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.

Stage-Independent Downregulation of the GSTA1/GSTA3 Gene Cluster on Chromosome 6p12 as a Metabolic Hallmark of Nasopharyngeal Carcinoma

Maria-Jose Oliach
Independent Researcher. Tarragona. Spain
Coia Romeu
Otolaryngology Department. Sant Joan Reus Hospital. University Rovira i Virgili. Tarragona. Spain
Antonio R. Romeu
antonioramon.romeu@iubilo.urv.cat

University Rovira i Virgili. Tarragona. Spain

Research Article

Keywords: Nasopharyngeal carcinoma (NPC), GSTA1, GSTA3, Differential gene expression, limma, Bioinformatics

Posted Date: June 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10112864/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.

Stage-Independent Downregulation of the GSTA1/GSTA3 Gene Cluster on Chromosome 6p12 as a Metabolic Hallmark of Nasopharyngeal Carcinoma

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

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

Abstract

Background: Nasopharyngeal carcinoma (NPC) is an epithelial malignancy uniquely characterized by strong associations with Epstein-Barr virus infection and environmental carcinogens. Despite its clinical relevance, the early metabolic alterations driving NPC pathogenesis remain incompletely understood.

Methods: This bioinformatic study investigated the transcriptional landscape of detoxification systems in NPC. Utilizing custom R scripts and the limma package, we performed differential expression analysis on the public GEO dataset GDS3341 (GSE12452), comparing tumor tissues across all clinical stages against non-malignant controls.

Results: We identified a robust, statistically significant, and stage-independent downregulation of both GSTA1 (Gene ID: 2938) and GSTA3 (Gene ID: 2940) in NPC tissues. Crucially, this transcriptional suppression was uniformly maintained from early to advanced clinical stages. Given that GSTA1 and GSTA3 are physically co-localized within the same gene cluster on chromosome 6p12, their simultaneous silencing strongly points toward a coordinated regional epigenetic or regulatory disruption.

Conclusions: The systemic collapse of the GSTA1/3 detoxification barrier occurs early during oncogenesis and is a fundamental metabolic hallmark of NPC, rather than a secondary consequence of tumor progression. This enzymatic deficit establishes an oxidative, hypermutable microenvironment that likely facilitates clonal evolution. As part of an ongoing, broader investigation into NPC biology, these findings unveil a critical metabolic vulnerability that could be exploited through pro-oxidant therapies or targeted stratified interventions.

Keywords

Nasopharyngeal carcinoma (NPC); GSTA1; GSTA3; Differential gene expression; limma; Bioinformatics.

Introduction

Nasopharyngeal carcinoma (NPC) is a distinct epithelial malignancy of the head and neck [1-3] that arises predominantly in the fossa of Rosenmüller [4]. While rare in Western countries, NPC is highly endemic in Southern China, Southeast Asia, and parts of North Africa [5]. The etiology of NPC is multifactorial, involving a strong causal link with latent Epstein-Barr virus (EBV) infection [6,7], genetic susceptibility [8,9], and dietary factors such as volatile nitrosamines [10]. Due to its deep anatomical location and non-specific early symptoms, most patients are diagnosed at locally advanced stages [1,11]. Although concurrent chemoradiotherapy is the primary treatment, local recurrence and distant metastasis remain the main causes of therapeutic failure [12]. Therefore, identifying the molecular mechanisms underlying NPC progression is essential to discover new diagnostic biomarkers and therapeutic targets.

Metabolic reprogramming and enhanced antioxidant defenses are established hallmarks of cancer that enable tumor survival in adverse microenvironments [13]. Malignant cells continuously generate high levels of reactive oxygen species (ROS) [14]; consequently, they must upregulate their detoxification systems to maintain redox homeostasis and prevent apoptosis [15]. This cellular defense network consists of several major enzymatic and non-enzymatic components. These include the glutathione (GSH) system -comprising glutathione S-transferases (GSTs) for toxin conjugation and glutathione peroxidases for lipid peroxide reduction- [16] and the superoxide dismutase (SOD) family, which catalyzes the dismutation of superoxide radicals into oxygen and hydrogen peroxide [17]. Additionally, the network includes catalase, the thioredoxin (TXN) system [18-20], and Phase I/II xenobiotic-metabolizing enzymes such as the cytochrome P450 (CYP) superfamily [21]. The coordinated activation of these components protects tumor cells from both endogenous oxidative stress and exogenous chemotherapy.

To investigate the specific transcriptional alterations governing these detoxification pathways, a high-throughput bioinformatic approach is required. The aim of this work is to characterize the impact of NPC on the detoxifying profile of tumor cells by analyzing public transcriptomic data. We utilized the GDS3341 dataset [22-24] from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database [25,26]. This dataset contains genome-wide expression profiles obtained via laser-capture microdissection, ensuring high cellular purity by isolating malignant epithelial cells from the surrounding stroma. By evaluating the differential expression of the complete repertoire of detoxification genes -including the GSH system, SODs, and xenobiotic metabolism- this study identifies key metabolic adaptations and potential therapeutic vulnerabilities in NPC cells.

This study represents the second installment of a broader research project dedicated to the molecular characterization of nasopharyngeal carcinoma. The initial phase of this investigation was previously published as a preprint on Research Square [27]. That preliminary study established the baseline transcriptomic profiling of highly expressed genes in NPC, demonstrated how these profiles stratify disease progression stages, and highlighted dysregulation in cellular transport and iron homeostasis. The current work extends those findings by focusing specifically on the functional evaluation of the cellular detoxification machinery. By integrating antioxidant and metabolic pathway analyses with our previous data, this study provides a more comprehensive understanding of NPC biology and helps identify targets to overcome drug resistance.

Methodology

Data Retrieval and Platforms. Transcriptomic profiles were obtained from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) repository. Specifically, the dataset GDS3341 (associated with series GSE12452) was retrieved, which comprises gene expression data derived from the Affymetrix Human Genome U133 Plus 2.0 Array platform (GPL570). This dataset includes 41 distinct samples categorized into healthy controls and various clinical stages of nasopharyngeal carcinoma (NPC), providing a robust framework for assessing stage-specific molecular alterations.

Data Processing Strategy and Rationale. The analytical rationale focused on reconstructing a cohesive expression matrix integrating platform annotations and sample-level data. Computational workflows were implemented within the R environment (RStudio) utilizing base R and custom scripts. The dataset was structured into a functional data frame encompassing 54,675 rows—corresponding to the total probe coverage of the GPL570 platform—and 42 columns (the unique probe identifier plus the 41 samples). Experimental metadata from GDS3341 were cross-referenced to define the clinical variables (V1–V41), establishing a structured matrix detailed in Table 1. Crucially, the hybridization intensity signals were pre-normalized into a non-linear scale spanning a dynamic range of 2 to 14. Because there is no established biological threshold to unambiguously distinguish true probe-mRNA hybridization signals from technical background noise, a stringent filtering strategy was applied to mitigate downstream artifacts. The study explicitly focused on highly expressed genes capable of undergoing robust cancer-induced silencing or activation. To achieve this, the target gene selection was strictly restricted to probesets whose mean expression values in at least one experimental group fell within the upper upper quartile (third quartile, Q3 >= 75th percentil) of the overall group profile, as detailed in Table 2.

Selection of Target Detoxification Genes. To evaluate the metabolic defense mechanisms in NPC, a curated subset of 50 hallmark probesets involved in cellular detoxification and antioxidant response was defined. This selection targeted key molecular circuitries, including the Thioredoxin (TRX) system, the NRF2-KEAP1 axis as the master regulator, NADPH-dependent networks (Pentose Phosphate Pathway), Mitochondrial Superoxide Dismutase (SOD), catalase, and the Cytochrome P450 (CYP450) enzymatic family. To ensure precise genomic contextualization, comprehensive annotations—including chromosomal locations, genomic coordinates, and strand orientations—were retrieved from the National Center for Biotechnology Information (NCBI) database, specifically aligned to the GRCh38.p14 human reference genome assembly [28]. Importantly, the design of the Affymetrix GPL570 platform inherently features redundancy, meaning that a single canonical gene (mRNA) can be interrogated by multiple independent probesets to capture distinct transcript variants, and conversely, more than one gene (mRNA) can hybridize to one probeset. Consequently, all redundant probesets corresponding to the target pathways were retained for parallel downstream analysis. The final structured list, detailing probe identifiers, genomic coordinates, and functional pathway classifications, is comprehensively detailed in Table 1S and Table 2S.

Differential Expression Analysis and Statistical Rationale. The primary objective of this study was to identify stage-specific critical genes and quantify differential expression across each detoxification mechanism. This analysis was strictly focused on probesets operating within the dynamically filtered high-expression spectrum ($Q_3 \ge$ 75th percentile in at least one group). While custom scripts were utilized to compute mean values and preliminary significance, the core statistical workflow was executed using the Bioconductor package limma (Linear Models for Microarray Data). This approach is exceptionally robust for multi-group clinical designs; limma fits

a separate linear model for each probe and applies an Empirical Bayes (eBayes) moderation method. By implementing this conditional filtering framework, the study successfully retains genes that are baseline-dormant but highly induced by cancer progression, as well as those actively expressed in healthy tissue but subsequently silenced by the tumor. Multiple testing corrections across all clinical stages versus the healthy controls were strictly controlled using the Benjamini-Hochberg False Discovery Rate (FDR) algorithm with a significance threshold of adj.P.Val < 0.05. Crucially, to ensure biological relevance beyond statistical significance, a biological magnitude filter was superimposed, requiring an absolute $\log_2$ fold-change threshold of 1.0 (equivalent to a 2-fold change in linear expression), with a focus on high-magnitude switch genes displaying alterations up to 4 units in the $\log_2$ scale (>= 16-fold change). This combined mathematical and biological stringency isolates the key enzymatic drivers of nasopharyngeal carcinoma progression.

Results and dicussion

While high-density microarrays are crucial for biomarker discovery, transcriptomic profiles must be interpreted considering the inherent operational limits of gene expression. The single presence of a given mRNA level does not definitively determine the biological activity of its encoded protein, as cellular homeostatic processes are extensively governed by post-translational modifications, structural protein stability, and localized enzymatic activation loops. Consequently, the lack of large-scale, high-magnitude transcriptional alterations observed across the clinical stages of nasopharyngeal carcinoma (NPC) in this cohort does not imply functional quiescence. Instead, it suggests a finely tuned regulatory architecture where critical pathways might achieve specialized defensive or pathological states without forcing an immediate acceleration of the core transcriptional machinery.

Intriguingly, the overarching finding of this study demonstrates that the global transcriptional expression of the core genes involved in detoxication systems and remain largely unaffected by NPC pathogenesis across all clinical stages. At first glance, this lack of differential regulation may appear paradoxical, given that malignant cells typically require highly active and finely tuned detoxification machinery to survive oxidative stress and sustain rapid proliferation. However, as previously noted, a neutral effect at the transcriptomic level does not preclude the functional efficacy of the system(s). The baseline expression observed here may still yield proteins that operate with optimal catalytic efficiency, or alternatively, the actual enzymatic activity could be tightly modulated via post-translational modifications, allosteric regulation, or localized substrate availability rather than de novo transcription. Consequently, these results suggest that while specific components like the GSTA1/3 cluster are selectively silenced, the broader detoxification network maintains a stable, constitutive transcriptional state that does not oscillate with advanced tumor burden.

Differential expression analysis using the `limma` package revealed a robust and statistically significant downregulation of both GSTA1 (Gene ID: 2938) and GSTA3 (Gene ID: 2940) in NPC tissues compared to non-malignant controls. Notably, this transcriptional suppression was not stage-dependent but remained consistently uniform and highly significant (Tsble 3) across all clinical stages of the disease. Given that GSTA1 and GSTA3 are physically co-localized within the same gene cluster on chromosome 6p12, their simultaneous and sustained repression strongly suggests a generalized, coordinated collapse of the Phase II detoxification barrier in the nasopharyngeal epithelium, occurring early during oncogenesis and persisting throughout tumor progression.

The constitutive downregulation of GSTA1 and GSTA3 across all anatomical stages of NPC highlights a pivotal metabolic reprogramming event in this malignancy. Under physiological conditions, these alpha-class glutathione S-transferases maintain cellular redox homeostasis and genomic integrity by conjugating electrophilic xenobiotics and reactive oxygen species (ROS) with glutathione. In the context of NPC -a malignancy intrinsically linked to environmental carcinogens (such as dietary nitrosamines) and Epstein-Barr virus (EBV) infection- the systemic loss of GSTA1 and GSTA3 expression unveils a critical vulnerability. Because these two genes reside adjacent to one another within the same genomic cluster on chromosome 6p12, their concurrent downregulation points toward a shared upstream regulatory disruption or regional epigenetic silencing. This joint enzymatic deficit deprives the nasopharyngeal epithelium of its primary defense against chemical and biological oncogenic insults, suggesting a hypermutable, oxidative microenvironment that drives both tumor initiation and clonal evolution.

Crucially, because this downregulation is sustained throughout all clinical stages, it underscores that the loss of GSTA1/3-mediated detoxification is a fundamental hallmark of NPC (Figure 1) pathogenesis rather than a secondary consequence of advanced tumor burden. While elevated levels of GST alpha enzymes are traditionally associated with chemoresistance and poor prognosis in various solid tumors, their profound suppression in NPC aligns with a phenotype susceptible to profound oxidative stress. This continuous transcriptional repression could be exploited therapeutically. For instance, the resulting compromised antioxidant capacity may render NPC cells selectively vulnerable to pro-oxidant therapies, platinum-based regimens, or novel ferroptosis-inducing agents, offering a potential stratifying biomarker for targeted interventions.

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

Supplementary Materials

Table 1S. Curated subset of hallmark detoxification and antioxidant response genes.

Overview of the target probes selected for stage-specific expression profiling in NPC. Genes are categorized by their primary metabolic or antioxidant circuitry, including the Thioredoxin (TRX) system, the NRF2-KEAP1 master regulatory axis, NADPH-dependent networks, Mitochondrial Superoxide Dismutase (SOD), catalase, and the Cytochrome P450 (CYP450) family. Probe IDs correspond to the Affymetrix Human Genome U133 Plus 2.0 Array (GPL570) nomenclature validated via the Reactome and MSigDB databases.

Table 2S. Precise genomic contextualization of the subset of hallmark genes

Comprehensive annotations—including chromosomal locations, genomic coordinates, and strand orientations—were retrieved from the National Center for Biotechnology Information (NCBI) database, specifically aligned to the GRCh38.p14 human reference genome assembly.

References

1. Yu-Pei Chen, Anthony T C Chan, Quynh-Thu Le, Pierre Blanchard, Ying Sun, Jun Ma. Nasopharyngeal carcinoma. Lancet 394(10192):64-80. 2019. PMID: 31178151. doi: 10.1016/S0140-6736(19)30956-0.
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. 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
4. Chiazo Amene, Maura Cosetti, Sudheer Ambekar, Bharat Guthikonda, Anil Nanda. Johann Christian Rosenmüller (1771-1820): A Historical Perspective on the Man behind the Fossa. J Neurol Surg B Skull Base 74(4):187–193.2013. PMID: 24436911. doi: 10.1055/s-0033-1342927.
5. Freddie Bray, Mathieu Laversanne, Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Isabelle Soerjomataram, Ahmedin Jemal. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74(3):229-263.2024. PMID: 38572751. doi: 10.3322/caac.21834.
6. 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.
7. 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.
8. 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.
9. Sai Wah Tsao, Yim Ling Yip, Chi Man Tsang, Pei Shin Pang, Victoria Ming Yi Lau, Guitao Zhang, Kwok Wai Lo. Etiological factors of nasopharyngeal carcinoma. Oral Oncol 50(5):330-8. 2014. PMID: 24630258. doi: 10.1016/j.oraloncology.2014.02.006.
10. 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.
11. Melvin L K Chua, Joseph T S Wee, Edwin P Hui, Anthony T C Chan. Nasopharyngeal carcinoma. Lancet 387(10022):1012-1024.2016. PMID: 26321262. doi: 10.1016/S0140-6736(15)00055-0.
12. Yunming Tian, Wei-Zeng Hua.g, Lei Zeng, Li Bai, Fei Han, Yuhong Lan. The Failure Patterns of Nasopharygeal Carcinoma After Intensity-Modulated Radiotherapy and Implications for Surveillance. Cancer Manag Res Oct 5:14:2813-2823.2022. PMID: 36221292. doi: 10.2147/CMAR.S347864.
13. Douglas Hanahan. Hallmarks of Cancer: New Dimensions. Cancer Discov 12(1):31-46. 2022. PMID: 35022204. doi: 10.1158/2159-8290.
14. Sharmin Akter, Rajesh Madhuvilakku, Anik Kumar Kar, Irin Sultana Nila, Pengda Liu, Hiroyuki Inuzuka, Wenyi Wei, Yonggeun Hong. Reactive oxygen species (ROS) in cancer:

from mechanism to therapeutic implications. Signal Transduct Target Ther 11(1):111.2026. PMC13018652. doi: 10.1038/s41392-026-02583-x.
15. Chiara Gorrini, Isaac S Harris, Tak W Mak. Modulation of oxidative stress as an anticancer strategy. Nat Rev Drug Discov 12(12):931-47.2013. PMID: 24287781. doi:
16. Ganesaratnam K Balendiran, Rajesh Dabur, Deborah Fraser. The role of glutathione in cancer. Cell Biochem Funct 22(6):343-52. 2004. PMID: 15386533. doi: 10.1002/cbf.1149.
17. Shuai Li, Lanyan Fu, Tian Tian, Liwen Deng, Huangbin Li, Weixiong Xia, Qing Gong. Disrupting SOD1 activity inhibits cell growth and enhances lipid accumulation in nasopharyngeal carcinoma. Cell Commun Signal 16(1):28.2018. PMID: 29891006. doi: 10.1186/s12964-018-0240-3.
18. John D Hayes, Albena T Dinkova-Kostova. The Nrf2 regulatory network provides an interface between redox and intermediary metabolism. Trends Biochem Sci 39(4):199-218. 2014. PMID: 24647116. doi: 10.1016/j.tibs.2014.02.002.
19. Tina Oberacker, Leonie Kraft, Moritz Schanz, Jörg Latus, Severin Schricker. The Importance of Thioredoxin-1 in Health and Disease. Antioxidants (Basel) 12(5):1078. 2023. PMID: 37237944. doi: 10.3390/antiox12051078.
20. Rebecca Seitz, Deniz Tümen, Claudia Kunst, Phillip Heumann, Stephan Schmid, Arne Kandulski, Martina Müller, Karsten Gülow. Exploring the Thioredoxin System as a Therapeutic Target in Cancer: Mechanisms and Implications. Antioxidants (Basel) Sep 4;13(9):1078. 2024. PMID: 39334737. doi: 10.3390/antiox13091078.
21. Fabio Hecht, Marco Zocchi, Fatemeh Alimohammadi, Isaac Spencer Harris. Regulation of antioxidants in cancer. Mol Cell. 84(1):23–33. 2023. PMID: 38029751. doi: 10.1016/j.molcel.2023.11.001
22. 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.
23. 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.
24. 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
25. Ron Edgar, Michael Domrachev, Alex E Lash. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30(1):207-10.2002. PMID: 11752295. doi: 10.1093/nar/30.1.207.
26. Tanya Barrett, Dennis B Troup, Stephen E Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F Kim, Maxim Tomashevsky, Kimberly A Marshall, Katherine H Phillippy, Patti M Sherman, Rolf N Muertter, Michelle Holko, Oluwabukunmi Ayanbule, Andrey Yefanov, Alexandra Soboleva. NCBI GEO: archive for functional genomics data sets--10 years on. Nucleic Acids Res 39(Database issue):D1005-10.2011. PMID: 21097893. doi: 10.1093/nar/gkq1184.
27. Coia Romeu, Maria-Jose Oliach, Antonio R. Romeu et al. Transcriptomic Profiling of Highly Expressed Genes Stratifies Nasopharyngeal Carcinoma Progression and Highlights Transferrin

Receptor Dysregulation: A Bioinformatic Approach, 08 June 2026, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-9935130/v1].
28. NCBI Homo sapiens Annotation Release 110. Accessed 19/06/2026. https://www.ncbi.nlm.nih.gov/refseq/annotation_euk/Homo_sapiens/110/

Table 1. Experimental design and sample stratification for dataset GDS3341/GSE12452.

Sample	Comment[1]	Stage[1]	Variable[2]
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]: 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).

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

Table 2. Quantile distribution of mean expression values across experimental groups

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

Descriptive statistical breakdown of the computed mean hybridization signals for each clinical and control subgroup. The matrix defines the boundary thresholds for the minimum values (0th percentile), first quartile (Q1, 25th percentile), median (Q2, 50th percentile), third quartile (Q3, 75th percentile), and maximum intensity values (100th percentile). These metrics mathematically define the dynamic baseline threshold (Q3 >=75th percentile) employed to filter out technical background noise and isolate biologically active, highly expressed transcriptomic candidates across nasopharyngeal carcinoma (NPC) progression. Because groups T3N0 and T3N1 only had one Sample (Table 1), they could not be included in the statistical analysis. Control (n=10), T1N0 (n=6), T1N1 (n=10), T2N1 (n=5), T2N2 (n=6) and T3N2 (n=2)

Table 3. Comprehensive genomic annotation and differential expression profile for critical detoxification genes in the NPC stage groups

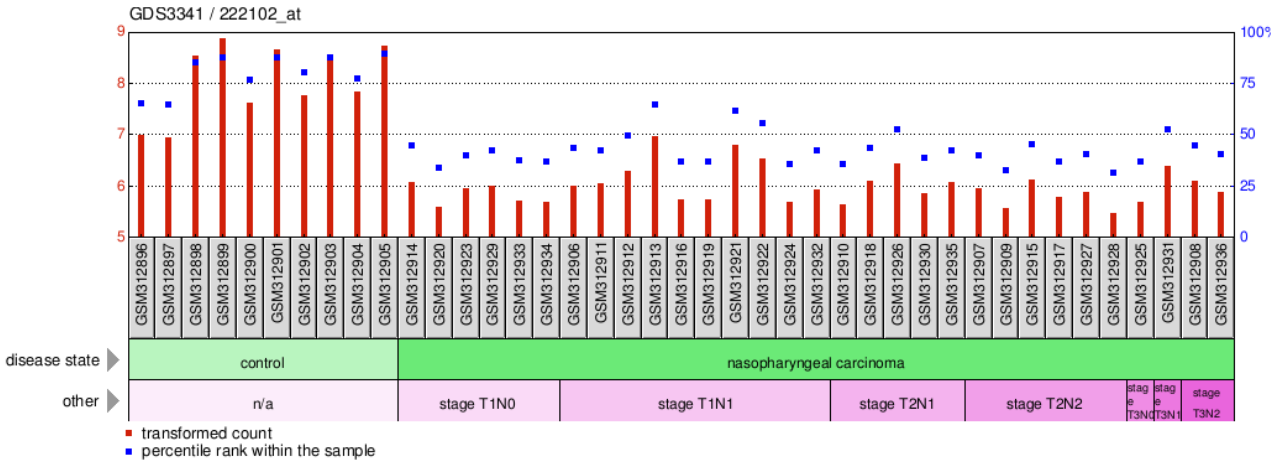
Probe ID	Control	NPC	Diff. Exp.	logFC	AveExpr	t	P.Value	adj.P.Val	B
		T1N0		**T1N0**	**T1N0**	**T1N0**	**T1N0**	**T1N0**	**T1N0**
222102_at	8.0590	5.8541	Down	-2.2050	6.5301	-9.1862	6.83E-11	6.90E-09	14.7388
203924_at	12.6189	7.5990	Down	-5.0199	9.5044	-7.5235	7.71E-09	3.89E-07	10.0223
200736_s_at	10.9542	10.0014	Down	-0.9528	10.5298	-4.4681	7.78E-05	2.62E-03	0.8939
		T1N1		**T1N1**	**T1N1**	**T1N1**	**T1N1**	**T1N1**	**T1N1**
222102_at	8.0590	6.1906	Down	-1.8684	6.5301	-8.9882	1.18E-10	1.19E-08	14.2061
203925_at	9.1127	7.4862	Down	-1.6265	8.0746	-6.5314	1.49E-07	7.51E-06	7.0913
203924_at	12.6189	9.5355	Down	-3.0834	9.5044	-5.3361	5.65E-06	1.90E-04	3.4875
		T2N1		**T2N1**	**T2N1**	**T2N1**	**T2N1**	**T2N1**	**T2N1**
222102_at	8.0590	6.0309	Down	-2.0281	6.5301	-7.9662	2.12E-09	2.14E-07	11.3919
203924_at	12.6189	8.0476	Down	-4.5713	9.5044	-6.4594	1.85E-07	9.34E-06	6.9614
203925_at	9.1127	7.5126	Down	-1.6001	8.0746	-5.2461	7.43E-06	2.50E-04	3.3154
		T2N2		**T2N2**	**T2N2**	**T2N2**	**T2N2**	**T2N2**	**T2N2**
222102_at	8.0590	5.8140	Down	-2.2450	6.5301	-9.3529	4.33E-11	4.38E-09	15.1837
203924_at	12.6189	7.7312	Down	-4.8876	9.5044	-7.3253	1.38E-08	6.98E-07	9.4274
201146_at	11.1895	10.2777	Down	-0.9117	10.7696	-4.6440	4.60E-05	1.55E-03	1.3924
201470_at	10.4919	11.3759	Up	0.8840	10.9135	4.3349	1.16E-04	2.92E-03	0.4918
		T3N2		**T3N2**	**T3N2**	**T3N2**	**T3N2**	**T3N2**	**T3N2**
222102_at	8.0590	6.0065	Down	-2.0526	6.5301	-5.7008	1.86E-06	1.88E-04	4.8975
203924_at	12.6189	8.4541	Down	-4.1647	9.5044	-4.1613	1.93E-04	9.74E-03	0.4080

Probe ID	GeneID	Symbol	Gene Title	Chr.	Start	End	Strand
222102_at	2940	GSTA3	glutathione S-transferase alpha 3	6	52,,896646	52,909,698	-
203924_at	2938	GSTA1	glutathione S-transferase alpha 1	6	52,791,371	52,803,816	-
200736_s_at	2876	GPX1	glutathione peroxidase 1	3	49,357,176	49,358,353	-
203925_at	2730	GCLM	glutamate-cysteine ligase modifier subunit	1	93,885,199	93,909,430	-
209531_at	2954	GSTZ1	glutathione S-transferase zeta 1	14	77,321,036	77,331,597	+
201146_at	4780	NFE2L2	nuclear factor erythroid 2-like 2	2	177,230,303	177,264,727	-
201470_at	9446	GSTO1	glutathione S-transferase omega 1	10	104254173	104267455	+

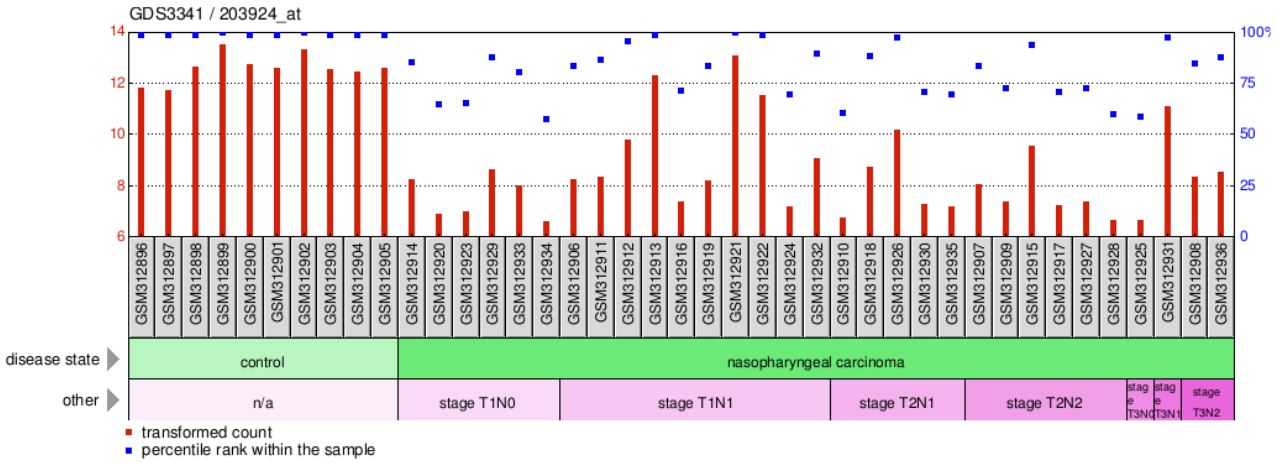
Summary of the prioritized critical detoxification probesets identified throughout linear modeling. Genomic features, including Chromosome, Start/End coordinates, and Strand orientation, were mapped using the NCBI GRCh38.p14 reference assembly. Mean Control and Mean NPC gropu represent the calculated average hybridization intensities in $\log_2$ scale. Statistical parameters were generated via the Bioconductor limma package utilizing Empirical Bayes moderation: logFC indicates $\log_2$ fold-change; AveExpr denotes the overall mean expression across all microarray samples; t represents the moderated *t*-statistic; P.Value corresponds to the unadjusted nominal significance; adj.P.Val reflects the Benjamini-Hochberg False Discovery Rate (FDR) adjusted *p*-value (threshold applied: adj.P.Val < 0.05); and B defines the log-odds of differential expression (derived directly from eBayes). B > 0: the gene is more likely to be dysregulated. B < 0: the gene is more likely to be technical noise or not have any real biological changes.

Figure 1. Transcriptional downregulation of GSTA1 and GSTA3 in Nasopharyngeal Carcinoma (NPC).

GSTA3, glutathione S-transferase alpha 3 (Gene ID, 2940)
Chromosome 6 (52,896646-52,909,698) strand -
Probe ID 222102_at



GSTA1, glutathione S-transferase alpha 1 (Gene ID, 2938)
Chromosome 6 (52,791,371-52,803,816) strand -
Probe ID 203924_at



Representative expression profiles extracted from the NCBI GEO dataset GDS3341 demonstrating the significant suppression of (A) GSTA1 (Gene ID: 2938) and (B) GSTA3 (Gene ID: 2940) transcripts across all clinical stages of NPC compared to non-malignant nasopharyngeal tissue controls. These visual profiles substantiate the uniform, stage-independent transcriptional silencing of the chromosome 6p12 gene cluster observed in our differential expression analysis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1S.xlsx
- Table2S.xlsx