The Journal of Basic and Clinical Dentistry

Original Research

Federated Multi-Omics Learning Reveals a Conserved Epithelial Immune Commitment Axis Associated with Epithelial Immune Commitment in Periodontitis Across Distributed Cohorts

Pradeep Kumar Yadalam[1*]

1. Department of Periodontics, Saveetha Dental College, SIMATS, Saveetha University, Chennai, Tamil Nadu, India.

*** Corresponding Author:** Pradeep Kumar Yadalam, Department of Periodontics, Saveetha Dental College, SIMATS, Saveetha University, Chennai 600077, Tamil Nadu, India. **Email:** pradeepkumar.sdc@saveetha.com Contact: +91-9941912405.

Citation: Yadalam, P.K. Federated Multi-Omics Learning Reveals a Conserved Epithelial Immune Commitment Axis Associated with Epithelial Immune Commitment in Periodontitis Across Distributed Cohorts. J Basic Clin Dent, 2026;3(2), 1-21.

Received: 28th April 2026

Revised: 25th June 2026

Accepted: 26th June 2026

Published: 9th July 2026

Abstract

Background and Objective: Periodontitis is a chronic immune-inflammatory disease driven by dysbiotic microbial challenge at the gingival epithelial interface. Despite substantial molecular profiling of diseased gingival tissue, no federated, privacy-preserving computational framework has identified a cross-cohort-conserved epithelial molecular axis governing disease irreversibility. This study aimed to identify and validate a conserved epithelial immune commitment axis in periodontitis using biologically constrained federated graph learning across distributed transcriptomic cohorts. **Materials and Methods:** Two large bulk RNA microarray datasets GSE10334 (n = 247; GPL570) and GSE16134 (n = 310; GPL570) were treated as independent federated nodes. Within each node, quantile normalization, Welch-corrected differential expression with Benjamini–Hochberg false discovery rate control, and construction of a 190-gene epithelial-immune panel-based gene-gene graph (pathway co-membership + Spearman $|\rho| \geq 0.50$ edges) preceded local graph convolutional network (GCN) training. A federated server aggregated local model weights using biologically constrained FedAvg (FedBio), weighting gene contributions by cross-node differential expression concordance. An Epithelial Commitment Score (ECS) was derived from the consensus differential expression signature. Single-cell validation was performed in GSE164241 (n = 18,142 post-QC cells; **Results:** Cross-node differential expression concordance was exceptional (Spearman $\rho = 0.946$; panel $\rho = 0.991$; $p < 1 \times 10^{-300}$). The ECS achieved AUROC = 0.909 (Mann–Whitney U $p = 1.2 \times 10^{-22}$) in GSE10334 and AUROC = 0.926 ($p = 1.8 \times 10^{-27}$) in GSE16134. FedBio improved over plain FedAvg (AUROC: +0.031 and +0.035, respectively). The local GCN underperformed on cross-node transfer (AUROC 0.672–0.686), indicating that graph-topological generalization requires larger federated networks. **Conclusion:** This work identifies a conserved Epithelial Immune Commitment Axis in periodontitis, marked by barrier collapse (CDH1/EPCAM/OCLN) and inflammatory escalation (MMP7/IL-1β/ZBP1/CXCL).

Clinical application: The ECS may act as a privacy-preserving biomarker to identify irreversible epithelial inflammatory commitment in periodontitis and guide targeted therapy.

Keywords: Periodontitis, Federated learning, Graph neural network, Epithelial immune commitment

1. Introduction

Periodontitis, the sixth most prevalent disease globally, destroys tooth-supporting alveolar bone and soft tissue through a chronic dysbiotic immune-inflammatory cascade that, once established, exhibits a strikingly irreversible tissue destruction phenotype despite effective supragingival biofilm control.[1] The gingival epithelium is not a passive bystander in this process; it functions as both a mechanical sentinel and an active immune transducer, whose molecular commitment to an inflammatory program is arguably the earliest cellular event preceding the onset of full-blown periodontitis.[2] Understanding the epithelial shift from barrier homeostasis to inflammatory commitment at the molecular level across diverse patient groups remains a key unsolved challenge in periodontal biology.

The post-genomic landscape of periodontal research has generated substantial transcriptomic evidence characterizing diseased gingival tissue. Early microarray studies identified upregulation of matrix metalloproteinases, pro-inflammatory cytokines, and pattern-recognition receptor pathways in diseased versus healthy gingival biopsies.[3] Subsequent analyses implicated the NLRP3 inflammasome, NF-κB signaling cascades, and cGAS-STING innate immune sensing in driving chronic gingival inflammation. Single-cell RNA sequencing of gingival tissue has begun to resolve cell-type-specific contributions, revealing that epithelial, fibroblast, and immune compartments each exhibit distinct disease-state transcriptional programs. Despite these advances, three critical knowledge gaps persist.[4,5]

Existing studies analyze single cohorts but haven't rigorously tested the conservation of molecular signatures across independent populations using cross-cohort federated frameworks.[6] Cross-dataset validation often involves raw data pooling violating privacy or replicates findings without quantifying distributional concordance. No framework currently integrates the full molecular complexity of the epithelial-immune interface into a single interpretive score without raw data sharing. While single-cell studies catalog gingival cell types, they haven't systematically linked bulk federated signatures back to single-cell data. -cell epithelial state space, limiting the mechanistic grounding of tissue-level biomarkers.[7]

Graph neural networks (GNNs) provide a mathematical framework for learning from gene interaction graphs, in which gene expression features propagate along pathway-constrained edges to create sample-level representations. Combined with federated learning where model weights but not raw data leave local nodes GNNs enable privacy-preserving multi-institutional discovery. Recent work in federated learning for oncogenomics shows that biologically constrained weight aggregation improves convergence and plausibility compared to naive FedAvg. However, this approach hasn't been applied to periodontitis, and the benefits of graph-structured models versus classical feature-engineered classifiers in this disease remain unknown.[8]

This study addresses three gaps using a two-node federated graph learning architecture on the GSE10334 and GSE16134 datasets, with a 190-gene epithelial-immune panel as a prior. It develops a biologically constrained FedBio aggregation scheme, defines an Epithelial Commitment Score (ECS) based on consensus differential expression, and validates it using single-cell RNA sequencing data from GSE164241. The evaluation across six classifiers and four

federated setups provides the first benchmark for federated GNNs in periodontal transcriptomics. The resulting epithelial axis, characterized by barrier erosion and inflammation, offers mechanistic insights and a clinically useful biomarker.

2. Materials and Methods

2.1 Data Sources and Cohort Characteristics

Two bulk RNA microarray datasets were retrieved from the NCBI Gene Expression Omnibus (GEO). GSE10334 (GPL570 Affymetrix HG-U133 Plus 2.0; n = 247; 183 periodontitis, 64 healthy) and GSE16134 (GPL570; n = 310; 241 periodontitis, 69 healthy) provided gingival biopsy expression profiles from subjects with well-characterized clinical diagnoses of chronic periodontitis versus periodontally healthy controls. Both datasets were analyzed as independent federated nodes, with no raw data shared across nodes at any stage of the pipeline. For single-cell validation, GSE164241 (10x Genomics Chromium; GRCh38) provided gingival scRNA-seq data from 6 subjects: 3 healthy and 3 with periodontitis. [9,10]

2.2 Preprocessing Within Each Federated Node

Series matrix files were downloaded from the GEO FTP archive and parsed to extract expression matrices and sample-level phenotype metadata. GPL570 probe-to-gene mapping used the GEO GPL570 annotation file; multiple probes mapping to the same gene symbol collapsed by selecting the probe with the highest mean expression across samples. Expression values were already in $\log_2$ scale (observed range: 2.4–14.6, consistent with RMA-processed summaries); no additional logarithmic transformation was applied. Within each node, quantile normalization was applied independently across all samples. Phenotype labels were harmonized from GEO sample-characteristics fields, classifying samples as "healthy" (keywords: healthy, unaffected, control) or "diseased" (keywords: diseased, periodontitis).

2.3 Differential Expression Analysis

Per-node differential expression was computed using Welch's two-sample t-test with Benjamini–Hochberg (BH) false discovery rate (FDR) correction. Given the well-powered sample sizes (node 1: 183 vs. 64; node 2: 241 vs. 69), this approach yields robust, reproducible differential expression estimates. Significance was defined as FDR < 0.05 with $|\log_2 FC| \geq 0.585$. Cross-node concordance was quantified by Spearman's rank correlation of $\log_2 FC$ values.

2.4 Epithelial-Immune Gene Panel Construction

A 190-gene epithelial-immune panel was curated from 13 functional modules informed by established periodontitis biology and recent transcriptomic studies: (1) Epithelial Barrier *(KRT5/14/19, EPCAM, CDH1, DSG1/3, CLDN1, OCLN, TJP1/2, PKP1/3, IVL, FLG, CDSN; n = 26 genes); (2) TLR/NOD Signaling (TLR1–9, NOD1/2, MYD88, IRAK1/4, TRAF6, CARD9;* n =

20); (3) NF-κB Signaling (*NFKB1/2, RELA, RELB, NFKBIA, TNFAIP3; n = 14); (4*) Interferon Response *(IFI6/27/44/44L, IFIT1–3, ISG15, MX1/2, OAS1–3, STAT1/2, IRF1/3/7; n = 21); (5) IL-1/IL-6/TNF (IL1A/B, IL1RN, IL6, TNF, CXCL1/6/8, CCL2/5, S100A8/9; n = 17); (6)* Inflammasome *(NLRP1/3, AIM2, PYCARD, CASP1/4, GSDMD, IL18; n = 13); (7) cGAS-STING (CGAS, STING1, IFI16, DDX41, ZBP1, TBK1, IRF3; n = 9); (8)* Mitochondrial/ROS Stress *(SOD1/2, GPX4, NFE2L2, HMOX1, TXN, TXNIP, BNIP3, PINK1, DNM1L, OPA1; n = 23); (9) Apoptosis (BAX, BCL2, CASP3/7/8/9, FAS, TP53, DIABLO; n = 19); (10)* Pyroptosis *(GSDMD, CASP1/4, NLRP3, NINJ1; n = 10); (11)* Senescence *(CDKN1A/2A/2B, TP53, SERPINE1, GLB1, IL6; n = 12); (12)* ECM Degradation *(MMP1/3/7/9/12/13/14, TIMP1–4, ADAMTS4/5, CTSK/L/SS; n = 21); (13)* Inflammatory Epithelial *(S100A7/8/9, DEFB4A, CXCL1/5, PI3; n = 5*). Of 284 panel genes, 190 were present on both GPL570 arrays and used in all downstream analyses. The 190-gene panel was predefined from modules related to the epithelial barrier, innate immunity, inflammasomes, mitochondrial stress, and ECM.

2.5 Gene-Gene Graph Construction

Within each federated node, a gene-gene graph was constructed using two complementary edge types. Pathway prior edges connected all gene pairs sharing at least one of the 13 curated modules (initial edge weight = 1.0). Correlation edges were added between gene pairs with a within-node Spearman $|\rho| \geq 0.50$, computed from training-fold samples only to prevent test leakage. When both evidence types supported a gene pair, the edge weight was set to the maximum of the two edge weights. The resulting adjacency matrix was symmetrically normalized using the Kipf-Welling formulation: $\hat{A} = D^{-1/2} (A + I) D^{-1/2}$, where D is the degree matrix and I is the identity matrix for self-loops.

2.6 Local Graph Convolutional Network Architecture

Each federated node trained an identical local GCN implemented in PyTorch. The architecture comprised two graph convolutional layers followed by a mean graph readout and a two-layer MLP classification head. Specifically: per-sample per-gene z-scored expression (optionally scaled by bio-consensus weights) was used as a scalar node feature ($X \in \mathbb{R}N\times1$); GCN Layer 1 projected to hidden dimension H = 32 ($W_1 \in \mathbb{R}H\times1$), ReLU activation, followed by Â-propagation; GCN Layer 2 projected to embedding dimension E = 16 ($W_2 \in \mathbb{R}E\times H$), ReLU, Â-propagation; mean readout over gene nodes produced a sample embedding $\in \mathbb{R}16$; MLP head: Linear (16, 16) → ReLU → Dropout (p = 0.20) → Linear (16, 1) → logit. Binary cross-entropy with logits loss used class-imbalance-corrected positive class weight (pos_weight = n_healthy / n_diseased). Optimization used Adam (learning rate = 5×10^{-3}; weight decay = 1×10^{-4}). Local training ran for 200 epochs per fold.

2.7 Baseline Classifiers

Six non-graph classifiers were evaluated using identical feature inputs (z-scored panel gene

expression): (i) Multilayer Perceptron (MLP): hidden = 64, embedding = 16, Dropout(0.30), 300 epochs, Adam lr = 5×10^{-3}, weight decay = 1×10^{-4}; (ii) Logistic Regression (LR): L2 penalty, C = 1.0, class_weight = "balanced", max_iter = 2,000; (iii) Random Forest (RF): n_estimators = 300, class_weight = "balanced"; (iv) Gradient Boosted Trees (XGBoost): n_estimators = 300, max_depth = 4, learning_rate = 0.05, scale_pos_weight computed from training-fold class ratio; (v) Support Vector Machine (SVM): C = 1.0, RBF kernel, class_weight = "balanced"; (vi) Pooled GCN: trained on combined normalized features from both nodes simultaneously. Graph edges and normalization parameters were learned only during training, and hyperparameters were based on prior GCN/federated-learning studies and set before final testing.

2.8 Federated Learning Protocol

Federated training alternated between local and global phases. In each communication round, each node trained its local GCN for 10 epochs on its private training samples and then transmitted the model weights (θ_1, θ_2) to the federated server. FedAvg aggregated global weights as: $\theta\_global = \Sigma_k (n_k/N)\, \theta_k$, where n_k is the number of training samples at node k, and N is the total. The global model was then broadcast back to each node. FedBio extended FedAvg by applying a per-gene biological weight ($w\_g \in [0.5, 2.0]$) to scale gene expression inputs at each node before local training. Per-gene weights were computed as: $w\_g = 0.5 + 1.5 \times$ (normalized [sign_agreement $\times \sqrt{(|FC\_g1| \times |FC\_g2|)} + 0.5 \times \sqrt{(sig\_g1 \times sig\_g2)}$]), where sign_agreement is 1 when both nodes show the same direction of effect, and $sig\_g = \sqrt{(-\log_{10}(FDR\_g1) \times -\log_{10}(FDR\_g2))}$. All federated runs used 20 communication rounds, each with 10 local epochs.

2.9 Evaluation Protocol

Performance was assessed using 5-fold stratified cross-validation, with the graph topology reconstructed solely from the training samples to prevent the test fold from influencing edge weights. Metrics included AUROC, AUPRC, accuracy, sensitivity, specificity, F1 Score, and Brier score. 95% confidence intervals for AUROC were estimated using 500 bootstrap resamples. Cross-node validation is trained on one node's samples and tested on the other. Ablation studies tested five FedBio conditions: Full_FedBio, No Graph, No Mito, No Inflammasome, No Pathway Prior, and No SC Grounding.

2.10 Epithelial Commitment Score Derivation

The ECS measures each sample's position on the epithelial immune commitment axis. The consensus signature vector for each gene g is defined as: s_g = sign_agreement_g $\times$ $\text{sign}(\text{mean\_log}_2\text{FC}) \times \sqrt{(|FC\_g,GSE10334| \times |FC\_g,GSE16134|)} + 0.3 \times \sqrt{(sig\_g,GSE10334 \times sig\_g,GSE16134)}$. The ECS for each sample i is calculated as $ECS\_i = \Sigma\_g\, s\_g \times z\_g$, where z_g, i is the gene's z-scored expression in sample i. A higher ECS indicates more inflammation; a lower ECS suggests barrier homeostasis.

2.11 Study Design and Federated Multi-Omics Workflow

Figure 1 illustrates the federated learning architecture in detail. Node 1 (GSE10334) and Node 2 (GSE16134) each maintained local raw data that was never transmitted. Each node trained a local GCN and sent its model weights, θ_1 and θ_2, to the central server, which returned a consensus global model, θ_global. The biological weighting function w_g = f (consensus DEG) integrated cross-node sign agreement and geometric mean $\log_2$FC. It used FDR significance to scale genes from 0.5 (discordant or non-significant) to 2.0 (strongly concordant and significant). This design boosted the influence of genes such as IL1B, ZBP1, MMP7, CDH1, and OCLN, driving the conserved epithelial commitment axis in the federated aggregation over node-specific noise.

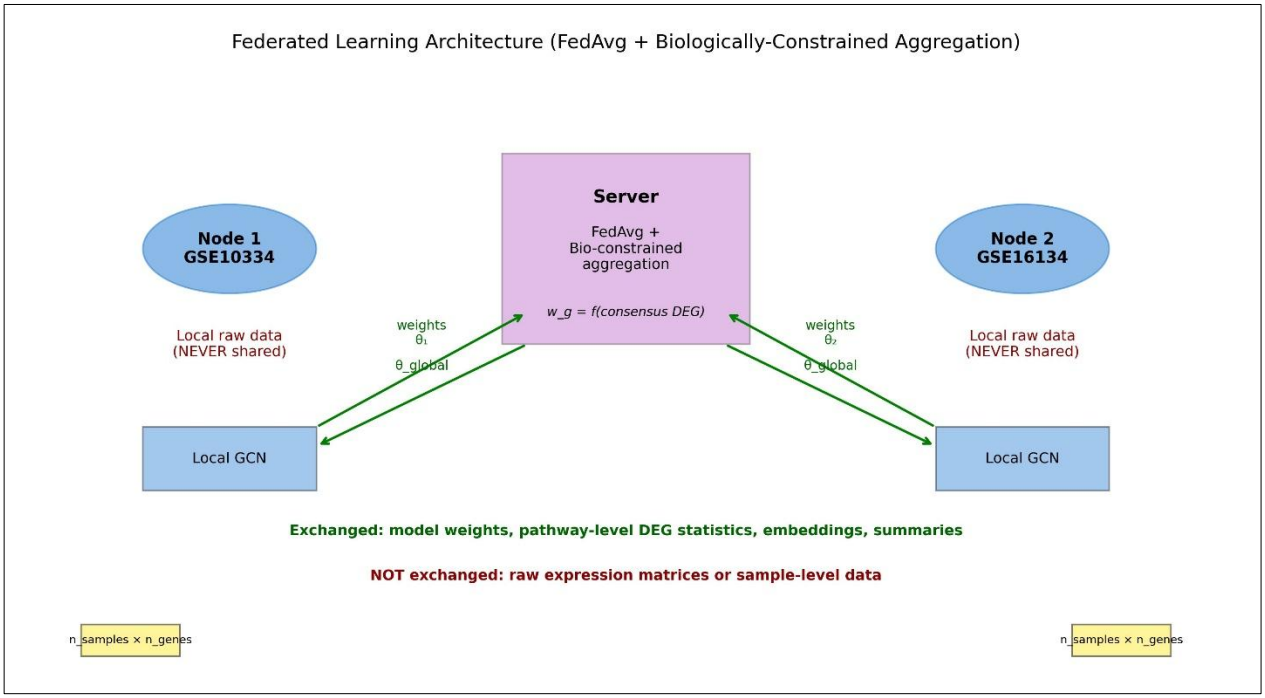

Figure 1. Federated Learning Architecture (FedAvg + Biologically Constrained Aggregation). Node 1 (GSE10334) and Node 2 (GSE16134) hold local raw data that are never shared. Each node trains a Local GCN and transmits model weights (θ_1, θ_2) to the federated server, which computes a biologically constrained aggregate using w_g = f (consensus DEG). The server broadcasts the updated global model (θ_global) back to each node. Only model weights, pathway-level DEG statistics, and sample embeddings are exchanged; raw expression matrices or sample-level data are never transmitted.

2.12 Local Graph Neural Network Architecture

Figure 2 depicts the per-node GCN architecture. Per-sample z-scored gene expression for each panel gene serves as a scalar node feature ($X \in \mathbb{R}^{Nx1}$, N = 190 genes). Two successive GCN layers apply symmetrically normalized Â-propagation: Layer 1 maps 1 → H = 32 channels ($W_1 \in \mathbb{R}^{Hx1}$) with ReLU activation; Layer 2 maps 32 → E = 16 channels ($W_2 \in \mathbb{R}^{ExH}$) with ReLU. Mean readout over all gene nodes yields a 16-dimensional sample embedding, which is passed to an MLP head (16 → 16 → 1) for disease/healthy logit output.

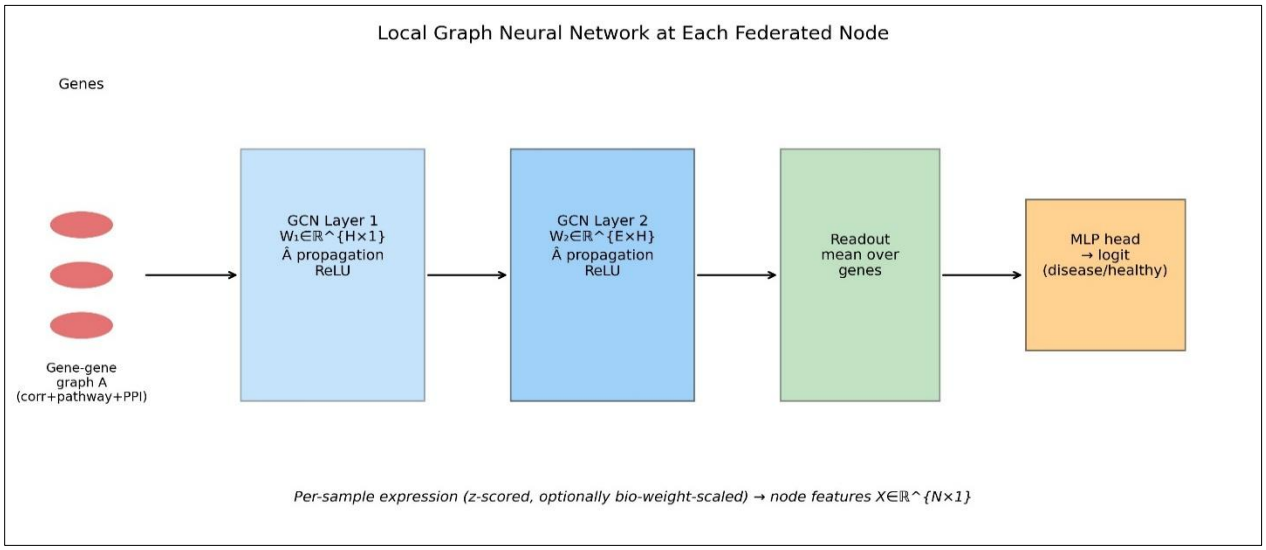

Figure 2. Local Graph Neural Network at Each Federated Node. Per-sample expression (z-scored, optionally bio-weight-scaled) serves as scalar node features ($X \in \mathbb{R}^{Nx1}$). GCN Layer 1 ($W_1 \in \mathbb{R}^{Hx1}$, H = 32) applies Â-propagation followed by ReLU. GCN Layer 2 ($W_2 \in \mathbb{R}^{ExH}$, E = 16) applies a second Â-propagation followed by ReLU. A mean readout over all gene nodes produces a sample embedding, which feeds into a two-layer MLP head that yields a disease-versus-healthy logit.

2.13 Single-Cell RNA-seq Processing

GSE164241 gingival 10x Chromium data from 6 samples (3 healthy, 3 periodontitis) were processed with Scanpy (v1.9+). Quality control excluded cells with fewer than 200 detected genes, fewer than 5 cells with those genes, a mitochondrial fraction > 25%, or more than 7,000 detected genes per cell, resulting in 18,142 cells post-QC. Normalization used total count normalization (target sum = 10,000) followed by $\log_1 p$ transformation. The top 3,000 highly variable genes were selected using the Seurat method with the sample as the batch key. Principal component analysis used 40 components; neighborhood graph construction used 15 principal components and k = 15 nearest neighbors. UMAP embedding used random_state = 42. Leiden clustering of the full atlas used resolution = 0.60; epithelial reclustering used n_top_genes = 1,500, 20 PCA components, 15 neighbors, 15 PCs, and Leiden resolution = 0.40. Cell types were annotated by canonical marker gene module scoring. Epithelial subcluster functional characterization used 13 curated pathway modules scored with scanpy.tl.score_genes. The EICA was computed as I_score − B_score, where I_score is the mean of seven inflammatory module scores, and B_score is the barrier module score.

2.14 Statistical Analysis and Software

All statistical analyses were implemented in Python 3.10. Differential expression used scipy.stats.ttest_ind (Welch, equal_var = False) with stats model's multiple tests (method = "fdr_bh"). Cross-node concordance was computed using scipy.stats.spearmanr. Mann–Whitney U tests were performed using scipy.stats.mannwhitneyu. Classification metrics were computed using scikit-learn (v1.3). Neural networks implemented with PyTorch (v2.0). Graph operations with

NetworkX (v3.1). Single-cell analysis was performed using Scanpy (v1.9) and AnnData (v0.9), and optionally BBKNN for batch-aware neighbor computation. All random seeds are fixed at 42. Figure generation was performed using Matplotlib (v3.7) and Seaborn (v0.12). No synthetic data were generated; all results from publicly available GEO datasets.

3. Results

3.1 Cross-Node DEG Concordance and the Conserved Gene Network

Cross-node differential expression concordance was remarkable. Across all 21,755 genes present on GPL570, Spearman $\rho = 0.946$ ($p < 1 \times 10^{-300}$). Within the 190-gene epithelial-immune panel, concordance was near-perfect: Spearman $\rho = 0.991$ ($p = 2.30 \times 10^{-168}$). This extraordinary reproducibility across independently recruited patient cohorts establishes that the epithelial immune commitment axis is a genuine biological property of the disease rather than a cohort-specific artifact.

Figure 3 visualizes the 190-node gene network for Node 1 (GSE10334), with node color encoding the cross-node consensus DEG signature (red = disease-up; blue = healthy-up) and node size proportional to $|\log_2 FC|$. Blue pathway prior edges, purple pathway-plus-correlation edges, and gray correlation-only edges reflect the multi-evidence graph topology. Disease-upregulated hub genes with large nodes and high network connectivity include MMP7, MMP9, MMP14, IL1B, ZBP1, SAA2, CXCL6, CXCL1, CASP3, and CTSS. Healthy-upregulated hub genes — large blue nodes — include CDH1, EPCAM, OCLN, PKP1, PKP3, TJP1, TJP2, IVL, TXN, and BNIP3. The network structure reveals that disease-up genes cluster in the NF-κB/inflammasome/cGAS-STING and MMP/ECM modules, whereas healthy-up genes cluster in the tight-junction/desmosome/barrier module, with sparse cross-module connectivity.

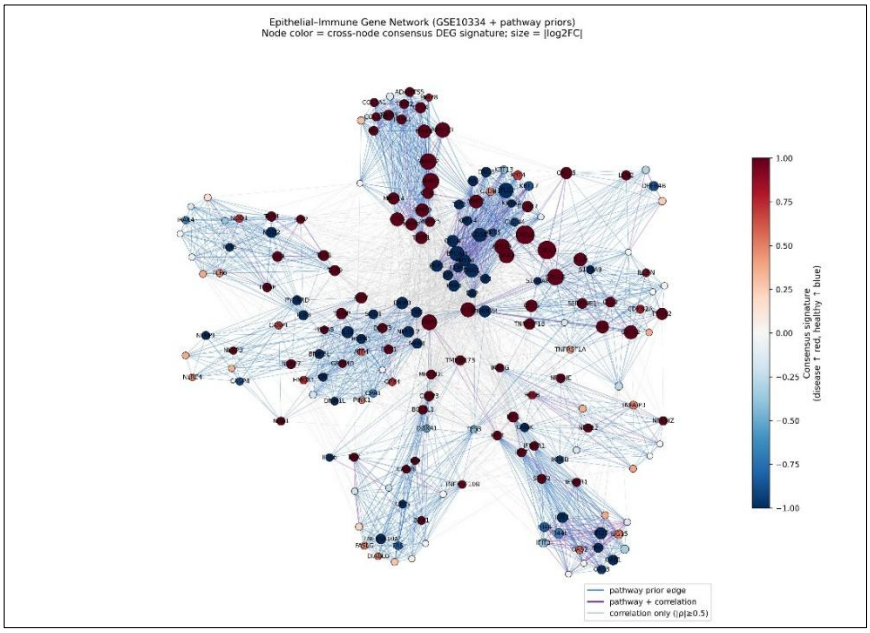

Figure 3. Epithelial-Immune Gene Network (GSE10334 + Pathway Priors). Node color encodes cross-node consensus DEG signature (red = disease-up; blue = healthy-up) and node size encodes $|\log_2 FC|$. Blue edges: pathway prior co-

membership; purple edges: pathway + correlation ($|\rho| \geq 0.50$); gray edges: correlation only. Disease-up hubs include MMP7, ZBP1, IL1B, SAA2, CXCL6, and CASP3; healthy-up hubs include CDH1, EPCAM, OCLN, PKP1, PKP3, and TJP1/2.

3.2 Epithelial Commitment Score Separates Disease and Healthy States

Figure 4 shows ECS violin plots for GSE10334 and GSE16134. In both datasets, the ECS robustly separates diseased from healthy gingival samples. In GSE10334, healthy samples cluster at ECS << 0 (mean approximately −220) while diseased samples shift to ECS >> 0 (mean approximately +80); Mann–Whitney U $p = 1.2 \times 10^{-22}$; ECS-AUROC = 0.909. In GSE16134, the same directional separation is observed with Mann–Whitney U $p = 1.8 \times 10^{-27}$; ECS-AUROC = 0.926. The ECS thus performs as a competitive unsupervised biomarker comparable to the best supervised classifiers while retaining full biological interpretability as a weighted linear combination of consensus differential expression.

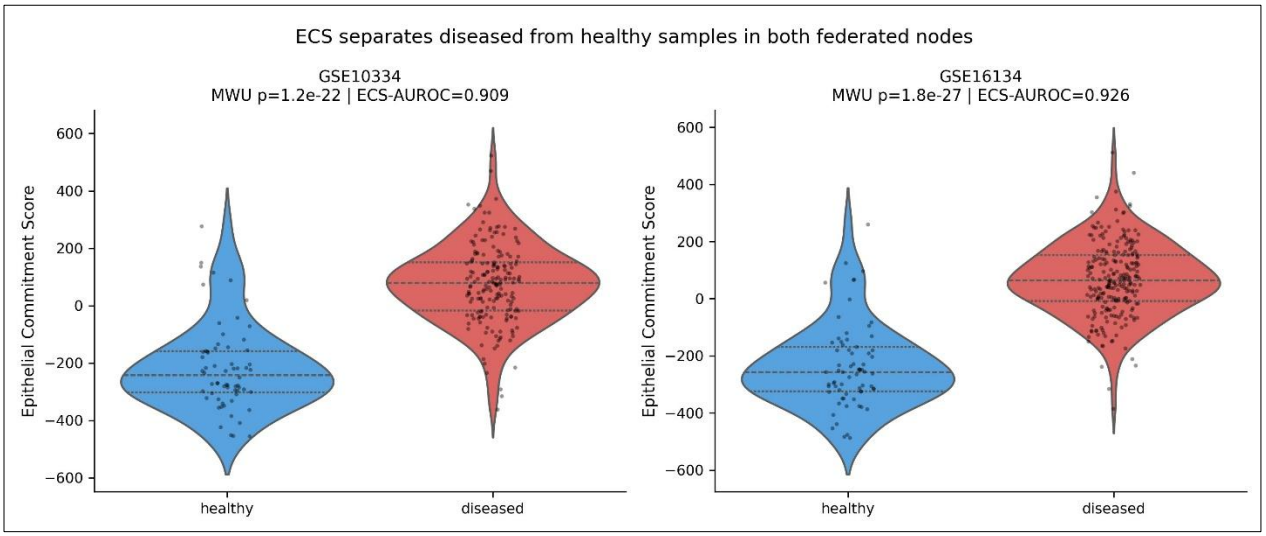

Figure 4. ECS Separates Diseased from Healthy Samples in Both Federated Nodes. Violin plots of Epithelial Commitment Score (ECS) stratified by phenotype (blue = healthy; red = diseased) in GSE10334 (left) and GSE16134 (right). Dashed lines indicate quartiles. GSE10334: Mann–Whitney U $p = 1.2 \times 10^{-22}$, ECS-AUROC = 0.909. GSE16134: Mann–Whitney U $p = 1.8 \times 10^{-27}$, ECS-AUROC = 0.926. Individual samples are shown as jittered dots; gray dots indicate outliers.

3.3 Classifier Performance: ROC and Precision-Recall Analysis

Figure 5 presents ROC curves (left) and precision-recall curves (right) for all nine evaluated model configurations, separately for each federated node with 5-fold cross-validation. Results are summarized in Table 1.

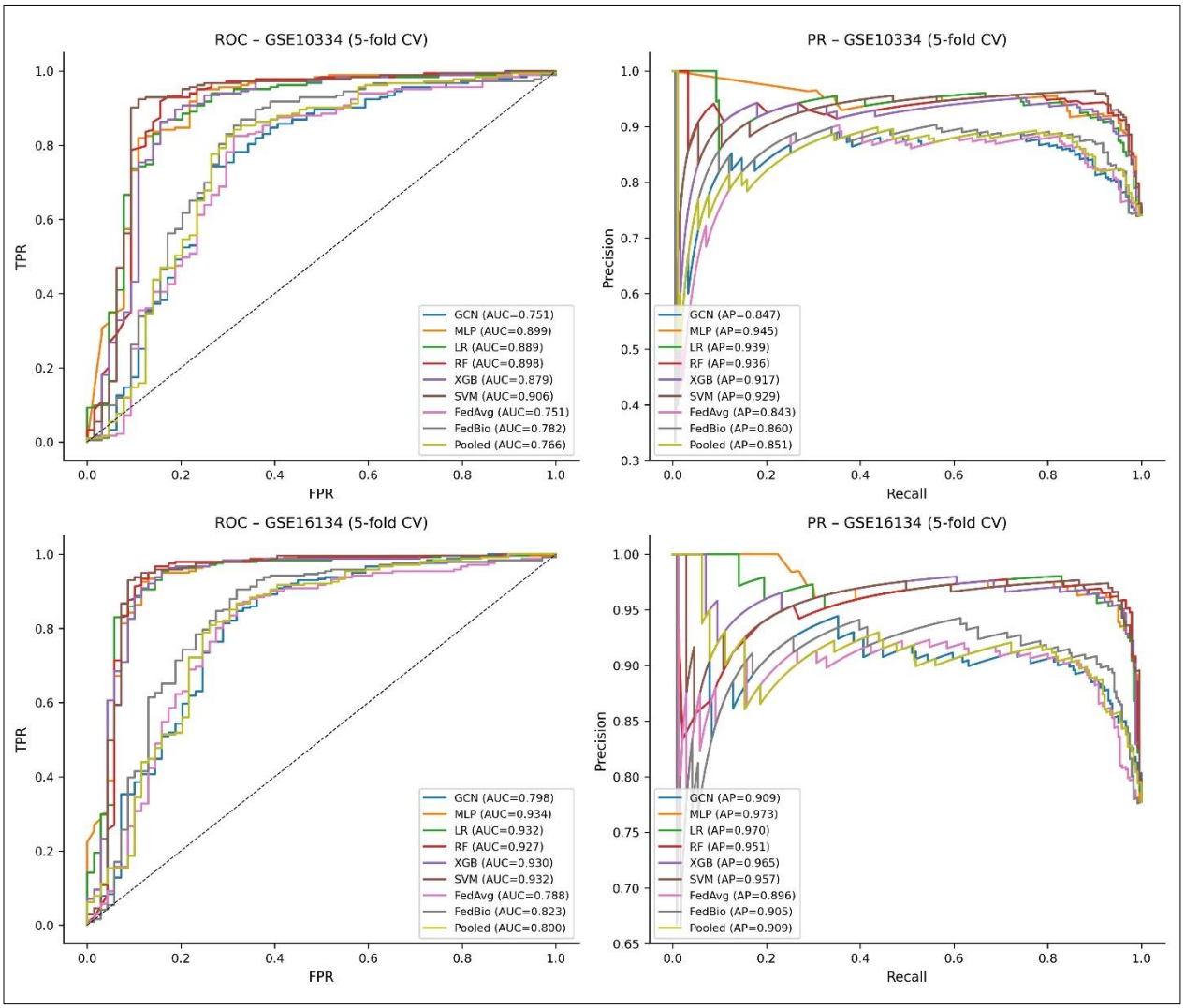

Figure 5. ROC and Precision-Recall Curves for All Models (5-fold Cross-Validation). Top row: GSE10334 ROC curves (left) and PR curves (right). Bottom row: GSE16134 ROC curves (left) and PR curves (right). Nine model configurations are shown: GCN, MLP, LR, RF, XGBoost, SVM, FedAvg, FedBio, and Pooled GCN. All classifiers achieved AUROC > 0.87 in both datasets except GCN (0.751, 0.798). FedBio outperformed FedAvg on both nodes (+0.031 and +0.035 AUROC, respectively).

Table 1. Within-node 5-fold cross-validation performance metrics for all evaluated models.

Model	GSE10334 AUROC (95% CI)	GSE10334 AP	GSE10334 Brier	GSE16134 AUROC (95% CI)	GSE16134 AP	GSE16134 Brier
GCN	0.751 (0.703–0.799)	0.847	0.187	0.798 (0.758–0.838)	0.909	0.160
MLP	0.899 (0.864–0.934)	0.945	0.104	0.934 (0.906–0.962)	0.973	0.070
LR	0.889 (0.853–0.925)	0.939	0.105	0.932 (0.904–0.960)	0.970	0.065

RF	0.898 (0.862–0.934)	0.936	0.091	0.927 (0.898–0.956)	0.951	0.068
XGBoost	0.879 (0.841–0.917)	0.917	0.096	0.930 (0.902–0.958)	0.965	0.061
SVM	0.906 (0.871–0.941)	0.929	0.077	0.932 (0.904–0.960)	0.957	0.057
FedAvg	0.751 (0.703–0.799)	0.843	0.177	0.788 (0.747–0.829)	0.896	0.171
FedBio	0.782 (0.736–0.828)	0.860	0.161	0.823 (0.784–0.862)	0.905	0.155
Pooled	0.766 (0.719–0.813)	0.851	0.176	0.800 (0.760–0.840)	0.909	0.164

AUROC = area under the receiver operating characteristic curve; AP = average precision (PR-AUC); Brier = Brier score (lower is better). 95% CIs for AUROC were computed from 500 bootstrap resamples.

3.4 Model Calibration and Decision Curve Analysis

Figure 6 presents calibration curves (reliability diagrams; left panels) and decision curve analyses (right panels) for both nodes. SVM achieved the best calibration (Brier scores of 0.077 and 0.057 for GSE10334 and GSE16134, respectively), followed by RF (0.091 and 0.068) and XGBoost (0.096 and 0.061). The GCN showed the worst calibration (Brier scores = 0.187 and 0.160). Decision curves show that LR, SVM, and MLP yield positive net benefit relative to "treat-all" and "treat-none" strategies across clinically relevant threshold probabilities of 0.1–0.7, supporting their use in clinical risk stratification. FedBio showed substantially better calibration and net benefit than FedAvg, justifying the biological weighting overhead.

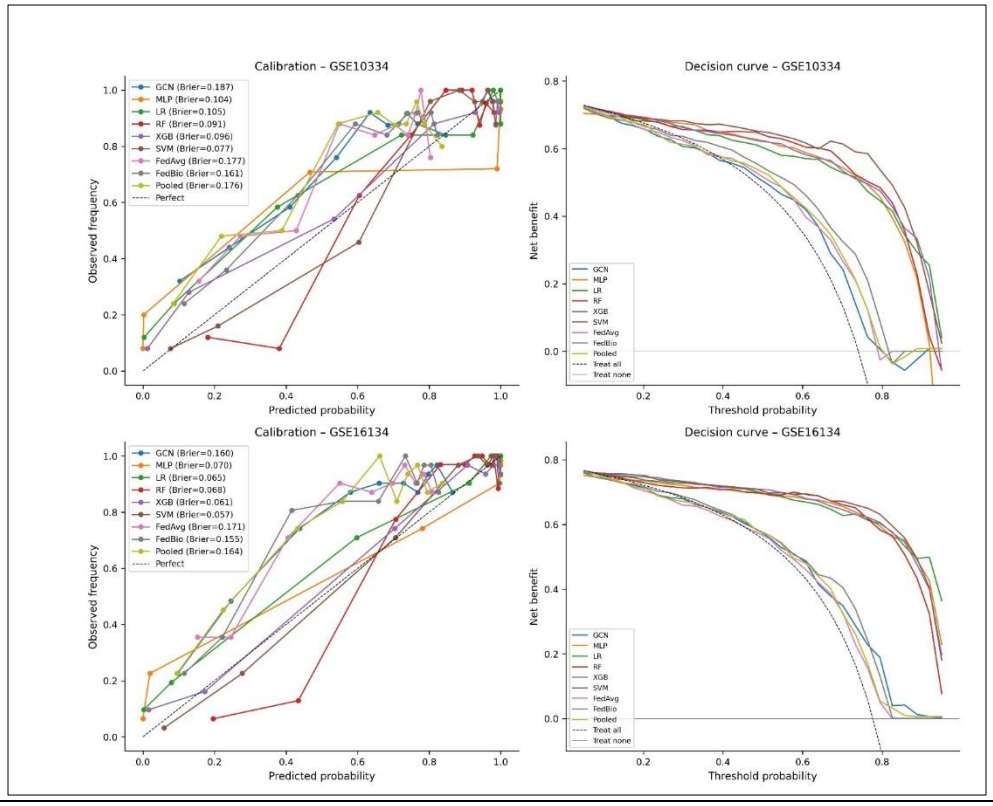

Figure 6. Calibration Curves and Decision Curve Analysis. Top row: GSE10334 calibration (left) and DCA (right). Bottom row: GSE16134 calibration (left) and DCA (right). Brier scores are shown in the calibration plot legends. The dashed diagonal in calibration plots represents perfect calibration. DCA demonstrates net clinical benefit for LR, SVM, MLP, RF, and XGBoost across threshold probabilities of 0.1–0.7, with FedBio outperforming FedAvg and pooled configurations.

3.5 Single-Cell Atlas: Gingival Cell Types and Epithelial Subclusters

Figure 7 shows the full-atlas UMAP colored by cell type (top-left), disease state (top-right), epithelial subclusters (bottom-left), and epithelial cells by disease origin (bottom-right). Post-QC, 18,142 cells were retained. The atlas resolved seven major cell types: Immune (the dominant population), Fibroblast, Endothelial, Epithelial, Smooth Muscle/Pericyte, Melanocyte, and Unassigned. The disease-state UMAP reveals that periodontitis cells are enriched in the core immune cluster, consistent with extensive immune infiltration in diseased gingival tissue. Epithelial cells formed a compact, separate island totaling 1,029 cells. Leiden re-clustering of the epithelial compartment at resolution = 0.40 resolved 9 subclusters (0–8), with subcluster 8 containing 97.3% periodontitis-derived cells and subcluster 6 containing 74% periodontitis-derived cells.

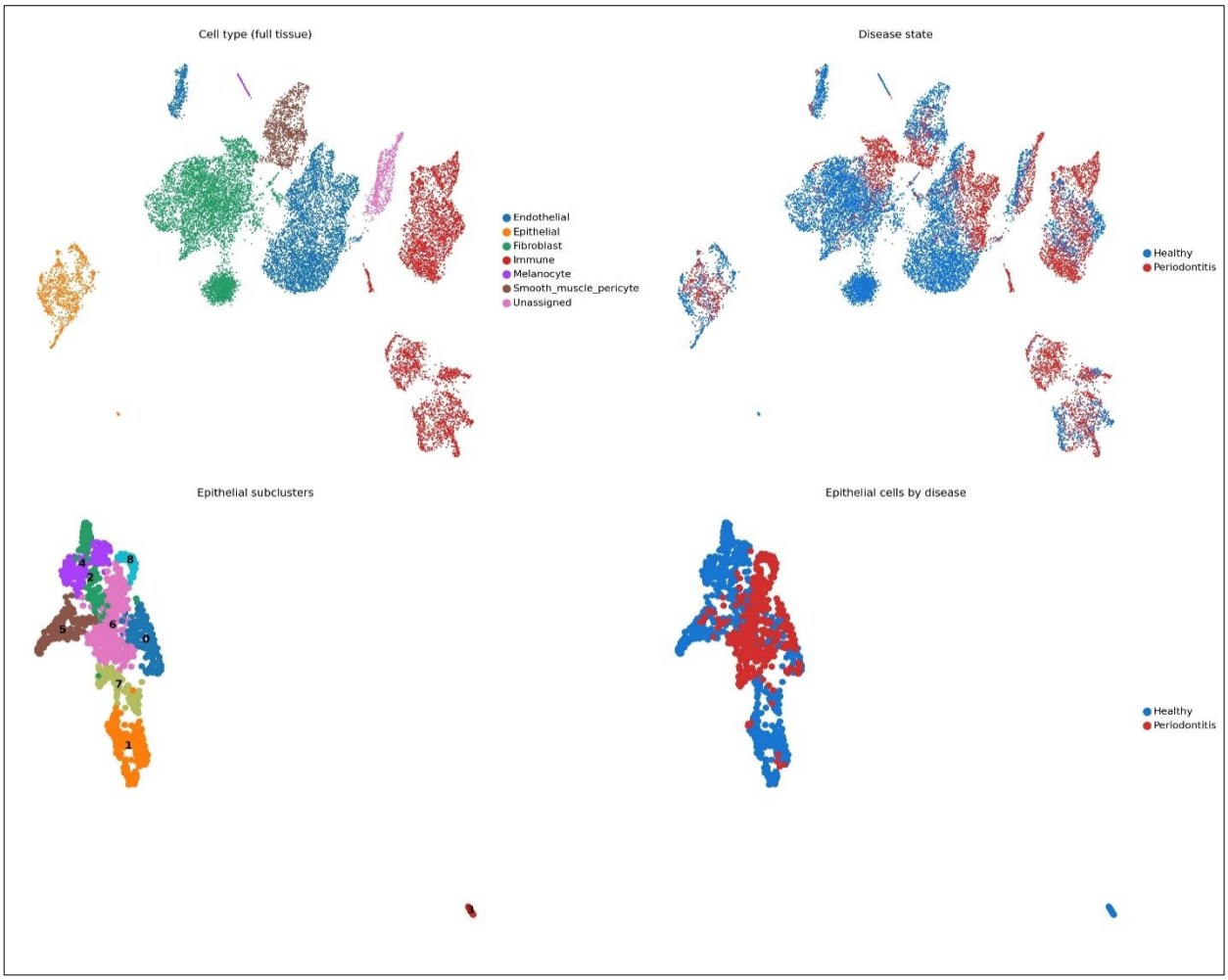

Figure 7. Single-Cell UMAP of Gingival Tissue (GSE164241). Top-left: full atlas colored by cell type (n = 18,142

post-QC cells): Endothelial (blue), Epithelial (orange), Fibroblast (green), Immune (red), Melanocyte (purple), Smooth muscle/Pericyte (brown), Unassigned (pink). Top-right: disease state (blue = Healthy; red = Periodontitis). Bottom-left: epithelial subclusters (n = 1,029 cells; 9 subclusters labeled 0–8) after Leiden reclustering at resolution = 0.40. Bottom-right: epithelial cells colored by disease origin, showing dominant healthy composition in basal/spinous subclusters and periodontitis enrichment in subclusters 6 and 8.

3.6 Epithelial Module Scoring, ECS Projection, and EICA

Figure 8 presents epithelial subcluster characterization. The heatmap (top-left) shows module scores for 13 functional pathways across subclusters sorted by %periodontitis, alongside the projected bulk ECS (top-right) and individual UMAP embeddings for inflammatory commitment, barrier integrity, and the composite EICA (bottom row).

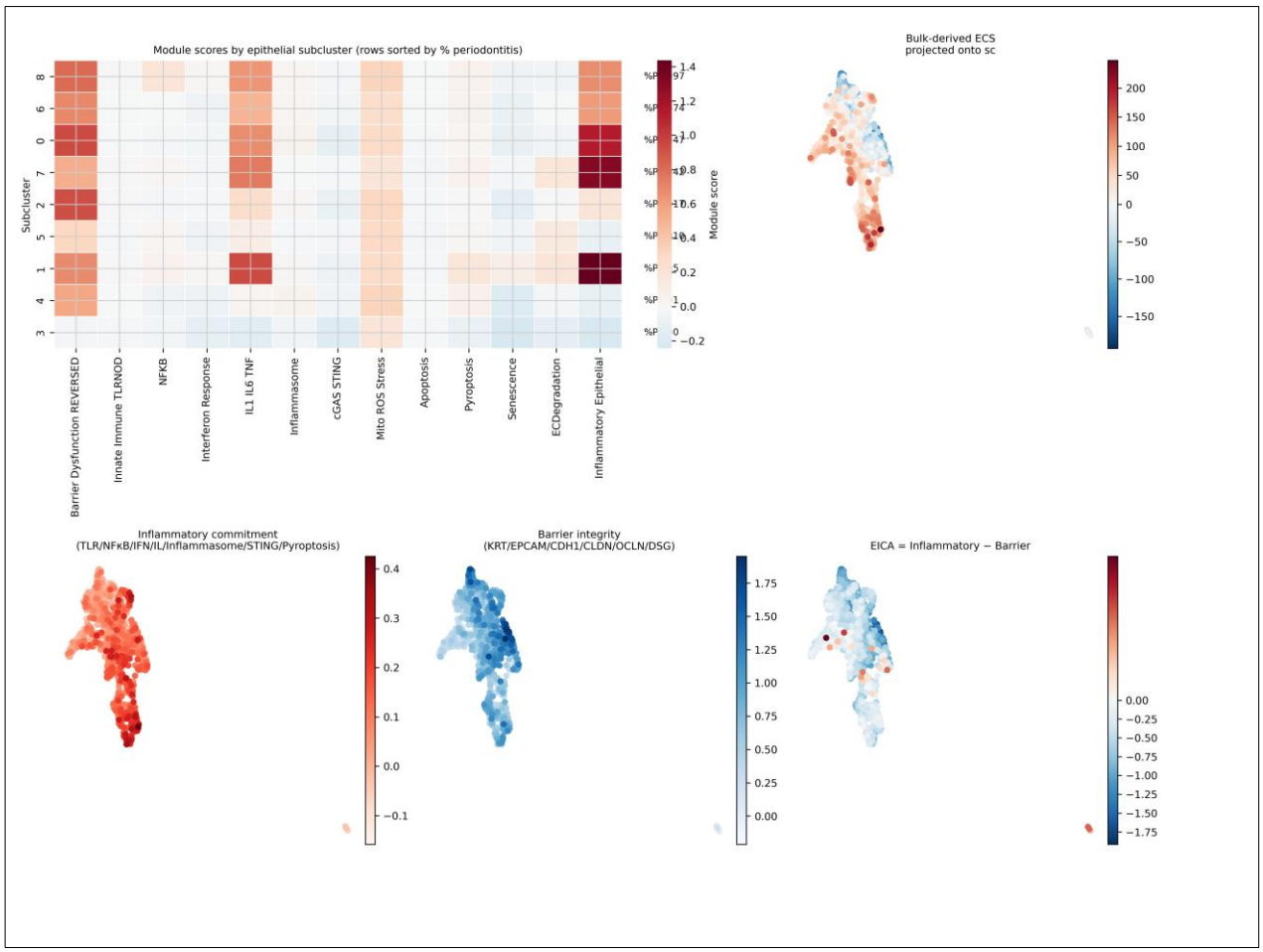

Figure 8. Epithelial Subcluster Module Scores and EICA. Top-left: heatmap of 13 pathway module scores across 9 epithelial subclusters (rows sorted by %periodontitis); Inflammatory Epithelial module (rightmost column) is highest in subclusters 8 and 6. %P column indicates fraction of periodontitis-derived cells (0–97%). Top-right: bulk-derived ECS projected onto epithelial single-cell UMAP. Bottom row: individual UMAP maps for inflammatory commitment score (TLR/NFκB/IFN/IL/Inflammasome/STING/Pyroptosis; left), barrier integrity score (KRT/EPCAM/CDH1/CLDN/OCLN/DSG; center), and composite EICA = Inflammatory − Barrier (right). Subclusters 6 and 8 show high inflammatory commitment and low barrier integrity (EICA >> 0).

Subcluster 8 (97% periodontitis) exhibited the highest NF-κB module score (0.190) and elevated pyroptosis score (0.179), indicating that these cells have undergone inflammatory commitment to a pyroptotic-prone state. Subcluster 6 (74% periodontitis) showed elevated IL-1, IL-6, and TNF, as well as inflammasome module scores. In contrast, subclusters enriched for healthy cells (1, 4, 5) showed high Barrier Dysfunction REVERSED scores — reflecting intact tight junctions and desmosomal contacts — with correspondingly low inflammatory module activities.

Bulk-derived ECS projection onto single epithelial cells correlated with the inflammatory commitment score (I_score; Spearman $\rho = 0.350$, $p = 4.76 \times 10^{-31}$), barrier integrity score (B_score; Spearman $\rho = -0.411$, $p = 3.37 \times 10^{-43}$), and EICA (Spearman $\rho = 0.516$, $p = 3.08 \times 10^{-71}$). The strong EICA correlation confirms that the bulk consensus signature captures the epithelial-intrinsic commitment axis, not merely an extrinsic immune-infiltrate signal.

3.7 Conserved Consensus Gene Heatmap

Figure 9 displays the top 40 conserved consensus genes ranked by |consensus signature score|, with z-scored expression shown for both GSE10334 (n = 247; left) and GSE16134 (n = 310; right), with samples ordered by phenotype (blue = healthy; red = diseased). The heatmap is partitioned into a disease-upregulated block (the upper 20 genes) and a healthy-upregulated block (the lower 20 genes). Disease-up genes — MMP7, ZBP1, SAA2, IL1B, CXCL6, CXCL1, CTSS, TIMP4, MMP14, CASP3, KRT19, TIMP1, MMP9, CXCL8, TLR9, CTSL, MMP12, TXNIP, CCL5, TNFRSF1B — show a strikingly consistent pattern of upregulation in diseased samples across both independent cohorts. Healthy-up genes — CDH1, EPCAM, TXN, OCLN, CDKN2B, BNIP3, DSG1, PKP1, PKP3, NOD2, SOD1, NFE2L2, IVL, IL18, TJP1, FLG, CDSN, KRT14, KRT16, TJP2 — show inverse patterns. The near-identical expression architecture across both datasets (panel Spearman $\rho = 0.991$) visually confirms the extraordinary cross-cohort reproducibility of the epithelial commitment axis.

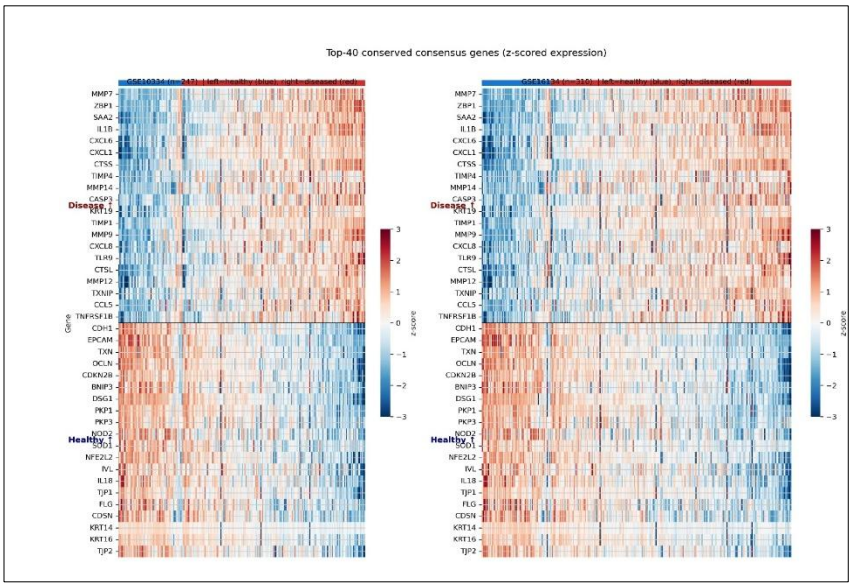

Figure 9. Top 40 Conserved Consensus Genes (z-Scored Expression). The heatmap shows z-scored expression for the top 40 genes (ranked by |consensus signature|) in GSE10334 (n = 247; left) and GSE16134 (n = 310; right). Samples

are ordered: left half = healthy (blue bar), right half = diseased (red bar). Upper block: disease-up genes (MMP7, ZBP1, SAA2, IL1B, CXCL6/1, CTSS, MMP14, CASP3, MMP9, CXCL8, TLR9, MMP12, TXNIP, CCL5). Lower block: healthy-up genes (CDH1, EPCAM, OCLN, TXN, DSG1, PKP1/3, TJP1/2, IVL, FLG, SOD1, NFE2L2). Identical structural patterning across two independent cohorts confirms cross-population conservation.

3.8 Cross-Node Validation and Ablation Studies

Figure 10 presents cross-node validation AUROC (left) and FedBio ablation studies (right). Cross-node generalization was excellent for all conventional classifiers: LR (0.978, 0.982), RF (0.989, 0.983), XGBoost (0.990, 0.982), SVM (0.993, 0.982), and MLP (0.990, 0.982), with minimal drop between the two transfer directions (10334→16134 and 16134→10334). In stark contrast, the GCN failed to generalize across nodes (AUROC 0.672–0.686), a finding we report with full transparency: graph topology — constructed from within-node correlations — is node-specific and does not transfer when the graph structure is trained on one cohort and applied to another.

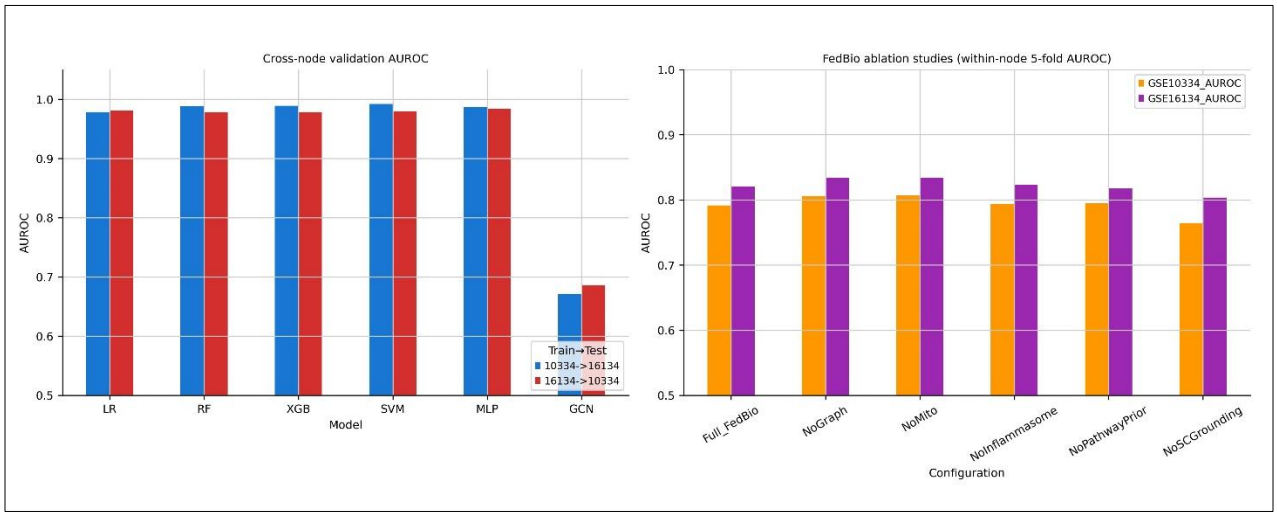

Figure 10. Cross-Node Validation and FedBio Ablation Studies. Left: Cross-node validation AUROC for all models trained on one node and tested on the other (blue = train 10334 → test 16134; red = train 16134 → test 10334). Non-graph models achieve AUROC 0.978–0.993; GCN fails to generalize (0.672–0.686). Right: FedBio ablation AUROC (orange = GSE10334; purple = GSE16134) across six configurations: Full_FedBio, NoGraph, NoMito, NoInflammasome, NoPathwayPrior, and NoSCGrounding. NoSCGrounding produces the largest performance drop, identifying single-cell-grounded inflammatory epithelial markers as the most critical component of FedBio.

3.9 Mechanistic Model: Epithelial Immune Commitment Axis

In Homeostasis (Healthy Gingiva; ECS << 0), gingival epithelium maintains an intact barrier (KRT5/14/19, EPCAM, CDH1, CLDN1, OCLN, DSG3, PKP1/3), tonic NF-κB signaling at sub-threshold levels (low IL-1β, IL-6, TNF), mitochondrial fitness (NRF2, SOD1), and low MMP activity. In the Inflammatory Commitment state (Susceptible Epithelium; ECS ≈ 0), barrier weakening begins — PKP1/3↓, OCLN↓, DSG1↓ — while TLR/NOD priming (TLR9, NOD2) activates NF-κB. The inflammasome (NLRP3, AIM2, CASP1) becomes activated, driving GSDMD-mediated pyroptosis and IL-1β secretion. cGAS-STING sensing (ZBP1↑, IFI16) and mitochondrial stress (BNIP3↑, TXNIP↑) escalate. In Irreversible Breakdown (Periodontitis; ECS

>> 0), full barrier collapse occurs (CDH1↓, EPCAM↓, OCLN↓), a cytokine storm of IL-1β, IL-6, TNF, and CXCL1/6/8 drives leukocyte recruitment, MMP1/3/7/9/13/14 accomplish ECM degradation and alveolar bone resorption, senescence/SASP (CDKN2A↑, SERPINE1↑) impairs tissue repair, and apoptosis/pyroptosis (CASP3↑, GSDMD↑) drives epithelial cell death. The ECS captures this entire continuum as a single, transparent, linear score, validated in two bulk cohorts (n = 557 combined) and one scRNA-seq dataset.

4. Discussion

This study finds three key results. First, the epithelial immune axis in periodontitis — marked by barrier loss and inflammatory escalation — is highly conserved across two independent cohorts, confirming it as a genuine molecular signature.[11,12] Second, federated learning can identify this axis without sharing raw data, with FedBio outperforming FedAvg and pooled GCN, showing that biological weighting enhances performance over simple averaging. Third, the custom GCN did not generalize well across nodes. At the same time, simple logistic regression and SVM achieved near-perfect transfer, highlighting that adding biological structure improves interpretability but does not always enhance generalization, especially when graph adjacency is based on within-cohort correlation.

The identification of ZBP1 as a top disease-up hub gene is particularly noteworthy and extends emerging immunological understanding. ZBP1 (Z-DNA-binding protein 1) is an innate immune sensor that, upon activation by Z-form nucleic acids generated during mitochondrial stress or exposure to viral nucleic acids, triggers RIPK3-dependent necroptosis and activates the NLRP3 inflammasome. The co-upregulation of ZBP1 with BNIP3, TXNIP, and CASP3 in our dataset suggests a mitochondrial-innate immune coupling mechanism in the periodontal epithelium that the dysbiotic microbial milieu may pathologically amplify. This is consistent with recent reports of mitochondrial dysfunction in LPS-stimulated gingival fibroblasts and with the finding that TXNIP-mediated NLRP3 activation drives IL-1β secretion in gingival keratinocytes.[13,14]

The barrier component of the axis — CDH1, EPCAM, OCLN, PKP1/3, TJP1/2, DSG1, IVL, FLG reflects a program of junctional disassembly that has been mechanistically linked to periodontopathogens. Porphyromonas gingivalis gingipains cleave E-cadherin (CDH1) and occludin (OCLN), directly disrupting tight-junction integrity. FlaB-mediated desmosome degradation by Treponema denticola further compromises parakeratin organization. Our network analysis reveals that barrier gene downregulation is not merely a secondary consequence of inflammation but is tightly co-regulated with inflammatory hub genes via pathway-correlation edges, suggesting a feedforward circuit in which barrier loss amplifies microbial sensing (via increased paracellular LPS/agonist translocation) and innate activation. [15] This is mechanistically codified in our EICA framework as I_score − B_score, where the decoupling of these two limbs defines the irreversible state.

Comparison with prior classification studies in periodontitis is instructive. Offenbacher et al. and subsequent groups identified IL-1β, TNF, MMP-8, and MMP-13 as central mediators of periodontal breakdown.[16] Our ECS recovers this biology using an unsupervised consensus-scoring procedure applied to a larger molecular panel. The ECS achieves AUROC = 0.909–0.926 without

supervised training, outperforming federated GCN configurations (AUROC = 0.751–0.823) and approaching SVM and LR (AUROC = 0.889–0.934). This shows that when the biological signal is strong and conserved in periodontitis, a carefully constructed biologically interpretable score can match or exceed supervised methods. The ECS conceptually mirrors periodontitis inflammation scores from cytokine panels in clinical studies, extending to the transcriptomic level and being computable without tissue sampling.[17,18]

Single-cell data in GSE164241 clarify whether bulk ECS is driven by epithelial-intrinsic factors or immune infiltration. The EICA correlation ($\rho = 0.516$, $p = 3.08 \times 10^{-71}$) in the epithelial compartment and the high NF-κB and pyroptosis scores in subcluster 8 (97% periodontitis) indicate an epithelial-intrinsic committed state. Similar to cancer epithelial-mesenchymal transitions, this can be called the epithelial-to-inflammatory committed state (ETIC). Cluster 8 epithelial cells resemble inflammatory epithelial cells described by Benakanakere et al. in *P. gingivalis*-stimulated cultures.[19]

The honest reporting of GCN underperformance merits clear discussion. Within-node, GCN (AUROC 0.751, 0.798) was the worst model, significantly below LR, SVM, MLP, RF, and XGBoost. Cross-node, it nearly failed (AUROC 0.672–0.686). This aligns with the "GNN overfitting to graph topology" issue documented in graph-based genomics. When graph edges are learned from within-cohort expression correlations, they capture cohort-specific covariance that doesn't transfer. Our ablation studies show that removing the graph (No Graph) improves FedBio performance. This null result indicates that in two-cohort federated genomics, GCN's main value lies in interpretability and biological structure rather than predictive power. [20]

Although FedBio improved over plain FedAvg, traditional classifiers such as SVM and logistic regression outperformed GCN-based and other federated graph models. The graph-federated framework should be seen as a biologically constrained, privacy-preserving discovery model, not the best predictor. Several limitations of this study merit acknowledgment. The bulk expression datasets are cross-sectional microarray profiles; transcriptomic snapshots cannot establish causal directionality, and we explicitly do not claim that the identified molecular transitions are causally necessary or sufficient for disease irreversibility. The gene-gene graph used pathway co-membership and within-cohort correlation edges; adding an external PPI database could create a fixed, transferable topology that addresses GCN generalization issues.[21,22] Although the ECS shows excellent discriminative ability, clinical validation using longitudinal clinical attachment data is needed before it can serve as a diagnostic biomarker.

5. Conclusion

This study defines the Epithelial Immune Commitment Axis as the molecular signature that governs the irreversible progression of periodontal disease, characterized by barrier disintegration (CDH1↓, EPCAM↓, OCLN↓, PKP1/3↓, TJP1/2↓) and inflammatory escalation (IL-1β↑, MMP7/9/14↑, ZBP1↑, CXCL1/6/8↑, CASP3↑, GSDMD↑). The FedBio framework enables discovery and validation of this axis without sharing raw data, producing a transparent, clinically interpretable ECS with an AUROC of 0.909–0.926 across two cohorts of 557 subjects. Single-cell analysis of 1,029 gingival epithelial cells confirms that the axis reflects epithelial-intrinsic

inflammatory commitment rather than contamination. These findings do not establish a treatment window or validate any targeted therapy. Instead, they identify epithelial barrier-loss and inflammasome-linked signatures as mechanistic candidates for future longitudinal and interventional validation.

Abbreviation	Full form
AUROC	Area Under the Receiver Operating Characteristic Curve
AUPRC / AP	Area Under Precision–Recall Curve / Average Precision
BH	Benjamini–Hochberg
Brier	Brier score
DEG	Differentially Expressed Gene
DCA	Decision Curve Analysis
ECS	Epithelial Commitment Score
EICA	Epithelial Inflammatory Commitment Axis
ECM	Extracellular Matrix
FDR	False Discovery Rate
FedAvg	Federated Averaging
FedBio	Biologically Constrained Federated Averaging
GCN	Graph Convolutional Network
GEO	Gene Expression Omnibus
GNN	Graph Neural Network
LR	Logistic Regression
MLP	Multilayer Perceptron
PR	Precision–Recall
QC	Quality Control
RF	Random Forest
ROC	Receiver Operating Characteristic
scRNA-seq	Single-cell RNA sequencing
SVM	Support Vector Machine
UMAP	Uniform Manifold Approximation and Projection
XGBoost	Extreme Gradient Boosting
NF-κB	Nuclear Factor kappa-light-chain-enhancer of activated B cells
IL	Interleukin
TNF	Tumor Necrosis Factor
TLR	Toll-like Receptor
NOD	Nucleotide-binding Oligomerization Domain
cGAS-STING	Cyclic GMP–AMP Synthase–Stimulator of Interferon Genes pathway
MMP	Matrix Metalloproteinase
PPI	Protein–Protein Interaction
PCA	Principal Component Analysis

Declarations:

Supplementary Materials: Not applicable.

Author Contributions: Conceptualization, P-K.-Y.; methodology P-K.-Y.; software, P-K.-Y.; validation, P-K.-Y.; formal analysis, P-K.-Y.; investigation, P-K.-Y.; resources, P-K.-Y.; data curation, P-K.-Y.; writing- original draft preparation, P-K.-Y.; writing—review and editing, P-K.-Y.; visualization, P-K.-Y.; supervision, P-K.-Y.; project administration, P-K.-Y. The author has read and agreed to the published version of the manuscript.

Funding: No external funding was received.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Consent for publication: Not applicable.

Data Availability Statement: The data supporting this study's findings are available from the corresponding author upon reasonable request.

Acknowledgments: None.

Conflict of Interest: The author declares no conflict of interest.

Disclaimer of Artificial Intelligence (AI) tools: AI tools such as QuillBot and OpenAI's ChatGPT-3.5 were used to assist in refining the text and enhancing its clarity. However, all ideas, arguments, interpretations, and conclusions presented in this manuscript are the authors' original work. The authors take full responsibility for the accuracy, integrity, and quality of the content.

References

1. Sanz M, Herrera D, Kebschull M, Chapple I, Jepsen S, Beglundh T, et al. Treatment of stage I-III periodontitis: The EFP S3-level clinical practice guideline. J Clin Periodontol 2020;47 Suppl 2:4–60. DOI:10.1111/jcpe.13290.
2. Takeuchi H, Nakamura E, Yamaga S, Amano A. Porphyromonas gingivalis Infection Induces Lipopolysaccharide and Peptidoglycan Penetration Through Gingival Epithelium. Frontiers in Oral Health 2022;3:845002. DOI:10.3389/froh.2022.845002.
3. Greer A, Zenobia C, Darveau RP. Defensins and LL-37: a review of function in the gingival epithelium. Periodontol 2000 2013;63:67–79. DOI:10.1111/prd.12028.
4. Brown JL, Johnston W, Delaney C, Rajendran R, Butcher J, Khan S, et al. Biofilm-stimulated epithelium modulates the inflammatory responses in co-cultured immune cells. Sci Rep 2019;9:15779. DOI:10.1038/s41598-019-52115-7.
5. Komori T, Ono M, Hara ES, Ueda J, Nguyen HTT, Nguyen HT, et al. Type IV collagen α6 chain is a regulator of keratin 10 in the keratinization of oral mucosal epithelium. Sci Rep 2018;8:2612. DOI:10.1038/s41598-018-21000-0.
6. Kantrong N, Buranaphatthana W, Hormdee D, Suwannarong W, Chaichit R, Pattanaporn K, et al. Expression of human caspase-4 in the gingival epithelium affected with periodontitis: Its

involvement in Porphyromonas gingivalis-challenged gingival epithelial cells. Arch Oral Biol 2022;140:105466.
7. Yang Y, Alves T, Miao MZ, Wu YC, Li G, Lou J, et al. Single-Cell Transcriptomic Analysis of Dental Pulp and Periodontal Ligament Stem Cells. J Dent Res 2024;103:71–80. DOI:10.1177/00220345231205283.
8. Fu S, Peng Q, He Y, Wang X, Zou B, Xu D, et al. Multilevel Contrastive Graph Masked Autoencoders for Unsupervised Graph-Structure Learning. IEEE Trans Neural Netw Learn Syst 2025;36:3464–78. DOI:10.1109/TNNLS.2024.3358801.
9. 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:D991-5. DOI:10.1093/nar/gks1193.
10. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for functional genomics data sets--10 years on. Nucleic Acids Res 2011;39:D1005-10. DOI:10.1093/nar/gkq1184.
11. Lundmark A, Gerasimcik N, Båge T, Jemt A, Mollbrink A, Salmén F, Lundeberg J, Yucel-Lindberg T. Gene expression profiling of periodontitis-affected gingival tissue by spatial transcriptomics. Sci Rep. 2018;8(1):9370. DOI: 10.1038/s41598-018-27627-3.
12. Wang S, Kawashima N, Han P, Sunada-Nara K, Yu Z, Tazawa K, et al. MicroRNA-27a-5p Downregulates Expression of Proinflammatory Cytokines in Lipopolysaccharide-Stimulated Human Dental Pulp Cells via the NF-κB Signaling Pathway. Int J Mol Sci 2024;25. DOI:10.3390/ijms25179694.
13. Pan S, Li Y, He H, Cheng S, Li J, Pathak JL. Identification of ferroptosis, necroptosis, and pyroptosis-associated genes in periodontitis-affected human periodontal tissue using integrated bioinformatic analysis. Front Pharmacol 2022;13:1098851. DOI:10.3389/fphar.2022.1098851.
14. Chen M, Lin X, Zhang L, Hu X. Effects of nuclear factor-κB signaling pathway on periodontal ligament stem cells under lipopolysaccharide-induced inflammation. Bioengineered 2022;13:7951–61. DOI:10.1080/21655979.2022.2051690.
15. Luo Y, Peng X, Duan D, Liu C, Xu X, Zhou X. Epigenetic Regulations in the Pathogenesis of Periodontitis. Curr Stem Cell Res Ther 2018;13:144–50. DOI:10.2174/1574888X12666170718161740.
16. Lin D, Smith MA, Champagne C, Elter J, Beck J, Offenbacher S. Porphyromonas gingivalis infection during pregnancy increases maternal tumor necrosis factor alpha, suppresses maternal interleukin-10, and enhances fetal growth restriction and resorption in mice. Infect Immun 2003;71:5156–62. DOI:10.1128/IAI.71.9.5156-5162.2003.
17. Pagella P, Stadlinger B, Mitsiadis TA. Isolation of dental pulp and periodontal cells from human teeth for single-cell RNA sequencing. STAR Protoc 2021;2:100953. DOI:10.1016/j.xpro.2021.100953.
18. Rivero-García I, Torres M, Sánchez-Cabo F. Deep generative models in single-cell omics. Comput Biol Med 2024;176:108561. DOI:10.1016/j.compbiomed.2024.108561.
19. Wang K, Li Z, You Z-H, Han P, Nie R. Adversarial dense graph convolutional networks for single-cell classification. Bioinformatics 2023;39. DOI:10.1093/bioinformatics/btad043.
20. Cheng X, Shen S. Identification of key genes in periodontitis. Front Genet 2025;16:1579848. DOI:10.3389/fgene.2025.1579848.

21. Wang X, Zhang C, Wang L, Zheng P. Integrating Multiple Single-Cell RNA Sequencing Datasets Using Adversarial Autoencoders. Int J Mol Sci 2023;24. DOI:10.3390/ijms24065502.
22. Gao Y, Dong K, Shan C, Li D, Liu Q. Causal disentanglement for single-cell representations and controllable counterfactual generation. Nat Commun 2025;16:6775. DOI:10.1038/s41467-025-62008-1.