Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Transcriptomic Profiling of Highly Expressed Genes Stratifies Nasopharyngeal Carcinoma Progression and Highlights Transferrin Receptor Dysregulation: A Bioinformatic Approach

Coia Romeu

Otolaryngology Department. Sant Joan Reus Hospital. University Rovira i Virgili. Tarragona. Spain.

Maria-Jose Oliach

Independent Researcher. Tarragona. Spain

Antonio R. Romeu

antonioramon.romeu@iubilo.urv.cat

University Rovira i Virgili. Tarragona. Spain.

Research Article

Keywords: Nasopharyngeal Carcinoma, GDS3341, Bioinformatic Stratification, Transferrin Receptor, Iron Metabolism, PCA

Posted Date: June 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9935130/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Transcriptomic Profiling of Highly Expressed Genes Stratifies Nasopharyngeal Carcinoma Progression and Highlights Transferrin Receptor Dysregulation: A Bioinformatic Approach

Coia Romeu[1], Maria-Jose Oliach[2], Antonio R. Romeu[3,4]

1. Otolaryngology Department. Sant Joan Reus Hospital. University Rovira i Virgili. Tarragona. Spain.
2. Independent Researcher. Tarragona. Spain
3. University Rovira i Virgili. Tarragona. Spain.
4. Corresponding author. Email: antonioramon.romeu@iubilo.urv.cat

Abstract

Background: Nasopharyngeal carcinoma (NPC) is a malignant epithelial tumor with complex pathogenesis and high metastatic potential. Disruption of iron metabolism is a recognized hallmark of cancer progression, yet its stage-specific transcriptomic landscape in NPC requires deeper elucidation. This study aims to identify highly expressed gene signatures that classify NPC progression and characterize the dysregulation of the transferrin receptor (TFRC) pathway.
Methods: Publicly available microarray expression data from the NCBI GEO DataSet Record GDS3341 —comprising laser-captured microdissected NPC tissues across different developmental stages and healthy controls— were analyzed. Using using custom R scripts highly expressed and differentially expressed genes were identified. Principal Component Analysis (PCA) and hierarchical clustering (dendrograms) were performed to evaluate sample stratification. Targeted expression analysis was conducted on the TFRC and associated iron-homeostasis genes.
Results: PCA and hierarchical clustering using the identified high-expression gene signatures clearly segregated healthy controls from NPC samples. Crucially, the TFRC exhibited significant, dysregulation highlighting an enhanced iron-acquisition phenotype in tumor cells. Paradoxically, the remaining genes driving the endosomal transport cycle maintained remarkably low and unaltered expression profiles across all stages.
Conclusion: Our bioinformatic analysis demonstrates that stage-specific transcriptomic signatures can precisely classify NPC development. Our findings reveal a distinct molecular uncoupling between TFRC and downstream transport machinery in NPC. This asymmetric phenotype suggests that the basal endosomal apparatus possesses sufficient functional recycling capacity to handle increased receptor trafficking without transcriptional upregulation.

Keywords: Nasopharyngeal Carcinoma; GDS3341; Bioinformatic Stratification; Transferrin Receptor; Iron Metabolism; PCA.

Inrouction

In 1964 Michael Anthony Epstein and Yvonne Barr discovered the Epstein-Barr virus (EBV) in Burkitt lymphoma cells [1]. It is one of the most common viruses in humans [2,3].with high prevalence in Southeast Asia [4] ; it is estimated that more than 95% of the world's adult population live with the EBV asymptomatically, it is seropositive carrying the virus for life The key points of EBV biology could be summariced as: (*i*) Once the primary infection occurs, the virus persists in the body in a latent (dormant) state, it is permanent (lifelong) [5,6] (*ii*) EBV primarily infects the epithelial cells of the oropharynx and, fundamentally, establishes latency in memory B lymphocytes, remaining hidden from the immune system as episome in the nucleus of these lymphocytes [7,8]; and (*iv*) the ctive replication (lytic cycle) occurs in the epithelial cells of the oropharynx (throat), however, based on a "dual tropism" cycle [9,10] the epithelial nasopharyngeal tissue remains healthy [11,12].

The EBV is associated with the development of certain types of cancer, such as Burkitt lymphoma and nasopharyngeal carcinoma (NPC) [2,13-18]. It is extremely important to note the enormous contrast between the prevalence of the virus and the incidence of the NPC. Based on the most recent epidemiological data (2022-2026), NPC is a rare cancer worldwide, but with a highly uneven geographical distribution. Current Global Statistics estimated 130,000 new cases of nasopharyngeal cancer occur annually worldwide [19].

Applying Bayes' Theorem is the best way to understand the relationship between the EBV and NPC. That is, that the virus alone does not explain cancer. The theorem deals with contrasts like these, for example the following: (*i*) given an elephant, the probability that it has four legs is high, but given a four-legged animal, the probability that it is an elephant is low; or (*ii*) if it rains, the probability of the ground being wet is high, but if the ground is wet, the probability that it has rained is not always high.

Here's how the "four-legged" logic fits into the NPC clinical reality, let (P) be the conditional probability and the following events:

- **Event A**, having the EBV.
- **Event B**, developing NPC

P(A|B) (the probability of **A** given that **B** has already happened) is very high: having the carcinoma (**B**), the probability that the virus is present (**A**) is almost 100%.

P(B|A) (the probability of **B** given that **A** has already happened) is extremely low: having the virus (**A**) —like 95% of humanity— the probability of developing the cancer (**B**) is minuscule (approximately 1 in 100,000 people per year).

Going on with this bayesian reasoning, for the conditional probability **P(A|B)** to increase, other factors must come into play. EBV only becomes oncogenic when combined with: (*i*) genetics, it is very common in populations of southern China and Southeast Asia; (*ii*) diet, consuming salted fish (with nitrosamines) irritates the mucous membrane; and (*iii*) environment, smoking or inhalation of chemical dusts. The relationship EBV and NPC is a classic example of how a common infectious agent can act as a necessary but not sufficient factor for the development of cancer [20-24]

In NPC, the virus and tumor cells develop together. The EBV undergoes an "abortive lytic cycle" and viral replication does not end with cell’s burst [25-27]. In EBV replication, the virus activates

only the first genes of the lytic cycle that allows the cancer cell to survive, but with the benefits of proliferation and immune evasion provided by the expression of viral lytic proteins [28,29]. These viral genes, instead of destroying the cell, send chemical signals that stimulate cell proliferation and the growth of blood vessels (angiogenesis) stimulating tumor growth and the virus spreads with it [30-33].

Within this theoretical framework, it is very valuable the NCP gene expression record (GDS3341) of the NCBI Gene Expression Omnibus (GEO) free database. It shows arrays data with complete coverage of the Human Genome of laser-captured, microdissected NPC tissue samples from Taipei, Taiwan, patiens and healthy nasopharyngeal tissue samples [34-36]. Based on a bioinformatic approach we were interested in exploring the genomic analysis further to discover new genes or gene expression associations to add to the current NPC knowledge. The aim of this work is to show, a longitudinal study of the highly expressed genes in key stages of tumor development, focusing on Gene Ontology (GO) enrichment analysis to elucidate functional roles of the NPC differentially expressed genes.

Methodology

The source of information was the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database, DataSet Record GDS3341 titled “Nasopharyngeal carcinoma”, Series GSE12452 the overall design consist in total RNA extracted from laser-captured epithelium from 31 nasopharyngeal carcinomas and 10 normal healthy nasopharyngeal tissue specimens. Platforms GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array for mRNA expression profiling. The GPL570 platform has 54,675 probe sets with complete coverage of the Human Genome.

Data acquisition and processing were conducted using custom R scripts (R version 4.6.0, 2026-04-24 ucrt), packages: base, hash and FactoMineR through the open-source Integrated Development Environment (IDE) Rstudio version: 2026.04.0+526. A custom R pipeline was developed to automate the download the Series GSE12452 samples of mRNA expression profiling extracted from laser-captured epithelium from 31 NPCs and 10 normal healthy nasopharyngeal tissue specimens. Table 1 shows the experimental design of the DataSet Record GDS3341 and the definition of variables used in this work. Table 2 shows the The American Joint Committee on Cancer (AJCC) 8th Edition Cancer Staging Manual standard PNC classification based on the primary tumour (T) and the spread to lymph nodes (N) extent.

The rationale of this work was based on the following tasks:

- Download all the data from the platform GPL570, in order to access gene expression data from the platform probes or from gene lists. The working tables were constructed by compiling information on the probe identifier and its corresponding associated gene in terms of NCBI Entrez id, Gene Symbol and Gene Title.
- Download the 41 individual tables of the Series GSE12452 samples. Expression units are normalised ($\log_2$) values that stand for the hybridisation signal between the RNA content of the applied sample and the platform probe.
- Creation of a dataset of all GDS3341 record. It was a *data.frame* wehre the rows number was the GPL570 platform probes (54,675) and as columns number are variables (V1-V41). From this *data.frame* all the partial *data.frames* were are generated to address the objectives of this work.

- Clustering analysi. Principal Component Analysis (PCA) and Hierarchical clustering analysis (dendrogram) to determine if gene expression patterns clearly differentiate between control and cancer groups.
- Starting from a list of genes, a functional analysis was performed using the PANTHER tool for Gene Ontology (GO) [37,38] to identify the three main GO categories (Biological Process, Molecular Function, Cellular Component), highlighting novel ontologies in NPC groups.
- The parametric Student's t-test is used to compare the means of two independent groups (control and cancer) and determine if the differences are statistically significant ($p < 0.05$). In R, the t.test() function (base package) assumed by default that variances were unequal.

Affymetrix GPL570 Platform key points: (*i*) the intensity of the hybridisation signal is proportional to the amount of mRNA present in the sample; (*ii*) if the amount of mRNA is extremely high, the probe can become "saturated"; (*iii*) not all probes hybridise with the same strength, the content of cytosine and guanine influences the intensity (two genes with the same amount of mRNA may show different intensity values); (*iv*) each probe cell contains millions of identical copies of a 25-base oligonucleotide; thus preventing early saturation; (*v*) each gene is usually represented by more than one probe set, targeting different regions of the same mRNA and increasing the statistical robustness; and (*vi*) It is possible that the value of a probe is not the result of actual hybridization with mRNA, but is an artifact (cross-hybridization, background noise, mismatches).

Results and dicussion

Based on the experimental design of the DataSet Record GDS3341, the working groups described the NPC stages development: Control (n=10), T1N0 (n=6), T1N1 (n=10), T2N1 (n=5), T2N2 (n=6) and T3N2 (n=2). Numeric variables were the NCBI GEO dataset Samples (Table 1). Total 39 variables (V1-V39). The cases were the 54,675 GPL570 Platform probe sets.

For each group and for each probe set the mean expression value was calculated. Table 3 shows the mean quantiles. The expression values range from approximately 2 to 14 in $\log_2$ arbitrary units . The minimum and maximum values were the 0% and 100% percentiles, respectively. The Q1 percentile was the expression value below which 25% of the population lies. The Q2, the 50%. The Q3, the 75%. The p50 percentile (Q2) was also the median. Because groups T3N0 and T3N1 only had one Sample (Table 1), they could not be included in the statistical analysis.

At this point, It is important to keep in mind that cell nucleus contains all the genes encoded in the genome, but the cell only expresses the genes it needs. Like the cell nucleus, the GPL570 Platform also has full coverage of the human genome, but in a given biological sample applied to the Platform doesn't contain all the genes either (like in a cell). However, due to artifacts, all the platform's probes, however weak, produce expression value, in most cases this is not due to actual hybridisation with mRNA. There was no threshold probe value that marks the boundary between real gene expression or artifact. Because high expressed gene contribute to high amounts of transcript mRNA to the cell cytoplasm, it is in all likelihood, that high probe values will be actual gene expression rates. To minimize the probability of artifacts, in this work we focused only on highly expressed genes. So that, the selected probes for each group were those whose expression value was greater or equal (>=) the mean 95th percentile (95%) (Table 3).

Based on the probe value of the high expressed genes the PCA and the hierarchical clustering analysis (dendrogram) clearly show the difference between the control group and each NPC

development stages groups (Figure 1 and 2). This strongly suggest a significant gene expression dysregulation in NPC. To elucidate the functional roles of the NPC differentially expressed genes Gene Ontology (GO) enrichment analysis was performed. Table 4 shows the lists of the top-50 most highly expressed genes of each experimental group that were used as the primary input dataset for subsequent Gene Ontology (GO) functional enrichment analysis.

Notably, thet analysis revealed a significant enrichment in the “Cargo Receptor Activity (GO:0038024)” Molecular Function Ontology in NPC groups, which was not observed in the control group. Genes associated with this ontology mediate endocytosis and vesicular transport. Several genes are associated with this ontology, but in the present GO enrichment analysis, the gene that gave rise to this “Cargo Receptor Activity (GO:0038024)” ontology was the TFRC (Transferrin Receptor, Gene ID 7037) Enhanced TFRC-mediated iron uptake is a well-documented hallmark of tumor progression, where malignant tumors have an abnormally high demand for iron to grow and proliferate [39-41].

Consistently, we found that TFRC gene expression values statistically significant increased in all NPC groups in comparison to the control group. Table 5 shows the detailed statistical analysis of the TFRC specific Affymetrix GPL570 Platform “208691_at” probe set expression values. Figure 4A highlights the significant increase of the TFRC mean expression mean values. At this point, it is important to emphasise the enormous increase of TFRC gene expression in NPC, as shown by the p-values (*e.g.,* 3.17e-06, in T1N0). The TFRC, expression values ranged 10-12. This numerical data are not linear; It is on a logarithmic (log_2) scale. Values 10-12 fall within the critical inflection point region. The increase in expression is enormous. To clarify, Figure 4B shows the gene expression values for the 54,675 probes on the platform highlighting the inflection point.

These results raise some questions The transferrin receptor does not act alone; it is embedded in a cycle in which the iron from transferrin is taken up by the carcinoma cell, and the transferrin, or apotransferrin, is returned to the bloodstream [42-44]. At this point, the question arises: are the other genes in the cycle also upregulated? The transferrin receptor would not be functional without the participation of all the elements of the cycle. Briefly, iron acquisition in NPC cells relies on a receptor-mediated endocytosis cycle. Transferrin binds to its receptor (TfR1), and dynamins facilitate the formation of a clathrin-coated vesicle. An ATP-dependent proton pump acidifies the endosomal compartment, promoting iron release (ATP6V, ATPase H+ transporting, Gene ID 523, 525, ...). Subsequently, the iron is reduced (*STEAP3*, STEAP family member 3 metalloreductase, Gene ID 55240) and transported into the cytosol via the divalent metal transporter 1 (*SLC11A2*, solute carrier family 11 (proton-coupled divalent metal ion transporter) DMT1, Gene ID 4891) [45,46]. Detailed analysis of the NPC effect on the expression of these genes involved in this iron uptake cycle showed no significant differences between NPC groups and the control group. Table 6 shows the detailed results for each experimental group. These results contrast sharply with those obtained showing a significant effect of cancer on TFRC expression (Table 5). As further evidence, PCA analyses based on gene expression values of these transferrin iron uptake pathway genes, even including TFRC gene expression values, did not distinguish the NPC groups from the control group (Figure 5).

Our analysis revealed a robust and statistically significant overexpression of the TFRC across all developmental stages of NPC compared to healthy controls, a finding that perfectly aligns with its established role in satisfying the elevated iron demands of rapidly proliferating malignant cells [40,41,47]. Intriguingly, however, the expression levels of genes encoding core structural and functional components of the clathrin-mediated endosomal internalization machinery —including dynamins, clathrin, endosomal reductases, and vacuolar ATPases— remained remarkably low and

unchanged relative to control tissues. This apparent discrepancy suggests that the baseline endosomal apparatus in nasopharyngeal epithelial cells possesses sufficient operational and recycling capacity to accommodate increased receptor trafficking without requiring a coordinated transcriptional upregulation. Alternatively, this phenotypic divergence underscores that TFRC regulation in NPC may be tightly governed by post-transcriptional mechanisms, such as the Iron Regulatory Protein / Iron Responsive Element (IRP/IRE) network, which selectively stabilizes TFRC transcripts under conditions of metabolic rewiring, without altering the transcription of general vesicular trafficking genes [48]. Furthermore, these findings support the emerging paradigm that overexpressed TFRC exerts non-canonical, iron-independent oncogenic functions in solid tumors. Specifically, it has been demonstrated that TFRC acts as a signaling scaffold that directly drives malignant progression and epithelial-mesenchymal transition (EMT) in NPC through the hyperactivation of survival cascades, such as the PI3K/Akt/mTOR signaling pathway [49]. Thus, the isolated surge in TFRC expression observed across all NPC stages likely reflects both its post-transcriptional stabilization and its secondary role as an oncogenic driver independent of the baseline endosomal import cycle.

Conclusions

- PCA and Hierarchical Clustering Analysis of NPC development stages, based on high-expressed genes expression value, clearly differentiates the carcinoma groups from the control group, highlighting a widespread effect on global transcriptomic.
- The robust overexpression of the TFRC is a consistent, early-onset feature maintained across all developmental stages of NPC, validating its potential as a reliable biomarker.
- Unlike TFRC, the core genes driving clathrin-mediated endocytosis, vesicle acidification, and iron liberation do not undergo transcriptional upregulation in NPC.
- The phenotypic divergence between isolated TFRC surges and stable endosomal gene profiles strongly suggests that TFRC upregulation in NPC extends beyond metabolic iron assimilation.

Competing interests

The authors declare no competing interest

Acknowledgements

This work has not been awarded grants by any Institution. The authors acknowledges NCBI GEO database contributors

References

1. Michael A. Epstein, Yvonne Barr. Irus particles in cultured lymphoblasts from burkitt's lymphoma. Lancet, 1(7335):702-3. 1964. PMID: 14107961. doi: 10.1016/s0140-6736(64)91524-7.
2. Paul J Farrell. Epstein-Barr Virus and Cancer. Annu Rev Pathol 14:29-53. 2019. .PMID: 30125149. doi: 10.1146/annurev-pathmechdis-012418-013023.
3. Samantha S. Soldan, Paul M. Lieberman. Epstein-Barr virus and multiple sclerosis. Nat Rev Microbiol. 21(1):51-64, 2023. PMID: 35931816. doi: 10.1038/s41579-022-00770-5.

4. Yudi Xiong, Mengting Yuan, Zhigang Liu, Jing Huang, Jianping Bi, Guoliang Pi, Ying Li, Yanping Li, Hanping He, Vivek Verma, Suqing Tian, Guang Han. Long-Term Outcomes of Nasopharyngeal Carcinoma by Epstein–Barr Virus Status in the Chinese Population: A Multicenter Investigation. J Clin Med 12(8):3005. 2023. PMID: 37109338. doi: 10.3390/jcm12083005
5. S C Hung, M S Kang, E Kieff. Maintenance of Epstein-Barr virus (EBV) oriP-based episomes requires EBV-encoded nuclear antigen-1 chromosome-binding domains, which can be replaced by high-mobility group-I or histone H1. Proc Natl Acad Sci U S A 98(4):1865-70. 2001. PMID: 11172042 doi: 10.1073/pnas.98.4.1865.
6. Keith M. Haan, Ashok Aiyar, Richard Longnecker. Establishment of Latent Epstein-Barr Virus Infection and Stable Episomal Maintenance in Murine B-Cell Lines. J Virol. 75(6):3016–3020. 2001 .PMID: 11222728. doi: 10.1128/JVI.75.6.3016-3020.2001.
7. Blossom Damania, Shannon C Kenney, Nancy Raab-Traub. Epstein-Barr virus: Biology and clinical disease. Cell 185(20):3652-3670. 2022. PMID: 36113467 doi: 10.1016/j.cell.2022.08.026.
8. Koichi R. Katsumura, Seiji Maruo, Kenzo Takada. EBV lytic infection enhances transformation of B-lymphocytes infected with EBV in the presence of T-lymphocytes. J Med Virol 84(3):504-10. 2012. PMID: 22246839. doi: 10.1002/jmv.23208.
9. David A Thorley-Lawson. EBV Persistence—Introducing the Virus. Curr Top Microbiol Immunol. 2015;390(Pt 1):151–209. 2015. PMID: 26424647. doi: 10.1007/978-3-319-22822-8.
10. Britta S Möhl, Jia Chen, Karthik Sathiyamoorthy, Theodore S Jardetzky, Richard Longnecker. Structural and mechanistic insights into the tropism of Epstein-Barr virus. Mol Cells 39(4):286-91. 2016. PMID: 27094060. doi: 10.14348/molcells.2016.0066.
11. Claire Shannon-Lowe, Martin Rowe. Epstein Barr virus entry; kissing and conjugation. Curr Opin Virol Feb:4:78-84. 2014. PMID: 24553068. doi: 10.1016/j.
12. Lindsey M Hutt-Fletcher. The Long and Complicated Relationship between Epstein-Barr Virus and Epithelial Cells. J Virol. 91(1):e01677-16. 2016. PMID: 27795426. doi: 10.1128/JVI.01677-16.
13. D Liebowitz. Nasopharyngeal carcinoma: the Epstein-Barr virus association. Semin Oncol. 21(3):376-81. 1994. PMID: 8209269.12.
14. Lawrence S. Young, Alan B. Rickinson. Epstein-Barr virus: 40 years on. Nat Rev Cancer. 4(10):757-68. 2004. PMID: 15510157. doi: 10.1038/nrc1452.
15 . Margaret E. McLaughlin-Drubin and Karl Munger. Viruses Associated with Human Cancer. Biochim Biophys Acta. 1782(3): 127–150. 2008. PMID: 18190596. doi: 10.1016/j.bbadis.2007.12.005.
16. Lawrence S. Young, Christopher W. Dawson. Epstein-Barr virus and nasopharyngeal carcinoma. Chin J Cancer 33(12):581-90. 2014. PMID: 25418193 doi: 10.5732/cjc.014.10197.
17. Claire Shannon-Lowe, Alan Rickinson. The Global Landscape of EBV-Associated Tumors. Front Oncol 9:713. 2019. PMID: 31448229. doi: 10.3389/fonc.2019.00713. ECollection 2019.
18. Pok Man Hau, Hong Lok Lung, Man Wu, Chi Man Tsang, Ka-Leung Wong, Nai Ki Mak, Kwok Wai Lo. Targeting Epstein-Barr Virus in Nasopharyngeal Carcinoma. Front Oncol 10:600. 2020. PMID: 32528868. doi: 10.3389/fonc.2020.00600
19. Yuna Zhang, Shanshan Gu, Hongxia Deng, Zhisen Shen. Global epidemiological profile in nasopharyngeal carcinoma: a prediction study. BMJ Open 14(12):e091087. 2024. PMID: 39658299. doi: 10.1136/bmjopen-2024-091087.
20. M C Yu. Diet and nasopharyngeal carcinoma. Prog Clin Biol Res 346:93-105. 1990. PMID: 2197634. doi: 0.1016/0277-9536(90)90310-u.
21. Umesh T. Sankpal, Hima Pius, Moeez Khan, Mohammed I. Shukoor, Pius Maliakal, Chris M Lee, Maen Abdelrahim, Sarah F. Connelly, Riyaz Basha. Environmental factors in causing

human cancers: emphasis on tumorigenesis Tumour Biol 33(5):1265-1274. 2012. PMID: 22614680. doi: 10.1007/s13277-012-0413-4.
22. Ilona Argirion, Katie R. Zarins, Julie J. Ruterbusch, Patravoot Vatanasapt, Hutcha Sriplung, Erlene K. Seymour, Laura S Rozek. Increasing Incidence of Epstein-Barr Virus-Related Nasopharyngeal Carcinoma in the United States. Cancer. 2019 Sep 16;126(1):121–130. 2019. PMID: 31524955. doi: 10.1002/cncr.32517.
23. John Malcolm Nicholls, Victor Ho-Fun Lee, Sik-Kwan Chan, Ka-Chun Tsang, Cheuk-Wai Choi, Dora Lai-Wan Kwong, Ka-On Lam, Sum-Yin Chan, Chi-Chung Tong, Tsz-Him So, To-Wai Leung, Mai-Yee Luk, Pek-Lan Khong & Anne Wing-Mui Lee. Negative plasma Epstein-Barr virus DNA nasopharyngeal carcinoma in an endemic region and its influence on liquid biopsy screening programmes. British Journal of Cancer. 121:690–698. 2019. doi: 10.1038/s41416-019-0575-6.
24. Zhi-Gong Wei, Xiao-Lin Hu, Yan He, Hui Guan, Jing-Jing Wang, Ling He, Xiao-Li Mu, Zhe-Ran Liu, Rui-Dan Li, Xing-Chen Peng. Clinical and survival analysis of nasopharyngeal carcinoma with consistently negative Epstein-Barr virus DNA. Head Neck 43(5):1465-1475. 2021. PMID: 33421240. doi: 10.1002/hed.26608.
25. Marta M Gaglia, Karl Munger. More than just oncogenes: mechanisms of tumorigenesis by human viruses. Curr Opin Virol. Sep 27;32:48–59. 2018. PMID: 30268926. doi: 10.1016/j.coviro.2018.09.003.
26. Takayuki Murata, Atsuko Sugimoto, Tomoki Inagaki, Yusuke Yanagi, Takahiro Watanabe, Yoshitaka Sato, Hiroshi Kimura. Molecular Basis of Epstein–Barr Virus Latency Establishment and Lytic Reactivation. Viruses 13(12):2344. 2021. PMID: 34960613. doi: 10.3390/v13122344.
27. Xiaoting Xu,Nannan Zhu,Junming Zheng, Yingying Peng,Mu-Sheng Zeng,Kai Deng , Chaohui Duan, Yan Yuan. EBV abortive lytic cycle promotes nasopharyngeal carcinoma progression through recruiting monocytes and regulating their directed differentiation. PLOS Pathogens | January 11, 2024. PMID: 38206974. doi: 10.1371/journal.ppat.1011934
28. Sai Wah Tsao, Chi Man Tsang, Kwok Wai Lo. Epstein–Barr virus infection and nasopharyngeal carcinoma. Trans R Soc Lond B Biol Sci 372(1732):20160270. 2017. PMID: 28893937. doi: 10.1098/rstb.2016.0270.
29. Hao Li, Chih-Ying Lee, Henri-Jacques Delecluse. Epstein-Barr virus lytic replication and cancer. Curr Opin Virol Feb:70:101438. 2025. doi: 10.1016/j.coviro.2024.101438.
30. Chi Man Tsang, Wen Deng, Yim Ling Yip, Mu-Sheng Zeng, Kwok Wai Lo, Sai Wah Tsao. Epstein-Barr virus infection and persistence in nasopharyngeal epithelial cells. Chin J Cancer 33(11):549–555. 2014. PMID: 25223910. doi: 10.5732/cjc.014.10169.
31. Olivia L Hatton, Aleishia Arnold-Harris, Steven Schaffert, Sheri M Krams, Olivia M Martinez. The Interplay Between Epstein Barr Virus and B Lymphocytes: Implications for Infection, Immunity, Immunol Res. 2014 May;58(0):268–276. 2014. PMID: 24619311. doi: 10.1007/s12026-014-8496-1.
32. Quincy Rosemarie, Bill Sugden. Epstein-Barr Virus: How Its Lytic Phase Contributes to Oncogenesis. Microorganisms 8(11):1824. 2020. PMID: 33228078 doi: 10.3390/microorganisms8111824.
33. Janardhan Avilala, Hirotomo Dochi, Ramsy Abdelghani, David Becnel, Grace Maresh, Xin Zhang, Li Li, Mark Sides, Gilbert Morris, Hong Liu, Zhen Lin. Interleukin-17 directly triggers Epstein-Barr virus lytic reactivation in latently infected human B cells. bioRxiv 2025.12.19.695390; doi: 10.64898/2025.12.19.695390.
34. Lori E Dodd, Srikumar Sengupta, I-How Chen, Johan A. den Boon, Yu-Juen Cheng, William Westra, Michael A. Newton, Beth F. Mittl, Lisa McShane, Chien-Jen Chen, Paul Ahlquist, Allan Hildesheim. Genes involved in DNA repair and nitrosamine metabolism and those located on chromosome 14q32 are dysregulated in nasopharyngeal carcinoma. Cancer Epidemiol

Biomarkers Prev 15(11):2216-25. 2006. PMID: 17119049 doi: 10.1158/1055-9965.EPI-06-0455.
35. Srikumar Sengupta, Johan A. den Boon, I-How Chen, Michael A. Newton, David B. Dahl, Meng Chen, Yu-Juen Cheng, William H Westra, Chien-Jen Chen, Allan Hildesheim, Bill Sugden, Paul Ahlquist. Genome-wide expression profiling reveals EBV-associated inhibition of MHC class I expression in nasopharyngeal carcinoma. Cancer Res 66(16):7999-8006. 2006. PMID: 16912175. doi: 10.1158/0008-5472.
36. Wan-Lun Hsu, Ka-Po Tse, Sharon Liang, Yin-Chu Chien, Wen-Hui Su, Kelly J Yu, Yu-Juen Cheng, Ngan-Ming Tsang, Mow-Ming Hsu, Kai-Ping Chang, I-How Chen, Tzu-I Chen, Czau-Siung Yang, Alisa M Goldstein, Chien-Jen Chen, Yu-Sun Chang, Allan Hildesheim. Evaluation of human leukocyte antigen-A (HLA-A), other non-HLA markers on chromosome 6p21 and risk of nasopharyngeal carcinoma. PLoS One 7(8):e42767. 2012. PMID: 22880099 doi: 10.1371/journal.pone.0042767
37. Huaiyu Mi, Anushya Muruganujan, Xiaosong Huang, Dustin Ebert, Caitlin Mills, Xinyu Guo, Paul D Thomas. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat Protoc. 14(3):703-721. 2019. PMID: 30804569. doi: 10.1038/s41596-019-0128-8.
38. Gene Ontology. PAN-GO Human Functionome. Panther Classification System. Accessed 20/05/2026. https://functionome.geneontology.org/
39. Cheng Yang, Junqiang Li, Yongdong Guo, Dongxue Gan, Chao Zhang, Ronglin Wang, Lei Hua, Liaoliao Zhu, Peixiang Ma, Jingjie Shi, Shanshan Li, Haichuan Su. Role of TFRC as a Novel Prognostic Biomarker and in Immunotherapy for Pancreatic Carcinoma. Front Mol Biosci 9:756895. 2022. PMID: 35372510. doi: 10.3389/fmolb.2022.756895.
40. Ying Shen, Xin Li, Dandan Dong, Bin Zhang, Yanru Xue, Peng Shang. Transferrin receptor 1 in cancer: a new sight for cancer therapy. Am J Cancer Res 8(6):916–931. 2018. PMID: 30034931.
41. Shufen Wang, Xuyan He, Qian Wu, Li Jiang, Liyun Chen, Yingying Yu, Pan Zhang, Xin Huang, Jia Wang 3 , Zhenyu Ju, Junxia Min, Fudi Wang. Transferrin receptor 1-mediated iron uptake plays an essential role in hematopoiesis. Haematologica 105(8):2071-2082. 2020. PMID: 31601687. doi: 10.3324/haematol.2019.224899.
42. Juan Luis Steegmann-Olmedillas. The role of iron in tumour cell proliferation. Clin Transl Oncol 13(2):71-6. 2011. PMID: 21324793. doi: 10.1007/s12094-011-0621-1.
43. Elena Gammella, Paolo Buratti, Gaetano Cairo, Stefania Recalcati. The transferrin receptor: the cellular iron gate. Metallomics 9(10):1367-1375. 2017. PMID: 28671201. doi: 10.1039/c7mt00143f.
44. Stefania Forciniti, Luana Greco, Fabio Grizzi, Alberto Malesci, Luigi Laghi. Iron Metabolism in Cancer Progression- Int J Mol Sci 21(6):2257. 2020. PMID: 32214052. doi: 10.3390/ijms21062257.
45. A Dautry-Varsat. Receptor-mediated endocytosis: the intracellular journey of transferrin and its receptor. Biochimie 68(3):375-81. 1986. PMID: 2874839. doi: 10.1016/s0300-9084(86)80004-9.
46. Kristine M Mayle, Alexander M Le, Daniel T Kamei. The Intracellular Trafficking Pathway of Transferrin. Biochim Biophys Acta. 1820(3):264–281. 2011. PMID: 21968002. doi: 10.1016/j.bbagen.2011.09.009
47. Myriam Y Hsu, Erica Mina, Antonella Roetto, Paolo E Porporato. Iron: An Essential Element of Cancer Metabolism. Cells 9(12):2591. PMID: 33287315. doi: 10.3390/cells9122591.
48. Masaki Miyazawa, Alexander R Bogdan, Kazunori Hashimoto, Yoshiaki Tsuji. Regulation of transferrin receptor-1 mRNA by the interplay between IRE-binding proteins and miR-7/miR-141 in the 3′-IRE stem–loops. RNA 24(4):468–479. 2018. PMID: 29295890. doi: 10.1261/rna.063941.117.

49. Guofei Feng, Yasushi Arima, Kaoru Midorikawa, Hatasu Kobayashi, Shinji Oikawa, Weilin Zhao, Zhe Zhang, Kazuhiko Takeuchi, Mariko Murata. Knockdown of TFRC suppressed the progression of nasopharyngeal carcinoma by downregulating the PI3K/Akt/mTOR pathway Cancer Cell Int 23(1):185.PMID: 37644594. doi: 10.1186/s12935-023-02995-7.

Table 1. Experimental design of the NCBI GEO database DataSet Record GDS3341

Sample[1]	Comment[2]	Stage[2]	Variable[3]
GSM312896	normal nasopharyngeal tissue, specimen N1	control	V1
GSM312897	normal nasopharyngeal tissue, specimen N2	control	V2
GSM312898	normal nasopharyngeal tissue, specimen N3	control	V3
GSM312899	normal nasopharyngeal tissue, specimen N4	control	V4
GSM312900	normal nasopharyngeal tissue, specimen N5	control	V5
GSM312901	normal nasopharyngeal tissue, specimen N6	control	V6
GSM312902	normal nasopharyngeal tissue, specimen N7	control	V7
GSM312903	normal nasopharyngeal tissue, specimen N8	control	V8
GSM312904	normal nasopharyngeal tissue, specimen N9	control	V9
GSM312905	normal nasopharyngeal tissue, specimen N10	control	V10
GSM312906	nasopharyngeal carcinoma, specimen T1	T1N1	V11
GSM312907	nasopharyngeal carcinoma, specimen T2	T2N2	V12
GSM312908	nasopharyngeal carcinoma, specimen T3	T3N2	V13
GSM312909	nasopharyngeal carcinoma, specimen T4	T2N2	V14
GSM312910	nasopharyngeal carcinoma, specimen T5	T2N1	V15
GSM312911	nasopharyngeal carcinoma, specimen T6	T1N1	V16
GSM312912	nasopharyngeal carcinoma, specimen T7	T1N1	V17
GSM312913	nasopharyngeal carcinoma, specimen T8	T1N1	V18
GSM312914	nasopharyngeal carcinoma, specimen T9	T1N0	V19
GSM312915	nasopharyngeal carcinoma, specimen T10	T2N2	V20
GSM312916	nasopharyngeal carcinoma, specimen T11	T1N1	V21
GSM312917	nasopharyngeal carcinoma, specimen T12	T2N2	V22
GSM312918	nasopharyngeal carcinoma, specimen T13	T2N1	V23
GSM312919	nasopharyngeal carcinoma, specimen T14	T1N1	V24
GSM312920	nasopharyngeal carcinoma, specimen T15	T1N0	V25
GSM312921	nasopharyngeal carcinoma, specimen T16	T1N1	V26
GSM312922	nasopharyngeal carcinoma, specimen T17	T1N1	V27
GSM312923	nasopharyngeal carcinoma, specimen T18	T1N0	V28
GSM312924	nasopharyngeal carcinoma, specimen T19	T1N1	V29
GSM312925	nasopharyngeal carcinoma, specimen T20	T3N0	V30
GSM312926	nasopharyngeal carcinoma, specimen T21	T2N1	V31
GSM312927	nasopharyngeal carcinoma, specimen T22	T2N2	V32
GSM312928	nasopharyngeal carcinoma, specimen T23	T2N2	V33
GSM312929	nasopharyngeal carcinoma, specimen T24	T1N0	V34
GSM312930	nasopharyngeal carcinoma, specimen T25	T2N1	V35
GSM312931	nasopharyngeal carcinoma, specimen T26	T3N1	V36
GSM312932	nasopharyngeal carcinoma, specimen T27	T1N1	V37
GSM312933	nasopharyngeal carcinoma, specimen T28	T1N0	V38
GSM312934	nasopharyngeal carcinoma, specimen T29	T1N0	V39
GSM312935	nasopharyngeal carcinoma, specimen T30	T2N1	V40
GSM312936	nasopharyngeal carcinoma, specimen T31	T3N2	V41

[1]**:** Series GSE12452

[2]: The AJCC 8th Edition TNM staging system of the NPC defineing the extent of the primary tumour (T) and the spread to lymph nodes (N) (Table 2).

[3]**:** Definition of the variables used in this work.

Table 2. The American Joint Committee on Cancer (AJCC) 8th Edition Cancer Staging Manual standard PNC classification based on the primary tumour (T) and the spread to lymph nodes (N) extent

T	Defines the extent of the primary tumor (T)
N	Defines the spread to lymph nodes (N).

T1	The tumor is confined to the nasopharynx, or has extended to the oropharynx (throat behind the mouth) and/or the nasal cavity, without involving the parapharyngeal space.
T2	The tumor has extended into the parapharyngeal space (fatty tissue adjacent to the throat) and/or involves adjacent soft tissues, such as the medial pterygoid, lateral pterygoid, or prevertebral muscles.
T3	is defined as a tumor that has caused unequivocal infiltration of bony structures.

N0	No cancer cells are found in the nearby lymph nodes.
N1	Unilateral (one side of the neck) metastasis in cervical lymph node(s) and/or unilateral or bilateral metastasis in retropharyngeal (behind the throat) lymph node(s), 6 cm or less in greatest dimension, and above the lower border of the cricoid cartilage.
N2	Bilateral (both sides of the neck) metastasis in cervical lymph node(s), 6 cm or less in greatest dimension, above the lower border of the cricoid cartilage.

T1N0	No cancer cells are found in the nearby lymph nodes.
T1N1	Unilateral (one side of the neck) metastasis in cervical lymph node(s) and/or unilateral or bilateral metastasis in retropharyngeal (behind the throat) lymph node(s), 6 cm or less in greatest dimension, and above the lower border of the cricoid cartilage.
T1N2	Bilateral (both sides of the neck) metastasis in cervical lymph node(s), 6 cm or less in greatest dimension, above the lower border of the cricoid cartilage.
T2N1	(Stage II): The tumor has extended into the parapharyngeal space or adjacent soft tissues, with node spread on one side of the neck or behind the throat.
T2N2	(Stage III): The tumor has extended into the parapharyngeal space or adjacent soft tissues, with node spread on both sides of the neck.
T3N0	(Stage III): The tumor has invaded the skull base/vertebrae/sinuses (T3), with no lymph node metastasis (N0).
T3N1	(Stage III): The tumor has invaded the skull base/vertebrae/sinuses (T3), with unilateral nodes above the cricoid (N1).
T3N2	(Stage III): The tumor has invaded the skull base/vertebrae/sinuses (T3), with bilateral nodes above the cricoid (N2).

Table 3. Quantile and 95th percentile distribution of mean gene expression values across GPL570 platform probes for the study groups.

Group	p0 (0%)	p25 (25p%) Q1	p50 (50%) Q2	p75 (75%) Q3	p95 (95%)	p100 (100%)
Control	2.983216	5.258903	6. 354983	7. 504058	9.479666	14.465510
T1N0	3.003892	5.210514	6.312997	7.515814	9.628229	14.454268
T1N1	2.988557	5.213469	6.315020	7.521362	9.613064	14.472801
T2N1	3.003166	5.240060	6.337178	7.509865	9.493534	14.526785
T2N2	2.966227	5.245108	6.345375	7.510474	9.524298	14.333972
T3N2	2.913693	5.196829	6.312606	7.526112	9.668062	14.538544

Quantiles and the 95th percentile were computed using the quantile() function in the R base package. Values represent the distribution of mean gene expression levels calculated across all 54,675 probes of the GPL570 platform for both control and nasopharyngeal carcinoma (NPC) experimental groups.

Table 4. Top 50 highly expressed genes across experimental groups utilized as input for Gene Ontology (GO) enrichment analysis

Control	T1N0	T1N1	T2N1	T2N2	T3N2
ND3	ND3	ND3	ND3	ND3	ND4
RPS3	ND4	ND4	ND4	RPS3	ND3
RPS29	RPS3	RPS3	RPS3	ND4	RPS3
UBB	CSTA	RPS23	RPS23	RPS23	COX2
ND4	RPS23	UBB	RPS29	RPS29	RPS23
RPS23	UBB	RPS29	UBB	UBB	KRT5
CSTA	RPS29	RPL34	COX2	RPL34	CFL1
AQP3	RPL8	RPL8	DDX17	COX2	UBB
RPL34	RPL34	DDX17	RPL34	RPL8	SLC25A5
DDX17	CFL1	CFL1	DDX5	RPS27A	RPS29
ANXA1	EIF4G2	RPS27A	CFL1	RPS5	HLA-DRA
GSTA1	RPS27A	DDX5	RPS27A	GNB2L1	ATP5B
RPL8	COX2	RPS5	RPL8	CFL1	RPL8
RPS27A	RPS5	EIF4G2	CCL20	RPL18	DDX5
SLPI	DDX17	RPL18	EIF4G2	RPL19	RPL34
RPL18	DDX5	CSTA	CSTA	CSTA	CLDN1
HSP90AA1	ATP5B	HLA-DRA	KRT5	EIF3L	EIF4G2
RPL19	RPL19	RPL19	RPS5	PARK7	DDX17
COX2	CCL20	PARK7	HLA-DRA	DDX17	RPS27A
EIF3L	PARK7	HLA-DPA1	HLA-DPA1	RPS16	PARK7
PPP1CB	TFRC	SRP14	RPL19	RPL24	RPS5
CFL1	KRT5	ATP5B	PARK7	DDX5	GNB2L1
DDX5	RPL18	KARS	RAC1	EIF4G2	CLTC
CSTB	RAC1	GNB2L1	ITGAV	SLC25A5	KRT15
EIF4G2	RPL24	SLC25A5	RPL18	PSMB4	RPL19
RPS5	PPP1CB	RPL24	KRT15	HLA-DPA1	KARS
HLA-DRA	SLC25A5	HMGN1	TFRC	SRP14	SOD1
HLA-DPA1	SRP14	EIF3L	EIF3L	CCT4	MIR612
GNB2L1	CLDN1	CCT4	PPP1CB	RPL36A	RPL24
MIR612	CCT4	TFRC	RPS16	HSP90AA1	MORF4L2
SLC25A5	EIF3L	CLTC	CLTC	KARS	CYCS
PRDX1	RPS16	PPP1CB	GNB2L1	CLTC	HSPA1A
SRP14	GNB2L1	PSMB4	SLC25A5	RPL11	PPP1CB
RPS16	CLTC	RPS16	ATP5B	USMG5	RPL18
PARK7	CYCS	RAC1	RPL24	ATP5B	RAC1
CLTC	ODC1	HSP90AA1	CD74	RPL35	COX7A2
KRT19	HLA-DPA1	VDAC3	MIR612	SSR4	RPS16
RAC1	COX7A2	MORF4L2	KARS	HMGN1	HMGN1
PLAC8	HLA-DRA	PTGES3	VDAC3	RAC1	SRP14
KARS	HMGN1	COX7A2	HMGN1	PPP1CB	CCT4
MYL12B	IDH1	SOD1	PSMB4	HLA-DRA	EIF3L
HSPA1A	PSMB4	CYCS	CCT4	VDAC3	PSMB4
SEPT9	VDAC3	DEK	CXCL10	GDI2	HSP90AB1
RPL24	DSP	DSP	IFI30	COX7A2	CSTA
HMGN1	USMG5	DHX15	CLDN1	CD63	SLPI
RPL11	H3F3A	RPL36A	SRP14	H2AFZ	USMG5
DYNLL1	KRT15	HSP90AB1	SEPT9	DEK	KRT19
CD63	ITGAV	H3F3A	CD63	DHX15	VDAC3
COX7A2	PTGES3	COX6A1	COX7A2	TFRC	RPL11
KRT5	SOD1	GDI2	DSP	MORF4L2	ITGAV

Genes were ranked based on their mean expression values calculated within each experimental group (control and NPC). The top 50 most highly expressed genes from each list were selected and used as the primary input dataset for subsequent Gene Ontology (GO) functional enrichment analysis.

Table 5. Student's t-test and descriptive statistics for transferrin receptor gene expression in control and cancer groups.

Variable	Control	Variable	T1N0	Variable	T1N1	Variable	T2N1	Variable	T2N2	Variable	T3N2
V1	10.5860	V19	12.4190	V11	12.1219	V15	12.5032	V12	11.1895	V13	11.7111
V2	11.6703	V25	12.2459	V16	11.7855	V23	12.8859	V14	11.4659	V41	12.1304
V3	11.3757	V28	12.1729	V17	12.1023	V31	11.2581	V20	11.9459		
V4	10.87460	V34	12.6723	V18	11.8018	V35	11.9240	V22	12.0385		
V5	10.9107	V38	12.5742	V21	11.7188	V40	11.8735	V32	11.8575		
V6	11.0194	V39	12.7374	V24	12.7424			V33	12.2089		
V7	9.5538			V26	12.8661						
V8	10.8930			V27	11.2821						
V9	11.1513			V29	12.9425						
V10	11.0614			V37	12.8564						
Mean	**10.9096**		**12.4703**		**12.2220**		**12.0890**		**11.7844**		**11.9207**
P-Value			**3.17e-06**		**7.60e-05**		**0.0086**		**0.0025**		**0.0392**

Individual group variables and mean expression values were compared using the parametric two-sample Student's t-test via the t.test() function in the R base package, assuming unequal variances. Statistical significance is defined as ($p < 0.05$).

Table 6. Mean expression values of genes involved in the NPC endosomal iron uptake cycle across experimental groups.

Data table header descriptions

- ID — Affymetrix GPL570 Platform Probe Set Identifier
- Gene_ID — NCBI Entrez Gene Database UID (Unique Identifier) Gene Identifier
- Symbol — A Gene Symbol from NCBI UniGene Database
- Gene Title — Title of the Gene represented by the probe set
- Control — Mean Gene Expression Value of the control group
- NPCs — Mean Gene Expression Value of the NPG groups
- Significant — * denotes p-value <= 0.05 rejecting the null hypothesis, statistically significant difference

ID	Gene_ID	Symbol	Gene Title	Control	T1N0	T1N1	T2N1	T2N2	T3N2
208691_at	7037	TFRC	transferrin receptor	10.9097	12.4703 *	12.2220 *	12.0890 *	11.7844 *	11.9208 *
209921_at	23657	SLC7A11	solute carrier family 7 (DMT1)	7.6220	8.3046	7.9694	8.2227	7.7606	8.4787
218424_s_at	55240	STEAP3	STEAP family member 3 metalloreductase	8.3585	8.4648	8.5355	8.7559	8.4185	9.2809
211043_s_at	1212	CLTB	clathrin light chain B	9.2219	9.1757	9.0236	8.9072	9.1844	9.2111
201972_at	523	ATP6V1A	ATPase H+ transporting, subunit A	10.6882	10.7952	10.7616	10.7244	10.6701	10.9640 *
1554847_at	525	ATP6V1B1	ATPase H+ transporting, subunit B1	5.4552	5.2213	5.2282	5.3077	5.3427	5.3149
36994_at	527	ATP6V0C	ATPase H+ transporting, subunit c	10.1098	9.8376	9.8877	9.9306	10.0095	10.2275
202872_at	528	ATP6V1C1	ATPase H+ transporting, subunit C1	7.4843	8.1836 *	8.2207 *	7.8167	8.0856	8.0457 *
214923_at	51382	ATP6V1D	ATPase H+ transporting, subunit D	6.9274	6.7903	6.8715	6.8855	6.9165	6.8670

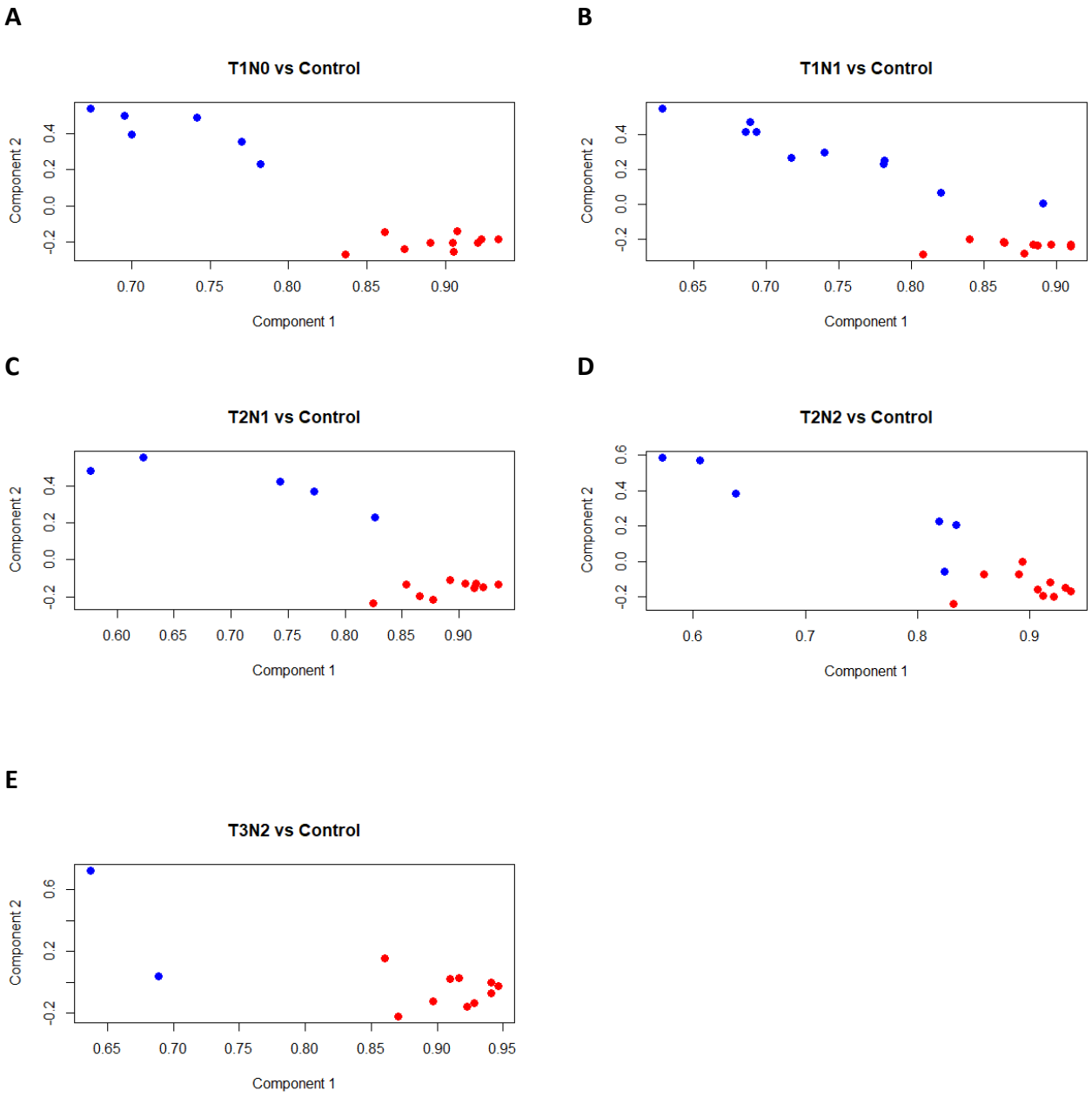


Figure 1. Principal component analysis of nasopharyngeal carcinoma development stages based on high-expression microarray profiles

Principal Component Analysis (PCA) of nasopharyngeal carcinoma (NPC) groups across the five developmental stages. Panels **(A–E)** represent individual developmental stages. The PCA was performed using the FactoMineR R package, with the x-axes and y-axes corresponding to Principal Component 1 (PC1) and Principal Component 2 (PC2), respectively. Control and Cancer serve as the categorical grouping variables, where red dots represent the healthy control group (n = 10) and blue dots denote the NPC group. Analysis was based on gene expression profiles derived from Affymetrix microarrays on the GPL570 platform (NCBI GEO accession: GDS3341). Dimensionality reduction was restricted to highly expressed genes, defined as probes with expression levels at or above the 95th percentile across samples, yielding a final subset of 2,734 probes.

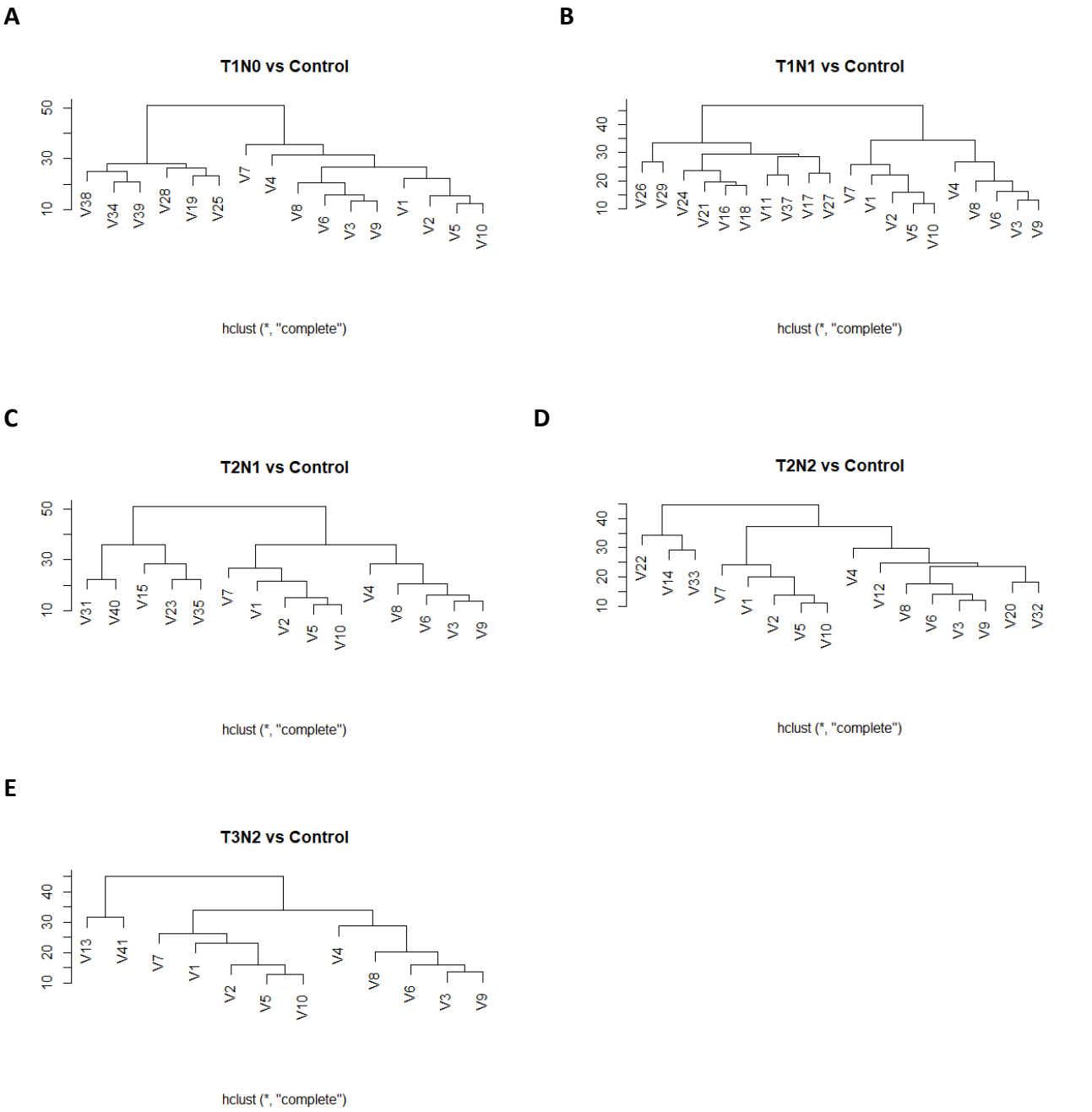
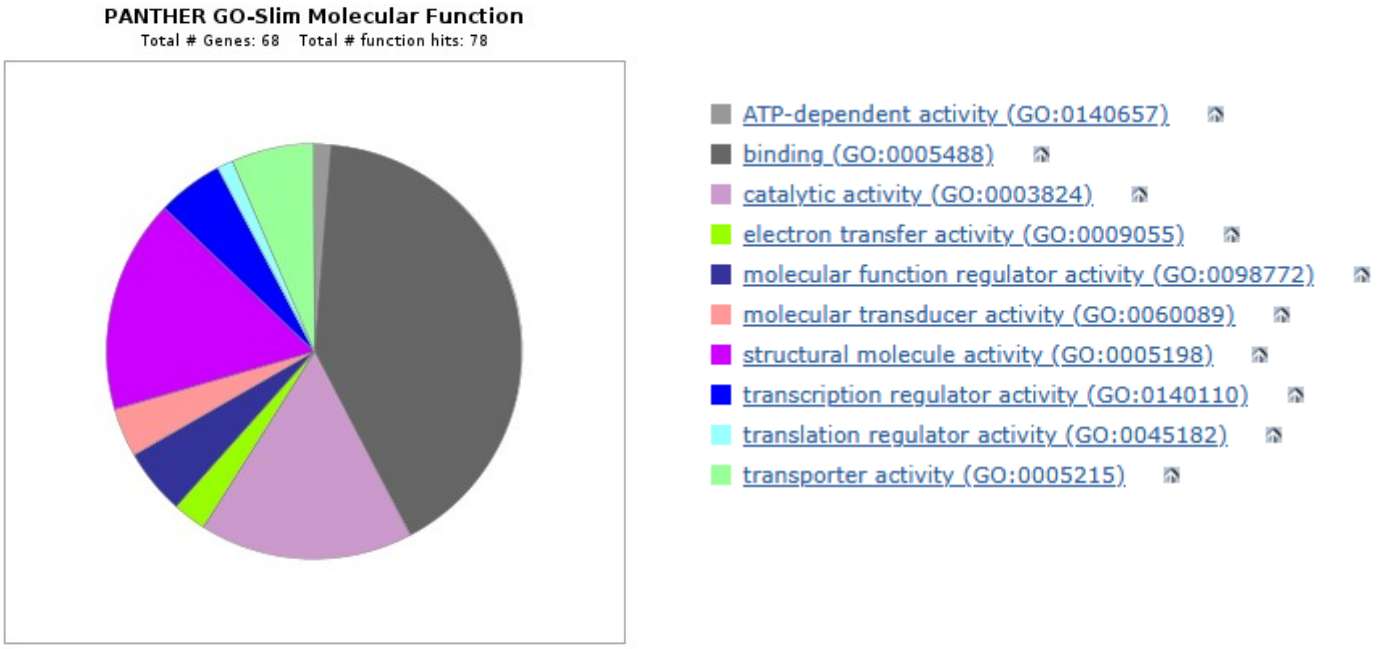


Figure 2. Hierarchical clustering analysis of nasopharyngeal carcinoma developmental stages based on high-expression microarray profiles.

Hierarchical clustering analysis of nasopharyngeal carcinoma samples across five developmental stages. Panels **(A–E)** represent individual developmental stages. Agglomerative hierarchical clustering was performed using the hclust function from the R base package. Node distances are indicated by the scale on the vertical sidebar of each dendrogram. Each leaf corresponds to a sample dataset variable; labels V1–V10 denote the healthy control group, whereas the remaining cancer sample variables are defined in Table 1. Dendrograms are based on gene expression profiles from Affymetrix microarrays on the GPL570 platform (NCBI GEO accession: GDS3341), restricted to highly expressed probes at or above the 95th percentile (2,734 probes).

A



B

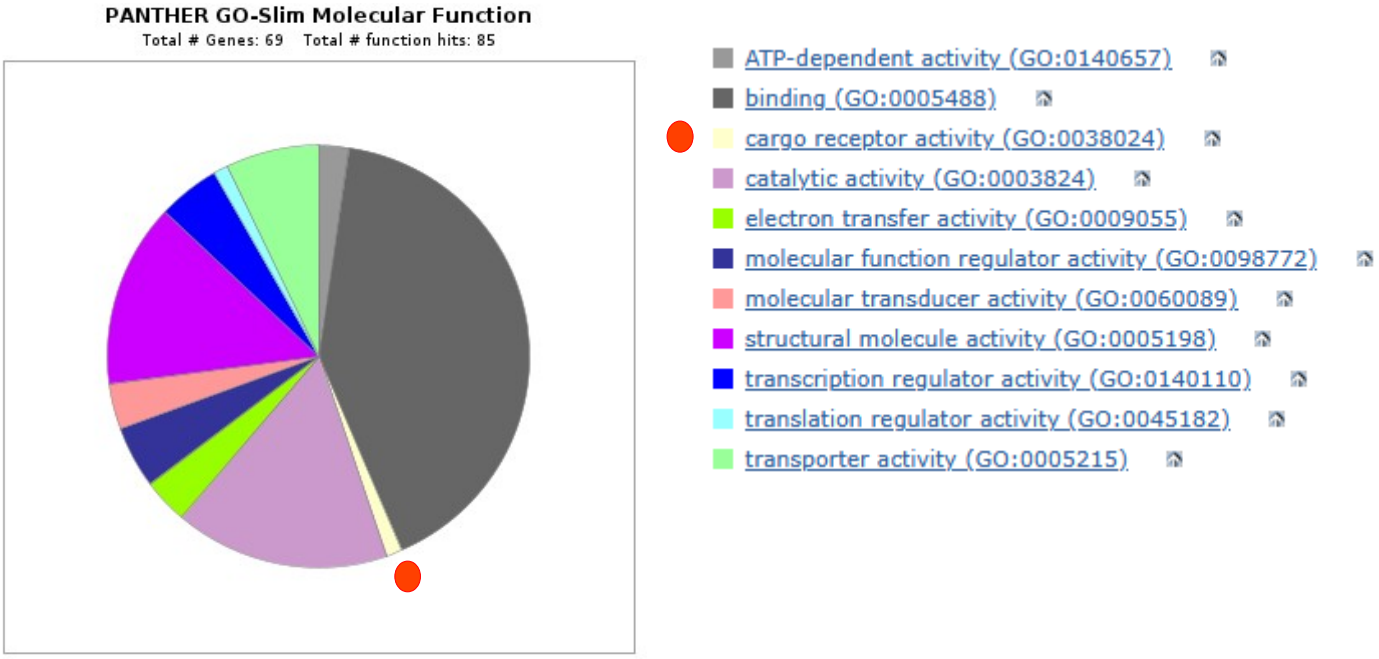
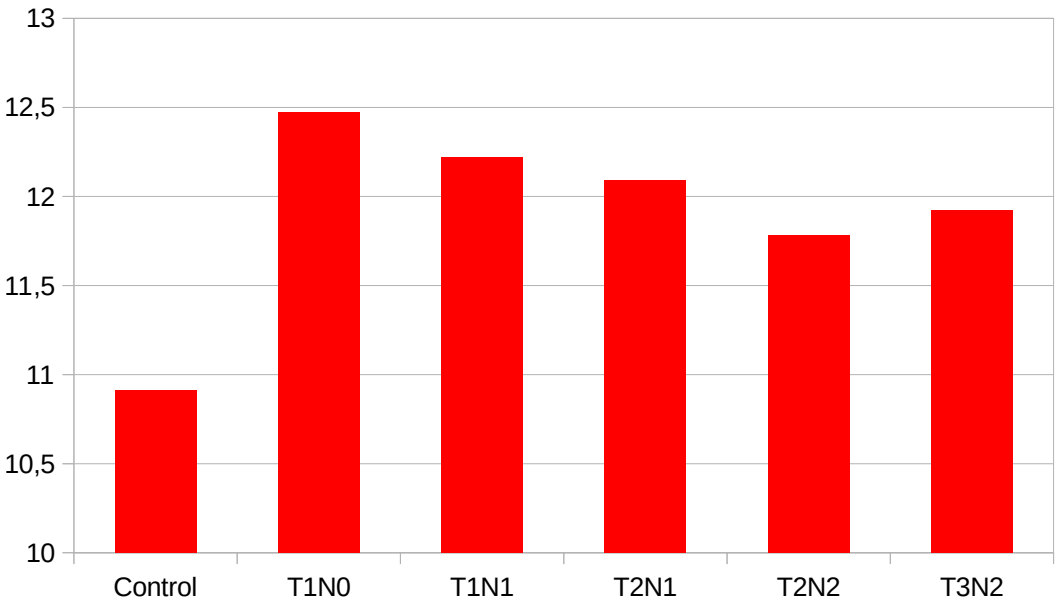
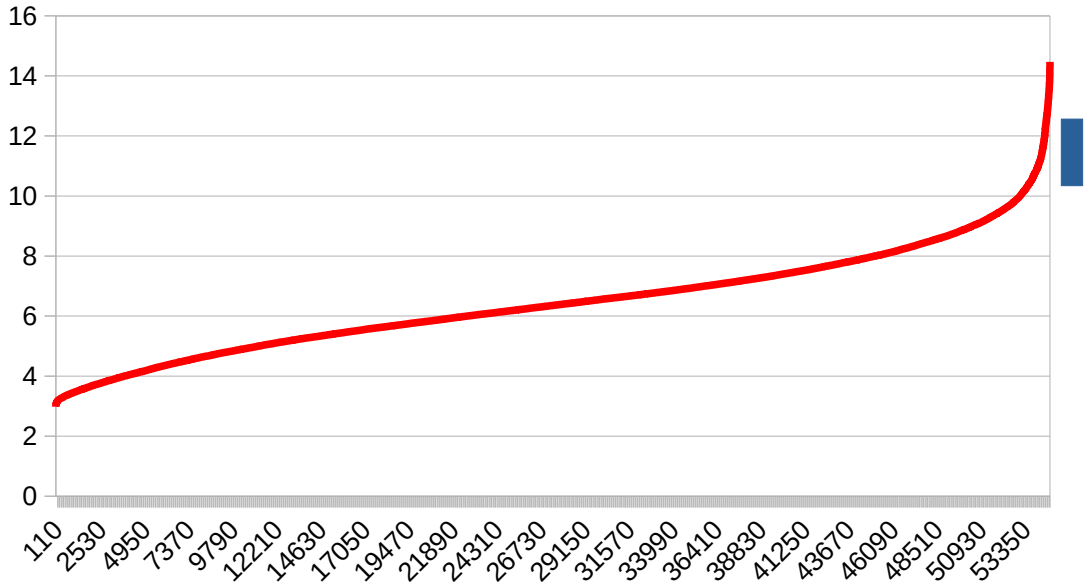


Figure 3. Identification of unique "Cargo Receptor Activity" ontology in advanced nasopharyngeal carcinoma via PAN-GO functional enrichment.
Gene Ontology (GO) enrichment analysis of "Molecular Function" categories using the PAN-GO Human Functionome platform. **(A)** Pie chart representing the functional classification for the healthy control group. **(B)** Pie chart representing the functional classification for the T2N2 nasopharyngeal carcinoma (NPC) group, serving as a representative example of the NPC cohorts. The specific input gene lists used for each group's analysis are provided in Table 4. The associated color-coding panels define the GO categories corresponding to each chart sector. The external red indicator highlights the cancer-specific emergence of the "Cargo Receptor Activity" ontology (GO:0038024), a functional category uniquely present in the tumor profile.

A


13
12,5
12
11,5
11
10,5
10
Control
T1N0
T1N1
T2N1
T2N2
T3N2

B


16
14
12
10
8
6
4
2
0
110
2530
4950
7370
9790
12210
14630
17050
19470
21890
24310
26730
29150
31570
33990
36410
38830
41250
43670
46090
48510
50930
53350

C

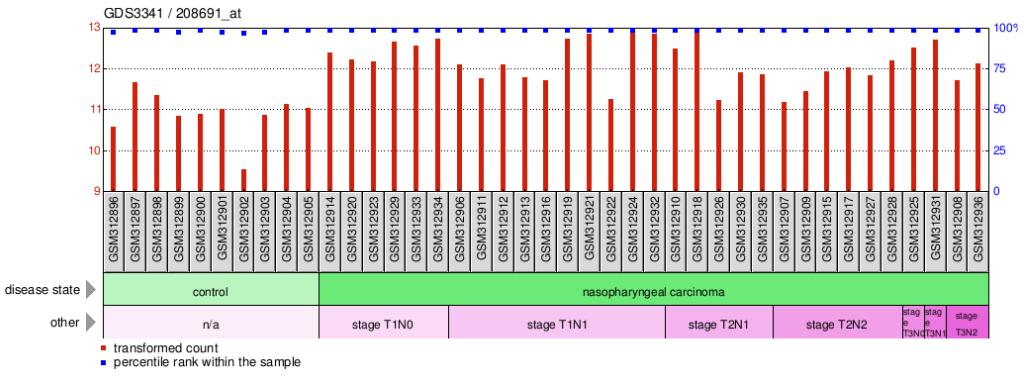

GDS3341 / 208691_at
13
12
11
10
9
100%
75
50
25
0
GSM312896
GSM312897
GSM312898
GSM312899
GSM312900
GSM312901
GSM312902
GSM312903
GSM312904
GSM312905
GSM312914
GSM312920
GSM312923
GSM312929
GSM312933
GSM312934
GSM312906
GSM312911
GSM312912
GSM312913
GSM312916
GSM312919
GSM312921
GSM312922
GSM312924
GSM312932
GSM312910
GSM312918
GSM312926
GSM312930
GSM312935
GSM312907
GSM312909
GSM312915
GSM312917
GSM312927
GSM312928
GSM312925
GSM312931
GSM312908
GSM312936
disease state
control
nasopharyngeal carcinoma
other
n/a
stage T1N0
stage T1N1
stage T2N1
stage T2N2
stage T3N0
stage T3N1
stage T3N2
transformed count
percentile rank within the sample

Figure 4. Expression profile of the transferrin receptor gene in nasopharyngeal carcinoma and verification of signal threshold levels.

(A) Bar plot illustrating the mean expression levels of the transferrin receptor gene across the study groups (control vs. NPC). Data are presented on a (log_2)-transformed scale (see Table 4). High statistical significance is indicated by p-values, where a minor nominal increase in the (log_2) value represents a non-linear, exponential surge in absolute gene expression. (**B)** Mean gene expression profile of the control group across all probes within the GPL570 platform (n = 54,675). The external blue bar highlights the inflection point of the distribution, demonstrating that the expression increments observed in panel A occur precisely at this critical baseline transition. **(C)** Representative supplementary data showing a screenshot of the specific transferrin receptor gene probe expression values retrieved from the NCBI GEO dataset GDS3341.

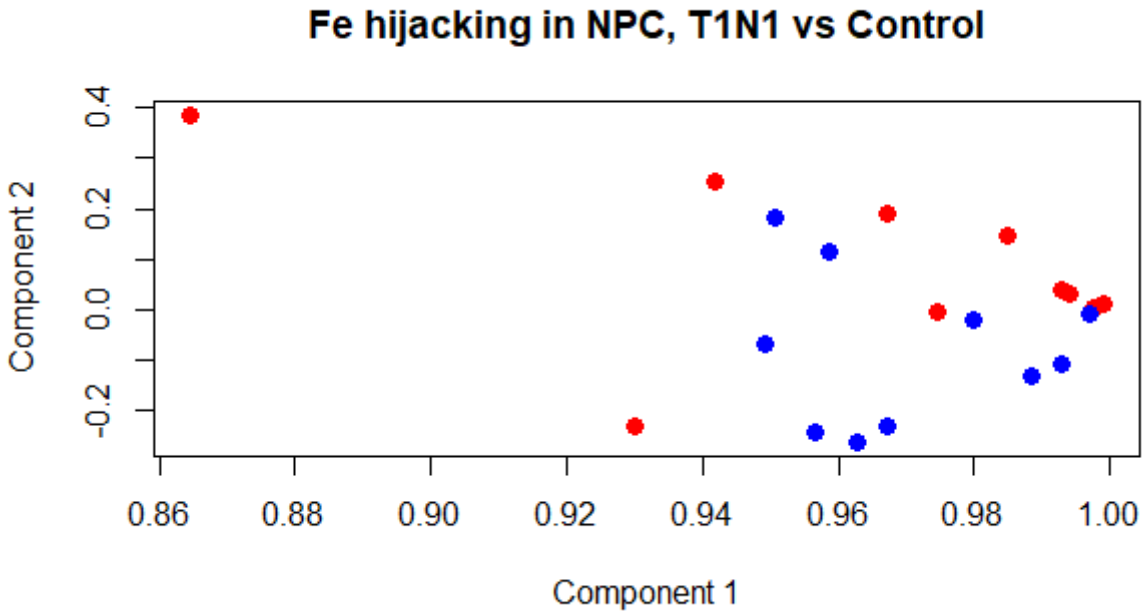


Figure 5. Selective overexpression of the transferrin receptor gene and its impact on sample clustering during endosomal iron uptake pathway analysis.

Comparative analysis of the endosomal iron uptake machinery demonstrates a stark contrast between components; while the transferrin receptor gene (TFRC) exhibits massive overexpression in nasopharyngeal carcinoma (NPC), the remaining genes driving endosomal iron internalization and trafficking show low baseline expression levels that remain unaltered by malignancy, as verified in the GDS3341 dataset. Representative two-dimensional Principal Component Analysis (PCA) score plot of the T1N1 experimental group (n = 10) versus healthy controls (n = 10), computed using the FactoMineR package in R. This specific subgroup, which encompasses the largest number of clinical variables, fails to achieve distinct cluster separation between the cancer group (blue circles) and the control group (red circles) across the first two principal components (PC1 and PC2). This lack of segregation underscores that the broader endosomal iron pathway signature remains homogeneous despite the isolated, highly localized surge in transferrin receptor expression.