle regulation.[18] Its mutations can disrupt normal cell cycle control, contributing to cancer development and progression.

The Cancer Genome Atlas (TCGA) plays a pivotal role in advancing our understanding of NPC by providing comprehensive genomic data and facilitating research into its molecular underpinnings. Although NPC was not one of the primary cancers studied in the original TCGA project, the data and methodologies developed by TCGA have significantly influenced NPC research.[19] TCGA's approach to characterising cancer genomes has been applied to NPC, leading to the identification of key genetic alterations and pathways involved in this cancer. Alterations in genes such as *TP53* and *PIK3CA* in NPC, which are critical for understanding the disease's course and possible treatment targets have been found.[20] TCGA's integration of genomic, transcriptomic and epigenomic data has been mirrored in NPC studies, providing a holistic view of the disease. This integrated approach helps in identifying potential biomarkers for early detection and targets for therapy. To fully analyse and interpret NPC-related genomic data, researchers frequently use a combination of tools and databases. The different bioinformatics tools used in many facets of NPC research are shown in Table I.[7,9-12,14,15]

Microarray Analysis for NPC

Hundreds of DEGs implicated in different signalling networks, molecular functions and biological processes can be identified using gene microarrays, which are high-throughput platforms for the investigation of gene expression. The efficiency and accuracy of analysis are improved when microarray technologies and bioinformatics tools are combined.[21] GSE12452 dataset was the most used by researchers (n = 13), followed by GSE64634 (n = 7) and GSE34573 (n = 7), respectively (Table II). Other datasets were GSE53819 and GSE13597 (n = 4), GSE52068, GSE62336, GSE32960 and GSE36682 (n = 2). Sengupta et al used GSE12452 dataset to measure mRNA expression levels for essentially all human genes and all latent Epstein-Barr virus (EBV) genes in nasopharyngeal carcinoma tissue samples and normal nasopharyngeal tissues.[22] Nine of the ten core genes *(FN1, MMP1, MMP3, PLAU, PLAUR, SERPINE1, SPP1, COL8A1, COL10A1)* were shown to be potentially valuable as NPC diagnostic biomarkers by Guo et al.[23] using GSE12452.

The two most used platforms in microarray technology were Affymetrix platform (GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) and IIlumina platform (Human Methylation450 BeadChip).[24] Affymetrix arrays, which are often referred to as high-density oligonucleotide arrays or oligonucleotide arrays, use a collection of 11–20 Perfect Match and Mismatch oligonucleotides to represent each gene. One of the most sophisticated methods for microarray-based gene expression analysis is Illumina technology. It has a minuscule feature size, dense features and the capacity to analyse several samples simultaneously. With the use of microarray technology, it has become possible to analyse over 10,000 genes at once and uncover genetic anomalies in tumours on a genome-wide scale by analysing the genetic profiles of biological samples. DNA-to-RNA and DNA-to-DNA strands are examples of complementary nucleotide chains that hybridise to create the foundation of microarray technology.[25] Because of this technology, it is now feasible to analyse tumour growth, disease progression, cellular response to stimuli and therapeutic target identification by connecting physiological cell states to gene expression patterns. Table III shows the common microarray datasets and platform used for GEO database analysis.

Identification of DEGs

The online programme GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r) was used in most investigations to process the raw microarray data of the datasets acquired from the GEO database and identify genes that are expressed differently in NPC tissues compared to normal nasopharyngeal mucosa.[26] GEO2R is an online tool allowing users to compare different data sets in a GEO series in the identification of DEGs across experimental conditions. GEO2R presents a simple interface that allows users to perform sophisticated analysis of GEO data to help identify and visualise DEGs.[7] Unlike GEOs and other dataset analysis tools, GEO2R does not rely on curated records and cross-examines original submitter-supplied data directly. This expands the utility of the database to a much wider audience, allowing a greater proportion of GEO data to be analysed in a timely manner and with more flexibility in determining comparison of groups of samples and type of analysis to perform.

The GEO database includes microarray datasets, such as GSE12452 and GSE53819. These datasets are used to find and analyse DEGs between diseased and healthy individuals. These datasets highlight the efficacy and scale of microarray technology for bioinformation collection.

CircRNA_0000285 has been suggested by Shuai et al.[27] as a potential new biomarker for NPC radiosensitivity. The integrated analysis of three transcriptome profiling datasets allowed the study to identify DE of lncRNAs, circRNAs and mRNAs between NPC and chronic nasopharyngitis (CNP) tissues simultaneously. A total of 50 mRNAs were found to be the last functioning genes in the circRNA miRNA mRNA network. Environmental stress, cell motility and migration, cytoskeleton, antiproliferative activity, regulation of voltage-gated calcium channels, cell proliferation and apoptosis, desmosome formation, annexin, B lymphocyte antigen receptor, bimodal regulator of epidermal growth factor receptor and mitogen activated protein kinase signalling, extracellular matrix protein, and chromosome segregation are among the physiological and pathological mechanisms linked to the functions of these target genes.

Chen et al.[28] identified aberrantly methylated DEGs or DEGs regulated by differentially methylated miRNA in a study that integrated methylation and miRNA expression patterns. They also created a pipeline for aggregating consensus DEGs from diverse datasets on several platforms, including microarray and RNA sequencing. Six mRNA expression datasets and two methylation datasets were specified for further study. They established robust consensus DEGs by merging several types of datasets and analysing them using bioinformatics. The study found that most DCGs were downregulated and likely to lose connection in cancer. 98% of differentially co-expressed partners (DCPs) that lost positive correlations with PAX5 were upregulated, while 85% of DCPs that lost negative correlations were downregulated.

Actively expressed EBV genes, including *EBNA1*, were co-expressed with upregulated DNA methyltransferase and PolyComb Group proteins. This suggests that EBV interacts with the host genome and modifies host genome methylation, leading to NPC.

1,218 (555 upregulated and 663 downregulated), 1,232 (348 upregulated and 884 downregulated) and 1,301 (553 upregulated and 748 downregulated) genes were found after analysis of the GSE12452, GSE34573, and GSE64634 datasets by Zhu et al.29 The cluster analysis of DEGs indicated significant changes between the normal nasopharyngeal mucosa and NPC specimens.

GO Terms Enrichment and KEGG Pathway Analysis

A variety of GO tools were available for extracting statistically meaningful findings from the investigation of the GO database. The GO resource (http://www.geneontology.org) is a contemporary biological database that created a set of organised, controlled vocabularies to explain gene function analysis.[9] The GO analysis was used to classify genes into molecular function, cellular component and biological process types. The pathway analysis of KEGG (www.genome.jp/kegg/pathway.html) was utilised to determine DEGs at the level of biological function.[10] KEGG is a knowledge repository that connects genomic data to higher level functional information for a methodical examination of gene function. It is frequently used to locate groups of co-expressed genes that have a common route. It stores graphical representations of biological processes, including metabolism, membrane transport, signal transduction and the cell cycle. The 'clusterProfiler' package (http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) in R software was used in most of the studies to investigate the biological function of DEGs using GO and KEGG pathway enrichment analysis.[30]

Gene Functional and Pathway Enrichment Analysis of DEGs

The DAVID database (http://david.abcc.ncifcrf.gov) is a web-based bioinformatics enrichment tool that allows for thorough high-throughput gene functional annotation analyses.[15] It incorporates a variety of public gene and protein annotation sources, providing information on over 1.5 million genes from over 65,000 species. DEGs were primarily engaged in cilium movement and the drug metabolism-cytochrome P450 pathway, which has been shown to be essential for the treatment of cancer according to a functional and pathway enrichment analysis conducted by Ye et al.31 using the dataset GSE64634. *DNALI1, RSPH4A, RSPH9, DNAI2* and *ALDH3A1* were also found to be hub DEGs in the study, whereas PPI module analysis showed that these hub genes interacted closely and may have a role in the pathway and biological processes linked to NPC.

PPI Network Construction and Hub Genes Identification

The connection between protein molecules is known as the PPI, and it is used to explore correlations related to genetic networks, biochemistry and signal transduction. Cytoscape software (version 3.4.0; http://www.cytoscape.org/) was used to design and visualise a PPI network, and the online search tool STRING (http://string-db.org) to investigate the relative interaction of the DEGs.[11] The ten most important hub genes were found in the study by Zhu et al.[29] using PPI network design. These were *DNALI1, DNAI2, RSPH9, RSPH4A, NDC80, TYMS, CCDC39, DNAH5, CALM*1 and *CCDC114*. The hub gene with the greatest level of linkage was DNALI1. Using the PPI network, Lu et al.32 discovered additional molecules associated with vimentin, revealing the linked genes to be *AURKB, AKT1, TPM4, DMD, TTN, TPM1, CASP7, CASP3, CASP6*, and *CASP8*. The findings further showed that vimentin engages in the p53 and tumour necrosis factor (TNF) signalling pathways and plays a critical role in the epithelial-mesenchymal transition process.

Validation of Expression and Clinical Analysis of Hub Genes

The GEPIA (http://gepia.cancer-pku.cn) and Oncomine databases (https://www.oncomine.org) were tools for comparing the expression levels of important genes in normal nasopharyngeal tissues and NPC.[12,33] The expression of 25 core genes was confirmed in NPC tissues by Guo et al.[23] by uploading the 25 core genes to the GEPIA database. Ten of the core genes—*FN1, MMP1, MMP3, PLAU, PLAUR, SERPINE1, SPP1, COL8A1, COL10A1,* and *COL17A1*—were found to be significantly overexpressed in NPC tissues when compared to normal nasopharyngeal tissues. The results aligned with the GSE40290 and GSE53819 datasets. Table IV shows some of the most highly expressed genes in NPC.[34-41]

Application of Bioinformatics in the Diagnosis and Management of NPC

By employing RNA sequencing and bioinformatics analysis to analyse and integrate the DEGs between radioresistant and radiosensitive NPC tissue samples, Sun et al.[42] discovered three key genes, *DOCK4, MCM9* and *POPDC3*, that may be implicated in the radioresistance of NPC. The results of this investigation offer fresh insights into the process underlying NPC radioresistance, and more experimental research focusing on these key genes is necessary. A work by Xue et al.[43] offered a theoretical foundation for the clinical application of *LRRC46, PCDP1* and *c9orf24* in the therapy of NPC in the future. The results of this study demonstrate that *c9orf24, PCDP1* and *LRRC46* are identified as putative gene markers for NPC using a variety of methodologies, such as microarray-based analysis, WGCNA analysis, GO and KEGG functional enrichment studies, and PPI network analysis. The identification of *LRRC46, c9orf24* and *PCDP1* may be useful in identifying NPC and offer novel therapeutic approaches.

In a work by Guo et al.[23], bioinformatic techniques utilising GSE40290 and GSE53819 were used to determine DEGs and possible pathways for NPC. There were 314 DEGs in all, and it was found that nasopharyngeal carcinogenesis could be linked to a number of biological processes and signalling pathways, such as those related to the extracellular matrix, the NF-κB signalling pathway, cancer pathways, the B cell receptor signalling pathway and the interaction between the extracellular matrix and receptors. Further investigation identified the main genes with significant diagnostic value for NPC as *FN1, MMP1, MMP3, PLAU, PLAUR, SERPINE1, SPP1, COL8A1 and COL10A1.*

Using three data sets (GSE68799, GSE12452, and GSE53819), Huang et al.[44] used WGCNA to identify and validate hub

Table I: Common bioinformatics instruments applied to different elements of the analysis of nasopharyngeal cancer

Bioinformatic tool	Description/usage
Genome analysis toolkit (GATK)	Variant discovery and genotyping in NPC genomes
Spliced Transcripts Alignment to a Reference (STAR)	RNA-seq alignment and transcript quantification
DESeq2	Differential gene expresson analysis in NPC RNA-seq data
Cufflinks	Transcript assembly and differential expression analysis
STRING	Protein-protein interaction networks in NPC
DAVID	Functional enrichment analysis of NPC-related genes
Integrative Genomics Viewer (IGV)	Visualisation of genomic data including NPC mutations and CNVs
MutSigCV	Identification of significantly mutated genes in NPC
UCSC Genome Browser	Genome annotation and visualisation tool
GenePattern, Broad Institute	Access to various tools for gene expression analysis, clustering, etc.
OncoPrint	Visualisation of oncogenic alterations in NPC
TCGA Data Portal	Access to NPC genomic data from the Cancer Genome Atlas
COSMIC	Catalogue of somatic mutations in NPC and other cancers
ENCODE	Exploration of regulatory elements and gene expression in NPC

NPC: Nasopharyngeal carcinoma

Table II: The common GSE dataset used[7]

Differentially expressed Genes	Numbers	Title/status
GSE40290	1	Molecular classification of non-keratinizing nasopharyngeal carcinoma using mRNA expression profiling
GSE53819	4	Genome-wide expressing profiling of nasopharyngeal carcinoma primary tumours versus non-cancerous nasopharyngeal tissues
GSE64634	7	mRNA expression profiling of nasopharyngeal carcinoma
GSE12452	13	mRNA expression profiling of nasopharyngeal carcinoma
GSE34573	7	A global view of the oncogenic landscape in nasopharyngeal carcinoma: an integrated analysis at the genetic and expression levels
GSE48501	1	Gene expression data from human radioresistant and radiosensitive nasopharyngeal carcinoma cells
GSE13597	4	Expression data from biopsies of nasopharyngeal carcinoma and non-malignant controls
GSE52068	2	Genome-wide analysis of DNA methylation between nasopharyngeal carcinoma tissues and normal nasopharyngeal epithelial tissues
GSE62336	2	Methylome study in nasopharyngeal carcinoma
GSE32960	2	microRNA profile of human nasopharyngeal tissues: nasopharyngeal carcinoma tissues vs. normal nasopharyngeal tissues
GSE36682	2	MicroRNA signatures correlate with diagnosis, distant metastasis and prognosis in nasopharyngeal carcinoma
GSE68799	1	RNASeq identified human transcriptome alterations in Chinese nasopharyngeal carcinoma
GSE43039	1	microRNA expression profiling of human nasopharyngeal carcinoma tissues and chronic nasopharyngitis specimens
GSE70970	1	Comprehensive microRNA profiling of nasopharyngeal carcinoma specimens
GSE95166	1	Designated experiment to screen for lncRNAs whose expressions are dysregulated in nasopharyngeal carcinoma tissue
GSE126683	1	Highlight its biologic importance, and suggest a therapeutic role for inhibitors of NF-KB pathway activation in the treatment of Waldenström's macroglobulinemia

genes. Hub genes with predictive values were screened using a different data set, GSE102349. Functional tests verified that *IGSF9* expression boosted NPC cell invasion, migration and proliferation in vitro. Patients with NPC had high expression levels of *IGSF9*, which was a suitable prognostic gene. As a member of the immunoglobulin superfamily, *IGSF9* is essential in suppressing the formation of synapses by controlling calmodulin-like activity. In addition to IGSF9 potentially causing NPC cell metastasis via the phosphatidylinositol 3-kinase/protein kinase B (Akt) signalling pathway, calmodulin is implicated in cancer metastasis.

An assortment of newly discovered methylation genes and miRNA were found by Wang et al.[45] and may be useful as possible biomarkers for NPC prognosis. The unfavourable prognosis of NPC may be associated with hypomethylation of *SRC, SMAD3, YWHAZ* and *HSPA4,* as well as hypermethylation of *miR129-2*. The capacity of miRNA biology to target numerous genes is its biggest advantage when applied to the clinical care of patients with NPC. However, before this four-miRNA signature can be successfully implemented clinically, more research is needed to understand the regulatory roles of the four miRNAs connected to p53 signalling or other important signalling pathways in the genesis and development mechanism of NPC and the specific regulatory mechanisms.

Zheng et al.[46] looked for changes in the regulators of the NF-κB pathway in NPC using whole-exome sequencing. Numerous genes involved in the regulation of the NF-κB pathway, such as *TNFAIP3, CYLD* and *NFKBIA,* were discovered to have many loss-of-function mutations. The investigation also revealed a noteworthy occurrence of

Table III: Microarray datasets and platform used for GEO databases[7]

Differentially expressed Genes	Platform	Nasopharyngeal	Normal carcinoma	Key findings
GSE40290 GSE53819	GPL8380 GPL6480	8 18	25 18	Core genes, including FN1, MMP1, MMP3, PLAU, PLAUR, SERPINE1, SPP1, COL8A1, COL10A1, had high diagnostic value for NPC
GES64634 GSE12452 GSE34573	GPL570 GPL570 GPL570	12 31 16	4 10 4	ARMC4, SERPINB3, MUC4 etc. have a close relationship with NPC
GSE48501	GPL570	14	Not stated	Identified three core genes, DOCK4, MCM9, and POPDC3 among 12 common downregulated genes
GSE34573 GSE12452 GSE64634 GSE13597 GSE53819 SRP058243 GSE52068 GSE62336 GSE32960 GSE36682 GSE36682	GPL570 GPL570 GPL570 GPL96 GPL6480 Illumina Hiseq 2000 Illumina Illumina GPL14722 GPL15311 GPL20699	15 31 12 25 18 41 24 25 312 62 246	4 10 4 3 18 4 24 25 18 6 17	Promoter hypermethylation, expression up-regulation, and association with overall survival, genes such as SCUBE2, PRKCB, IKZF1, MAP4K1, and GATA6 could be promising novel diagnostic biomarkers
GSE12452 GSE53819	GPL570 GPL6480	49	28	The expression levels of CDK1, CDC45, RSPH4A, and ZMYND10 probably affected survival of NPC patients according to GEPIA database, and identification of small-molecule compounds which may be efficacious in the treatment of NPC
GSE53819	Agilent-014850	18	18	c9orf24, PCDP1, and LRRC46 are identified as potential gene for NPC
GSE68799 GSE12452 GSE53819 GSE102349	Not stated	46	Not stated	Three hypoxia signatures (99-gene, 26-gene and 15-gene) have prognostic value in NPC patients
GSE12452 GSE13597 GSE43039 GSE70970 GSE95166 GSE126683	GPL570 [HG-U133A] Affymetrix nCounter® Arraystar V4 Agilent-045997 Arraystar V3	10 3 20 17 4 3	31 25 20 246 4 3	lncRNA-miRNA-mRNA networks play significant roles in the development and progression of tumors
circRNA_0000285	Arraystar, Inc	42	42	KIAA0101 ranked top overexpressed gene lists in GSE6631 dataset. highly expressed in NPC tissues and cell lines
GSE64634 GSE53819 GSE12452	GPL6480 GPL570 GPL570	14 18 31	4 18 10	A total of 306 DEGs and 13 hub genes were identified and may be regarded as diagnostic biomarkers for NPC
GSE12452 GSE34573 GSE64634	GPL570 GPL570 GPL570	31 16 12	10 3 4	Upregulated DEGs were significantly enriched in biological processes, including 'cell adhesion', 'cell division', 'mitosis' and 'mitotic cell cycle'
GSE12452 GSE34573	GPL570	31 13	10 3	DEGs were mostly enriched in ciliummovement, antimicrobial humoral response, O-glycan processing, mucosal immune response, hormone and neurotransmitter biosynthetic process, and drugmetabolism-cytochrome P450 pathway
GSE64634 GSE12452	GPL570	12 31	4 10	Critical node proteins were identified in the network, including CDK1, SMC4, KNTC1, KIF23, AURKA and ATAD2 involved in NPC
GSE12452 GSE13597 GSE34573 GSE53819 GSE64634	GPL570 GPL96 GPL570 GPL6480 GPL570	50	22	Vimentin promotes the negative biological behaviours of NPC, including its occurrence, malignancy, and poor prognosis
GSE32960	GPL570	312	18	Identification of NPC patients with a four-miRNA signature may increase the prognostic value and provide reference information for precision medicine

Table III: Microarray datasets and platform used for GEO databases[7]

Differentially expressed Genes	Platform	Nasopharyngeal	Normal carcinoma	Key findings
GSE12452 GSE34573	GPL570	44	13	Module's analysis revealed that cyclin-dependent kinase 1 and exportin 1 were involved in the pathway of Epstein Barr virus infection
GSE52068 GSE62336	Illumina	24 25	12 25	Identified crucial genes that were indicted to be hypomethylated, instead of hypermethylated, in the NPC samples, including SRC, SMAD3, YWHAZ and HSPA4
GSE12452 GSE13597	GPL96	56	13	Up-regulated genes were significantly involved in cell cycle, oocyte meiosis, DNA replication and p53 signalling pathway

GEO: Gene Expression Omnibus, NPC: Nasopharyngeal carcinoma, DEG: Differentially expressed genes

Table IV: The genes and mRNAs that are significantly expressed in nasopharyngeal carcinoma, together with their corresponding pathways, target genes and role in the pathogenesis

Gene/mRNA	Target Genes/proteins	Pathways	Role in NPC pathogenesis
EBER34	EBV proteins (LMP1, LMP2A)	NF-κB, JAK/STAT	Immune evasion, cell growth and survival, oncogenic potential
LMP134	TRAF proteins, IKK complex	NF-κB, JNK, PI3K/AKT	Cell proliferation, inhibition of apoptosis, metastasis
BART34	Various cellular genes	miRNA regulatory pathways	Regulation of apoptosis, immune response, epithelial-mesenchymal transition (EMT)
BCL235	Pro-apoptotic proteins (BAX, BAK)	Intrinsic apoptotic pathway	Anti-apoptotic, tumour survival and growth
STAT336	Cyclin D1, BCL2, VEGF	JAK/STAT	Cell proliferation, survival, angiogenesis
EGFR37	RAS, PI3K, PLCγ	EGFR, MAPK, PI3K/AKT	Cell proliferation, survival, migration, angiogenesis
PD-L138	PD-1 receptor on T cells	Immune checkpoint	Inhibition of T cell function, immune evasion, tumour progression
VEGF39	VEGFRs	VEGF signalling pathway	Angiogenesis, tumour blood supply, growth, and metastasis
MMP940	Extracellular matrix components	ECM remodelling	Breakdown of the extracellular matrix, tumour invasion, and metastasis
CXCL1241	CXCR4 receptor	Chemokine signalling pathway	Tumour cell migration, invasion, metastasis

NPC: Nasopharyngeal carcinoma

mutations in other genes associated with DNA repair, apoptosis and cell division. The researchers found that increased NF-κB pathway activity in NPC was associated with mutations in the regulators of the NF-κB pathway. IκBα inhibitors are among the possible therapeutic targets in the NF-κB pathway that the researchers identified. They also discovered a connection between worse clinical outcomes for NPC patients and mutations in the regulator of the NF-κB pathway.

DISCUSSION

In this review, we summarised integrated bioinformatics techniques that allow for the discovery of important gene functions and related biological mechanisms linked with NPC pathogenesis. We examined and compared their essential information and features of OMIC databases, array technology platforms, and main outcomes (Table III). These tools enable researchers to quickly examine a large number of datasets from complex data platforms, discover genes, proteins, gene alterations or mutations associated with patient survival, ask specific questions and test hypotheses. Each bioinformatics tool has distinct advantages. The most comprehensive source of knowledge on the roles played by genes is the GO knowledge base. Genes were classified into several different categories for the GO analysis, and functional and metabolic pathways were discovered using the KEGG database. Some databases offer additional features, such as the top differential gene display function in GEPIA, which enables clinicians and researchers to choose potential target genes for diagnosis or treatment.

Over the years, novel approaches based on scientific knowledge from cancer bioinformatics, such as gene therapy and molecular-targeted therapies, have contributed to remarkable achievements and clinical benefits.[47] Based on these technologies, a few possible RNA and protein biomarkers have been discovered. Precision therapy and prognosis relied on verified biomarkers with improved screening, diagnosis, and monitoring capabilities. Based on this, gene chip, RNA sequencing, and bioinformatics analysis have emerged, providing a thorough screening of tumour biomarkers as well as a means for understanding the significance of detailed biomarkers in cancer pathology. The pathogenesis of cancer, including gene mutation, transcriptional regulation, protein synthesis, and metabolic alterations, is a systematic mechanism.[48] Based on these relationships, integrating two or more types of omics research and using machine learning methods to perform association analysis on molecules at multiple levels could compensate for

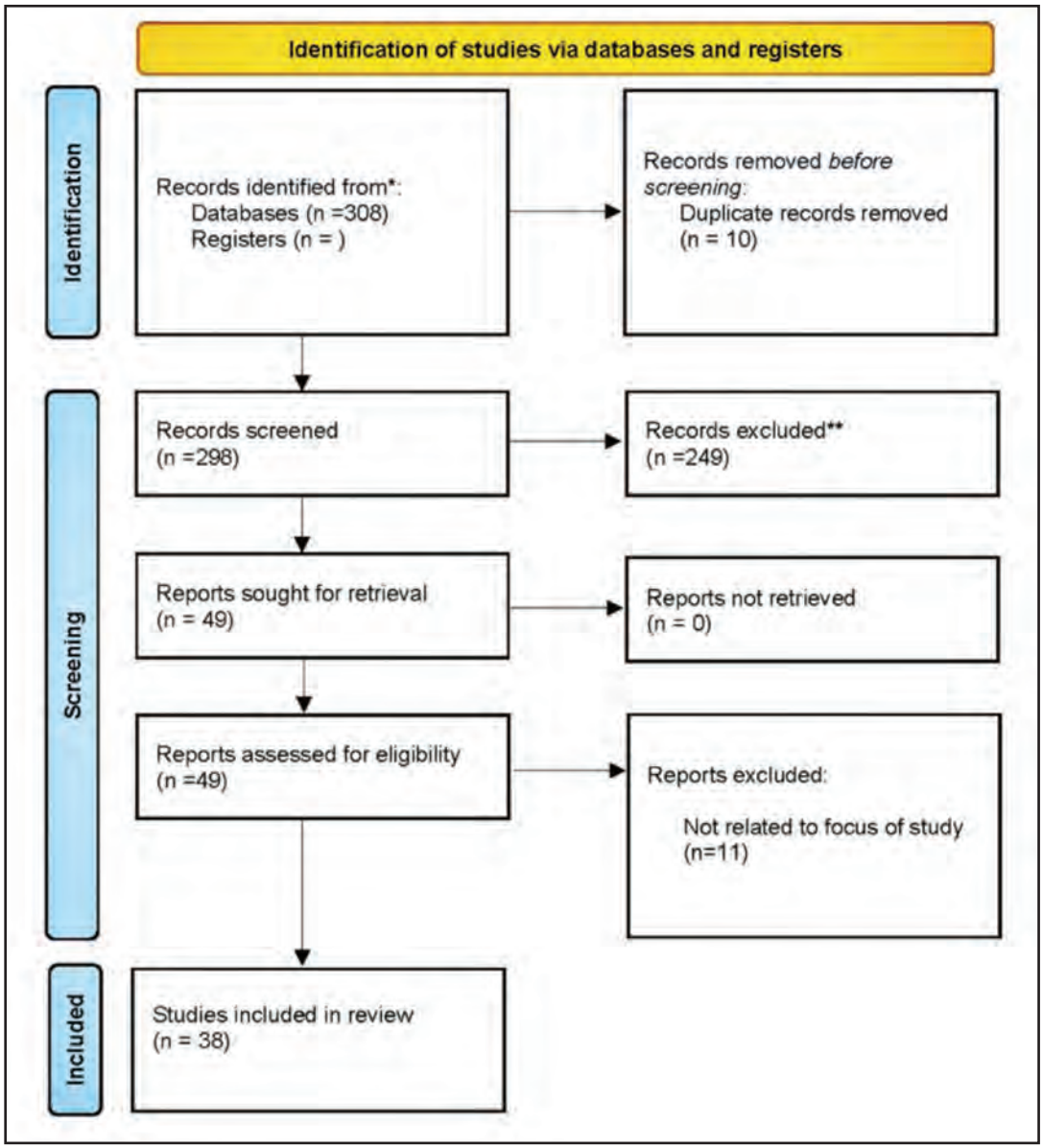

Fig. 1: PRISMA flow diagram of study selection.

the lack of data caused by single omics analysis and reduce the likelihood of false positive results. To advance the study and development of tumour biomarkers, multi-omics integration has emerged as a new trend.

Understanding the genetic profile of nasopharyngeal cancer by the related cancer gene atlas has been instrumental in identifying potential biomarkers and therapeutic targets, which can ultimately improve diagnosis and treatment outcomes for patients. In addition to TCGA, there are several other initiatives and resources that focus on cancer genomics and may provide insights specifically related to NPC. International Cancer Genome Consortium (ICGC) is an international collaboration that aims to catalogue genomic abnormalities in various types of cancers, including NPC.[49] It facilitates data sharing and collaborative research among scientists worldwide. Catalogue of Somatic Mutations in Cancer (COSMIC) is a database that catalogues somatic mutations and other genetic alterations in human cancers, including NPC.[50] It provides detailed information on mutations, copy number variations and gene fusions found in cancer samples. GEO is a public repository that archives and freely distributes high-throughput gene expression and molecular abundance data, including data related to NPC.[7] Researchers can access datasets to explore gene expression patterns, biomarkers, and potential therapeutic targets in NPC. Although not specific to cancer, Encyclopaedia of DNA Elements (ENCODE) provides valuable genomic data on regulatory elements and gene expression across different tissues, which can be informative for understanding gene regulation in NPC.[51]

A majority of the research did, however, point out some limits of the bioinformatics tool in addition to its clear benefits in the identification of particular biomarkers. First, the small sample size decreased statistical power, particularly due to

the tiny patient populations in some subgroups, which led to variable results for various subgroups.[47] Second, there were no body fluid-related data included in the study to reflect vimentin's diagnostic significance as a non-invasive marker for NPC and no animal tests to confirm and investigate the biological activities of vimentin in NPC.[32] Third, a thorough and systematic investigation of gene expression and epigenetic modifications was still absent.[28] This might be beneficial i0n offering insight on the progression of cancer and identifying possible biomarkers, particularly for NPC. Because of the substantial infiltration of immune and stromal cells into the tumour, analysis without sufficient control of tumour composition may result in bias or false negative. Because most studies lack this information, future studies should specify the tumour analysis based on the section of tissue used in order to standardise the percentage of cancer cells being analysed.

While many bioinformatics tools are versatile and can be applied to multiple cancer types, some tools are more specialised or more frequently used in certain cancers due to the specific needs of those cancers. Tools like Genome Analysis Toolkit (GATK) and MuTect2 are fundamental for variant analysis in NPC as well as other cancers.[52] GATK facilitates the identification of mutations, insertions and deletions in the genome, whereas MuTect2, a GATK tool, aids in the identification of somatic point mutations and indels in tumour-normal pairs that are utilised in NPC genomic investigations. The choice of tools for transcriptomic and proteomic analysis might vary depending on the specific cancer type and the focus of the study. Comprehensive databases like TCGA provide a wealth of data for multiple cancers, facilitating cross-cancer comparisons and integrative analyses. TCGA provides comprehensive genomic data for various cancers such as breast, lung, and colorectal cancers. Tools like cBioPortal and OncoLnc make it easier to access and analyse this data.[53,54] In contrast, clinical trial databases and survival analysis tools are universally important across all cancer types for evaluating treatment outcomes and patient prognosis. The application of bioinformatic tools in NPC allows for the detection of gene fusions (which are important in NPC because they may influence tumour development), the identification of recurrent copy number variations, the provision of access to NPC genomic data from large-scale studies such as TCGA for thorough analysis, and the analysis of survival data based on clinical factors and molecular profiles in NPC patients. In other cancer types, bioinformatic tools identify somatic mutations (which are important for comprehending the burden of mutations and possible driver mutations), analyse protein-protein interaction networks (which are vital for comprehending intricate interactions in signalling pathways), and provide regulatory element data across a range of tissues, which helps comprehend gene regulation and epigenetic modifications.

Although these integrative bioinformatics tools may provide better knowledge and understanding, several key aspects of these tools need further verification. While multidimensional analysis and data comparison can be achieved through the selection of datasets and their sources, variations in the split points and datasets collected may lead to substantially distinct conclusions. There may be limitations stemming from the limited utilisation of chosen databases from GEO, in addition to the variability in sample sizes among researchers. Since the genetics of NPC is poorly known, improved treatment strategies like targeted therapy and immunotherapy are desperately needed.[28,55] Diverse NPC subtypes exhibited distinct genomic modifications, with undifferentiated NPC exhibiting a high incidence of mutations. Confirming the expression and function of the identified hub DEGs in NPC requires additional research. Computational cancer research may benefit greatly from the traditional methods of statistics and bioinformatics for the analysis of biological sequences, large-scale OMIC data sets, the genome and protein three-dimensional structure.[56] It is envisioned that advances in systems biology and cancer bioinformatics will enhance therapeutic design, prevention and diagnostics.

Although bioinformatics has revolutionised our understanding of many aspects of NPC, there are several limitations and challenges in using bioinformatics to fully elucidate its pathogenesis. NPC pathogenesis involves intricate interactions between genetic, epigenetic and environmental factors. Integrating data from multiple omics platforms (genomics, transcriptomics, proteomics) to capture this complexity is challenging. Bioinformatics tools often struggle with integrating diverse datasets to provide a comprehensive understanding of NPC biology. Access to large, well-annotated datasets of NPC samples is limited. Small sample sizes and heterogeneity among patient populations (e.g., ethnic diversity, Epstein-Barr virus status) can complicate bioinformatics analyses and limit the generalisability of findings. Bioinformatics often identifies associations or correlations between molecular features and NPC pathogenesis. However, validating these findings experimentally (in vitro or in vivo) to understand functional relevance and causality is resource-intensive and time-consuming. NPC pathogenesis is dynamic, involving evolving interactions between tumour cells and the microenvironment over time. Bioinformatics analyses capture snapshots of molecular profiles, but understanding temporal changes and dynamic interactions requires sophisticated computational models and longitudinal data. NPC pathogenesis involves not only genomic alterations but also epigenetic modifications, protein interactions, metabolic changes and immune responses. Bioinformatics tools primarily focus on genomics and transcriptomics, potentially overlooking other crucial aspects of NPC biology. NPC incidence varies significantly across populations, with distinct genetic and environmental risk factors. Bioinformatics analyses may not fully account for ethnic or geographic variability, limiting the applicability of findings to diverse populations. NPC pathogenesis involves complex networks of molecular interactions. Bioinformatics approaches, while powerful, often simplify these networks into linear pathways or regulatory modules, potentially oversimplifying the true complexity of NPC biology.

CONCLUSION

Bioinformatics methods have been extensively employed in the research of NPC to characterise genomic changes, signalling networks, differentially expressed genes (DEGs)

and cellular heterogeneity. These findings have improved our understanding of the molecular mechanisms behind NPC and have the potential to stimulate the development of novel therapeutic strategies. The application of bioinformatics tools has significantly improved patient outcomes, made it simpler to assess complex data, find potential biomarkers and therapy targets. The development of bioinformatics technology has facilitated the search for potential prognostic biomarkers for NPC. With a greater understanding of the complexity of tumour biology and potential molecular pathways provided by the bioinformatics integration analysis, improved strategies for early detection, outcome prediction, disease recurrence detection and therapeutic approaches for NPC will be possible.

ACKNOWLEDGMENTS

This work was funded by Universiti Sains Malaysia (USM) Research University Grant with project code: RUI/1001/PPSP/8012245.

REFERENCES

1. Chua MLK, Wee JTS, Hui EP, Chan ATC. Nasopharyngeal carcinoma. Lancet 2016; 387: 1012-24.
2. Lo KW, Huang DP. Genetic and epigenetic changes in nasopharyngeal carcinoma. Semin Cancer Bio 2002; 12: 451-62.
3. Brenner C. Applications of bioinformatics in cancer. Cancers 2019; 11: 1630.
4. Olsen LR, Campos B, Barnkob MS, Winther O, Brusic V, Andersen MH. Bioinformatics for cancer immunotherapy target discovery. Cancer Immunol Immunother 2014; 63: 1235-49.
5. Bayat A. Science, medicine, and the future: Bioinformatics. BMJ 2002; 324: 1018–22. https://doi.org/10.1136/bmj.324.7344.1018.
6. Xia X. Bioinformatics and drug discovery. Curr Top Med Chem 2017; 17: 1709-26.
7. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res 2012; 41: D991-5.
8. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007; 23: 1846-7.
9. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 2004; 32: 258D-261.
10. Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000; 28: 27-30.
11. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43: D447-52.
12. Gene Expression Profiling Interactive Analysis. [cited December 2022] Accessed from: http://gepia.cancer-pku.cn
13. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009; 151(4): 264-9.
14. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008; 9: 559. https://doi.org/10.1186/1471-2105-9-559.
15. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003; 4: P3.
16. Holt ME, Mittendorf KF, LeNoue-Newton M, Jain NM, Anderson I, Lovly CM, et al. My cancer genome: coevolution of precision oncology and a molecular oncology knowledgebase. JCO Clin Cancer Inform 2021: 995-1004.
17. Chen F, Zhang Y, Li X, Gao J, Ma H, Wang X, et al. The prognostic value of deficient mismatch repair in Stage II–IVa nasopharyngeal carcinoma in the era of IMRT. Sci Rep 2020; 10: 9690.
18. Bruce JP, To K-F, Lui VWY, Chung GTY, Chan Y-Y, Tsang CM, et al. Whole-genome profiling of nasopharyngeal carcinoma reveals viral-host co-operation in inflammatory NF-κB activation and immune escape. Nat Commun 2022; 13: 4353.
19. Chen H, Shi X, Ren L, Wan Y, Zhuo H, Zeng L, et al. Screening of core genes and prediction of ceRNA regulation mechanism of circRNAs in nasopharyngeal carcinoma by bioinformatics analysis. Pathol Oncol Res 2023; 29: 1610960.
20. Lin DC, Meng X, Hazawa M, Nagata Y, Varela AM, Xu L, et al. The genomic landscape of nasopharyngeal carcinoma. Nat Genet 2014; 46: 866-71.
21. Jiang X, Feng L, Dai B, Li L, Lu W. Identification of key genes involved in nasopharyngeal carcinoma. Braz J Otorhinolaryngol 2017; 83: 670-6.
22. Sengupta S, Den Boon JA, Chen I-H, Newton MA, Dahl DB, Chen M, et al. Genome-wide expression profiling reveals EBV-associated inhibition of MHC class I expression in nasopharyngeal carcinoma. Cancer Res 2006; 66: 7999-8006.
23. Guo W, Zheng X, Hua L, Zheng X, Zhang Y, Sun B, et al. Screening and bioinformatical analysis of differentially expressed genes in nasopharyngeal carcinoma. J Cancer 2021; 12: 1867-83.
24. Gibson G. Microarrays in ecology and evolution: a preview. Molecul Ecol 2002; 11: 17–24.
25. Janvilisri T. Omics-based identification of biomarkers for nasopharyngeal carcinoma. Dis Markers 2015; 2015: 1-10.
26. Pan X, Liu J-H. Identification of four key biomarkers and small molecule drugs in nasopharyngeal carcinoma by weighted gene co-expression network analysis. Bioengineered 2021; 12: 3647-61.
27. Shuai M, Hong J, Huang D, Zhang X, Tian Y. Upregulation of circRNA_0000285 serves as a prognostic biomarker for nasopharyngeal carcinoma and is involved in radiosensitivity. Oncol Lett 2018; 16(5): 6495-501.
28. Chen Y, Zhou C, Li H, Li H, Li Y. Identifying key genes for nasopharyngeal carcinoma by prioritized consensus differentially expressed genes caused by aberrant methylation. J Cancer 2021; 12: 874-84.
29. Zhu H, Fei Q, Qian L, Liu B, He X, Yin L. Identification of key pathways and genes in nasopharyngeal carcinoma using bioinformatics analysis. Oncol Lett 2019; 17(5): 4683-94.
30. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012; 16: 284-7.
31. Ye Z, Wang F, Yan F, Wang L, Li B, Liu T, et al. Identification of candidate genes of nasopharyngeal carcinoma by bioinformatical analysis. Arch Oral Biol 2019; 106: 104478.
32. Lu W, Luo J, Wu M, Hou J, Yang X, Chen G, et al. Expression of vimentin in nasopharyngeal carcinoma and its possible molecular mechanism: a study based on immunohistochemistry and bioinformatics analysis. Pathol Res Pract 2019; 215: 1020-32.
33. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004; 6: 1-6.
34. Tsao SW, Tsang CM, To KF, Lo KW. The role of Epstein-Barr virus in epithelial malignancies. J Pathol 2015; 235(2): 323-33.
35. Kouvidou C, Kanavaros P, Papaioannou D, Stathopoulos E, Sotsiou F, Datseris G, et al. Expression of bcl-2 and p53 proteins in nasopharyngeal carcinoma. Absence of correlation with the presence of EBV encoded EBER1-2 transcripts and latent membrane protein-1. Molecul Pathol 1995; 48: M17-22.
36. Si Y, Xu J, Meng L, Wu Y, Qi J. Role of STAT3 in the pathogenesis of nasopharyngeal carcinoma and its significance in anticancer therapy. Front Oncol 2022; 12: 1021179.

37. Chen X, Liang R, Lai L, Chen K, Zhu X. Prognostic role of EGFR/p-EGFR in patients with nasopharyngeal carcinoma: a meta-analysis. Front Oncol. 2021; 11: 697369.
38. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. Lancet 2019; 394(10192): 64-80.
39. Wang F, Peng L, Wang Y, Liu X. A meta-analysis of vascular endothelial growth factor for nasopharyngeal cancer prognosis. Front Oncol 2018; 8: 486.
40. Lo KW, To KF, Huang DP. Focus on nasopharyngeal carcinoma. Cancer Cell. 2004; 5(5): 423-8.
41. Ma J, Chen R, Xu Y, Du X, Liu N, Li Y, et al. CXCL12 genetic variants as prognostic markers in nasopharyngeal carcinoma. Onco Targets Ther 2015: 8: 2835-42.
42. Sun Z, Wang X, Wang J, Wang J, Liu X, Huang R, et al. Key radioresistance regulation models and marker genes identified by integrated transcriptome analysis in nasopharyngeal carcinoma. Cancer Medicine 2021; 10: 7404-17.
43. Xue K, Cao J, Wang Y, Zhao X, Yu D, Jin C, et al. Identification of potential therapeutic gene markers in nasopharyngeal carcinoma based on bioinformatics analysis. Clinical Translational Sci 2020; 13: 265-74.
44. Huang D, Liu Q, Zhang W, Huang C, Zheng R, Xie G, et al. Identified IGSF9 association with prognosis and hypoxia in nasopharyngeal carcinoma by bioinformatics analysis. Cancer Cell Int 2020; 20: 498.
45. Wang Y, Zhao Q, Lan N, Wang S. Identification of methylated genes and miRNA signatures in nasopharyngeal carcinoma by bioinformatics analysis. Mol Med Report 2018; 4909-16.
46. Zheng H, Dai W, Cheung AKL, Ko JMY, Kan R, Wong BWY, et al. Whole-exome sequencing identifies multiple loss-of-function mutations of NF-κB pathway regulators in nasopharyngeal carcinoma. Proc Natl Acad Sci USA 2016; 113: 11283-8.
47. Singer J, Irmisch A, Ruscheweyh H-J, Singer F, Toussaint NC, Levesque MP, et al. Bioinformatics for precision oncology. Brief Bioinform 2019; 20: 778-88.
48. Deng JL, Xu Y, Wang G. Identification of potential crucial genes and key pathways in breast cancer using bioinformatic analysis. Front Genet 2019; 10: 695.
49. Wang Y, He Y, Duan X, Pang H, Zhou P. Construction of diagnostic and prognostic models based on gene signatures of nasopharyngeal carcinoma by machine learning methods. Transl Cancer Res 2023; 12(5): 1254-69.
50. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue of somatic mutations in cancer. Nucleic Acids Res 2019; 47: D941–7.
51. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489(7414): 57-74.
52. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20: 1297-303.
53. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013; 6.
54. Zhao Q, Li H, Li W, Guo Z, Jia W, Xu S, et al. Identification and verification of a prognostic signature based on a miRNA–mRNA interaction pattern in colon adenocarcinoma. Front Cell Dev Biol 2023; 11: 1161667.
55. Abdullah B, Alias A, Hassan S. Challenges in the management of nasopharyngeal carcinoma: a review. Malays J Med Sci 2009; 16: 50-4.
56. Wu D, Rice CM, Wang X. Cancer bioinformatics: A new approach to systems clinical medicine. BMC Bioinformatics 2012; 13: 71.