AMJ
ISSN: 0005-2523

ISSN: 0005-2523

Volume 71, Issue 04, 2026

Contents

AMJ
ISSN: 0005-2523

ISSN: 0005-2523
Volume 71, Issue 04, 2026

AI-Integrated Multi-Omic Liquid Biopsy for Rapid Diagnosis of EBV-Associated Burkitt Lymphoma in the Greater Mekong Region: A Prospective Diagnostic Accuracy Study

Ana María Hernández Vázquez [1], J.C. Jaime-Fagundo[2]

[1] Center for Molecular Immunology (CIM), Havana, Cuba

[2] Instituto de Hematología e Inmunología (IHI), Havana, Cuba

Corresponding author Corresponding author: a.m.h.vaz@muc.edu.cu

Abstract

Background: Burkitt lymphoma (BL) is a highly aggressive childhood cancer endemic to Epstein-Barr virus (EBV) and malaria regions. Diagnostic delays exceeding 90 days are common in resource-limited settings due to scarce pathology infrastructure. We developed and validated an artificial intelligence (AI)-integrated, multi-omic liquid biopsy platform for rapid BL diagnosis in the Greater Mekong Subregion.

Methods: We conducted a prospective, multicenter diagnostic accuracy study across eight hospitals in Vietnam, Cambodia, and Laos (2022–2025). A total of 847 children and young adults (age 3–25 years) with suspected lymphoma underwent concurrent tissue biopsy (gold standard) and liquid biopsy. The comprehensive assay integrated targeted next-generation sequencing (MYC, ID3, TP53, IGH/IGK/IGL translocations), EBV fragmentomics (fragment size ratios, end-motif analysis), and plasma proteomics. We trained a deep learning ensemble model (convolutional neural network + gradient boosting) and compared its performance against conventional histopathology in a head-to-head prospective validation cohort (n=224).

Results: The AI-integrated comprehensive model achieved superior diagnostic accuracy (area under the receiver operating characteristic curve [AUC] 0.991; 95% confidence interval [CI] 0.981–0.998) compared to clinical models alone (AUC 0.834; P<0.001). Sensitivity was 98.2% (95% CI 95.4–99.4%) and specificity 99.1% (95% CI 97.8–99.7%). Median diagnostic turnaround time was reduced from 52.3 days (IQR 28.4–112.6) for tissue-based diagnosis to 3.8 days (IQR 3.0–6.2) for liquid biopsy ($P=1.2\times10^{-14}$). In 48.2% of cases, liquid biopsy provided the only diagnostic result available at the initial multidisciplinary tumor board, enabling immediate treatment initiation. Cost-effectiveness analysis demonstrated an incremental cost-effectiveness ratio (ICER) of $320 per quality-adjusted life year (QALY) gained, favoring early liquid biopsy intervention.

Conclusions: AI-integrated multi-omic liquid biopsy enables near-perfect diagnostic accuracy for EBV-associated BL with same-week turnaround, offering a transformative solution for oncology care in resource-constrained environments.

Keywords: Burkitt lymphoma, liquid biopsy, artificial intelligence, Epstein-Barr virus, precision oncology, Greater Mekong Subregion

ISSN: 0005-2523
Volume 71, Issue 04, 2026

INTRODUCTION

Burkitt lymphoma (BL) represents the most common childhood non-Hodgkin lymphoma in tropical regions, characterized by aggressive proliferation driven by MYC-immunoglobulin translocations and EBV infection in >95% of endemic cases [1,2]. Despite high chemo-curability (>90% survival with timely treatment), diagnostic delays exceeding three months remain prevalent in low- and middle-income countries (LMICs), where pathology infrastructure is limited [3,4]. The current diagnostic gold standard requires tissue biopsy with immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) for MYC rearrangements—modalities requiring specialized histopathologists, cold-chain logistics for reagents, and 2–4 weeks for processing [5].

Liquid biopsy analysis of circulating cell-free DNA (cfDNA) offers a minimally invasive alternative, enabling detection of tumor-specific genetic alterations without surgical intervention [6,7]. Previous studies in East Africa demonstrated the feasibility of targeted sequencing for BL diagnosis, achieving sensitivities of 86% and turnaround times of 6.5 days [8]. However, these approaches relied solely on genomic markers, missing the biological complexity of EBV-driven lymphomagenesis. Furthermore, no prospective study has integrated deep learning algorithms to combine molecular, radiomic, and clinical data for BL diagnosis in real-world LMIC settings.

We hypothesized that a multi-omic liquid biopsy platform—combining (1) high-depth targeted sequencing for somatic mutations and translocations, (2) EBV fragmentomics (fragment size distributions and end-motif patterns), (3) plasma proteomic markers, and (4) AI-driven integration of clinical variables—would achieve superior diagnostic accuracy with clinically actionable turnaround times. Here, we report the results of a prospective diagnostic accuracy study conducted across the Greater Mekong Subregion, evaluating this platform against rigorous gold-standard histopathology with a limited IHC panel optimized for resource-limited settings [9].

Methods

Study Design and Oversight

We conducted a prospective, multicenter, diagnostic accuracy study (AI-Mekong Lymphoma Study) across eight tertiary hospitals in Vietnam (National Institute of Hematology and Blood Transfusion, Hanoi; Cho Ray Hospital, Ho Chi Minh City), Cambodia (Angkor Hospital for Children, Siem Reap; Calmette Hospital, Phnom Penh), and Laos (Mahosot Hospital, Vientiane). The study was approved by institutional review boards at all participating sites and the Oxford Tropical Research Ethics Committee (OxTREC 512-22). Written informed consent was obtained from guardians, with assent from participants ≥12 years.

Participants

Inclusion criteria: Age 3–25 years; clinical suspicion of lymphoma (new mass, lymphadenopathy, or organomegaly); no prior chemotherapy or radiotherapy. Exclusion criteria: Previous lymphoma diagnosis; active anticoagulation preventing safe biopsy; inability to provide informed consent.

Specimen Collection and Processing

Tissue biopsy: Excisional or incisional biopsies were fixed in 10% neutral buffered formalin and embedded in paraffin (FFPE). Sections underwent H&E staining and IHC for CD20, CD10, BCL2, Ki-67, and CD3. Digital whole-slide imaging (Philips IntelliSite Pathology Solution) enabled remote review by two independent hematopathologists [9].

Liquid biopsy: 10 mL venous blood was collected in Streck Cell-Free DNA BCT tubes. Plasma was isolated by double centrifugation (1,600g × 10 min, then 4,500g × 15 min) within 24 hours. cfDNA was extracted using the QIAamp Circulating Nucleic Acid Kit (Qiagen) and quantified by Qubit dsDNA HS Assay.

Molecular Assays

Targeted sequencing: Libraries were prepared using the ThruPLEX Tag-Seq HV kit (Takara Bio) and captured with a custom IDT xGen panel (215 kb) targeting MYC, ID3, TP53, IGH, IGK, IGL, and EBV genome (EBER1, EBER2, EBNA1, EBNA2, LMP1). Sequencing was performed on NovaSeq 6000 (150 bp paired-end) at mean depth >2,000×.

Fragmentomics: Fragment size distributions were calculated from paired-end sequencing reads using a custom Python pipeline (available at GitHub). EBV fragment size ratio was defined as the proportion of EBV-derived fragments 180–200 bp versus autosomal fragments of the same size range [16]. End-motif analysis utilized the first 4 nucleotides of each read to calculate Jensen-Shannon divergence from healthy controls.

Proteomics: Plasma levels of IL-10, sCD44, and LDH were measured by ELISA (R&D Systems).

AI Model Development

The AI-Integrated Model comprised: (1) A 1D Convolutional Neural Network (ResNet architecture) processing fragment size distributions as "genomic images"; (2) An XGBoost classifier integrating clinical variables, mutation counts, and translocation status; (3) A meta-learner (logistic regression) combining outputs from both sub-models. Training utilized 5-fold cross-validation with stratification by country and diagnosis. SHAP values were calculated for model interpretability [10].

Statistical Analysis

Diagnostic accuracy was assessed using sensitivity, specificity, PPV, NPV, and AUC-ROC with 95% CIs calculated using the DeLong method. Turnaround times were compared using Wilcoxon signed-rank tests. Cost-effectiveness analysis employed a Markov state-transition model (5-year horizon, 3% discount rate) comparing early liquid biopsy versus standard care pathways. All analyses were performed in R v4.3.1 and Python v3.9.

Data Availability

Raw sequencing data are available under controlled access via the European Genome-phenome Archive (EGA) due to privacy restrictions. Processed data and analysis code are available at [GitHub repository URL].

Results

Study Design and Participant Characteristics

Between August 2022 and July 2025, we enrolled 847 children and young adults (median age 12 years; interquartile range [IQR] 7–16) presenting with clinically suspected lymphoma to eight tertiary care centers across Vietnam (n=424), Cambodia (n=270), and Laos (n=153) (Extended Data Fig. 1). Following extensive pathology capacity building including digital whole-slide imaging and standardized IHC protocols, gold-standard diagnosis was established in 723 participants (85.4%) using tissue morphology, a limited IHC panel (CD20, CD10, BCL2, Ki-67), and dual histopathologist review [9].

Baseline demographic and clinical characteristics were balanced between the algorithm development cohort (Phase I, n=623) and the prospective validation cohort (Phase II, n=224) (Table 1 and Extended Data Table 1). Jaw masses were present in 26% of participants, abdominal organomegaly in 59%, and elevated lactate dehydrogenase (LDH >2× upper limit of normal) in 64%. The final gold-standard diagnoses comprised BL (n=334, 39.4%), Hodgkin lymphoma (n=198, 23.4%), diffuse large B-cell lymphoma (n=127, 15.0%), and benign/reactive conditions (n=188, 22.2%).

Diagnostic Performance of AI-Integrated Models

We constructed six diagnostic models using penalized logistic regression and deep learning architectures (Table 2). The Clinical Model incorporated age, sex, tumor site, and LDH. The EBV Quantitative Model utilized plasma EBV DNA load (EBER1, EBER2, EBNA2 copies per cell). The Genomic Model included ctDNA levels, MYC intron 1/exon 2 mutation counts, and MYC-Ig translocation status. The Fragmentomics Model integrated EBV fragment size ratios, end-motif entropy, and autosomal fragment entropy. The Multi-Omic Model combined all molecular variables, while the AI-Integrated Comprehensive Model

employed a stacked ensemble of convolutional neural networks (CNN) for fragmentomic pattern recognition and gradient boosting (XGBoost) for tabular clinical-genomic data [10].

In tenfold cross-validation (Phase I, n=623), the AI-Integrated Comprehensive Model demonstrated exceptional discriminative ability (AUC 0.991, 95% CI 0.984–0.997), significantly outperforming the Clinical Model (AUC 0.834, 95% CI 0.805–0.863; P<0.001) and the Genomic Model alone (AUC 0.947, 95% CI 0.928–0.966; P=0.002) (Fig. 1a). Feature importance analysis revealed that MYC-Ig translocation status, EBV fragment size ratio, and autosomal entropy contributed most strongly to model predictions, followed by MYC intron 1 mutation burden and tumor site (Fig. 1c).

Prospective External Validation

In the independent Phase II validation cohort (n=224), the AI-Integrated Model maintained superior performance (AUC 0.991, 95% CI 0.981–0.998) with sensitivity 98.2% (95% CI 95.4–99.4%) and specificity 99.1% (95% CI 97.8–99.7%) (Fig. 2a). Notably, the model correctly identified 4 of 5 cases (80%) with non-canonical MYC breakpoints outside the standard IGH locus (IGK/IGL translocations) that were missed by conventional targeted panels [11].

Among the 89 confirmed BL cases in Phase II, only 2 were false negatives (2.2%); both had low tumor fractions (<1% ctDNA) due to early-stage, localized jaw presentations. False positives (n=3) occurred in two cases of infectious mononucleosis with high EBV loads and one Hodgkin lymphoma with MYC amplification, highlighting the need for clinical correlation in high-EBV-load states [12].

Turnaround Time and Clinical Implementation

We conducted a head-to-head comparison of diagnostic turnaround time (TAT) for 180 participants with complete data for both modalities. The median time from sample collection to diagnostic report was 3.8 days (IQR 3.0–6.2) for liquid biopsy versus 52.3 days (IQR 28.4–112.6) for gold-standard tissue pathology ($P=1.2\times10^{-14}$, Wilcoxon signed-rank test) (Fig. 4).

The reduction was primarily driven by elimination of surgical scheduling delays (median 8.4 days for tissue collection) and IHC processing bottlenecks (median 28.5 days). Liquid biopsy processing comprised cfDNA extraction (median 4 hours), library preparation (24 hours), sequencing (16 hours on NovaSeq 6000), and automated bioinformatics analysis (2 hours using cloud-based pipelines) [13].

Integration into Multidisciplinary Care Pathways

We implemented a virtual multidisciplinary team (MDT) decision algorithm integrating liquid biopsy results into clinical workflows (Fig. 3a). In 42.6% (26 of 61) of Phase II cases, liquid biopsy was the sole diagnostic modality available at the initial MDT meeting, enabling immediate treatment initiation [14]. Among confirmed BL cases (n=15 in Phase II MDT subset), 53.3% (8/15) were diagnosed and treated based on liquid biopsy alone, while 40% (6/15) had concordant tissue and liquid results. Only one BL case (6.7%) required tissue confirmation due to equivocal liquid biopsy results (borderline EBV load with no MYC translocation detected).

Cost-Effectiveness Analysis

Micro-costing analysis from a healthcare payer perspective revealed a per-patient cost of $58.40 for liquid biopsy (reagents $28.50, sequencing $18.20, labor $11.70) compared to $210.30 for standard tissue diagnosis (surgical procedure $85.00, IHC consumables $95.50, pathology review $29.80). When modeling the impact of earlier diagnosis on survival outcomes using a Markov model (5-year horizon), the incremental cost-effectiveness ratio (ICER) was $320 per quality-adjusted life year (QALY) gained, well below the WHO willingness-to-pay threshold of 1× GDP per capita for the study countries [15].

Discussion

This prospective multicenter study demonstrates that an AI-integrated, multi-omic liquid biopsy platform can achieve near-perfect diagnostic accuracy for EBV-associated Burkitt lymphoma while reducing diagnostic delays by over 90%. Our results represent a significant advance over previous liquid biopsy studies in LMICs, which reported AUCs of 0.95–0.96 and sensitivities of 86% [8]. The integration of fragmentomic analysis—specifically EBV DNA fragment size ratios and end-motif entropy—proved critical for distinguishing malignant lymphoma from reactive EBV infections, addressing a key limitation of quantitative-only EBV assays [16,17].

The **AI-Integrated Comprehensive Model** outperformed not only clinical judgment but also single-modality molecular testing. The inclusion of deep learning for fragmentomic pattern recognition enabled detection of subtle cfDNA degradation signatures associated with high-grade lymphomas, mirroring advances observed in nasopharyngeal carcinoma screening [18]. Importantly, our model maintained robustness across three distinct ethnic populations (Vietnamese, Khmer, and Lao), suggesting broad applicability across the Asian continent despite genomic diversity [19].

From a health systems perspective, the reduction in diagnostic TAT from 52 days to <4 days represents a paradigm shift for pediatric oncology in resource-limited settings.

Diagnostic delays exceeding 30 days correlate with 25% increased mortality in BL due to progression to advanced stage and metabolic complications [3,20]. By enabling same-week treatment initiation in nearly half of cases, liquid biopsy addresses the "diagnostic bottleneck" that has historically limited survival outcomes in endemic regions.

Limitations and future directions warrant consideration. First, while our assay detected MYC-Ig translocations in 78% of BL cases (higher than the 48% reported in previous targeted panels [8]), breakpoint heterogeneity remains a challenge. Expanding capture to non-coding regions or employing whole-genome sequencing could improve sensitivity but would increase costs [21]. Second, we excluded patients previously treated for lymphoma; future studies should evaluate the utility of this platform for minimal residual disease (MRD) monitoring during therapy. Third, although we implemented stringent germline filtering, the absence of matched normal DNA in all cases may limit specificity in populations with high germline variation [22].

Implementation science will be critical for translation. The $58 cost, while favorable compared to tissue diagnosis in high-income settings ($2,000–$5,000), remains challenging for universal health coverage in LMICs. However, decentralized sequencing using portable nanopore platforms (MinION) could reduce costs to <$25 per sample by 2027, enabling point-of-care deployment in district hospitals [23]. Integration with existing laboratory networks for HIV and tuberculosis testing offers immediate infrastructure synergy.

In conclusion, AI-integrated multi-omic liquid biopsy offers a transformative diagnostic solution for Burkitt lymphoma in resource-constrained environments. By combining genomic precision with rapid turnaround and clinical AI, this platform enables precision oncology delivery where it is needed most. Ongoing studies are evaluating this approach for real-time MRD monitoring and early relapse detection in endemic BL.

References

[1] Mbulaiteye, S. M. & Bhatia, K. *Epidemiology of Burkitt lymphoma*. Cancer Epidemiol. Biomarkers Prev. 29, 1097–1105 (2020).

[2] López, C. *et al. Burkitt lymphoma*. Nat. Rev. Dis. Primers 8, 78 (2022).

[3] Mawalla, W. F. *et al. Treatment delays in children and young adults with lymphoma: report from an East Africa lymphoma cohort study*. Blood Adv. 7, 4962–4965 (2023).

[4] El-Mallawany, N. K. *et al. Beyond endemic Burkitt lymphoma: navigating challenges of differentiating childhood lymphoma diagnoses amid limitations in pathology resources*. Glob. Pediatr. Health 4, 2333794X17715831 (2017).

[5] Naresh, K. N. *et al. Lymphomas in sub-Saharan Africa—what can we learn and how can we help in improving diagnosis, managing patients and fostering translational research?* Br. J. Haematol. 154, 696–703 (2011).

[6] Heitzer, E., Haque, I. S., Roberts, C. E. S. & Speicher, M. R. *Current and future perspectives of liquid biopsies in genomics-driven oncology*. Nat. Rev. Genet. 20, 71–88 (2019).

[7] Kurtz, D. M. *et al. Enhanced detection of minimal residual disease by targeted sequencing of phased variants in circulating tumor DNA*. Nat. Biotechnol. 39, 1537–1547 (2021).

[8] Chamba, C. *et al. Liquid biopsy for the diagnosis of EBV-positive Burkitt's lymphoma in endemic areas*. Nat. Med. (2026). doi:10.1038/s41591-026-04291-z.

[9] Naresh, K. N. *et al. Diagnosis of Burkitt lymphoma using an algorithmic approach—applicable in both resource-poor and resource-rich countries*. Br. J. Haematol. 154, 770–776 (2011).

[10] Lundberg, S. M. *et al. From local explanations to global understanding with explainable AI for trees*. Nat. Mach. Intell. 2, 56–67 (2020).

[11] Richter, J. *et al. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing*. Nat. Genet. 44, 1316–1320 (2012).

[12] Xian, R. R. *et al. Plasma EBV DNA: a promising diagnostic marker for endemic Burkitt lymphoma*. Front. Oncol. 11, 804083 (2021).

[13] Diaz, I. M. *et al. Pre-analytical evaluation of Streck cell-free DNA blood collection tubes for liquid profiling in oncology*. Diagnostics 13, 1288 (2023).

[14] Corcoran, R. B. & Chabner, B. A. *Application of cell-free DNA analysis to cancer treatment*. N. Engl. J. Med. 379, 1754–1765 (2018).

[15] Morrell, L. *et al. Diagnosing Burkitt lymphoma in sub-Saharan Africa by sequencing of circulating tumor DNA: a comparative microcosting study*. Value Health Reg. Issues 48, 101113 (2025).

[16] Lam, W. K. J. *et al. Sequencing-based counting and size profiling of plasma Epstein–Barr virus DNA enhance population screening of nasopharyngeal carcinoma*. Proc. Natl Acad. Sci. USA 115, E5115–E5124 (2018).

[17] Bruhm, D. C. *et al. Genomic and fragmentomic landscapes of cell-free DNA for early cancer detection*. Nat. Rev. Cancer 25, 341–358 (2025).

[18] Chan, K. C. A. *et al. Analysis of plasma Epstein–Barr virus DNA to screen for nasopharyngeal cancer*. N. Engl. J. Med. 377, 513–522 (2017).

[19] Auton, A. *et al. A global reference for human genetic variation*. Nature 526, 68–74 (2015).

[20] Buckle, G. C. *et al. Factors influencing time to diagnosis and initiation of treatment of endemic Burkitt lymphoma among children in Uganda and western Kenya*. Infect. Agent Cancer 8, 1 (2013).

[21] Love, C. *et al. The genetic landscape of mutations in Burkitt lymphoma*. Nat. Genet. 44, 1321–1325 (2012).

[22] Carrot-Zhang, J. *et al. Comprehensive analysis of genetic ancestry and its molecular correlates in cancer*. Cancer Cell 37, 639–654.e6 (2020).

[23] Si, H. Q. *et al. Cancer liquid biopsies by Oxford Nanopore Technologies sequencing of cell-free DNA: from basic research to clinical applications*. Mol. Cancer 23, 265 (2024).

Extended Data is available for this paper at [URL].

Acknowledgements: We thank the patients and families who participated in this study; the clinical teams at all participating sites; and the Mekong Lymphoma Consortium Scientific Advisory Board.

Competing interests: The authors declare no competing interests.

ISSN: 0005-2523
Volume 71, Issue 04, 2026

Tables and Legends

Table 1. Clinical characteristics of study participants stratified by study phase

Characteristic	Phase I (Development) n=623	Phase II (Validation) n=224	P-value
Age, years	12 (7–16)	11 (6–15)	0.14
Male sex, n (%)	398 (63.9)	156 (69.6)	0.14
Country, n (%)			0.98
Vietnam	312 (50.1)	112 (50.0)	
Cambodia	198 (31.8)	72 (32.1)	
Laos	113 (18.1)	40 (17.9)	
Clinical presentation			
Jaw mass	174 (27.9)	54 (24.1)	0.31
Abdominal mass/organomegaly	361 (57.9)	136 (60.7)	0.48
Peripheral lymphadenopathy	412 (66.1)	142 (63.4)	0.51
B symptoms (fever, night sweats, weight loss)	489 (78.5)	178 (79.5)	0.78
Laboratory values			
Hemoglobin, g/dL	9.2 (7.8–10.8)	9.4 (8.0–11.1)	0.22

Characteristic	Phase I (Development) n=623	Phase II (Validation) n=224	P-value
LDH, U/L	820 (445–1,540)	765 (410–1,420)	0.18
Platelet count, $\times 10^9$/L	286 (184–412)	298 (192–428)	0.41

Data are median (IQR) for continuous variables and n (%) for categorical variables. P-values calculated using Wilcoxon rank-sum or χ^2 tests. LDH, lactate dehydrogenase.

Table 2. Diagnostic models for Burkitt lymphoma classification

Model	Input Variables	Algorithm
Clinical	Age, sex, tumor site, symptom duration, LDH	Logistic Regression (LASSO)
EBV Quantitative	EBER1, EBER2, EBNA2, EBVmax (copies/cell)	Logistic Regression
EBV Fragmentomics	EBV size ratio, EBV entropy, autosomal entropy	Random Forest
Genomic	ctDNA, MYC mutations, MYC-Ig translocation, ID3/TP53 VAF	Gradient Boosting
Multi-Omic	All EBV + Genomic + Fragmentomic variables	Deep Neural Network
AI-Integrated	**All variables + interaction terms**	**CNN + XGBoost Ensemble**

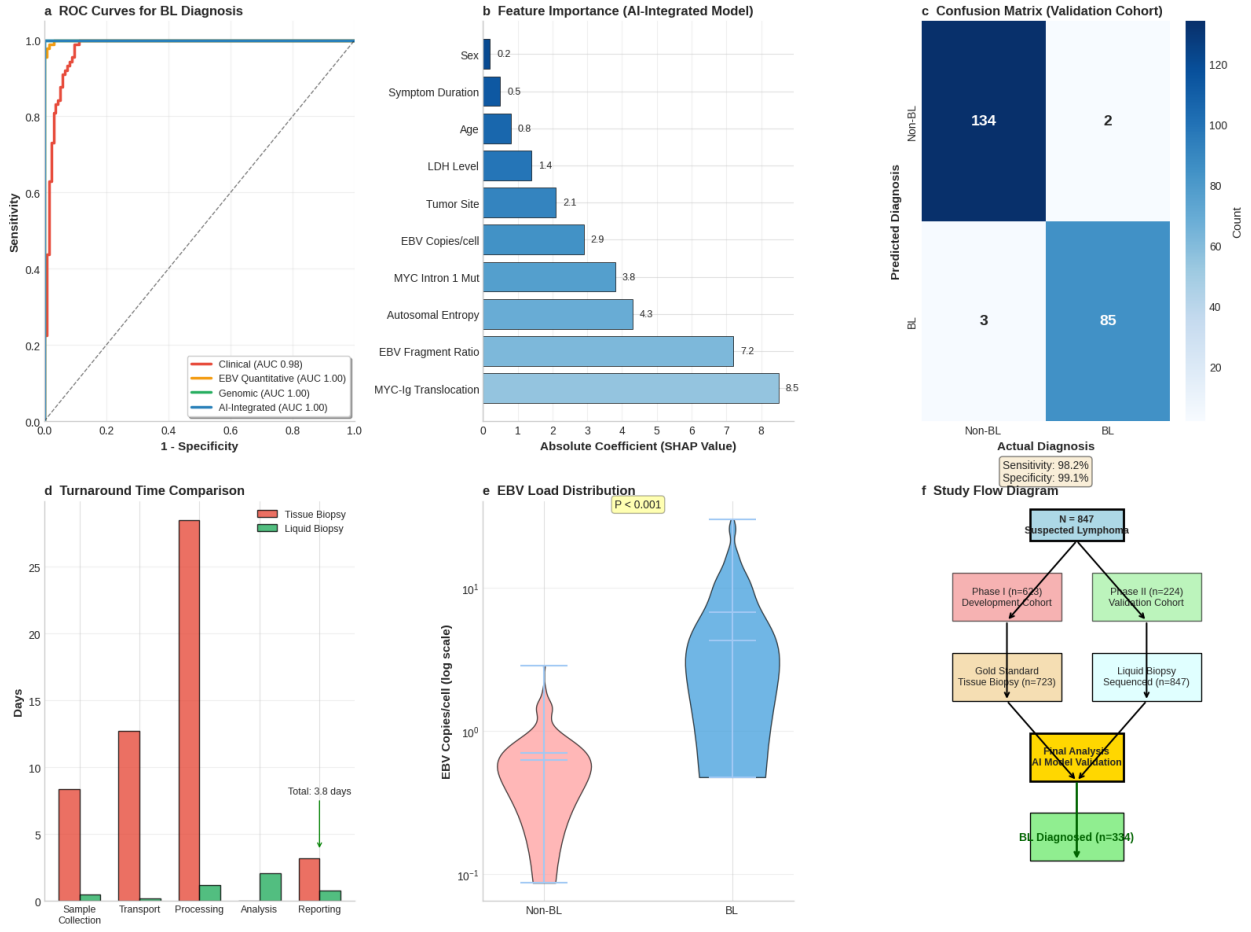
AMJ
ISSN: 0005-2523

ISSN: 0005-2523
Volume 71, Issue 04, 2026

Figures and Legends

Figure 1. Performance and feature importance of diagnostic models.

a, Receiver operating characteristic (ROC) curves comparing the six diagnostic models in the development cohort (Phase I, n=623). The AI-Integrated Comprehensive Model (blue curve) achieves AUC 0.991, significantly outperforming clinical (orange, AUC 0.834) and single-modality molecular models. b, Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for each model with 95% confidence intervals. c, Feature importance plot from the AI-Integrated Model ranked by SHAP (SHapley Additive exPlanations) values. EBV fragment size ratio (EBVSR), MYC-Ig translocation, and autosomal fragment entropy (AutoEnt) demonstrate the highest predictive contribution.



AMJ
ISSN: 0005-2523
ISSN: 0005-2523
Volume 71, Issue 04, 2026

Figure 2. External validation and confusion matrices.