

Supplementary Table S1. Summary of key parameter settings for analytical procedures in this study.

Analysis Module	Parameter Name	Parameter Value / Setting	Note / Justification
Immune Cell Deconvolution	Permutations	[1000]	Number of CIBERSORTx permutations.
	Sample Filtering Threshold	$p < 0.05$	Only samples with a CIBERSORTx *p*-value < 0.05 were retained for subsequent analysis.
Random Forest (RF)	Model Type	Classifier	The outcome variable was disease state (BE/EAC vs. Normal).
	ntree (Number of Trees)	500	Number of trees in the forest.
	mtry (Features per Split)	2.44949	$\lfloor \sqrt{6} \rfloor = 2$ (Default for classification, where *p*=6 is the number of candidate genes).
	Random Seed	315	Ensures the results are reproducible.
	Variable Importance Metric	MeanDecreaseGini	Measure based on the total decrease in node impurity (Gini Index).
	OOB Error Estimate	be 7.7% eac 5.3%	Out-of-bag error rate for the final model.
LASSO Regression	Chosen λ (lambda) Value	be 0.01911284 eac 0.01242307	The value of λ that gives the minimum mean cross-validated error.
	λ Selection Rule	lambda.min	Use the λ value that minimizes the cross-validated error.
Train/Validation Split	Split Ratio	70% / 30%	70% of samples for training, 30% for validation.
	Splitting Method	Stratified Random Sampling	Splitting was stratified by disease state to preserve class proportions.
	Random Seed	1321	Ensures the results are reproducible.

Supplementary Table S2. Detailed performance metrics of diagnostic models based on internal validation sets.

Metric	BE vs. Normal Model	EAC vs. Normal Model
AUC (95% CI)	0.94 (0.858 - 1)	0.98 (0.9571-1)
Predefined threshold	0.5	0.5
Sensitivity (95% CI)	0.87	0.91
Specificity (95% CI)	1	1
Calibration slope	1.36	2.20
Calibration intercept	0.56	2.23

Supplementary Table S3. Class counts and associated metrics.

Confusion Matrix	Predicted: Control	Predicted: Disease
Actual: Control	4 (True Negative, TN)	2 (False Positive, FP)
Actual: Disease	0 (False Negative, FN)	20 (True Positive, TP)
Total (Predicted)	4	22

Performance Metrics	Value (95% CI)	Formula
Confusion Matrix	Predicted: Control	Predicted: Disease
Sensitivity	100.0% (83.2 – 100.0)	$TP / (TP + FN)$
Specificity	66.7% (22.3 – 95.7)	$TN / (TN + FP)$
PPV	90.9% (70.8 – 98.9)	$TP / (TP + FP)$
NPV	100.0% (39.8 – 100.0)	$TN / (TN + FN)$

Supplementary Table S4. Summary of Gene Lists from Bioinformatics Analysis Pipeline .

Analysis Step Name	Genome name	Gene Count	Gene List
Differentially Expressed Genes Analysis	SDEGs	101	<i>DSG3, HSPB8, LYPD3, CYP4B1, CST6, AHNK2, S100A7, PI3, PPP1R3C, BNIP3, SYNPO2L, BEX4, CDA, CEACAM7, ESPL1, CRYAB, TPD52L1, HSPA2, ULBP2, PTN, LY6G6C, ALOX15B, SUSL4, MUC4, SCNN1B, HLF, EFS, KRT23, ATP6V0A4, VLDLR, C4BPB, GSDMB, MDK, DDC, RARRES1, HOXB7, SLC26A3, ZG16, TMEM47, ATP4A, KCNE2, TMEM176A, GCNT3, COL4A5, PRKAR2B, FOXF1, SRPX, MYLK, ADRA2A, CXCL3, TGFB111, SLC6A20, ALDOB, DCN, CLCA1, DGKD, OTC, GPRC5C, CRISPLD2, MMP1, MTIM, CPS1, FILIP1L, MISP, F5, ABCA8, TRIM31, CA9, FMOD, MUC2, ATP4B, TM4SF5, PDGFD, RCAN2, ACTG2, AKR7A3, VILI, RGS2, SLC6A14, SST, FMO5, TMEM176B, AZGP1, TM4SF20, MEPIA, ALDH1A1, VNN1, TAGLN, CAPN9, SPINK4, CDH17, MUC13, PBLD, EPCAM, SI, AOC1, PGC, TM4SF4, OLFM4, LIPF, GKN1</i>
PPI Network	Hub Genes	10	<i>VILI, CDH17, MUC2, OLFM4, MUC13, ATP4B, GKN1, LIPF, ATP4A, PGC</i>
Enrichment Analysis & PPI Network	Core Genes	18	<i>MUC4, MDK, GCNT3, ADRA2A, MUC2, VILI, SST, CAPN9, MUC13, PBLD, SI, PGC, GKN1, CDH17, OLFM4, ATP4B, LIPF, ATP4A</i>
Immune Infiltration Analysis	Key Immune Genes	40	<i>DSG3, HSPB8, LYPD3, CYP4B1, CST6, AHNK2, PPP1R3C, BNIP3, SYNPO2L, BEX4, CDA, CRYAB, TPD52L1, HSPA2, LY6G6C, SCNN1B, EFS, ATP6V0A4, VLDLR, C4BPB, GSDMB, DDC, HOXB7, ZG16, GCNT3, ADRA2A, CXCL3, OTC, CPS1, TRIM31, CA9, TM4SF5, TM4SF20, MEPIA, VNN1, MUC13, PBLD, SI, TM4SF4, OLFM4</i>
Intersection screening candidate genes	Shared Key Genes	6	<i>GCNT3, ADRA2A, MUC13, PBLD, SI, OLFM4</i>
Machine Learning	Final Feature Genes	2	<i>MUC13, PBLD</i>