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.

Causal Structural Equation Modeling Links the Epigenetic Landscape to the Warburg Effect in Nasopharyngeal Carcinoma

Maria-José Oliach

Independent Researcher. Tarragona. Spain

Antonio Romeu

antonioramon.romeu@iubilo.urv.cat

University Rovira i Virgili https://orcid.org/0000-0002-3898-061X

Research Article

Keywords: Nasopharyngeal carcinoma; Epigenetic landscape; Warburg effect; Structural equation modeling; H19; Transcriptomics; GSE12452.

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11767626/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.

Causal Structural Equation Modeling Links the Epigenetic Landscape to the Warburg Effect in Nasopharyngeal Carcinoma

Running Title: *Epigenetic Landscape and Warburg Effect in NPC*

Maria-Jose Oliach[1], Antonio R. Romeu[2]*

[1] Independent Researcher. Tarragona. Spain.
[2]. University Rovira i Virgili. Tarragona. Spain.
**Correspondence: antonioramon.romeu@iubilo.urv.cat*

Abstract

Background: Nasopharyngeal carcinoma (NPC) is an Epstein-Barr virus (EBV)-associated epithelial malignancy characterized by profound epigenetic dysregulation and aerobic glycolysis (the Warburg effect). Although differential expression analyses consistently identify concurrent alterations across chromatin modifiers and glycolytic enzymes, standard correlative frameworks cannot test directional hierarchies between these biological hallmarks. Here, we applied structural equation modeling (SEM) to test whether the coordinated epigenetic landscape causally drives the Warburg effect in NPC.

Methods: Laser-capture microdissected transcriptomic profiles from 41 samples (31 NPC tumors and 10 normal nasopharyngeal controls; NCBI GEO GSE12452 / GDS3341, Affymetrix HG-U133 Plus 2.0) were analyzed using limma, hierarchical clustering, and principal component analysis (PCA). Eleven differentially expressed biomarkers were assigned to two latent constructs: the Epigenetic Landscape (DNMT1, EZH2, HDAC2, SMARCA4, BRD4, and the long non-coding RNA H19) and the Warburg Effect (GAPDH, LDHA, CS, CYC, and TIMMDC1). Standardized transcriptomic scores were fitted via SEM in R (lavaan v0.6-21) using Satorra-Bentler robust maximum likelihood estimation.

Results: All 11 indicators were significantly dysregulated in NPC (adjusted $p < 0.001$), featuring marked downregulation of H19 (log2FC = -1.8595) alongside coordinated upregulation of chromatin regulators and metabolic enzymes (log2FC = 0.4411 to 1.4546). PCA separated tumor from control tissues (73.7% cumulative variance) and resolved functional gene clusters along PC1 (94.4% variance). The structural model demonstrated acceptable global fit (scaled chi-square = 72.790, df = 41, $p = 0.002$; robust CFI = 0.915; SRMR = 0.080) with highly significant factor loadings ($p < 0.001$; H19 lambda = -0.861; positive indicators lambda = 0.655 to 0.942). Crucially, the Epigenetic Landscape exerted a strong, direct causal effect on the Warburg Effect (standardized beta = 0.805, $p < 0.001$), accounting for 64.8% of its latent variance ($R^2 = 0.648$).

Conclusions: Epigenetic remodeling acts as an upstream driver of metabolic reprogramming in NPC, supporting dual epigenetic-metabolic therapeutic strategies.

Keywords: Nasopharyngeal carcinoma; Epigenetic landscape; Warburg effect; Structural equation modeling; H19; Transcriptomics; GSE12452.

1. Introduction

Nasopharyngeal carcinoma (NPC) is an Epstein-Barr virus (EBV)-associated epithelial malignancy endemic to Southeast Asia and Southern China [1, 2]. Because of its deep anatomical location and nonspecific early symptoms, most patients present at locoregionally advanced stages [3]. Current standard-of-care therapy relies on intensity-modulated radiotherapy (IMRT) combined with induction and concurrent gemcitabine-cisplatin chemotherapy, which has substantially improved locoregional control [4, 5]. Furthermore, plasma EBV DNA quantification guides clinical risk stratification [6], while anti-PD-1 immune checkpoint inhibitors have expanded options for recurrent or metastatic disease [7, 8]. Nevertheless, distant metastasis and treatment resistance remain major clinical hurdles [9].

Cancer progression is driven by interconnected biological capabilities conceptualized as the Hallmarks of Cancer, among which metabolic reprogramming and non-mutational epigenetic plasticity occupy central positions [10, 11]. In previous transcriptomic studies of NPC, our group stratified disease progression and delineated dysregulated regulatory networks across core oncogenic hallmarks [12–16]. Specifically, we identified a coordinated epigenetic regulatory hub consisting of five upregulated chromatin and DNA modifiers—DNA methyltransferase 1 (DNMT1), enhancer of zeste homolog 2 (EZH2), histone deacetylase 2 (HDAC2), the SWI/SNF catalytic subunit SMARCA4, and bromodomain-containing protein 4 (BRD4)—coupled with the pronounced downregulation of the imprinted long non-coding RNA H19 during regional tumor progression [15]. In parallel, we demonstrated that NPC exhibits a hybrid glycolysis/oxidative phosphorylation (OXPHOS) Warburg phenotype [14, 17, 18], characterized by the stage-specific upregulation of key glycolytic and mitochondrial effectors, namely glyceraldehyde-3-phosphate dehydrogenase (GAPDH), lactate dehydrogenase A (LDHA), citrate synthase (CS), cytochrome c (CYC), and the complex I assembly factor TIMMDC1 [14, 19]. Remarkably, H19 displays an opposing expression trajectory relative to both epigenetic activators and metabolic effectors, pointing to an antagonistic regulatory mechanism that merits dedicated investigation.

Although conventional differential expression pipelines such as linear models for microarray data (limma) reliably detect individual dysregulated genes and pairwise correlations [20], they cannot evaluate directional causal hypotheses among unobserved biological processes. Structural equation modeling (SEM) addresses this analytical gap by formalizing biological hallmarks as latent variables inferred from multiple observed transcriptomic indicators [21, 22]. Implemented in R via the lavaan package using maximum likelihood estimation with Satorra-Bentler robust scaling [23, 24], SEM

simultaneously assesses measurement model validity and directional structural paths between latent constructs while explicitly accounting for measurement error and residual indicator covariances. Nevertheless, applying SEM to transcriptomic datasets involves inherent methodological limitations: cross-sectional expression profiles capture steady-state mRNA abundance rather than longitudinal dynamics, and moderate clinical sample sizes require parsimonious model specification and cautious evaluation of global fit indices to prevent overfitting [22, 25].

Because primary NPC is treated predominantly with chemoradiotherapy rather than surgical resection, assembling large, homogeneous surgical tissue cohorts remains clinically challenging [26]. Consequently, curated public repositories such as the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) are indispensable for reproducible systems-level discovery in NPC [27]. In particular, dataset GSE12452 (curated as GDS3341) serves as a gold-standard clinical transcriptome resource [28, 29]. Generated on the Affymetrix Human Genome U133 Plus 2.0 array (GPL570), this cohort comprises 41 laser-capture microdissected nasopharyngeal tissue specimens, ensuring high epithelial tumor purity and minimizing stromal confounding [28, 30]. Interrogation of this rigorously annotated dataset enables robust modeling of coordinated transcriptional programs across normal nasopharyngeal epithelium and primary carcinomas.

Building directly upon our prior differential expression findings in GSE12452 [12–16], this bioinformatic study applies causal structural equation modeling to test the directional effect of the epigenetic regulatory hub on Warburg metabolic reprogramming in NPC. By integrating six epigenetic indicators (DNMT1, EZH2, HDAC2, SMARCA4, BRD4, and H19) and five metabolic indicators (GAPDH, LDHA, CS, CYC, and TIMMDC1) within a unified latent variable architecture, we seek to quantify the structural contribution of epigenetic dysregulation to the glycolytic phenotype of nasopharyngeal carcinoma.

2. Materials and Methods

2.1 Transcriptomic Dataset and Laser-Capture Microdissected Cohort

Publicly available microarray expression data were retrieved from the NCBI Gene Expression Omnibus (GEO) under Series accession GSE12452 and curated DataSet record GDS3341 [27–29]. The cohort comprises N = 41 human (Homo sapiens) nasopharyngeal tissue specimens, including 31 histologically confirmed nasopharyngeal carcinoma (NPC) samples spanning clinical TNM stages

(T1N0 to T3N2) and 10 normal nasopharyngeal epithelial control samples. All specimens were isolated via laser-capture microdissection prior to RNA extraction and hybridized onto the Affymetrix Human Genome U133 Plus 2.0 Array platform (GPL570), ensuring epithelial-specific transcriptomic measurements free of stromal or immune cell contamination [28, 30].

2.2 Biomarker Selection and Probe Mapping

Eleven candidate biomarkers were selected a priori based on our previous systems-level transcriptomic analyses of NPC hallmarks [14,15] and mapped to their representative Affymetrix GPL570 probesets. Five indicators represented the Warburg Effect latent construct: GAPDH (Glyceraldehyde-3-Phosphate Dehydrogenase; GeneID 2597; probe AFFX-HUMGAPDH/M33197_3_at), CS (Citrate Synthase; GeneID 1431; probe 208660_at), CYC (Cytochrome c; GeneID 54205; probe 208905_at), TIMMDC1 (Translocase of Inner Mitochondrial Membrane Domain Containing 1; GeneID 51300; probe 223004_s_at), and LDHA (Lactate Dehydrogenase A; GeneID 3939; probe 200650_s_at). Six indicators represented the Epigenetic Landscape latent construct: H19 (H19 Imprinted Maternally Expressed Transcript; GeneID 283120; probe 224348_s_at), DNMT1 (DNA Methyltransferase 1; GeneID 1786; probe 201697_s_at), EZH2 (Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit; GeneID 2146; probe 203358_s_at), SMARCA4 (SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin Subfamily A Member 4; GeneID 6597; probe 213720_s_at), HDAC2 (Histone Deacetylase 2; GeneID 3066; probe 201833_at), and BRD4 (Bromodomain Containing 4; GeneID 23476; probe 238612_at) (Table 1).

2.3 Differential Expression and Multivariate Exploratory Analyses

Normalized log2-transformed expression intensities across all 41 samples were evaluated for differential expression between NPC (n = 31) and normal control (n = 10) groups using empirical Bayes moderated t-statistics implemented in the R package limma [20]. For each probeset, log2 fold change (log2FC), average expression (AveExpr), moderated t-statistic, nominal p-value, Benjamini-Hochberg false discovery rate adjusted p-value (adj.P.Val), and log-odds B-statistic were computed. Statistical significance was defined as adj.P.Val < 0.05. Bivariate linear associations across the 11 genes were quantified using Pearson correlation coefficients. Unsupervised hierarchical clustering with Euclidean distance and complete linkage was visualized as a row-standardized heatmap annotated by experimental group, clinical TNM stage, and biological pathway. Principal component

analysis (PCA) was performed across both sample profiles and functional gene clusters to evaluate variance partitioning and group separability.

2.4 Structural Equation Modeling (SEM)

To test the directional causal hypothesis linking the Epigenetic Landscape (exogenous latent variable) to the Warburg Effect (endogenous latent variable), covariance-based structural equation modeling was conducted in R using the lavaan package (version 0.6-21) [23]. Prior to model estimation, all 11 transcriptomic indicators were standardized to z-scores (mean = 0, standard deviation = 1) to place observed variables on a common scale and facilitate comparison of standardized factor loadings (lambda). Latent variable scales were identified by fixing latent factor variances to unity (std.lv = TRUE).

Given the moderate sample size (N = 41) and potential multivariate non-normality typical of microarray distributions, parameters were estimated via maximum likelihood with Satorra-Bentler scaled test statistics and robust standard errors (estimator = 'MLM', NLMINB optimization) [23, 24]. Based on biological co-regulation within chromatin-modifying complexes and modification indices, two intra-construct residual error covariances were specified within the Epigenetic Landscape measurement block (EZH2 ~~ HDAC2 and HDAC2 ~~ H19). Global model fit was assessed using the Satorra-Bentler scaled chi-square statistic, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Standardized Root Mean Square Residual (SRMR), and Root Mean Square Error of Approximation (RMSEA) with 90% confidence intervals [25].

3. Results

3.1. Differential Transcriptomic Expression of Epigenetic Landscape and Warburg Effect Indicators

Differential expression analysis performed with limma on the laser-capture microdissected GSE12452 cohort (N = 41; 31 nasopharyngeal carcinoma [NPC] samples and 10 normal nasopharyngeal epithelium controls) confirmed significant transcriptional dysregulation across all 11 selected probesets representing the epigenetic landscape and the Warburg effect (Table 2). All five metabolic indicators of the Warburg effect were robustly upregulated in NPC compared with normal tissue: LDHA (log2FC = 1.0207, t = 5.5726, B = 5.1554, adjusted p = 2.45e-06), TIMMDC1 (log2FC = 0.9746, t = 5.8894, B = 6.2392, adjusted p = 1.35e-06), CYC (log2FC = 0.9329, t =

5.2207, B = 3.9663, adjusted p = 6.91e-06), GAPDH (log2FC = 0.9243, t = 5.7108, B = 5.6266, adjusted p = 1.89e-06), and CS (log2FC = 0.7621, t = 5.0887, B = 3.5254, adjusted p = 9.31e-06).

Within the epigenetic landscape construct, five chromatin and DNA methylation regulators exhibited coordinated upregulation in malignant epithelium: EZH2 displayed the strongest positive induction and highest statistical significance across the entire panel (log2FC = 1.4546, t = 6.9758, B = 10.0022, adjusted p = 8.71e-08), followed by HDAC2 (log2FC = 0.8126, t = 4.2804, B = 0.9110, adjusted p = 1.22e-04), SMARCA4 (log2FC = 0.7149, t = 4.1380, B = 0.4697, adjusted p = 1.71e-04), DNMT1 (log2FC = 0.6339, t = 3.5836, B = -1.1762, adjusted p = 7.89e-04), and BRD4 (log2FC = 0.4411, t = 3.6374, B = -1.0222, adjusted p = 7.37e-04). Conversely, the imprinted long non-coding RNA H19 underwent pronounced transcriptional repression in NPC tumors relative to control mucosa (log2FC = -1.8595, t = -6.1173, B = 7.0244, adjusted p = 9.05e-07), establishing an inverse transcriptional polarity within the epigenetic landscape.

3.2. Multivariate Sample Stratification and Bivariate Correlation Architecture

Unsupervised hierarchical clustering of z-scored expression profiles segregated normal control specimens from NPC tumor tissues across clinical stages T1N0 to T3N2, revealing a reciprocal expression pattern characterized by elevated H19 and suppressed metabolic and chromatin-modifying transcripts in controls versus concerted activation of Warburg effect and epigenetic regulators alongside H19 silencing in tumors (Figure 1). Principal component analysis (PCA) at the sample level demonstrated clear separation between experimental groups, with the first two principal components accounting for 73.7% of total variance (PC1 = 62.1%, PC2 = 11.6%; Figure 2, left panel). Feature-level PCA further confirmed functional co-clustering of the five Warburg effect genes (GAPDH, LDHA, CYC, CS, and TIMMDC1) and the five upregulated epigenetic regulators (EZH2, DNMT1, SMARCA4, HDAC2, and BRD4), whereas H19 occupied an opposite coordinate along PC1 (which explained 94.4% of gene-level variance; Figure 2B panel).

Pearson correlation analysis across all 41 samples corroborated strong internal coherence within and between both biological systems (Table 3). Pairwise correlations among Warburg effect indicators were consistently positive and strong, ranging from r = 0.5800 (CS vs. TIMMDC1) to r = 0.8859 (GAPDH vs. LDHA). Similarly, positive inter-correlations linked the upregulated epigenetic regulators, led by EZH2 with HDAC2 (r = 0.7150), EZH2 with SMARCA4 (r = 0.6581), and DNMT1 with SMARCA4 (r = 0.6149). Notably, H19 correlated negatively with every single epigenetic and metabolic indicator, showing its strongest inverse associations with EZH2 (r =

-0.7203), DNMT1 (r = -0.6615), GAPDH (r = -0.6472), SMARCA4 (r = -0.6386), and BRD4 (r = -0.6287).

3.3. Structural Equation Modeling of the Epigenetic Landscape-Warburg Effect Axis

The causal structural equation model estimated via Satorra-Bentler robust maximum likelihood (MLM) in lavaan v0.6-21 converged normally after 38 iterations across 25 free parameters and 41 observations (Figure 3; Table 4). Global goodness-of-fit indices supported an acceptable representation of the observed covariance structure given the sample size, yielding a Satorra-Bentler scaled chi-square of 72.790 (df = 41, p = 0.002; scaling correction factor = 1.008), a robust Comparative Fit Index (CFI) of 0.915, a robust Tucker-Lewis Index (TLI) of 0.886, and a Standardized Root Mean Square Residual (SRMR) of 0.080. As expected in low-degree-of-freedom models fitted to small sample sizes (N = 41), the robust Root Mean Square Error of Approximation (RMSEA) was elevated (0.138; 90% CI: 0.084-0.189).

In the measurement model, all 11 transcriptomic indicators loaded significantly onto their hypothesized latent constructs (all $p < 0.001$). For the exogenous Epigenetic Landscape construct, standardized factor loadings (lambda) were strongly positive for EZH2 (lambda = 0.853, R2 = 0.728), SMARCA4 (lambda = 0.728, R2 = 0.531), DNMT1 (lambda = 0.691, R2 = 0.477), HDAC2 (lambda = 0.658, R2 = 0.433), and BRD4 (lambda = 0.655, R2 = 0.429), whereas H19 exhibited a robust negative loading (lambda = -0.861, R2 = 0.742). For the endogenous Warburg Effect construct, standardized loadings were uniformly high across GAPDH (lambda = 0.942, R2 = 0.888), LDHA (lambda = 0.922, R2 = 0.850), CYC (lambda = 0.883, R2 = 0.780), CS (lambda = 0.873, R2 = 0.762), and TIMMDC1 (lambda = 0.733, R2 = 0.537). Two intra-construct residual error covariances within the epigenetic landscape reached statistical significance: EZH2 with HDAC2 (unstandardized estimate = 0.142, SE = 0.066, z = 2.157, p = 0.031; standardized r = 0.373) and HDAC2 with H19 (unstandardized estimate = 0.193, SE = 0.071, z = 2.730, p = 0.006; standardized r = 0.520).

In the structural model, the Epigenetic Landscape exerted a large, positive, and highly significant direct causal effect on the Warburg Effect (unstandardized beta = 1.356, SE = 0.301, z = 4.501, $p < 0.001$; standardized beta = 0.805). Overall, variation in the latent Epigenetic Landscape accounted for 64.8% of the variance in Warburg Effect reprogramming (R2 = 0.648; unexplained standardized residual variance = 0.352).

4. Discussion

While transcriptomic profiling studies in nasopharyngeal carcinoma have repeatedly cataloged concurrent alterations in chromatin-modifying enzymes and metabolic pathways, conventional differential expression and pairwise correlation frameworks cannot distinguish parallel co-regulation from directional hierarchy [1, 15, 20]. By embedding differential expression outputs within a latent-variable structural equation modeling (SEM) architecture, the present study demonstrates that coordinated remodeling of the epigenetic landscape acts as a primary upstream driver of the Warburg effect in NPC, explaining nearly two-thirds (R2 = 0.648; standardized beta = 0.805) of the variance in glycolytic and mitochondrial metabolic reprogramming. These findings advance our previous bioinformatic investigations of NPC progression [12–16] by establishing a quantitative causal bridge between two foundational hallmarks of cancer [10, 11].

A defining structural feature of the Epigenetic Landscape construct is its bipolar organization, in which transcriptional activation of five chromatin writers, erasers, readers, and remodelers (EZH2, DNMT1, HDAC2, SMARCA4, and BRD4) is counterbalanced by profound silencing of the imprinted lncRNA H19 (log2FC = -1.8595; lambda = -0.861). In Epstein-Barr virus (EBV)-associated NPC, latent viral oncoproteins such as LMP1 and EBNA1 actively recruit host DNA methyltransferases (DNMT1) and Polycomb repressive complex 2 (EZH2) to establish hypermethylated, repressive chromatin states across tumor-suppressive loci [31–33]. Consistent with epigenetic silencing of the H19/IGF2 imprinting control region reported in EBV-infected epithelial cells and NPC tissues [34, 35], loss of H19 removes a critical post-transcriptional and chromatin-level brake on oncogenic signaling. Simultaneously, upregulation of BRD4, HDAC2, and the SWI/SNF catalytic subunit SMARCA4 facilitates super-enhancer assembly and transcriptional elongation at pro-growth and metabolic gene promoters [36–38]. Furthermore, the significant intra-construct residual covariances detected between EZH2 and HDAC2 (r = 0.373) and between HDAC2 and H19 (r = 0.520) reflect direct physical and functional coupling—most notably the recruitment of histone deacetylases into EZH2-containing repressive complexes and lncRNA-guided chromatin scaffolds [39, 40]—beyond the shared variance captured by the global latent factor.

Downstream of this epigenetic perturbation, the endogenous Warburg Effect factor integrates both classical aerobic glycolysis (GAPDH, lambda = 0.942; LDHA, lambda = 0.922) and core mitochondrial tricarboxylic acid cycle and respiratory chain components (CYC, lambda = 0.883; CS, lambda = 0.873; TIMMDC1, lambda = 0.733). Rather than reflecting complete mitochondrial shutdown, modern cancer metabolism recognizes that the Warburg effect in rapidly proliferating

carcinomas relies on heightened glycolytic flux coupled with active mitochondrial anaplerosis and Complex I assembly (mediated by TIMMDC1) to supply biosynthetic precursors and redox equivalents [17, 19, 41]. Epigenetic mechanisms directly orchestrate this dual metabolic program: EZH2-mediated H3K27me3 deposition and DNMT1-driven promoter methylation silence negative regulators of glycolysis (such as FBP1 and VHL), while BRD4 and HDAC2 sustain MYC- and HIF-1alpha-dependent transcription of LDHA, GAPDH, and mitochondrial biogenesis genes [15, 42–44]. Our SEM parameter estimates confirm that these regulatory interactions operate as a unified, system-level causal axis rather than isolated gene-to-gene effects.

From a translational perspective, demonstrating that the epigenetic landscape quantitatively governs Warburg metabolic reprogramming carries direct therapeutic implications for advanced or recurrent NPC. Single-agent metabolic inhibitors targeting LDHA or glycolysis frequently encounter compensatory metabolic plasticity in solid tumors [45]. Conversely, pharmacologic disruption of the upstream epigenetic machinery using clinically advancing EZH2 inhibitors (e.g., tazemetostat), BET/BRD4 inhibitors, or class I HDAC inhibitors may simultaneously collapse both chromatin-mediated immune evasion and downstream Warburg bioenergetics [46–48]. Moreover, the strong factor loadings of EZH2, H19, GAPDH, and LDHA suggest that composite transcriptomic scores derived from these 11 indicators could serve as biomarker panels for stratifying patients likely to benefit from combined epigenetic-metabolic therapies.

Several methodological strengths and limitations warrant careful consideration. A major strength of this investigation is the use of laser-capture microdissected tissue profiles (GSE12452) [28, 30], which minimizes stromal and lymphocytic dilution and ensures that the inferred epigenetic-metabolic coupling reflects intrinsic epithelial tumor biology. Additionally, the application of Satorra-Bentler robust maximum likelihood estimation protects parameter standard errors and test statistics against mild departures from multivariate normality [24]. Nonetheless, three limitations should be acknowledged: first, the sample size (N = 41), although substantial for microdissected primary NPC biopsies, restricts statistical power and accounts for the elevated RMSEA (0.138), a well-documented small-sample behavior of RMSEA in models with modest degrees of freedom where CFI (0.915) and SRMR (0.080) provide more reliable indicators of global fit [25, 49]; second, cross-sectional transcriptomic data establish directional statistical causality under a theoretically specified model rather than temporal longitudinal precedence; and third, mRNA abundance does not capture post-translational modifications or enzymatic flux directly. Future validation in larger multi-center RNA-

seq cohorts and functional perturbation assays in EBV-positive NPC organoids will be essential to confirm causal kinetics.

5. Conclusions

In summary, structural equation modeling of microdissected nasopharyngeal carcinoma transcriptomes demonstrates that coordinated remodeling of the epigenetic landscape—driven by upregulation of EZH2, SMARCA4, DNMT1, HDAC2, and BRD4 alongside marked silencing of H19—exerts a strong, statistically significant causal effect on Warburg metabolic reprogramming (standardized beta = 0.805, $p < 0.001$, $R2 = 0.648$). By bridging differential expression analysis with latent-variable causal inference, this framework provides a rigorous quantitative paradigm for dissecting inter-hallmark dependencies in NPC and supports dual epigenetic-metabolic targeting as a rational therapeutic strategy.

Author Statements

Competing Interests

The authors declare no conflicts of interest.

Author Contributions

Conceptualization, methodology, bioinformatic and statistical analysis, writing—original draft preparation, and writing—review and editing were performed by the authors. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

The authors acknowledge the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database and its contributors for providing public access to the functional genomics datasets utilized in this study. This research received no external funding.

Use of AI

During the preparation of this work, the authors used an advanced AI assistant to support data analysis pipelines in R, refine the grammatical structure of the English manuscript, and ensure

compliance with formatting guidelines. The authors maintain full responsibility for the academic integrity and final validation of the scientific content.

Supplementary Material

Table S1. Normalized log2-transformed expression matrix of the 11 candidate probesets across all 41 laser-capture microdissected specimens in GSE12452.

References

1. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. Lancet. 2019;394(10192):64-80.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49.
3. Bossi P, Chan AT, Licitra L, Trama A, Orlandi E, Hui EP, et al. Nasopharyngeal carcinoma: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2021;32(4):452-65.
4. Zhang Y, Chen L, Hu GQ, Zhang N, Zhu XD, Yang KY, et al. Gemcitabine and Cisplatin Induction Chemotherapy in Nasopharyngeal Carcinoma. N Engl J Med. 2019;381(12):1124-35.
5. Blanchard P, Lee A, Marguet S, Leclercq J, Ng WT, Ma J, et al. Chemotherapy and radiotherapy in nasopharyngeal carcinoma: an update of the MAC-NPC meta-analysis. Lancet Oncol. 2015;16(6):645-55.
6. Chan KCA, Woo JKS, King A, Zee BCY, Lam WKJ, Chan SL, et al. Analysis of Plasma Epstein-Barr Virus DNA to Screen for Nasopharyngeal Cancer. N Engl J Med. 2017;377(6):513-22.
7. Mai HQ, Chen QY, Chen D, Hu C, Yang K, Wen J, et al. Toripalimab or placebo plus chemotherapy as first-line treatment in advanced nasopharyngeal carcinoma: a multicenter randomized phase 3 trial. Nat Med. 2021;27(9):1536-43.
8. Yang Y, Qu S, Li J, Hu C, Xu M, Li W, et al. Camrelizumab versus placebo in combination with gemcitabine and cisplatin as first-line treatment for recurrent or metastatic nasopharyngeal carcinoma (CAPTAIN-1st): a multicentre, randomised, double-blind, phase 3 trial. Lancet Oncol. 2021;22(8):1162-74.
9. Wong KCW, Hui EP, Lo KW, Lam WKJ, Johnson D, Li L, et al. Nasopharyngeal carcinoma: an evolving paradigm. Nat Rev Clin Oncol. 2021;18(11):679-95.
10. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12(1):31-46.

11. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. Science. 2017;357(6348):eaal2380.

12. Romeu C, Oliach MJ, Romeu AR. Transcriptomic Profiling of Highly Expressed Genes Stratifies Nasopharyngeal Carcinoma Progression and Highlights Transferrin Receptor Dysregulation: A Bioinformatic Approach. Research Square [Preprint]. 2026 [cited 2026 Nov 16]. Available from: https://doi.org/10.21203/rs.3.rs-9935130/v1

13. Oliach MJ, Romeu C, Romeu AR, et al. Stage-Independent Downregulation of the GSTA1/GSTA3 Gene Cluster on Chromosome 6p12 as a Metabolic Hallmark of Nasopharyngeal Carcinoma. Research Square [Preprint]. 2026 [cited 2026 Nov 16]. Available from: https://doi.org/10.21203/rs.3.rs-10112864/v1

14. Oliach MJ, Romeu AR. The Warburg Effect Hallmark in Nasopharyngeal Carcinoma: Stage-Specific Up-Regulation of LDHA, GAPDH, CS, CYC, MRPL3, and TIMMDC1 Drives a Hybrid Glycolysis/OXPHOS Profile during Regional Progression. Research Square [Preprint]. 2026 [cited 2026 Nov 16]. Available from: https://doi.org/10.21203/rs.3.rs-10225258/v1

15. Romeu A, Oliach MJ. An Epigenetic Regulatory Hub in Nasopharyngeal Carcinoma: Coordinated Action of DNMT1, EZH2, HDAC2, SMARCA4, BRD4, and H19 during Regional Progression. Research Square [Preprint]. 2026 [cited 2026 Nov 16]. Available from: https://doi.org/10.21203/rs.3.rs-10521122/v1

16. Romeu C, Oliach MJ, Romeu A. The Multi-Pathway Transcriptomic Footprint of Latent EBV-LMP1 Signaling in Advanced Nasopharyngeal Carcinoma: Host TRAF1 Induction as a Transcriptomic Surrogate of Invisible Episomal LMP1 Signaling. Research Square [Preprint]. 2026 [cited 2026 Nov 16]. Available from: https://doi.org/10.21203/rs.3.rs-10945214/v1

17. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science. 2009;324(5930):1029-33.

18. Liberti MV, Locasale JW. The Warburg Effect: How Does it Benefit Cancer Cells? Trends Biochem Sci. 2016;41(3):211-8.

19. Pavlova NN, Zhu J, Thompson CB. The hallmarks of cancer metabolism: Still emerging. Cell Metab. 2022;34(3):355-77.

20. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

21. Kline RB. Principles and Practice of Structural Equation Modeling. 5th ed. New York: Guilford Press; 2023.

22. Grace JB, Schoolmaster DR Jr, Guntenspergen GR, Little AM, Mitchell BR, Miller KM, et al. Guidelines for a graph-theoretic implementation of structural equation modeling. Ecosphere. 2012;3(8):73.

23. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. J Stat Softw. 2012;48(2):1-36.

24. Satorra A, Bentler PM. Corrections to test statistics and standard errors in covariance structure analysis. In: von Eye A, Clogg CC, editors. Latent Variables Analysis: Applications for Developmental Research. Thousand Oaks (CA): Sage Publications; 1994. p. 399-419.

25. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling. 1999;6(1):1-55.

26. Chua MLK, Wee JTS, Hui EP, Chan ATC. Nasopharyngeal carcinoma. Lancet. 2016;387(10022):1012-24.

27. 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. 2013;41(Database issue):D991-5.

28. Sengupta S, den Boon JA, Chen IH, 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(16):7999-8006.

29. Dodd LE, Sengupta S, Chen IH, den Boon JA, Cheng YJ, Westra W, et al. Genes involved in DNA repair and nitrosamine metabolism and those located on chromosome 14q32 are dysregulated in nasopharyngeal carcinoma. Cancer Epidemiol Biomarkers Prev. 2006;15(11):2216-25.

30. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, et al. Laser capture microdissection. Science. 1996;274(5289):998-1001.

31. Tsai CL, Li HP, Lu YJ, Hsueh C, Liang Y, Chen CL, et al. Activation of DNA methyltransferase 1 by EBV LMP1 Involves c-Jun NH(2)-terminal kinase signaling. Cancer Res. 2006;66(24):11668-76.

32. Luo X, Hong L, Cheng C, Li N, Zhao X, Shi F, et al. DNMT1 mediates metabolic reprogramming induced by Epstein-Barr virus latent membrane protein 1 and reversed by grifolin in nasopharyngeal carcinoma. Cell Death Dis. 2018;9(6):619.

33. Tong ZT, Cai MY, Wang XG, Kong LL, Mai SJ, Liu YH, et al. EZH2 supports nasopharyngeal carcinoma cell aggressiveness by forming a co-repressor complex with HDAC1/HDAC2 and Snail to inhibit E-cadherin. Oncogene. 2012;31(5):583-94.

34. Ng A, Tang JP, Goh CH, Hui KM. Regulation of the H19 imprinting gene expression in human nasopharyngeal carcinoma cells by promoter methylation. Int J Cancer. 2003;104(2):179-87.

35. Li X, Lin Y, Yang X, Wu X, He X. Long noncoding RNA H19 regulates EZH2 expression by interacting with miR-630 and promotes cell invasion in nasopharyngeal carcinoma. Biochem Biophys Res Commun. 2016;473(4):913-9.

36. Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013;153(2):320-34.

37. Kadoch C, Hargreaves DC, Hodges C, Elias L, Ho L, Ranish J, et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat Genet. 2013;45(6):592-601.

38. Donati B, Lorenzini E, Ciarrocchi A. BRD4 and Cancer: going beyond transcriptional regulation. Mol Cancer. 2018;17(1):164.

39. van der Vlag J, Otte AP. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. Nat Genet. 1999;23(4):474-8.

40. Luo M, Li Z, Wang W, Zeng Y, Liu Z, Qiu J. Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. Cancer Lett. 2013;333(2):213-21.

41. DeBerardinis RJ, Chandel NS. Fundamentals of cancer metabolism. Sci Adv. 2016;2(5):e1600200.

42. Dong C, Yuan T, Wu Y, Wang Y, Fan TW, Miriyala S, et al. Loss of FBP1 by Snail-mediated repression provides metabolic advantages in basal-like breast cancer. Cancer Cell. 2013;23(3):316-31.

43. Pang Y, Liu J, Li X, Xiao G, Wang H, Yang G, et al. EZH2 promotes metabolic reprogramming in glioblastomas through epigenetic repression of EAF2-HIF1α signaling. Oncotarget. 2016;7(29):45134-43.

44. Stine ZE, Walton ZE, Altman BJ, Hsieh AL, Dang CV. MYC, Metabolism, and Cancer. Cancer Discov. 2015;5(10):1024-39.

45. Boudreau A, Purkey HE, Hitz A, Robarge K, Peterson D, Labadie S, et al. Metabolic plasticity underpins innate and acquired resistance to LDHA inhibition. Nat Chem Biol. 2016;12(10):779-86.

46. Italiano A, Soria JC, Toulmonde M, Michot JM, Lucchesi C, Varga A, et al. Tazemetostat, an EZH2 inhibitor, in relapsed or refractory B-cell non-Hodgkin lymphoma and advanced solid tumours: a first-in-human, open-label, phase 1 study. Lancet Oncol. 2018;19(5):649-59.

47. Stathis A, Bertoni F. BET Proteins as Targets for Anticancer Treatment. Cancer Discov. 2018;8(1):24-36.

48. Hui KF, Yeung PL, Chiang AK. Induction of MAPK- and ROS-dependent autophagy and apoptosis in gastric carcinoma by combination of romidepsin and bortezomib. Oncotarget. 2016;7(4):4454-67.

49. Kenny DA, Kaniskan B, McCoach DB. The Performance of RMSEA in Models With Small Degrees of Freedom. Sociol Methods Res. 2015;44(3):486-507.

Table 1. Candidate transcriptomic biomarkers, functional annotations, latent construct assignments, and Affymetrix GPL570 probeset identifiers.

Latent	Sytem	GeneID	Gene.Symbol	Gene.Title	Epression	ProbeID
Warburg	Warburg Effect	2597	GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase	up	AFFX-HUMGAPDH/M33197_3_at
Warburg	Warburg Effect	1431	CS	Citrate Synthase	up	208660_at
Warburg	Warburg Effect	54205	CYC	Cytochrome c	up	208905_at
Warburg	Warburg Effect	51300	TIMMDC1	Translocase of Inner Mitochondrial Membrane Domain	up	223004_s_at
Warburg	Warburg Effect	3939	LDHA	Lactate Dehydrogenase A	up	200650_s_at
Epigenetics	Epigenetic Regulatory Hub	283120	H19	H19 Imprinted Maternally Expressed Transcript	down	224348_s_at
Epigenetics	Epigenetic Regulatory Hub	1786	DNMT1	DNA Methyltransferase 1	up	201697_s_at
Epigenetics	Epigenetic Regulatory Hub	2146	EZH2	Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit	up	203358_s_at
Epigenetics	Epigenetic Regulatory Hub	6597	SMARCA4	SWI/SNF related BAF chromatin remodeling complex subunit ATPase 4	up	213720_s_at
Epigenetics	Epigenetic Regulatory Hub	3066	HDAC2	Histone Deacetylase 2	up	201833_at
Epigenetics	Epigenetic Regulatory Hub	23476	BRD4	Bromodomain Containing 4	up	238612_at

Eleven differentially expressed biomarkers selected a priori from laser-capture microdissected nasopharyngeal carcinoma (NPC) transcriptomes (NCBI GEO GSE12452 / GDS3341) were mapped to their representative probesets on the Affymetrix Human Genome U133 Plus 2.0 Array (GPL570). Five indicators represent the endogenous Warburg Effect latent construct (GAPDH, CS, CYC, TIMMDC1, and LDHA), and six indicators represent the exogenous Epigenetic Regulatory Hub (Epigenetic Landscape) latent construct (H19, DNMT1, EZH2, SMARCA4, HDAC2, and BRD4). Expression direction indicates transcriptional upregulation (up) or downregulation (down) in primary NPC tumors (n = 31) relative to normal nasopharyngeal epithelial controls (n = 10).

Table 2. Differential expression analysis of Epigenetic Landscape and Warburg Effect indicators in nasopharyngeal carcinoma versus normal nasopharyngeal epithelium.

logFC	AveExpr	t	P.Value	adj.P.Val	B	ProbeID	Gene
0.9243	12.9946	5.7108	6.87E-07	1.89E-06	5.6266	AFFX-HUMGAPDH/M33197_3_at	GAPDH
0.7621	10.8700	5.0887	5.93E-06	9.31E-06	3.5254	208660_at	CS
0.9329	11.6786	5.2207	3.77E-06	6.91E-06	3.9663	208905_at	CYC
0.9746	10.4410	5.8894	3.67E-07	1.35E-06	6.2392	223004_s_at	TIMMDC1
1.0207	12.3991	5.5726	1.11E-06	2.45E-06	5.1554	200650_s_at	LDHA
-1.8595	5.1915	-6.1173	1.65E-07	9.05E-07	7.0244	224348_s_at	H19
0.6339	9.2471	3.5836	7.89E-04	7.89E-04	-1.1762	201697_s_at	DNMT1
1.4546	8.5500	6.9758	7.92E-09	8.71E-08	10.0022	203358_s_at	EZH2
0.7149	9.7224	4.1380	1.40E-04	1.71E-04	0.4697	213720_s_at	SMARCA4
0.8126	9.5335	4.2804	8.85E-05	1.22E-04	0.9110	201833_at	HDAC2
0.4411	7.1044	3.6374	6.70E-04	7.37E-04	-1.0222	238612_at	BRD4

Differential gene expression between laser-capture microdissected NPC tumors (n = 31) and normal nasopharyngeal epithelium controls (n = 10) in dataset GSE12452 was computed using empirical Bayes moderated t-statistics in the R package limma. Abbreviations: logFC, log2 fold change (NPC vs. normal control); AveExpr, average log2-transformed normalized expression intensity across all 41 specimens; t, moderated t-statistic; P.Value, nominal two-sided p-value; adj.P.Val, Benjamini-Hochberg false discovery rate (FDR)-adjusted p-value; B, log-odds B-statistic of differential expression. All 11 probesets exceeded the statistical significance threshold (adj.P.Val < 0.001).

Table 3. Bivariate Pearson correlation matrix across the 11 transcriptomic indicators in the GSE12452 cohort (N = 41).

	GAPDH	CS	CYC	TIMMDC1	LDHA	H19	DNMT1	EZH2	SMARCA4	HDAC2	BRD4
GAPDH	1,0000										
CS	0,8321	1,0000									
CYC	0,8020	0,8298	1,0000								
TIMMDC1	0,6577	0,5800	0,6562	1,0000							
LDHA	0,8859	0,7621	0,8222	0,7246	1,0000						
H19	-0,6472	-0,5128	-0,4669	-0,5522	-0,5522	1,0000					
DNMT1	0,5288	0,4837	0,4005	0,5019	0,4250	-0,6615	1,0000				
EZH2	0,6522	0,6582	0,6817	0,5503	0,6148	-0,7203	0,5966	1,0000			
SMARCA4	0,4550	0,5170	0,4011	0,4859	0,3543	-0,6386	0,6149	0,6581	1,0000		
HDAC2	0,6152	0,6589	0,7723	0,5615	0,5985	-0,3578	0,2989	0,7150	0,4440	1,0000	
BRD4	0,5424	0,5667	0,4638	0,3447	0,5198	-0,6287	0,3856	0,4999	0,5047	0,3190	1,0000

Lower-triangular matrix displaying pairwise Pearson linear correlation coefficients (r) computed from normalized log2-transformed expression intensities (z-score) across all 41 laser-capture microdissected nasopharyngeal samples (31 NPC tumors and 10 normal controls). Positive values indicate direct co-expression among Warburg Effect effectors (GAPDH, CS, CYC, TIMMDC1, LDHA) and upregulated chromatin regulators (DNMT1, EZH2, SMARCA4, HDAC2, BRD4), whereas negative values reflect the inverse transcriptional trajectory of the imprinted long non-coding RNA H19 relative to all 10 activated indicators.

Table 4. Global goodness-of-fit indices and parameter estimates from the causal structural equation model (SEM) linking the Epigenetic Landscape to the Warburg Effect in nasopharyngeal carcinoma (N = 41).

Panel A. Global Goodness-of-Fit Indices and Model Summary

Fit Index / Model Metric	Value	Reference / Notes
Sample size (N) / Free parameters	41 / 25	Converged in 38 iterations (NLMINB)
Estimator	MLM	Maximum likelihood with Satorra–Bentler correction
Satorra–Bentler scaled χ^2 (df, p-value)	72.790 (41, p = 0.002)	Standard χ^2 = 73.370 (p = 0.001); scaling factor = 1.008
Baseline model χ^2 (df, p-value)	424.506 (55, p < 0.001)	Standard baseline χ^2 = 434.385; scaling factor = 1.023
Robust Comparative Fit Index (CFI)	0.915	Acceptable global fit (≥ 0.90)
Robust Tucker–Lewis Index (TLI)	0.886	Marginal/acceptable fit in small-N models
Standardized Root Mean Square Residual (SRMR)	0.080	Acceptable global fit (≤ 0.08)
Robust RMSEA (90% CI)	0.138 (0.084–0.189)	p(RMSEA ≤ 0.05) = 0.008; elevated due to small N and low df
Information Criteria (AIC / BIC / SABIC)	957.730 / 1000.570 / 922.309	Log-likelihood (H0) = −453.865

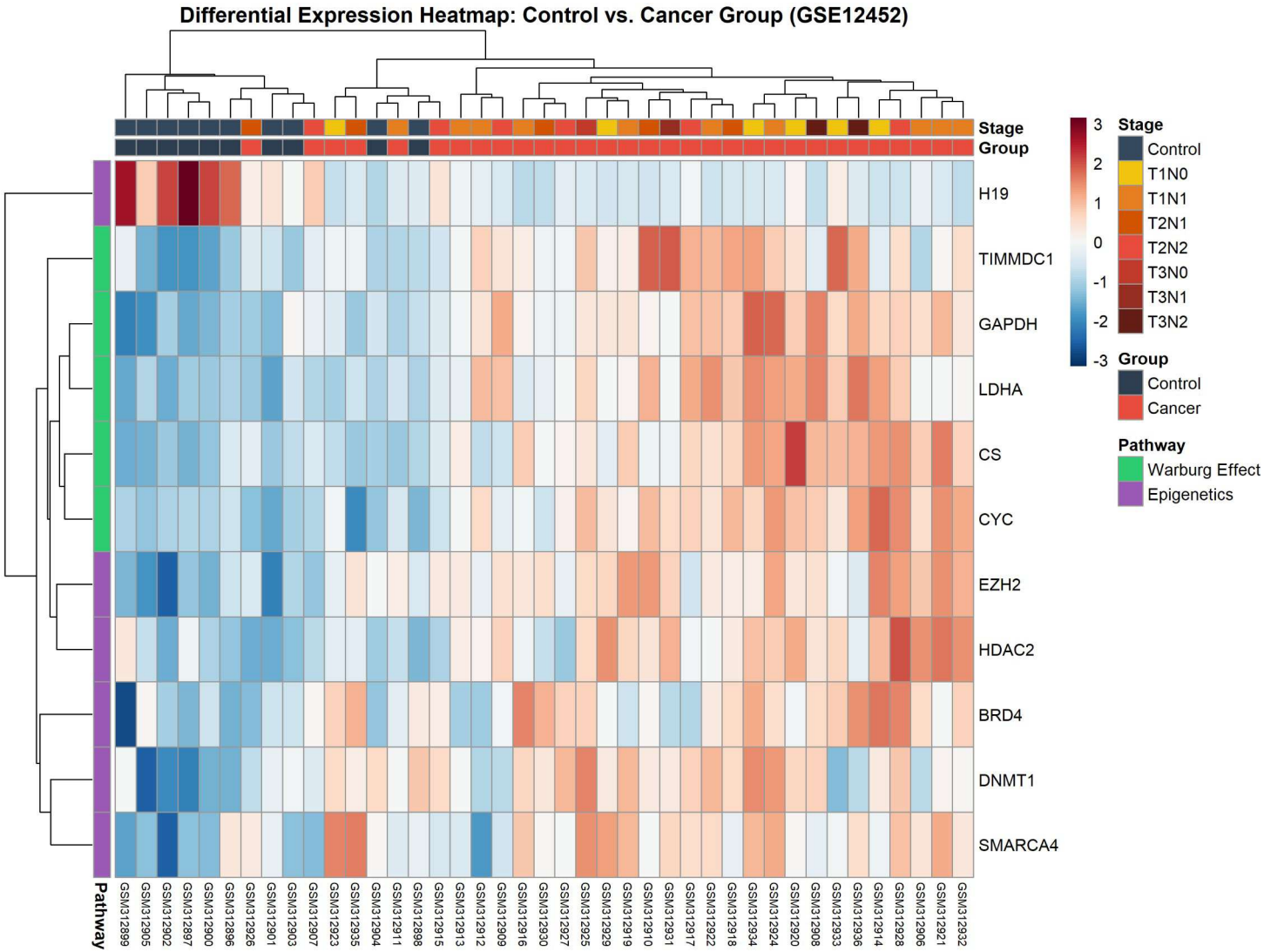
Panel B. Measurement, Structural, Residual Covariance, and Variance Parameter Estimates

Model Component / Parameter	Unstd. Est.	Robust SE	z-value	p-value	Std.all	R^2
Measurement Model: Epigenetic Landscape (λ)						
Epigenetics =~ EZH2	0.843	0.116	7.244	< 0.001	0.853	0.728
Epigenetics =~ SMARCA4	0.720	0.137	5.260	< 0.001	0.728	0.531
Epigenetics =~ DNMT1	0.682	0.149	4.578	< 0.001	0.691	0.477
Epigenetics =~ HDAC2	0.646	0.119	5.414	< 0.001	0.658	0.433
Epigenetics =~ BRD4	0.647	0.133	4.859	< 0.001	0.655	0.429
Epigenetics =~ H19	−0.851	0.162	−5.235	< 0.001	−0.861	0.742
Measurement Model: Warburg Effect (λ)						
Warburg =~ GAPDH	0.552	0.080	6.926	< 0.001	0.942	0.888

Model Component / Parameter	Unstd. Est.	Robust SE	z-value	p-value	Std.all	R^2
Warburg =~ LDHA	0.541	0.084	6.434	< 0.001	0.922	0.850
Warburg =~ CYC	0.518	0.088	5.895	< 0.001	0.883	0.780
Warburg =~ CS	0.512	0.082	6.265	< 0.001	0.873	0.762
Warburg =~ TIMMDC1	0.430	0.084	5.138	< 0.001	0.733	0.537
Structural Regression Path (β)						
Warburg ~ Epigenetics	**1.356**	**0.301**	**4.501**	**< 0.001**	**0.805**	**0.648**
Intra-Construct Residual Covariances (θ)						
.EZH2 ~~ .HDAC2	0.142	0.066	2.157	0.031	0.373	—
.HDAC2 ~~ .H19	0.193	0.071	2.730	0.006	0.520	—
Indicator Residual Variances (θ)						
.EZH2	0.265	0.066	4.031	< 0.001	0.272	—
.SMARCA4	0.458	0.115	3.981	< 0.001	0.469	—
.DNMT1	0.510	0.108	4.702	< 0.001	0.523	—
.HDAC2	0.546	0.101	5.410	< 0.001	0.567	—
.BRD4	0.557	0.093	6.002	< 0.001	0.571	—
.H19	0.252	0.075	3.376	0.001	0.258	—
.GAPDH	0.109	0.036	3.053	0.002	0.112	—
.LDHA	0.146	0.044	3.340	0.001	0.150	—
.CYC	0.215	0.044	4.841	< 0.001	0.220	—
.CS	0.232	0.071	3.273	0.001	0.238	—
.TIMMDC1	0.452	0.115	3.915	< 0.001	0.463	—
Latent Factor Variances (ψ)						
Epigenetics (exogenous variance, fixed)	1.000	—	—	—	1.000	—
.Warburg (unexplained disturbance, fixed)	1.000	—	—	—	0.352	—

Note: Parameter estimates were derived from covariance-based structural equation modeling in R (lavaan v0.6-21) on z-scored transcriptomic profiles across N = 41 nasopharyngeal tissue specimens (31 NPC tumors and 10 normal controls; GSE12452). Latent factor scales were identified by fixing latent variances/disturbances to unity (std.lv = TRUE). Standard errors and test statistics were computed using Satorra–Bentler robust maximum likelihood (estimator = "MLM"). Completely standardized estimates (Std.all) correspond directly to the path coefficients, residual correlations, and residual variances displayed in Figure 3 (rounded to two decimal places in the path diagram). Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; CI, confidence interval; df, degrees of freedom; RMSEA, Root Mean Square Error of Approximation; SABIC, sample-size adjusted BIC; SE, standard error; Unstd. Est., unstandardized estimate.

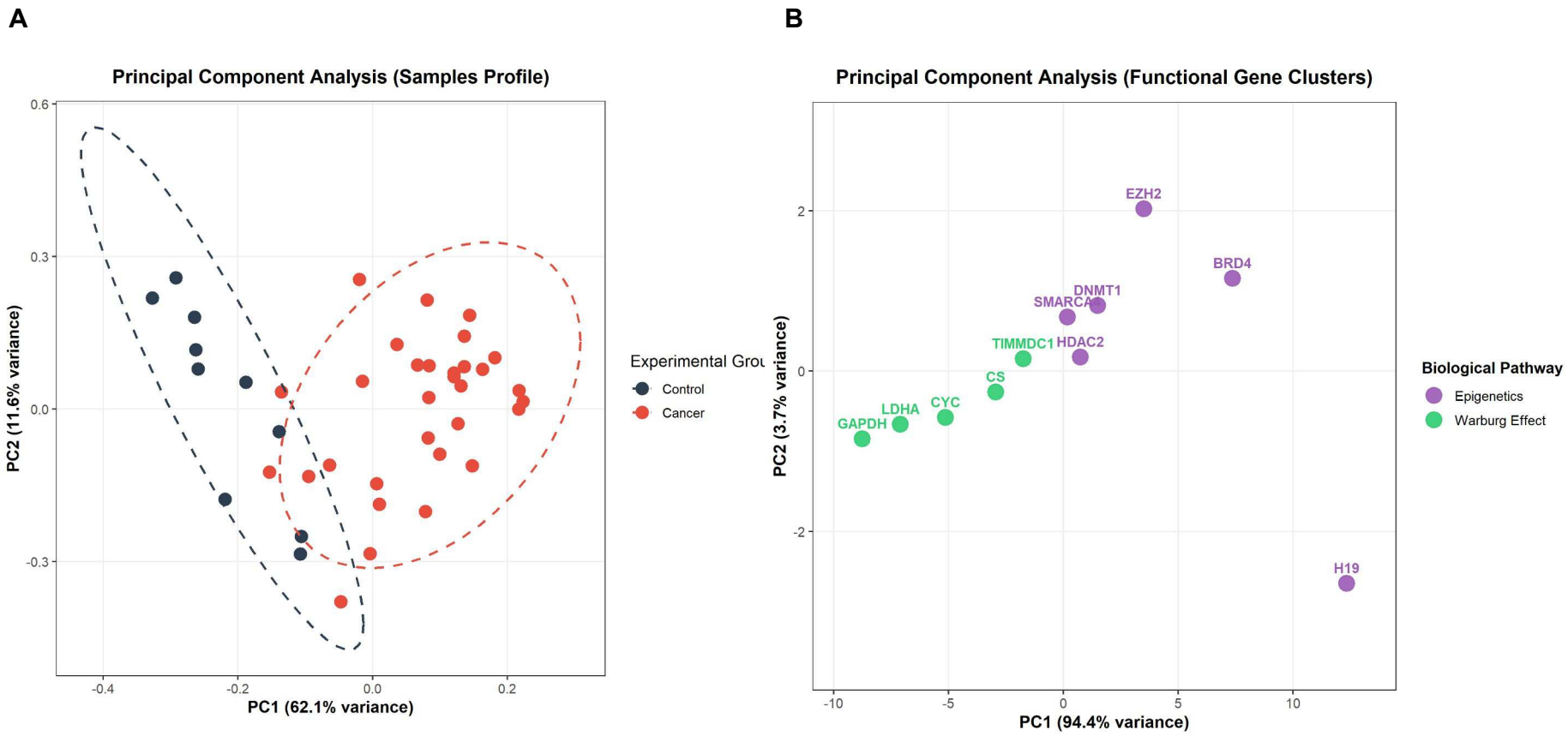
Figure 1. Unsupervised hierarchical clustering of Epigenetic Landscape and Warburg Effect transcriptomic indicators across normal nasopharyngeal epithelium and nasopharyngeal carcinoma stages.



Row-standardized (z-score) expression heatmap of the 11 candidate biomarkers across all 41 laser-capture microdissected specimens from NCBI GEO GSE12452 (10 normal nasopharyngeal controls and 31 primary NPC tumors spanning clinical TNM stages T1N0 to T3N2). Hierarchical clustering was

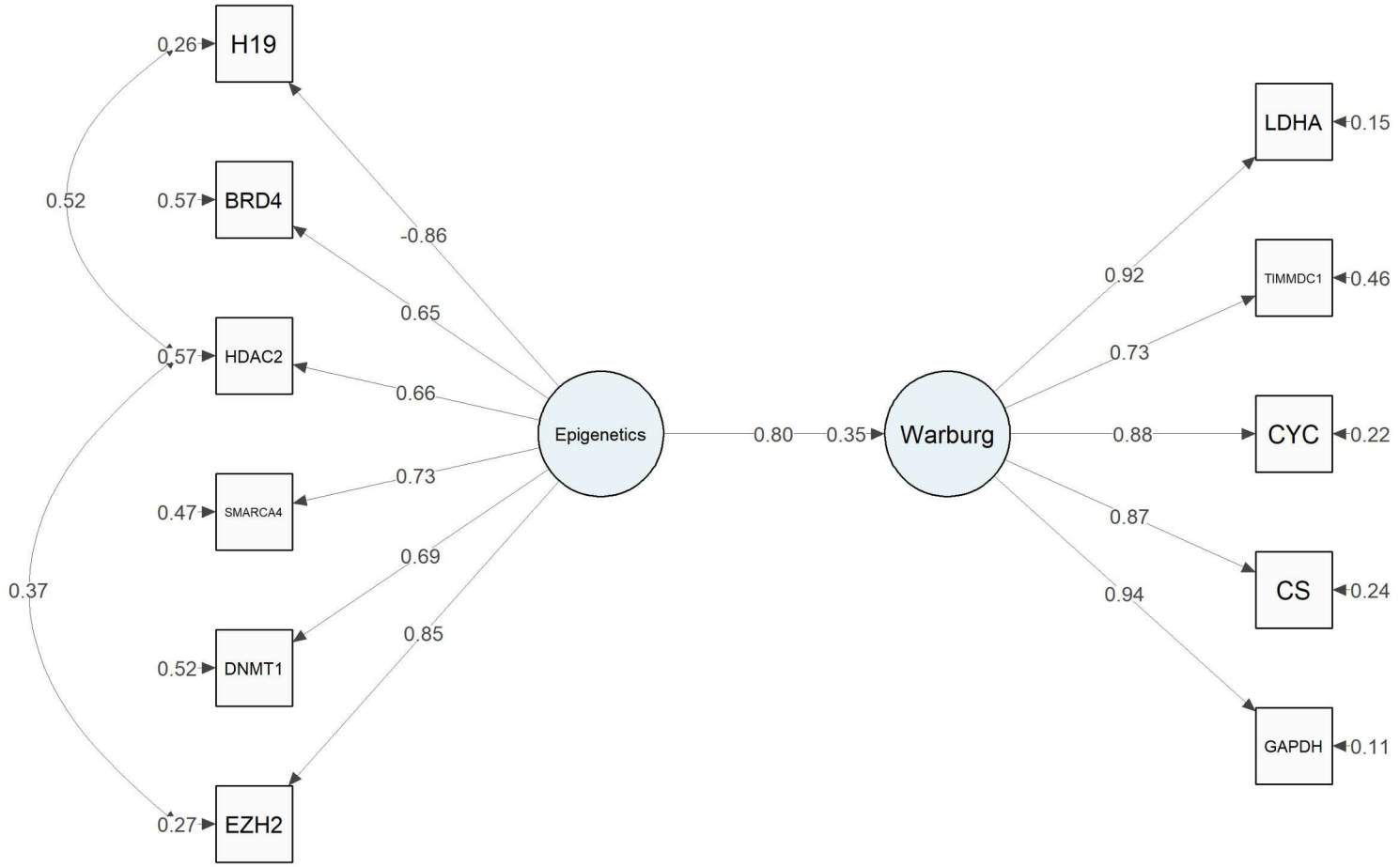
performed using Euclidean distance and complete linkage. Top annotation tracks indicate clinical TNM stage (Stage) and experimental condition (Group: Control vs. Cancer), while the left annotation bar denotes biological pathway assignment (Warburg Effect in green; Epigenetics in purple). Red and blue shading represent relative transcriptional upregulation and downregulation, respectively, illustrating the reciprocal silencing of H19 alongside coordinated activation of chromatin regulators and metabolic enzymes in NPC tumors.

Figure 2. Principal component analysis (PCA) of sample-level separation and functional gene clustering in GSE12452.



(A) Sample-level PCA projection of the 41 laser-capture microdissected nasopharyngeal specimens based on the 11-gene transcriptomic panel, showing clear segregation between normal epithelial controls (dark blue circles, n = 10) and primary NPC tumors (red circles, n = 31) with 95% confidence ellipses. The first two principal components account for 73.7% of total sample variance (PC1 = 62.1%, PC2 = 11.6%). **(B)** Feature-level PCA of functional gene clusters, demonstrating tight co-clustering of the five upregulated Warburg Effect effectors (GAPDH, LDHA, CYC, CS, and TIMMDC1; green) and the five activated epigenetic regulators (EZH2, DNMT1, SMARCA4, HDAC2, and BRD4; purple), contrasted against the opposing coordinate of the silenced imprinted lncRNA H19 along PC1 (94.4% of gene-level variance).

Figure 3. Path diagram of the causal structural equation model linking the Epigenetic Landscape to the Warburg Effect in nasopharyngeal carcinoma.



Completely standardized path diagram estimated via Satorra-Bentler robust maximum likelihood (MLM) in lavaan (N = 41). Circles denote unobserved latent variables (exogenous Epigenetics construct and endogenous Warburg construct), and rectangles denote z-score standardized transcriptomic indicators. Single-

headed straight arrows from latent constructs to observed indicators represent standardized factor loadings (lambda; all $p < 0.001$), confirming positive loadings for EZH2 (0.85), SMARCA4 (0.73), DNMT1 (0.69), HDAC2 (0.66), and BRD4 (0.65), a strong negative loading for H19 (-0.86), and uniformly high positive loadings across GAPDH (0.94), LDHA (0.92), CYC (0.88), CS (0.87), and TIMMDC1 (0.73). Double-headed curved arrows on the left denote statistically significant intra-construct residual error covariances between EZH2 and HDAC2 ($r = 0.37$, $p = 0.031$) and between HDAC2 and H19 ($r = 0.52$, $p = 0.006$). Short double-headed arrows on outer boxes and the endogenous latent factor indicate unexplained standardized residual variances (e.g., 0.35 for Warburg). The central directional structural path demonstrates a strong, direct causal effect of the Epigenetic Landscape on the Warburg Effect (standardized beta = 0.805 [rounded to 0.80 in diagram], $p < 0.001$), explaining 64.8% of its latent variance ($R^2 = 0.648$).

Supplementary Files

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

- ManuscriptMJOARTableS1.pdf