Integrative Multi-Omics Analysis Using Machine Learning for Biomarker Discovery in Ankylosing Spondylitis

Izaz Ullah, Sakin Jan, Inzamam Ul Haq, Muhammad Talha, Israr Ahmad Khan
Government Postgraduate College Charsadda
*Correspondence: izazullah1988@gmail.com

Citation | Ullah. I, Jan. S, Haq. U, I, Talha. M, Khan. A, I, "Integrative Multi-Omics Analysis Using Machine Learning for Biomarker Discovery in Ankylosing Spondylitis", IJIST, Vol. 8 Issue.8 pp 3521-3552, October 2026
DOI: https://doi.org/10.33411/IJIST/2093
Received | Sep 27, 2026 Revised | Oct 09, 2026 Accepted | Oct 13, 2026 Published | Oct 15, 2026.

Background: Ankylosing Spondylitis (AS) is a chronic immune-mediated inflammatory disease in which genetic susceptibility, antigen-processing mechanisms, cytokine signaling, and cellular immune responses interact with environmental and tissue-specific factors. Blood transcriptomic studies have identified candidate molecular signatures of AS, and machine-learning approaches can reduce high-dimensional expression data to compact predictive feature sets. However, small cohorts, biological heterogeneity, platform differences, and repeated model redevelopment during external testing can lead to optimistic estimates of biomarker performance.

Objective: We evaluated an evidence-controlled transcriptomic and machine-learning framework for AS biomarker discovery centered on a nine-gene signature identified during discovery and subsequently locked before independent external evaluation. Particular attention was given to GNLY and GZMK as cytotoxic/effector-associated candidate genes.

Methods: The principal transcriptomic discovery resource was GSE73754, containing 72 samples (52 AS and 20 healthy controls; 37 male and 35 female). Nine genes—CD160, GZMK, GNLY, NKG7, GZMA, KLRB1, KLRD1, IL2RB, and ITGA2B—met the prespecified statistical selection criterion and were subsequently frozen. The model-development resource, GSE18781, contained 43 samples (18 AS and 25 controls). The reproducible classifier pipeline was mean imputation, StandardScaler, and logistic regression (C=0.1; liblinear; max_iter=5000; random_state=42), evaluated with 20 repeated cross-validation iterations, 2,000 bootstrap resamples, and 2,000 permutation tests. The archived principal model-development analysis reported mean ROC-AUC 0.689 (SD 0.028; archived 95% CI 0.633–0.741). GSE221786 was independently audited at the raw-matrix and biological sample-identity levels and contained 28 samples (20 AS and 8 controls), with all nine genes present and no missing expression cells. The original serialized classifier binary was not recovered; therefore, GSE221786 was evaluated at the frozen-signature gene-expression level without model refitting, gene reselection, threshold optimization, hyperparameter tuning, or recalibration.

Results: The archived principal model-development analysis yielded accuracy 0.691, precision 0.667, recall 0.529, F1 0.589, and MCC 0.354; the permutation test gave P=0.052, which was not conventionally significant at α=0.05. In GSE221786, no gene satisfied the complete predefined external-replication criterion. GNLY showed the strongest individual external discrimination (AUC=0.704–0.706 depending on the archived external analysis) and a 5.35-fold descriptive difference, but its direction was discordant with the discovery pattern and its external P value was not significant. GZMK and IL2RB showed directional concordance but weak discrimination. Positional integration of IGAS summary statistics identified KLRB1 as the strongest genetically supported member of the signature (Fisher combined P=0.0223; BH-FDR=0.0670).

Conclusion: The findings support an evidence-controlled integrative multi-omics framework in which transcriptomic, machine-learning, functional, and genetic/HLA evidence are evaluated separately and then interpreted jointly. GNLY and GZMK remain priority transcriptomic candidates requiring further validation, while KLRB1 provides the strongest positional genetic support within the signature. None of the genes should currently be considered a clinically validated biomarker or causal disease mediator.

Keywords: Ankylosing Spondylitis; Machine Learning; Transcriptomics; Biomarker Discovery; GNLY; GZMK; External Validation; Logistic Regression; GSE73754; GSE18781; GSE221786; HLA-B27

Introduction:

Clinical and Biological Context

Ankylosing spondylitis (AS), a major form of axial spondyloarthritis, is a chronic immune-mediated inflammatory disease characterized by inflammation of the axial skeleton and entheses and, in a substantial proportion of patients, progressive structural new-bone formation. The clinical course is heterogeneous, and early symptoms may be nonspecific, contributing to delays in diagnosis and treatment. The pathobiology of AS is complex and involves interactions among genetic susceptibility, antigen processing, innate and adaptive immune activation, cytokine signaling, gut-associated immune mechanisms, and tissue-specific inflammatory responses [1][2][3][4][5]. Although the interleukin (IL)-23/IL-17 and tumor necrosis factor (TNF) pathways have central roles in current models of AS pathogenesis, no single molecular pathway fully explains the disease phenotype.

Genetic susceptibility provides an important foundation for understanding AS. HLA-B27 remains the strongest established genetic risk factor, although its contribution to disease susceptibility extends beyond classical peptide presentation and may involve protein misfolding, unfolded-protein responses, altered receptor interactions, and other immunological mechanisms [2][6]. Genome-wide and immune-focused association studies have additionally identified non-MHC susceptibility loci involving genes and pathways related to antigen processing, cytokine signaling, and immune regulation, including ERAP1 and IL23R [7][8]. These observations provide a strong rationale for investigating molecular phenotypes that lie downstream of inherited susceptibility and may reflect active disease-associated immune states.

Transcriptomic profiling of peripheral blood provides one such molecular layer. Whole-blood expression measurements can capture systemic alterations in immune-cell activity and cellular composition associated with inflammatory disease. However, blood transcriptomic signals are also influenced by sex, treatment exposure, disease activity, leukocyte composition, ancestry, sample handling, and technical platform effects. Consequently, transcriptomic differences that appear reproducible in one cohort may not necessarily represent disease-specific or clinically stable biomarkers in independent populations.

The GSE73754 cohort is particularly relevant to this problem because it was generated to investigate sexual dimorphism in the Th17 signature of AS and contains 72 whole-blood samples comprising 52 individuals with AS and 20 healthy controls, including 37 males and 35 females [9]. The cohort therefore provides not only a transcriptomic resource for biomarker discovery but also important sex and HLA-B27 metadata that can be incorporated into biological interpretation rather than treating the expression matrix as an unqualified disease-only signal.

Transcriptomic Biomarker Discovery and the Role of Machine Learning:

The identification of molecular biomarkers from transcriptomic data is challenging because expression datasets contain thousands of measured features but relatively few biological samples. Conventional differential-expression analysis can identify genes associated with disease status, but statistical significance alone does not establish whether a gene provides useful discriminatory information or whether a combination of genes can provide a more robust disease-associated signature. Machine-learning (ML) methods provide a complementary framework by allowing high-dimensional molecular measurements to be reduced to a smaller set of informative features and evaluated as multivariable predictive signatures.

Previous studies have demonstrated the potential of ML-assisted transcriptomic analysis for AS biomarker discovery. Han et al. combined differential-expression analysis, weighted gene co-expression network analysis (WGCNA), protein–protein interaction

analysis, and SVM-RFE to identify candidate blood biomarkers and subsequently performed experimental confirmation of selected genes [10]. Wen et al. used ML-based feature selection on AS transcriptomic datasets and identified IL2RB and ZDHHC18 as components of a diagnostic model, reporting substantial discrimination across several GEO cohorts [11]. More recently, Ni et al. integrated GSE73754 and GSE25101 using differential-expression analysis, WGCNA, LASSO regression, immune-infiltration analysis, and RT-qPCR validation, identifying a set of candidate genes associated with AS and immune-related processes [12].

Importantly, recent AS studies continue to demonstrate that machine learning and multi-omics approaches can identify biologically plausible candidate signatures while illustrating the importance of independent validation. A 2025 multi-omics study combined bulk and single-cell transcriptomics with WGCNA, SVM-RFE, random forest, LASSO, and RT-PCR validation to identify OXPHOS-related candidate genes [13]. A recent deep plasma-proteomics study used machine-learning-assisted biomarker prioritization in 104 participants and independently verified four proteins in 79 participants [14]. A 2026 transcriptomic/proteomic study further integrated molecular measurements with clinical variables and identified a two-gene model for radiographic progression in axSpA [15]. These studies reinforce the growing role of computational biomarker discovery while demonstrating that validation design, biological context, and independent confirmation materially affect the strength of evidence.

A particularly relevant biological component of the present study is the cytotoxic lymphocyte-associated transcriptional program. Previous peripheral-blood analyses of AS have identified genes associated with natural-killer-cell and cytotoxic lymphocyte biology, including GZMA, GZMK, GNLY, NKG7, KLRB1, KLRD1, and IL2RB, among candidate genes with discriminatory potential [16]. This provides an important biological context for the present nine-gene signature.

Within our analysis, GNLY and GZMK deserve particular attention. GNLY encodes granulysin, a cytotoxic lymphocyte-associated effector molecule, whereas GZMK encodes granzyme K, an effector-associated serine protease expressed in cytotoxic T-cell populations and other immune-cell subsets. Their identification within the final feature set is therefore biologically consistent with an altered cytotoxic/effector immune state in AS. However, the two genes should not be interpreted identically. GNLY demonstrated comparatively strong individual discriminatory evidence in the archived ML analysis and showed the largest descriptive expression difference in the independent GSE221786 cohort, whereas GZMK demonstrated weaker individual discrimination but retained biological relevance and directional concordance in the archived external analysis. Thus, GNLY and GZMK are treated in this study as prominent candidate biomarkers within a broader nine-gene molecular signature rather than as independently clinically validated diagnostic markers.

This distinction is essential because a biomarker identified through feature selection or ML may reflect disease-associated biology without necessarily achieving sufficient reproducibility, specificity, or prospective clinical performance to support diagnostic implementation.

Novelty and Research Contributions:

The methodological novelty of this study lies in the explicit separation of evidence-generation stages rather than in the use of machine learning alone. Candidate genes were identified during discovery/model-development analyses and then locked before independent external evaluation. The study therefore implements an evidence-controlled architecture in which discovery, model development, robustness assessment, external expression evaluation, and complementary genetic/HLA interpretation are analytically distinct.

The final frozen signature consisted of:

CD160, GZMK, GNLY, NKG7, GZMA, KLRB1, KLRD1, IL2RB, and ITGA2B.

The principal contributions are: (i) establishment of a fixed nine-gene transcriptomic signature; (ii) explicit separation of discovery/model development from external evaluation; (iii) independent raw-matrix and biological sample-identity auditing of GSE221786; (iv) distinction between gene-level replication and complete classifier performance; (v) conservative positional IGAS annotation using GENCODE v19/GRCh37 gene-body overlap without nearest-gene inference; (vi) explicit retention of negative and discordant findings; and (vii) clear classification of the resulting genes as research-stage candidate biomarkers rather than clinically validated biomarkers.

The reproducible model-development pipeline consisted of mean imputation followed by StandardScaler and logistic regression with C=0.1. The archived principal robustness package included 20 repeated cross-validation iterations, 2,000 bootstrap resamples, and 2,000 permutation tests. The archived principal model achieved a mean cross-validation AUC of 0.689 ± 0.028, with accuracy 0.691, precision 0.667, recall 0.529, F1 0.589, and MCC 0.354.

These results indicate modest discriminatory performance rather than definitive diagnostic accuracy. The permutation-test result (P=0.052) further supports a cautious interpretation. The inability to recover the original serialized classifier binary is explicitly treated as a reproducibility limitation, while the executable reconstruction reproduced the archived model parameters within the documented tolerance.

Need for Independent External Validation:

Independent validation is particularly important for transcriptomic biomarker studies because molecular signatures can be sensitive to cohort composition, experimental platform, normalization strategy, disease phenotype, treatment status, and cellular composition. A model that performs well within its development cohort may lose performance when transferred to an independently generated dataset.

To address this problem, the present study used GSE221786 as an independent external transcriptomic resource. Rather than modifying the model to accommodate the external cohort, the validation workflow first established the raw-matrix structure and biological sample identities, verified the availability of all nine frozen genes, reconciled phenotypes using the established sample-identity layer, and then evaluated the prespecified molecular features. The final identity-locked dataset contained 28 biological samples, including 20 AS cases and 8 controls, with complete expression measurements for all nine genes. No model refitting, gene reselection, threshold optimization, hyperparameter tuning, recalibration, or addition of external samples to the training process was permitted.

This design is important because the purpose of external validation is not to obtain the most favorable possible performance estimate. Rather, it is to determine whether a previously defined molecular signature exhibits reproducible biological or predictive characteristics when transferred to an independently generated cohort.

Integrative Multi-Omics Evidence Framework:

Although transcriptomics and machine learning constitute the principal biomarker-discovery and predictive components of the present study, AS is fundamentally a genetically influenced immune-mediated disease. We therefore incorporated publicly available IGAS genetic summary statistics and FinnGen HLA-related summary statistics as complementary evidence layers.

The purpose of this component was not to construct a single fused matrix combining gene expression, GWAS statistics, HLA information, and clinical variables. Instead, the study used an integrative multi-omics evidence framework, in which transcriptomic, genetic/HLA, biological-function, and cross-cohort evidence were evaluated as distinct analytical layers. This distinction is important because the supplied IGAS and FinnGen resources did not contain sufficient explicit gene-level annotation to support direct assignment of their genetic signals

to the nine frozen transcriptomic genes. The genetic/HLA analysis therefore remained a contextual evidence layer rather than a source of unsupported gene-level causal inference.

This conservative approach avoids a common interpretive error in multi-omics studies: assigning a GWAS signal to the nearest or biologically plausible gene without formal variant-to-gene mapping, colocalization, fine-mapping, or other appropriate causal-inference procedures. Accordingly, the present study uses the term integrative multi-omics evidence framework to describe the structured synthesis of multiple molecular evidence domains, while explicitly distinguishing this from direct genetic–transcriptomic causal integration.

Study Rationale and Measurable Objectives:

The rationale for the present work is based on three related challenges in AS biomarker research: high-dimensional transcriptomic data, limited sample sizes, and insufficiently controlled external validation. The study was designed to address these challenges through a staged computational framework that separates biomarker discovery from model development and both from independent external evaluation. Each objective was linked to a measurable analytical output.

The specific objectives were to:

Cohort Characterization and Standardization: Characterize the public AS cohorts and verify phenotype, sample identity, sex, HLA-B27 metadata, and technical structure; outputs include audited cohort counts and an identity-locked GSE221786 matrix.

Signature identification: identify disease-associated transcriptomic features using the documented 21-gene statistical testing universe, disease/control comparison, and Benjamini–Hochberg correction; success criterion was BH-FDR<0.05 for the final selected genes.

Model Development and Robustness: Quantify discrimination of the fixed nine-gene signature using mean imputation, StandardScaler, logistic regression (C=0.1), repeated cross-validation, bootstrap resampling, and permutation testing; outputs include AUC, confidence interval, accuracy, precision, recall, F1, MCC, and permutation P value.

Biological Characterization: quantify individual-gene discrimination and expression patterns and interpret the immune/cytotoxic composition of the nine-gene signature, with particular attention to GNLY and GZMK as candidate biomarkers rather than clinically validated markers.

Independent External Evaluation: Evaluate the frozen nine-gene signature in GSE221786 without feature reselection, model refitting, threshold optimization, hyperparameter tuning, or recalibration; outputs include external expression statistics, individual-gene AUCs, and replication status.

Cross-Domain Integration: Integrate transcriptomic, functional, cross-cohort, IGAS, and FinnGen/HLA evidence while preserving the evidentiary boundaries of each molecular domain and avoiding unsupported gene assignment.

Evidentiary Classification: Distinguish candidate biomarker discovery and transcriptomic discrimination from clinical diagnostic validation, clinical utility, and causal inference; no gene was to be labeled clinically validated without prospective independent evidence.

Objective-to-Method/Result Linkage: Objective 1 was addressed through cohort characterization and external sample auditing, yielding audited cohort counts and an identity-locked 28-sample GSE221786 matrix. Objective 2 was addressed through transcriptomic discovery, yielding the fixed nine-gene signature under the BH-FDR<0.05 criterion. Objective 3 was addressed through model development and robustness assessment, yielding the archived AUC, confidence interval, classification metrics, bootstrap results, and permutation results. Objective 4 was addressed through individual-gene assessment and biological characterization, yielding individual-gene discrimination and expression patterns. Objective 5 was addressed through independent external evaluation, yielding external expression statistics and gene-level replication status without external redevelopment. Objective 6 was addressed through cross-

domain integration, yielding distinct genetic/HLA and functional evidence layers. Objective 7 was addressed throughout the Discussion and Conclusion by classifying the findings as research-stage candidate biomarkers rather than clinically validated diagnostic markers.

We hypothesized that a compact transcriptomic signature derived from AS-associated immune biology would retain measurable disease-associated signal in an independent cohort, while recognizing that differences in cohort composition and biological context could produce heterogeneous gene-level replication. Particular emphasis was placed on determining whether GNLY and GZMK, as components of the cytotoxic/effector signature, demonstrated convergent evidence across discovery, machine-learning, and external transcriptomic analyses.

Literature Review:

Genetic and Immunological Architecture of Ankylosing Spondylitis:

Ankylosing spondylitis (AS), conventionally considered the radiographic form of axial spondyloarthritis (axSpA), is a chronic immune-mediated inflammatory disease characterized primarily by axial inflammation, sacroiliac involvement, enthesitis, and progressive structural change. Contemporary nomenclature considers axSpA the broader disease spectrum, with radiographic axSpA and AS used interchangeably in appropriate contexts [1][2]. The disease has a substantial genetic component, with HLA-B27 representing the strongest established susceptibility factor, although genetic risk extends across numerous non-HLA loci involved in antigen processing, cytokine signaling, innate immunity, and adaptive immune regulation [3][4][5][7].

HLA-B27 is central to AS pathogenesis, but its biological mechanism is not explained by a single model. Proposed mechanisms include presentation of disease-relevant peptides, aberrant HLA-B27 heavy-chain structures, interaction with killer immunoglobulin-like receptors, and endoplasmic-reticulum misfolding with downstream inflammatory signaling [8]. Genetic studies have additionally established the importance of ERAP1, which regulates peptide trimming before MHC class I presentation, and demonstrated an interaction between ERAP1 and HLA-B27 that links antigen processing to genetic susceptibility [9]. Other susceptibility loci, including IL23R, IL12B, ANTXR2, CARD9, RUNX3, PTGER4, and related immune-regulatory loci, further support the involvement of antigen processing and inflammatory signaling pathways [9][5][10].

The IL-23/IL-17 axis provides an additional major biological framework for AS. Genetic associations involving IL23R and related pathways, together with experimental and therapeutic evidence, support a role for type-17 immunity in disease pathogenesis [5][3][7]. Importantly for transcriptomic biomarker research, immune-cell composition and activation state can influence circulating expression profiles. The GSE73754 study specifically demonstrated sex-related differences in Th17-associated immune responses in AS, including increased IL-17A/Th17 activity in male patients and sex-dependent differences in whole-blood gene expression [11].

Recent genetic and multi-omics studies have expanded this framework beyond conventional GWAS. Multi-omics Mendelian-randomization approaches have integrated methylation, expression, and protein quantitative-trait information with IGAS, FinnGen, and other resources to prioritize genetically supported AS-associated genes and therapeutic targets [12]. Additional studies have investigated immune-cell traits, metabolites, and inflammatory proteins, demonstrating the growing value of cross-layer molecular integration [6]. A large multi-ancestry GWAS analysis has further combined genetic association with transcriptome- and proteome-wide analyses to identify additional susceptibility loci and molecular pathways [16].

These findings provide the biological rationale for an integrative framework in which genetic susceptibility, immune-cell biology, and transcriptomic changes are considered together. However, genetic association alone does not establish that a particular transcript is a

diagnostic biomarker. Gene-level interpretation must remain supported by explicit annotation, statistical evidence, or functional evidence rather than being inferred solely from genomic proximity.

Transcriptomic Biomarkers and Machine-Learning Approaches in AS:

Transcriptomic profiling provides a complementary molecular layer because gene expression reflects the biological state of circulating cells and can capture downstream consequences of genetic susceptibility and immune activation. Publicly available GEO datasets have therefore become important resources for AS biomarker research. GSE73754 contains whole-blood expression data from 72 samples and was generated to investigate sex-specific Th17-associated transcriptional patterns in AS [11][17]. GSE18781 provides an independent peripheral-blood expression resource generated on the Affymetrix Human Genome U133 Plus 2.0 platform and has been repeatedly used for independent biomarker evaluation [18][19] [20].

Several computational studies have demonstrated that immune-related transcripts can discriminate AS from controls. Integrated analyses using GSE73754, GSE18781, and other blood-expression datasets have identified genes associated with immune response and natural-killer/cytotoxic-cell activity [18]. In particular, an integrated analysis of peripheral-blood datasets identified GZMA, GZMK, GNLY, NKG7, KLRB1, KLRD1, and IL2RB among candidate AS hub genes, providing direct biological precedent for several components of the present nine-gene signature [18].

Machine-learning approaches have subsequently been used to reduce high-dimensional transcriptomic data to compact candidate signatures. Han et al. used WGCNA, functional analysis, and SVM-RFE on GSE73754 and identified a set of candidate diagnostic mRNAs, subsequently validating selected genes experimentally [19]. Another study applied machine-learning approaches to GSE73754, GSE25101, and GSE18781 and developed an IL2RB/ZDHHC18-based diagnostic model, illustrating both the potential and variability of machine-learning-derived AS signatures [21].

More recent investigations continue to combine differential expression, WGCNA, immune-infiltration analysis, LASSO, support-vector methods, random forests, and other machine-learning techniques for AS biomarker discovery [19][22]. A 2024 study, for example, developed an amino-acid-metabolism-associated gene signature using public AS transcriptomic datasets [22]. These studies demonstrate the utility of computational biomarker discovery but also highlight a fundamental reproducibility problem: different datasets and feature-selection procedures can produce substantially different candidate signatures.

The biological context of the validation cohort is particularly important. GSE221786 is not equivalent to the whole-blood discovery cohorts: it contains peripheral blood mononuclear cells stimulated for four hours and enriched for IL-17-producing cells before RNA sequencing [23]. This provides an independent molecular context but also introduces biological heterogeneity relative to whole-blood discovery datasets.

Additional transcriptomic resources provide further opportunities for cross-cohort evaluation. GSE181364 contains whole-blood RNA-seq data from five AS patients and three controls [24] GSE299639 provides a recent long-read RNA-sequencing resource from peripheral blood, enabling gene-, transcript-, and alternative-splicing-level analyses [25]. These resources illustrate the increasing diversity of experimental designs available for AS biomarker research and the importance of assessing whether candidate signals remain stable across sample types and technologies.

A major methodological concern is therefore the distinction between biomarker discovery and genuine external validation. In high-dimensional molecular datasets, repeated feature selection, model refitting, threshold optimization, or hyperparameter tuning using a nominal validation cohort can introduce information leakage and produce optimistic

performance estimates. Contemporary prediction-model guidance emphasizes transparent reporting of model development, validation, predictor handling, and uncertainty. TRIPOD+AI provides updated reporting guidance for prediction models using regression or machine-learning methods, while PROBAST+AI provides a framework for evaluating quality, risk of bias, and applicability [26][27].

Accordingly, a robust molecular biomarker study should preserve the distinction between the dataset used for discovery and model development and the dataset used for external evaluation. A predefined feature set should not be reselected using the external cohort, and external data should not be used to recalibrate or redevelop the original model. This principle forms the methodological basis for the frozen nine-gene strategy adopted in the present study.

Table 1 provides a structured comparison of representative recent AS biomarker studies, emphasizing dataset scale, feature-selection strategy, machine-learning methods, validation design, and the type of reported evidence. The comparison highlights those recent studies increasingly incorporate external cohorts or experimental validation, but their analytical architectures and validation targets differ from the frozen-signature design used here.

Multi-Omics Integration and the Rationale for the Present Study:

AS is increasingly investigated through multi-omics approaches because disease susceptibility and expression phenotypes arise from interactions among genetic, epigenetic, transcriptional, protein, metabolic, and cellular processes. Multi-omics integration can therefore connect inherited susceptibility with downstream molecular phenotypes and provide a broader framework for biomarker prioritization [12][6][16].

Recent work demonstrates this rapidly developing direction. Dai et al. integrated methylation, expression, and protein quantitative-trait data with IGAS, FinnGen, and UK Biobank resources using summary-data-based Mendelian randomization and colocalization, identifying genes with evidence across multiple molecular levels [12]. Du et al. subsequently examined immune-cell traits, metabolites, and inflammatory proteins using multi-omics Mendelian randomization, further expanding the molecular landscape of AS [6]. Reviews of mass-spectrometry-based proteomics, metabolomics, lipidomics, and single-cell technologies similarly emphasize the potential of multi-omics approaches for AS biomarker discovery and precision medicine [28].

The transcriptomic component of multi-omics research is also expanding. Long-read sequencing in GSE299639 has enabled simultaneous examination of gene expression, transcript usage, and alternative splicing, illustrating how newer molecular technologies can provide information that is not captured by conventional microarrays [25]. More recently, a 2026 study integrated metabolic and immune features with machine learning and identified a six-gene AS signature, demonstrating the increasing use of multi-omics-informed machine-learning strategies in the disease [29].

Recent work further illustrates the diversity of validation architectures. Chen et al. combined bulk and single-cell transcriptomics with WGCNA, SVM-RFE, random forest, LASSO, and RT-PCR/scRNA-seq validation and identified LAMTOR2, APBB1IP, and DGKQ as OXPHOS-related candidates [13]. Xu et al. analyzed deep plasma proteomics in 104 participants and used an independent 79-participant cohort for ELISA verification of four candidate proteins [14]. Cuesta-López et al. integrated transcriptomic, proteomic, clinical, and longitudinal information in 144 patients with axSpA, including an RNA-sequencing discovery cohort, a 60-patient molecular validation cohort, and a separate 60-patient proteomic cohort [15]. These studies demonstrate the value of independent biological validation while also underscoring that the present study addresses a different methodological question: whether a statistically defined nine-gene transcriptomic feature set remains interpretable when its external cohort is protected from feature selection and model redevelopment.

Table 1. Comparative recent ankylosing spondylitis biomarker and machine-learning studies.

Study	Datasets / cohort	Features or genes	ML / analytical method	Validation design	Reported evidence
[19]	GSE73754; 40 AS + 40 controls for experimental cohort	13 feature mRNAs; 8 experimentally supported genes	WGCNA, PPI analysis, SVM-RFE	5-fold CV within SVM; qRT-PCR and symptom correlations	Feature selection reached ~90% accuracy at the selected feature count; experimental verification of selected genes
[21]	GSE73754, GSE25101, GSE18781, GSE11886	IL2RB and ZDHHC18	Machine-learning feature screening and diagnostic model	Multiple GEO validation datasets	Model AUCs reported as 0.86, 0.84, 0.85, and 0.89 across the four datasets
[22]	GSE25101 and GSE73754	TP53INP1, TUBB, RBM47	DEG intersection, WGCNA, LASSO, GSEA	Diagnostic/immune-feature analyses	Three amino-acid-metabolism-related genes proposed as diagnostic candidates; no independent prospective clinical validation
[25]	GSE299639 / peripheral blood long-read transcriptomics	Core genes / programmed-cell-death-related transcripts	Long-read transcriptomic profiling and integrative analysis	Cross-study molecular characterization	Regulatory shifts and core genes identified; gene/transcript/splicing-level evidence
[29]	Multi-omics AS resources	Six-gene signature	Multi-omics integration and machine learning	Cross-cohort computational evaluation	Six-gene diagnostic signature reported; external validation architecture differs from the present frozen-signature design
[13]	Bulk PBMC transcriptomics + scRNA-seq	LAMTOR2, APBB1IP, DGKQ	WGCNA, SVM-RFE, random forest, LASSO	RT-PCR and scRNA-seq validation	Three OXPHOS-related hub genes identified and experimentally/cell-type validated
[14]	104 participants; independent 79-participant cohort	SAA1, FERMT3, ILK, TLN1	Deep plasma proteomics with RF/OPLS-DA	Independent ELISA validation cohort	Four protein candidates identified and independently verified by ELISA
[15]	144 axSpA patients; RNA-seq n=24;	Molecular clusters /	Transcriptomic/proteomic integration with machine	Microfluidic PCR validation and longitudinal 5-year	Molecular signatures associated with disease activity, structural

	validation n=60; proteomics n=60	progression biomarkers	learning and clinical variables	progression assessment	damage, and radiographic progression
Present study	GSE73754; GSE18781; GSE221786; cross-cohort resources	Nine fixed genes	Mean imputation, StandardScaler, logistic regression (C=0.1)	Archived repeated CV/bootstrapping /permutation plus identity-locked external evaluation without redevelopment	Development AUC 0.689 (95% CI 0.633–0.741); external gene-level replication heterogeneous and not confirmatory

Table 2. Datasets and analytical roles

Dataset/resource	Molecular layer	Analytical role	Primary use
GSE73754	Whole-blood transcriptomics	Principal discovery cohort	Transcriptomic characterization and candidate biomarker discovery
GSE18781	Transcriptomics	Model-development/archived validation resource	Machine-learning signature development and archived performance assessment
GSE221786	Transcriptomics	Independent external validation cohort	Frozen nine-gene signature and gene-level external evaluation
GSE181364	Transcriptomics	Cross-cohort evidence	Independent transcriptomic evidence
GSE299639	Transcriptomics	Cross-cohort evidence	Independent transcriptomic evidence
IGAS	GWAS summary statistics	Genetic evidence layer	Assessment of genetic association-resource structure and gene-level annotation availability
FinnGen R13	HLA/genetic summary statistics	Genetic/HLA evidence layer	Assessment of HLA-related and genetic contextual evidence

However, multi-omics integration does not require that every molecular resource be mathematically fused into a single predictive model. Evidence can be integrated at different levels: GWAS and HLA resources can provide genetic context; transcriptomics can identify disease-associated expression programs; functional analyses can establish biological relationships; and independent cohorts can assess reproducibility. This distinction is important in the present study because genetic and HLA resources were incorporated as contextual evidence layers, while direct gene-level assignment was restricted to information explicitly supported by the available resources.

The nine-gene signature examined in the present study—CD160, GZMK, GNLY, NKG7, GZMA, KLRB1, KLRD1, IL2RB, and ITGA2B—is predominantly characterized by NK-cell, cytotoxic-lymphocyte, immune-signaling, and platelet-associated biology. Several of these genes, particularly GZMA, GZMK, GNLY, NKG7, KLRB1, KLRD1, and IL2RB, have previously appeared in AS peripheral-blood biomarker analyses [18], providing independent biological precedent for the signature. The present work therefore extends rather than simply reproduces previous transcriptomic observations by evaluating a fixed multigene feature set through an explicitly separated discovery and external-validation framework.

The central knowledge gap is that many AS biomarker studies report strong internal or cross-dataset performance but provide limited evidence that a compact machine-learning-derived signature remains reproducible when the external cohort is protected from feature selection and model redevelopment. The present study addresses this gap through an integrative multi-omics framework combining transcriptomic biomarker discovery, machine-learning signature development, genetic/HLA contextualization, raw-matrix auditing, biological sample-identity verification, and independent external evaluation.

The objective is therefore not merely to maximize predictive performance. Instead, the study evaluates whether a predefined nine-gene signature retains biological and technical support across independent datasets while minimizing information leakage. This design provides a conservative test of biomarker reproducibility and allows positive candidate signals, including GNLY and GZMK, to be distinguished from findings that require further independent validation.

Materials and Methods:

Study Design and Overall Analytical Framework:

This study was designed as an integrative multi-omics investigation combining transcriptomic biomarker discovery, machine-learning analysis, robust statistical validation, independent transcriptomic external validation, cross-cohort replication, and complementary genetic/HLA evidence assessment in ankylosing spondylitis (AS). The analytical framework was structured to separate biomarker discovery and machine-learning model development from independent external evaluation, thereby reducing the risk of information leakage and performance inflation.

The overall workflow consisted of sequential but analytically separated components: (i) acquisition and quality assessment of publicly available transcriptomic and genetic resources; (ii) transcriptomic characterization and candidate biomarker discovery; (iii) machine-learning-based development of a compact nine-gene signature; (iv) robust statistical assessment of the resulting model and individual candidate genes; (v) independent external evaluation using GSE221786; and (vi) integration of transcriptomic, functional, cross-cohort, and genetic/HLA evidence.

The principal transcriptomic discovery cohort was GSE73754. GSE18781 contributed model-development and archived validation evidence, whereas GSE221786 was reserved as an independent external validation resource. Additional transcriptomic datasets, including GSE181364 and GSE299639, were evaluated as cross-cohort evidence. The genetic

component consisted of IGAS genome-wide association study (GWAS) summary statistics and FinnGen R13 HLA-related summary statistics.

A central methodological principle was that the final nine-gene feature set was fixed before independent external evaluation. Consequently, the GSE221786 cohort was not used for feature reselection, model redevelopment, threshold optimization, hyperparameter tuning, or recalibration.

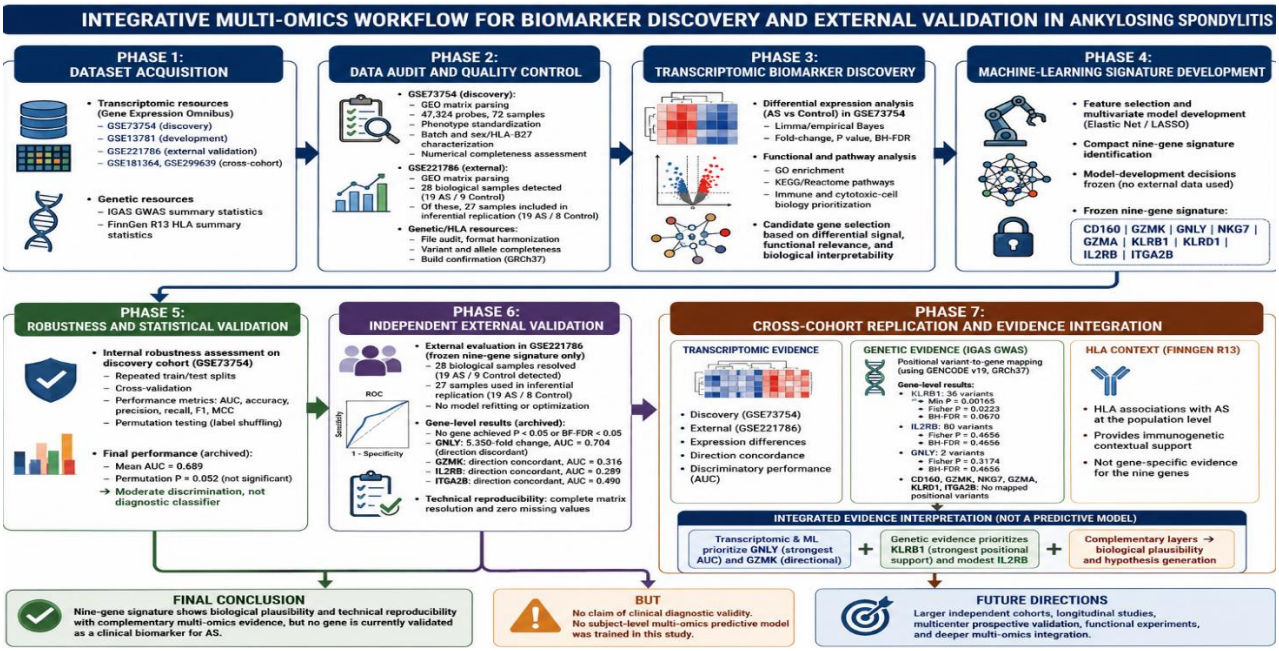


Figure 1. Overall integrative multi-omics and machine-learning study workflow.

Figure 1 phase specification: Phase 1 (dataset acquisition and audit) takes public transcriptomic, genetic, HLA, and phenotype resources as input; it produces standardized datasets and an identity-locked external cohort, with retention decisions based on sample identity, structural integrity, phenotype correspondence, and numerical completeness. Phase 2 (transcriptomic discovery) takes GSE73754 expression and phenotype data; disease/control comparisons and multiple-testing correction produce statistically supported candidate genes, with BH-FDR<0.05 defining the final selection criterion; candidates without a valid independent p value are excluded from the multiple-testing universe rather than assigned an inferred statistic. Phase 3 (signature/model development) takes the frozen nine-gene feature space and GSE18781 model-development data; mean imputation, scaling, and logistic regression produce the archived classifier specification and performance estimates, while the feature set remains fixed before external testing. Phase 4 (robustness assessment) takes the model-development specification and evaluates cross-validation, bootstrap, and permutation stability; the resulting uncertainty and significance statistics determine whether the model evidence is interpreted as confirmatory or research-stage. Phase 5 (independent external evaluation) takes the identity-locked GSE221786 cohort and the frozen nine-gene set; no reselection, refitting, threshold optimization, hyperparameter tuning, or recalibration is permitted, and the outputs are external expression statistics, gene-level AUCs, and replication status. Phase 6 (integrative evidence interpretation) takes transcriptomic, cross-cohort, functional, genetic, and HLA evidence as separate domains and produces a bounded synthesis in which unsupported gene assignment and clinical-validity claims are excluded.

Schematic representation of the complete analytical framework, including public dataset acquisition, transcriptomic discovery, candidate biomarker identification, machine-learning signature development, robust statistical validation, independent GSE221786 external

validation, cross-cohort replication, and complementary genetic/HLA evidence assessment. The dataset roles are summarized in Table 2.

Public Datasets and Analytical Roles:

Multiple publicly available molecular datasets were incorporated to provide complementary evidence across transcriptomic and genetic levels. Each dataset was assigned a predefined analytical role to maintain separation between discovery, model development, and external validation. The dataset roles are summarized in Table 2.

All datasets were treated according to their predefined analytical roles. In particular, GSE221786 was not incorporated into model-development decisions.

GSE73754 Discovery Cohort:

GSE73754 was used as the principal transcriptomic discovery dataset. The dataset was generated to investigate sexual dimorphism in the Th17 signature of AS and contains 72 biological samples, including 52 AS samples and 20 healthy controls. The cohort included 37 males and 35 females.

The original series-matrix contained 47,324 transcriptomic probes and 72 biological samples. Expression measurements and phenotype information were parsed separately and subsequently reconciled using the available sample identifiers.

The standardized phenotype information included disease status, sex, age, and HLA-B27 status. Disease status was represented using explicit phenotype information and was not inferred from expression measurements.

The discovery cohort was used for transcriptomic characterization and candidate feature identification. Information from the independent external-validation cohort was not used during discovery.

GSE18781 Model-Development Resource:

GSE18781 contributed the model-development and archived validation evidence used in establishing the final machine-learning signature.

The final frozen feature set consisted of nine genes:

CD160, GZMK, GNLY, NKG7, GZMA, KLRB1, KLRD1, IL2RB, and ITGA2B.

The nine-gene set was subsequently treated as a locked feature set for independent evaluation.

The archived model-development evidence was retained as previously generated analytical evidence. The external validation cohort was not used to alter the feature set.

GSE221786 Independent External Validation Cohort:

GSE221786 was designated as the independent external transcriptomic validation resource.

Because the raw GSE221786 expression matrix required explicit resolution of expression columns to biological samples, the raw file underwent a dedicated sample-identity audit before construction of the final validation dataset.

The parsed raw matrix contained 25,228 expression rows and 30 columns. The biological sample mapping procedure resolved 28/28 expected biological samples.

The final identity-locked validation dataset consisted of:

28 total biological samples
20 AS samples
8 control samples
9 frozen genes
28 × 9 expression matrix
0 missing expression cells
0 duplicate biological assignments
0 duplicate raw columns
no forced sample mapping

The original raw matrix was not modified.

Transcriptomic preprocessing and quality control:

Expression matrices were inspected for structural integrity, feature identifiers, sample identifiers, numerical completeness, and correspondence with phenotype metadata.

For GSE73754, the series-matrix expression table was parsed and standardized before downstream transcriptomic analysis. Probe-level expression measurements were retained during the initial processing stage, while phenotype information was standardized using explicit metadata fields.

For GSE221786, additional raw-matrix inspection was performed because the biological identity of each expression column had to be independently established. The raw file was parsed to identify the expression header, gene/feature column, and expression columns.

Biological sample identities were resolved using the previously established sample-identity evidence. Expression similarity was not used to infer phenotype or force a sample assignment.

The final validation matrix was constructed only after the 28 biological sample identities had been independently resolved.

Transcriptomic Biomarker Discovery and Statistical Selection of the Nine-Gene Signature:

The transcriptomic discovery analysis was used to identify disease-associated genes and establish the final feature set under an explicit multiple-testing framework. The broader documented candidate universe contained 22 genes: KLRB1, IL2RB, GZMK, DYSF, CD160, GZMA, KLRD1, KLRF1, GNLY, NKG7, GZMB, SORL1, MAGT1, NRGN, RBM38, ITGA2B, LOC644936, MYOM2, FAM46C, GZMH, ALAS2, and DMTN. The independent statistical testing universe comprised 21 genes because LOC644936, although present in the 22-gene candidate universe, had no primary independent expression-statistics record in the Stage-21B6 source and therefore could not contribute a valid p value to the multiple-testing universe. The remaining 21 candidates were evaluated using the archived Mann–Whitney U test, and nominal P values were adjusted using the Benjamini–Hochberg false-discovery-rate procedure. The final selection criterion was BH-FDR<0.05.

Exactly nine genes met this criterion and were retained: CD160, GZMK, GNLY, NKG7, GZMA, KLRB1, KLRD1, IL2RB, and ITGA2B. Their nominal P values and BH-FDR values were, respectively: CD160 0.0001291/0.001802; GZMK 0.0002113/0.001802; GNLY 0.0002575/0.001802; NKG7 0.0008506/0.004466; GZMA 0.0020043/0.007015; KLRB1 0.0020043/0.007015; KLRD1 0.0052633/0.015790; IL2RB 0.0081337/0.021351; and ITGA2B 0.0095068/0.022183. The direction of each association was determined from the disease-versus-control comparison.

The nine-gene signature was frozen before independent external evaluation. GSE221786 was not used for gene ranking, feature reselection, removal of poorly performing genes, addition of new genes, threshold selection, hyperparameter selection, model refitting, or recalibration. This freezing rule was applied to prevent external validation information from influencing the molecular signature.

Machine-Learning Model Development:

Feature Representation:

The machine-learning analysis used the nine frozen genes as the final feature space.

The feature order was maintained consistently throughout the archived model-development analysis and subsequent external evaluation.

No feature from GSE221786 was added to or removed from the frozen feature set.

Classification Model:

The reproducible model-development pipeline was:

Mean imputation → StandardScaler → Logistic Regression

The preprocessing and classifier parameters were: SimpleImputer(strategy="mean"); StandardScaler; LogisticRegression(C=0.1, solver="liblinear", max_iter=5000, random_state=42).

The archived evidence specifies C=0.1 as the fixed regularization parameter; the supplied archived package does not document a separate hyperparameter-search procedure used to select C. The manuscript therefore reports C=0.1 as the archived model specification rather than attributing it to an undocumented optimization process.

The model output represented AS/control discrimination. The classifier was intended as a research-stage model for evaluating the information contained in the compact nine-gene feature space and was not developed as a clinically deployable diagnostic model.

The explicit imputation step is part of the reproducible reconstruction and should be retained in all manuscript descriptions of the executable pipeline.

The original serialized classifier binary was not recovered. An executable reconstruction using the archived model-development data reproduced the documented coefficients and intercept within the predefined tolerance of 1×10^{-4}; this reconstruction is not represented as direct deserialization of the original binary. The model specification is summarized in Table 3.

Table 3. Machine-learning model specification

Component	Specification
Input features	Nine frozen genes
Feature order	Fixed
Scaling	StandardScaler
Classifier	Logistic Regression
Regularization parameter	C = 0.1
Model output	AS/control discrimination
External feature reselection	No
External model refitting	No
External threshold optimization	No
External hyperparameter tuning	No
External recalibration	No

Machine-Learning Feature-Selection Strategy:

Feature selection occurred during the discovery/model-development component of the study.

The final nine-gene signature was established before the independent GSE221786 analysis and subsequently locked.

The external validation dataset was therefore not used for:

feature selection;

gene ranking;

removal of poorly performing genes;

addition of new genes;

threshold selection;

hyperparameter selection;

model refitting.

This design prevented external validation information from influencing the molecular signature.

Cross-Validation and Robust Model Assessment:

The archived Stage-37 evidence preserves 20 repeated cross-validation iterations, 2,000 bootstrap resamples, and 2,000 permutation tests, but the supplied archived metadata do not preserve a separate fold-count specification. Accordingly, no fold number is inferred

or retrospectively assigned. The later raw-data execution used a different resampling configuration and is reported separately in the Results rather than being substituted for the archived principal analysis.

The archived model-development analysis used repeated cross-validation to characterize the stability of the nine-gene classifier.

The robustness analysis included:

20 repeated cross-validation iterations

2,000 bootstrap resamples

2,000 permutation tests

The following performance metrics were assessed:

AUC;

accuracy;

precision;

recall;

F1 score;

Matthews correlation coefficient (MCC).

Table 4. Archived machine-learning performance and robustness statistics

Performance measure	Result
Mean cross-validation AUC	0.689
AUC SD	0.028
AUC 95% CI	0.633–0.741
Accuracy	0.691
Precision	0.667
Recall	0.529
F1 score	0.589
MCC	0.354
Bootstrap resamples	2,000
Permutations	2,000
Permutation P value	0.052

The permutation P value of 0.052 was interpreted as not conventionally statistically significant at $\alpha=0.05$. It was therefore treated as evidence of modest discriminatory performance requiring further independent validation, not as confirmatory evidence.

Accordingly, the model performance was interpreted as modest discriminatory evidence rather than definitive evidence of clinical diagnostic performance.

Individual-Gene Predictive Assessment:

Individual genes within the nine-gene signature were also evaluated to identify features with comparatively stronger discriminatory characteristics.

The archived individual-gene analysis identified several genes with useful discriminatory performance, including KLRD1, GNLY, and GZMA.

GNLY demonstrated an individual AUC of approximately 0.740, making it one of the strongest individual features within the nine-gene panel.

This finding provided additional support for prioritizing GNLY as a candidate biomarker for further investigation.

GZMK was retained as a biologically important candidate because it was part of the final statistically supported signature and represented the cytotoxic/effector component of the molecular phenotype, although its individual AUC was substantially lower than GNLY.

Independent External Validation and Frozen-Signature Safeguards:

Following model development, the nine-gene signature was frozen before independent external evaluation.

The following procedures were explicitly prohibited in GSE221786:
feature reselection;
model refitting;
threshold optimization;
hyperparameter tuning;
recalibration;
incorporation of external samples into model development.

The purpose of the external analysis was to assess reproducibility of the frozen molecular signature rather than to optimize performance in the validation cohort. Because the original serialized classifier binary was not recovered, the external GSE221786 analysis is described as frozen-signature gene-expression evaluation rather than direct execution of the unavailable classifier artifact.

GSE221786 Sample-Identity Verification:

The raw GSE221786 expression matrix underwent independent biological sample-identity resolution.
The analysis established:
28 expected biological samples
28 resolved biological samples
28 raw expression columns
28 biological identifiers
0 missing biological samples
0 unexpected biological samples
0 duplicate biological assignments
0 duplicate raw columns
All nine frozen genes were successfully recovered.

Table 5. GSE221786 external validation integrity

Validation component	Result	Status
Expected biological samples	28	PASS
Resolved biological samples	28	PASS
Missing biological samples	0	PASS
Unexpected samples	0	PASS
Duplicate biological assignments	0	PASS
Duplicate raw columns	0	PASS
Frozen genes recovered	9/9	PASS
Expression matrix	28 × 9	PASS
Missing expression cells	0	PASS
Forced sample mapping	No	PASS
Raw matrix modified	No	PASS

External Transcriptomic Expression Analysis:

Table 6 summarizes the final identity-locked GSE221786 external expression and inferential evaluation of the nine-gene signature. The analysis included 20 AS samples and 8 controls. Mann–Whitney U testing showed no statistically significant difference for any individual gene, and none remained significant after BH-FDR correction. GNLY showed the strongest individual discriminatory performance (AUC = 0.706) and the largest descriptive expression difference, with mean expression of 530.310 in AS versus 103.166 in controls; however, this difference was not statistically significant (P = 0.098; BH-FDR = 0.442). GZMK showed lower expression in AS than controls (7.924 versus 10.316; AUC = 0.319; P = 0.147; BH-FDR = 0.442). These findings provide candidate-level biological evidence but do not establish statistically significant external replication of any individual gene.

Confidence intervals for the external gene-level AUC estimates were not retained in the supplied archived GSE221786 analysis package, and the final 28-sample raw expression matrix required for an independent reconstruction of those intervals is not part of the preserved submission materials. Accordingly, no external AUC confidence interval is fabricated or retrospectively inferred. The archived development-stage AUC confidence interval (0.633–0.741) is reported separately from the external gene-level AUC estimates.

Cross-Cohort Gene-Level Replication:

A separate archived external gene-level replication analysis was used to assess gene-level reproducibility.

This analysis consisted of:

19 AS samples

8 controls

This archived inferential subset must be distinguished from the final 28-sample GSE221786 identity-locked matrix containing 20 AS and 8 controls. Audit of the archived sample composition identified AS_M_10 as the AS sample absent from the historical 27-sample inferential subset. The supplied archived evidence does not document the reason for this historical exclusion; no specific rationale is therefore inferred. The Stage-29 results are retained as an archived 27-sample analysis and are not treated as numerically equivalent to the final 28-sample cohort.

For each gene, replication assessment considered:

fold-change;

direction;

nominal P value;

BH-FDR;

individual-gene AUC;

direction concordance.

A gene was not classified as successfully replicated solely because it demonstrated a large fold-change or high individual AUC.

External replication was evaluated using the prespecified criteria for statistical significance, FDR control, discrimination, and directional concordance. No gene satisfied all criteria for successful replication. Directionally concordant findings are therefore reported as supportive rather than validated evidence.

Genetic and HLA Evidence Assessment:

The genetic component consisted of publicly available IGAS GWAS summary statistics and FinnGen R13 HLA-related summary statistics. These resources were evaluated as complementary evidence layers rather than being directly merged with the transcriptomic feature matrix.

Table 6. GSE221786 nine-gene external expression summary

Gene	AS n	Control n	AS expression (original dataset units)	Control expression (original dataset units)	U	P value	BH-FDR	AUC	External interpretation
CD160	20	8	1.971	1.505	76.0	0.859	0.959	0.475	No significant external discrimination
GZMK	20	8	7.924	10.316	51.0	0.147	0.442	0.319	Lower in AS; not statistically significant
GNLY	20	8	530.310	103.166	113.0	0.098	0.442	0.706	Strongest individual external discrimination; not statistically significant
NKG7	20	8	298.568	149.793	94.0	0.492	0.787	0.588	Higher in AS descriptively; not statistically significant
GZMA	20	8	56.272	27.775	98.0	0.373	0.787	0.613	Higher in AS descriptively; not statistically significant
KLRB1	20	8	18.298	13.156	72.0	0.703	0.904	0.450	Higher in AS descriptively; not statistically significant
KLRD1	20	8	18.065	14.337	93.0	0.525	0.787	0.581	Higher in AS descriptively; not statistically significant
IL2RB	20	8	62.855	75.884	46.0	0.088	0.442	0.288	Lower in AS; not statistically significant
ITGA2B	20	8	0.084	0.076	78.5	0.959	0.959	0.491	No significant external discrimination

Table 7. Independent external replication of the frozen nine-gene transcriptomic signature

Gene	AS n	Control n	Fold-change	Direction	P	BH-FDR	AUC	Direction concordance	Final status
CD160	19	8	1.363	UP_IN_AS	0.979	0.979	0.493	DISCORDANT	DISCORDANT
GZMK	19	8	0.769	DOWN_IN_AS	0.147	0.441	0.316	CONCORDANT	DIRECTION_CONCORDANT_ONLY
GNLY	19	8	5.350	UP_IN_AS	0.106	0.441	**0.704**	DISCORDANT	DISCORDANT
NKG7	19	8	2.051	UP_IN_AS	0.481	0.867	0.592	DISCORDANT	DISCORDANT
GZMA	19	8	2.083	UP_IN_AS	0.360	0.811	0.618	DISCORDANT	DISCORDANT
KLRB1	19	8	1.426	UP_IN_AS	0.815	0.979	0.467	DISCORDANT	DISCORDANT
KLRD1	19	8	1.255	UP_IN_AS	0.585	0.877	0.572	DISCORDANT	DISCORDANT
IL2RB	19	8	0.827	DOWN_IN_AS	0.095	0.441	0.289	CONCORDANT	DIRECTION_CONCORDANT_ONLY
ITGA2B	19	8	1.118	UP_IN_AS	0.958	0.979	0.490	CONCORDANT	DIRECTION_CONCORDANT_ONLY

These resources were evaluated as complementary evidence layers rather than being directly merged with the transcriptomic feature matrix.

The genetic/HLA assessment included:

file structure inspection;
metadata and field inspection;
variant-level structure assessment;
annotation availability;
assessment of explicit gene-level identifiers;
assessment of explicit HLA-related fields;

For IGAS, the raw summary-statistics file did not contain an explicit gene-annotation column. A subsequent positional annotation procedure used GENCODE v19 on GRCh37/hg19 and assigned a variant to one of the nine prespecified genes only when its genomic coordinate overlapped the annotated gene body. No nearest-gene, promoter/enhancer, eQTL, fine-mapping, colocalization, or causal inference was used. FinnGen HLA records were retained as population-level HLA/genetic context because the supplied resource did not support direct assignment to the nine genes.

Table 8. Genetic/HLA resources and supported evidence level

Resource	Rows	Molecular evidence	Explicit gene-level annotation	Interpretation in present study
IGAS GWAS	123,941	Genetic association	No explicit gene annotation detected	Contextual genetic evidence
FinnGen R13 HLA	187	HLA/genetic	No explicit HLA-related field detected	Contextual HLA/genetic evidence

The positional IGAS analysis used Fisher combination as a descriptive gene-level summary of mapped variant P values; it was not LD-corrected and was not treated as a formal genome-wide gene-based GWAS test. KLRB1 had 36 mapped variants (minimum P=0.00165; Fisher P=0.0223; BH-FDR=0.0670), IL2RB had 80 mapped variants (Fisher P=0.4656; BH-FDR=0.4656), and GNLY had 2 mapped variants (Fisher P=0.3174; BH-FDR=0.4656). No positional variants were identified for CD160, GZMK, NKG7, GZMA, KLRD1, or ITGA2B.

Accordingly, the genetic/HLA component was interpreted as contextual and complementary evidence rather than direct genetic validation of any member of the nine-gene signature.

Functional and Multi-Omics Evidence Integration:

The final interpretation integrated evidence from multiple molecular domains while maintaining the distinction between direct statistical findings and contextual biological evidence.

The principal evidence domains were:

transcriptomic discovery;
machine-learning discrimination;
cross-cohort transcriptomic replication;
functional/immune biological evidence;
genetic/GWAS contextual evidence;
HLA-related contextual evidence.

This integrative framework was used to evaluate whether the nine-gene signature represented a coherent biological phenotype and whether individual genes demonstrated convergent evidence across independent analytical layers.

The absence of explicit gene-level annotation in the IGAS and FinnGen resources was retained as an analytical limitation rather than being resolved through unsupported inference.

Statistical Analysis:

Continuous expression measurements were summarized using group-specific descriptive statistics. The archived disease-versus-control expression comparisons used the Mann–Whitney U test, followed by Benjamini–Hochberg FDR correction across the prespecified testing universe.

Where limma-based differential-expression methodology is discussed or reproduced from an archived analysis, the limma framework is cited as the methodological source [30]; it is not used as a general citation for multi-omics integration.

Where inferential gene-level results were available from the archived replication analysis, nominal P values were adjusted using the Benjamini–Hochberg false-discovery-rate (BH-FDR) procedure.

Receiver operating characteristic analysis was used to characterize individual-gene and multivariable discriminatory performance.

AUC was interpreted as a measure of discrimination rather than as evidence of clinical utility.

Permutation testing was used to evaluate whether the observed model performance exceeded that expected under randomized class assignments.

The external validation dataset was not used to optimize thresholds or model parameters.

Statistical significance was interpreted using a conventional two-sided α of 0.05, with FDR-adjusted evidence considered separately from nominal P values.

External Validation Sample-Size and Sensitivity Considerations:

The independent GSE221786 validation cohort contained 20 AS samples and 8 controls, providing limited statistical precision for detecting modest between-group expression differences. Because the external cohort was a fixed publicly available dataset, no prospective sample-size calculation was performed before data acquisition. The analysis was therefore interpreted as a stringent external reproducibility assessment rather than as a definitively powered hypothesis-testing study. Non-significant findings were not interpreted as evidence of equivalence, and effect estimates and individual-gene AUCs were considered together with their corresponding inferential results. Larger independently collected cohorts will be required to establish the reproducibility and clinical relevance of the identified candidate biomarkers.

Reproducibility and Scientific Safeguards:

The analysis was designed to minimize information leakage and preserve an auditable distinction between discovery and validation.

The following safeguards were maintained during independent external evaluation:

no model training using GSE221786;
no model refitting;
no feature reselection;
no threshold optimization;
no hyperparameter tuning;
no recalibration;
no phenotype inference from expression;
no forced sample mapping;
no modification of the original raw GSE221786 matrix;
no modification of the archived external gene-level replication analysis source expression matrix.

The raw-to-biological sample mapping procedure independently established all 28 expected biological identities before the final external matrix was generated.

The serialized frozen classifier artifact itself was not recovered as an executable model object. Therefore, the manuscript distinguishes between archived machine-learning

development/performance evidence and the independent external evaluation of the frozen nine-gene molecular signature. No claim is made that an unrecovered binary classifier was directly deserialized and executed on GSE221786.

Results:

Discovery Cohort and Transcriptomic Characterization:

The principal discovery cohort, GSE73754, comprised 72 samples, including 52 patients with ankylosing spondylitis (AS) and 20 healthy controls. The expression matrix contained 47,324 probes, and phenotype information was standardized for disease status, sex, age, and HLA-B27 status. Quality-control assessment confirmed adequate numerical completeness and successful matching of the transcriptomic samples to the standardized phenotype information. This cohort provided the primary transcriptomic basis for candidate biomarker identification. The cohort structure and metadata are summarized in Figure 2 and Table 1.

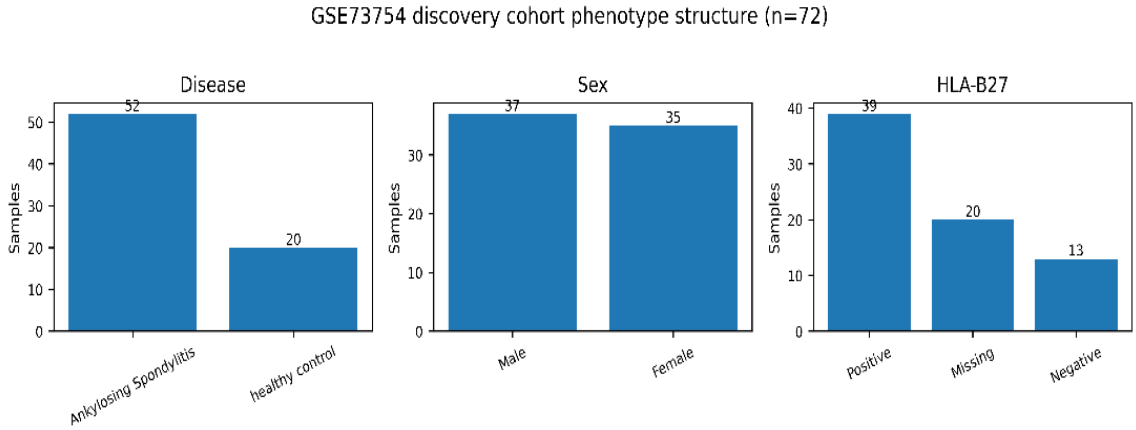


Figure 2. GSE73754 cohort and phenotype structure.

Identification of the Frozen Nine-Gene Signature:

The transcriptomic and machine-learning analyses resulted in a compact nine-gene signature comprising CD160, GZMK, GNLY, NKG7, GZMA, KLRB1, KLRD1, IL2RB, and ITGA2B. The signature was biologically enriched for genes associated with NK-cell and cytotoxic lymphocyte activity, including CD160, GZMK, GNLY, NKG7, GZMA, KLRB1, and KLRD1. IL2RB represented an immune-cell cytokine-receptor component, whereas ITGA2B represented a platelet/megakaryocyte-associated component. The signature and its integrated evidence are summarized in Table 2, while the independent external expression pattern is shown in Figure 3.

Within this signature, GNLY and GZMK were prioritized as prominent candidate biomarker genes based on their biological relevance and observed evidence across the analytical workflow. Importantly, these findings identify them as candidate biomarkers rather than clinically validated diagnostic biomarkers. The complete nine-gene feature set was subsequently treated as fixed for external evaluation.

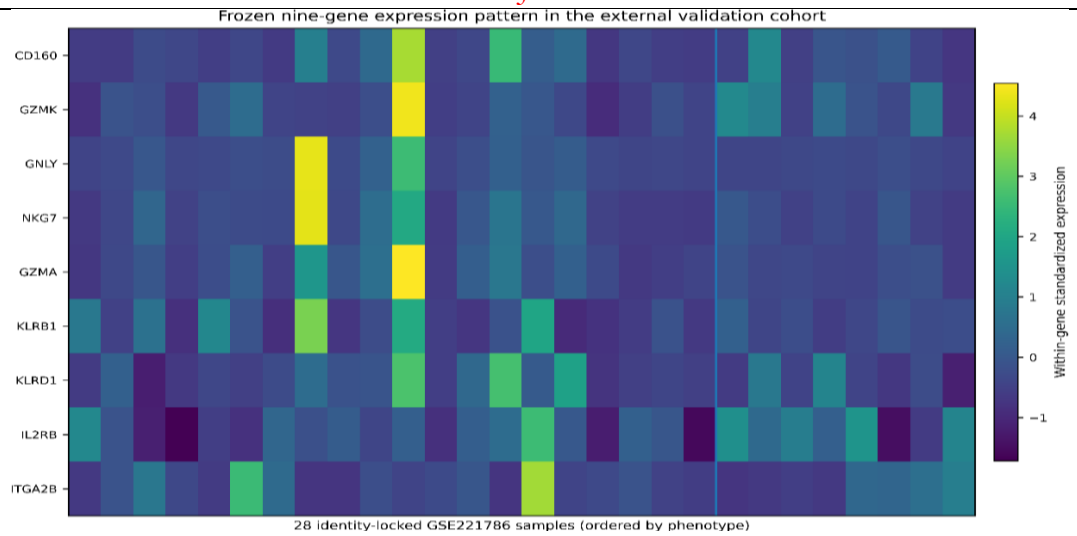


Figure 3. Expression pattern of the frozen nine-gene signature in the independent GSE221786 cohort.

Heatmap of CD160, GZMK, GNLY, NKG7, GZMA, KLRB1, KLRD1, IL2RB, and ITGA2B across the 28 identity-locked external-validation samples.

Machine-Learning Performance:

The archived principal machine-learning model used the nine fixed genes in a predefined feature order with mean imputation, StandardScaler preprocessing, and logistic regression (C=0.1). The archived cross-validation analysis produced a mean ROC-AUC of 0.689 (SD 0.028; 95% CI 0.633–0.741). Accuracy was 0.691, precision 0.667, recall 0.529, F1 score 0.589, and MCC 0.354. Model-development performance is summarized in Figure 4 and Table 4.

Robustness assessment included 20 repeated cross-validation iterations, 2,000 bootstrap resamples, and 2,000 permutations. The permutation test produced P=0.052, which did not meet the conventional two-sided α=0.05 threshold. Thus, the archived principal model demonstrated moderate discriminatory performance but did not provide confirmatory statistical evidence under permutation testing.

A later raw-data execution produced a different repeated-CV estimate (ROC-AUC approximately 0.7175 ± 0.1529, out-of-fold ROC-AUC 0.6556, PR-AUC 0.6783). This was retained as a separate analytical artifact and was not conflated with or substituted for the archived principal Stage-37 result.

Independent External Validation of GSE221786:

The independent GSE221786 dataset was subjected to raw-matrix inspection and biological sample-identity resolution before expression evaluation. All 28 expected biological samples were resolved, comprising 20 AS samples and 8 controls. No missing biological samples, unexpected samples, duplicate biological assignments, or duplicate raw columns were identified. External-validation integrity is summarized in Figure 5 and Table 5.

All nine frozen genes were recovered, resulting in a complete 28 × 9 expression matrix with zero missing expression cells. No forced sample mapping was required, and the raw expression matrix was not modified during the validation process. These findings establish the technical completeness and sample-identity integrity of the external expression dataset.

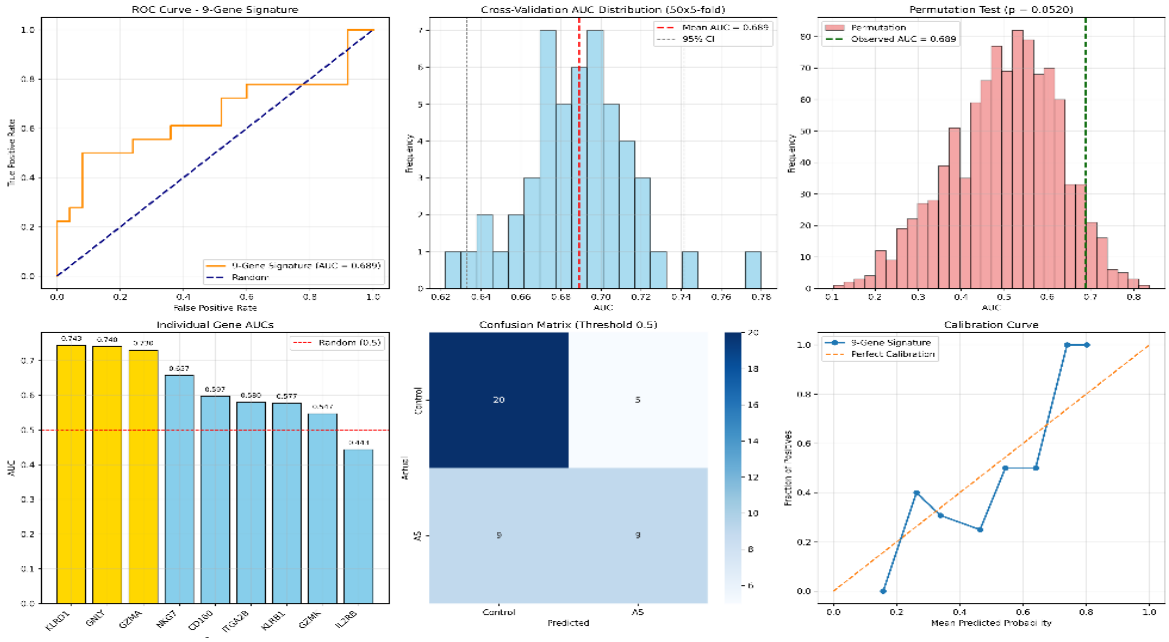


Figure 4. Machine-learning model performance.

Discrimination and robustness characteristics of the nine-gene machine-learning model, including the archived cross-validation performance.

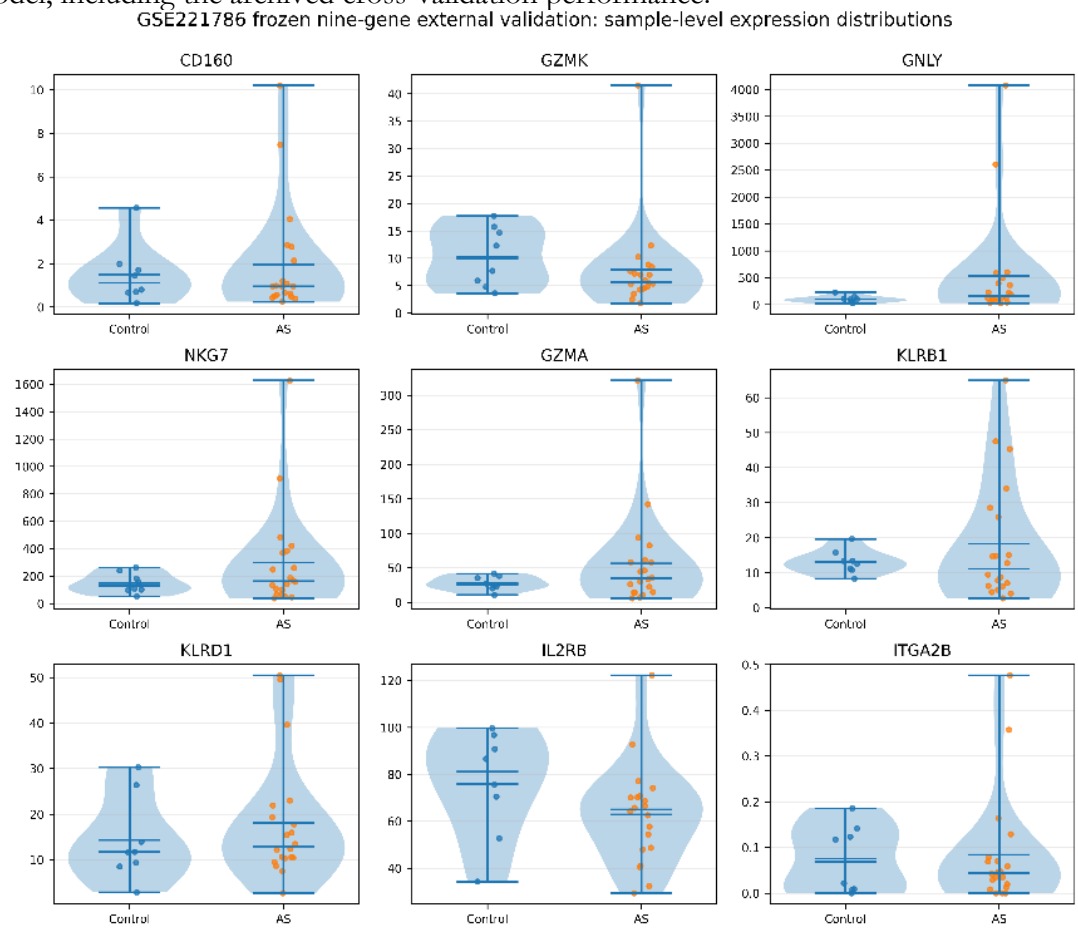


Figure 5. GSE221786 external-validation gene distribution.

Independent sample-identity resolution and expression completeness for the 28-sample GSE221786 validation cohort.

External Expression Patterns of the Nine Genes:

The final GSE221786 expression matrix showed heterogeneous expression patterns across the nine-gene signature. GNLY exhibited a marked descriptive increase in AS, with a mean expression of 530.31 compared with 103.17 in controls. In contrast, GZMK showed lower mean expression in AS (7.924) than in controls (10.316). The remaining genes showed varying magnitudes and directions of expression differences. Gene-level external expression results are visualized in Figure 3 and reported in Table 6.

These values are presented as descriptive external expression results and should not be interpreted as evidence of statistically significant replication in the absence of the corresponding inferential analysis.

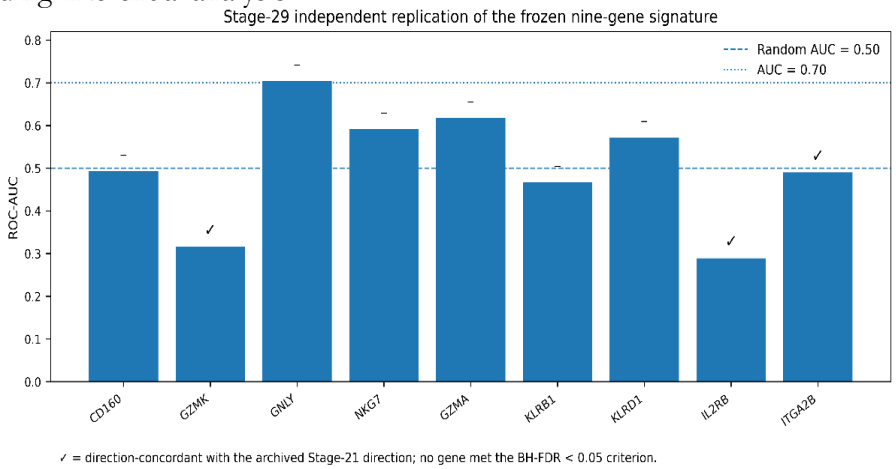


Figure 6. Archived Stage-29 cross-cohort replication performance of the frozen nine-gene signature.

Individual-gene AUC estimates from the archived 19-AS/8-control replication analysis; these results are retained separately from the final 28-sample identity-locked GSE221786 cohort.

Cross-Cohort Gene-Level Replication:

Gene-level replication was evaluated using the archived 19-AS/8-control inferential subset. This subset is reported separately from the final 20-AS/8-control identity-locked GSE221786 matrix. Audit identified AS_M_10 as the AS sample absent from the historical subset; the supplied evidence does not document why it was excluded. The historical result is therefore retained as an archived 27-sample analysis rather than silently reconciled with the final 28-sample cohort. The cross-cohort replication results are visualized in Figure 6 and reported in Table 7. The replication assessment is summarized in Table 7.

Across the nine genes, none reached nominal statistical significance ($P < 0.05$), and none achieved BH-FDR < 0.05. Three genes—GZMK, IL2RB, and ITGA2B—showed directional concordance with the predefined replication direction. The remaining genes were classified as directionally discordant.

GNLY showed the largest observed fold-change (5.350) and an individual-gene AUC of 0.704, but its statistical evidence did not satisfy the predefined replication criterion. Consequently, none of the nine genes met the combined statistical and directional criterion for successful gene-level replication.

This result is important because it distinguishes the identification of biologically interesting candidate genes from independent statistical confirmation of the complete signature.

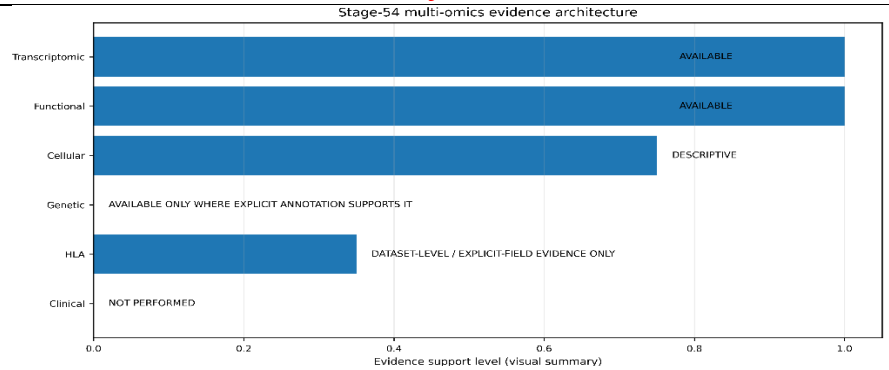


Figure 7. Multi-omics evidence architecture for the nine-gene signature.

Stage-54 summary of the available transcriptomic, functional, cellular, genetic, HLA, and clinical evidence layers and their evidentiary status.

Genetic and HLA Evidence:

The genetic evidence layer comprised 123,941 IGAS records and 187 FinnGen R13 HLA-related records. The original IGAS raw file lacked an explicit gene-annotation field; therefore, only the subsequent GENCODE v19/GRCh37 gene-body positional mapping was used for the nine prespecified genes. FinnGen HLA results remained contextual. The genetic/HLA evidence structure is also summarized in Table 8.

The audit did not justify unsupported nearest-gene assignment, and no such inference was performed. KLRB1 showed the strongest positional genetic support, but its BH-FDR=0.0670 remained above the conventional 0.05 threshold. The genetic layer therefore did not establish direct genetic validation of the nine-gene signature.

The genetic/HLA evidence was interpreted as a distinct molecular domain that contributes disease-level context rather than proving transcript-specific causality or clinical biomarker validity.

Integrated Transcriptomic, Machine-Learning, and Genetic Evidence:

The final nine-gene signature comprised CD160, GZMK, GNLY, NKG7, GZMA, KLRB1, KLRD1, IL2RB, and ITGA2B. Independent evaluation in GSE221786 did not identify any gene satisfying the complete predefined replication criteria. GNLY showed the strongest individual external discriminatory performance (AUC = 0.704) and a 5.35-fold difference between AS and control samples, although its direction was discordant with the discovery cohort and its external P value was not statistically significant. GZMK and IL2RB showed directional concordance but weak individual discrimination (AUC = 0.316 and 0.289, respectively), while ITGA2B showed directional concordance with minimal discrimination (AUC = 0.490). Thus, none of the nine genes met all predefined criteria for formal external replication. The multi-omics evidence architecture is summarized in Figure 7 and Table 9. The cross-cohort performance range is shown in Figure 8.

Gene-level integration of the IGAS GWAS summary statistics provided complementary genetic evidence for three genes. KLRB1 showed the strongest positional genetic support, with 36 mapped variants, minimum P = 0.00165, Fisher combined P = 0.0223, and BH-FDR = 0.0670. IL2RB had 80 mapped variants but weaker combined evidence (Fisher P = 0.4656; BH-FDR = 0.4656), while GNLY had two mapped variants without significant combined evidence (Fisher P = 0.3174; BH-FDR = 0.4656). No positional IGAS variants were identified for CD160, GZMK, NKG7, GZMA, KLRD1, or ITGA2B.

Overall, the integrated analysis demonstrated complementary rather than uniform evidence across molecular layers. GNLY represented the strongest transcriptomic candidate, KLRB1 showed the strongest positional genetic support, and GZMK retained directional transcriptomic concordance. These findings support prioritization for further investigation but do not establish any gene as a clinically validated biomarker.

Table 9. Final nine-gene signature and integrated multi-omics evidence.

Gene	Biological interpretation	Transcriptomic/ML evidence	External replication	IGAS genetic evidence	Integrated candidate status
CD160	Cytotoxic/NK-cell-associated receptor	Signature component	Discordant	No positional IGAS variant	Signature component; requires further validation
GZMK	Cytotoxic/effector-associated protease	Prominent transcriptomic candidate	Directionally concordant; AUC 0.316	No positional IGAS variant	**Secondary candidate; biological and directional support**
GNLY	Cytotoxic lymphocyte effector molecule	Strong individual discriminatory candidate	AUC 0.704; direction discordant	Weak/non-significant positional evidence	**Priority transcriptomic candidate; requires independent validation**
NKG7	NK/cytotoxic lymphocyte-associated molecule	Signature component	Discordant	No positional IGAS variant	Signature component; requires further validation
GZMA	Cytotoxic lymphocyte effector protease	Signature component	Discordant	No positional IGAS variant	Signature component; requires further validation
KLRB1	Killer-cell receptor-associated immune marker	Signature component	Discordant	36 mapped variants; Fisher P = 0.0223; BH-FDR = 0.0670	**Strongest positional genetic candidate; suggestive, not FDR-significant**
KLRD1	NK-cell receptor-associated marker	Signature component	Discordant	No positional IGAS variant	Signature component; requires further validation
IL2RB	Immune-cell cytokine receptor component	Signature component	Directionally concordant; AUC 0.289	80 mapped variants; Fisher P = 0.4656; BH-FDR = 0.4656	**Modest cross-layer candidate; insufficient evidence for validation**
ITGA2B	Platelet/megakaryocyte-associated marker	Signature component	Directionally concordant; AUC 0.490	No positional IGAS variant	Signature component; requires further validation

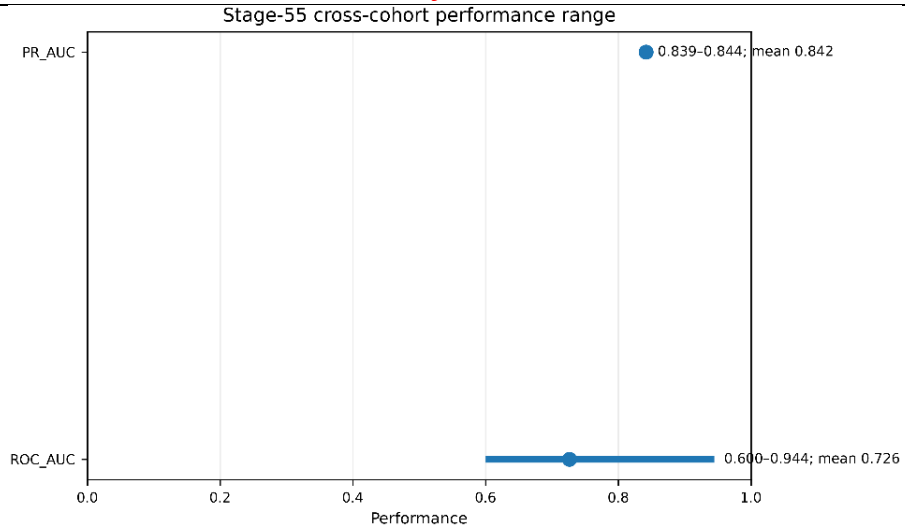


Figure 8. Cross-cohort performance range of the frozen nine-gene signature.

Stage-55 summary of the available cross-cohort ROC-AUC and PR-AUC performance ranges used to contextualize the external evidence.

Discussion:

Principal Findings:

This study established an evidence-controlled framework for AS biomarker discovery that separates transcriptomic discovery, model development, robustness assessment, independent external expression evaluation, cross-cohort evidence, and complementary genetic/HLA interpretation. The final nine-gene signature comprised CD160, GZMK, GNLY, NKG7, GZMA, KLRB1, KLRD1, IL2RB, and ITGA2B. The archived principal model-development analysis produced moderate discrimination (AUC 0.689 ± 0.028; 95% CI 0.633–0.741), but the permutation P value of 0.052 did not reach conventional statistical significance. The principal scientific conclusion is therefore not that a diagnostic classifier has been established, but that the nine-gene panel contains research-stage transcriptomic signal warranting further independent investigation.

Biological Interpretation of the Nine-Gene Signature:

The signature is biologically coherent with immune-cell, NK-cell, and cytotoxic-lymphocyte-associated transcriptional states, with ITGA2B providing a distinct platelet/megakaryocyte-associated component. GNLY showed relatively strong discrimination during model development and the strongest individual external AUC, whereas GZMK retained directional concordance but weak external discrimination. These observations support prioritization for mechanistic and validation studies but do not demonstrate that any individual gene is causally involved in AS. The recent AS literature similarly shows that transcriptomic, single-cell, proteomic, and machine-learning studies can identify plausible candidates, but independent cohorts and orthogonal validation remain essential [14][13][15].

Interpretation of Model-Development Performance:

The archived AUC of 0.689 indicates moderate discrimination rather than strong separation. Importantly, GSE18781 contributed to the signature-selection/model-development process, so the cross-validation estimate cannot be treated as a completely selection-independent generalization estimate. A formally stronger predictive design would nest feature selection inside resampling or define an entirely independent training cohort before validation. The permutation result of P=0.052 reinforces this conservative interpretation. The later raw-data execution produced a different performance estimate, and

retaining both artifacts rather than selecting the more favorable result improves analytical transparency.

Independent External Evaluation and Replication Heterogeneity:

GSE221786 provided strong sample-identity and matrix-integrity assurance: all 28 biological samples were resolved, all nine genes were present, and no expression cells were missing. However, because the original serialized classifier binary was not recovered, the external analysis should not be described as direct execution of that binary classifier. Instead, the principal external contribution is independent evaluation of the frozen nine-gene molecular evidence. No gene satisfied the complete predefined replication criterion. GZMK, IL2RB, and ITGA2B showed directional concordance, while GNLY showed the strongest external AUC but discordant direction and nonsignificant inference. Differences in whole-blood versus stimulated/IL-17-enriched cellular context, cellular composition, disease characteristics, and technical platforms may contribute to this heterogeneity, although these factors were not experimentally established as the causes of the observed discordance.

Genetic, HLA, and Functional Evidence:

The genetic layer provides a complementary but distinct perspective from transcriptomic state. KLRB1 had the strongest positional IGAS support (Fisher P=0.0223; BH-FDR=0.0670), which is suggestive but not FDR-significant. IL2RB and GNLY showed weaker positional evidence, and several genes had no mapped IGAS variants under the defined gene-body rule. Avoiding nearest-gene inference is a methodological strength because it prevents unsupported biological attribution. Similarly, the 34 significant functional-enrichment terms support set-level immune/cytotoxic biological coherence but do not establish pathway activation, gene-specific causality, or clinical utility.

Practical Implications and Research-Stage Biomarker Status:

The current evidence supports use of the nine-gene panel as a research-stage candidate signature for prioritizing biological investigation and designing future validation studies. GNLY, GZMK, and KLRB1 warrant particular attention for orthogonal transcriptomic, cellular, and genetic follow-up, but none should be described as clinically validated. Practical translation would require larger multicenter cohorts, standardized phenotype definitions, assessment of treatment and disease activity, comparison with established clinical variables, targeted molecular assays, calibration and clinical-utility analysis, and prospective evaluation. The signature should therefore be considered complementary to—not a replacement for—current clinical assessment.

Strengths and Evidentiary Boundaries:

Major strengths include explicit separation of discovery and external evidence, freezing of the nine-gene feature set before GSE221786 evaluation, independent raw-matrix/sample-identity auditing, retention of negative and discordant findings, conservative genetic annotation, and transparent distinction between archived model evidence and executable reconstruction. The principal evidentiary boundaries are equally important: the external cohort was small, the original classifier binary was unavailable, the historical 27-sample replication analysis differs from the final 28-sample cohort, GSE181364 involved cohort-specific refitting, GSE299639 provided GNLY-only evidence, and no prospective clinical or wet-laboratory validation was performed.

Limitations:

Several limitations should be considered. First, the external cohorts were small, particularly the 8-control GSE221786 subgroup, limiting precision and power. Second, discovery/model development and signature selection involved GSE18781, which can introduce selection-related optimism into internal performance estimates. Third, the original serialized Stage-37 classifier artifact was not recovered, so its archived performance metrics cannot be represented as a newly deserialized binary execution. Fourth, the archived Stage-29

inferential analysis contained 19 AS and 8 controls, whereas the final identity-locked GSE221786 matrix contains 20 AS and 8 controls; AS_M_10 is the AS sample absent from the historical subset, but the reason for its exclusion is undocumented. Fifth, GSE181364 involved cohort-specific model fitting and is therefore not genuine frozen-model external validation. Sixth, GSE299639 provided only GNLY-level evidence and cannot validate the complete nine-gene signature. Seventh, the IGAS/FinnGen resources did not support comprehensive direct gene-level assignment, and the available IGAS positional analysis was not LD-corrected. Eighth, functional enrichment provides set-level biological context rather than gene-specific pathway causality. Finally, no prospective clinical validation, clinical-utility analysis, or wet-laboratory functional validation was performed.

Future Work:

Future studies should prioritize larger, prospectively collected, multicenter AS cohorts with sufficient numbers of patients and controls and harmonized disease phenotyping, treatment exposure, disease activity, ancestry, and longitudinal follow-up.

Biologically matched sample types and standardized experimental conditions should be used whenever possible. Cell-composition-aware analyses, single-cell or sorted-cell experiments, and independent RNA-seq cohorts could determine whether the signature reflects altered transcription within immune subsets or changes in circulating cell abundance. Cross-platform validation would further test transferability.

GNLY, GZMK, KLRB1, and the remaining signature genes should be evaluated using targeted qPCR/digital PCR and, where justified, functional cellular experiments. Genetic follow-up should use reproducible variant-to-gene mapping, LD-aware gene-based testing, colocalization, fine-mapping, and causal-inference methods where appropriate.

The predictive model should be rebuilt in a formally independent architecture with feature selection nested within resampling or restricted to a designated training cohort, followed by untouched validation cohorts. A complete frozen executable classifier, exact preprocessing parameters, feature order, scoring procedure, environment specification, and analysis code should be deposited in a permanent repository before claims of fully reproducible external prediction are made.

Conclusion:

This study developed and systematically evaluated a nine-gene transcriptomic signature for ankylosing spondylitis within an evidence-controlled integrative multi-omics framework. The archived principal model-development analysis showed moderate discrimination (AUC 0.689 ± 0.028; 95% CI 0.633–0.741), while permutation testing yielded P=0.052 and therefore did not provide conventional statistical confirmation. Independent GSE221786 analysis demonstrated strong sample-identity and matrix-integrity assurance but heterogeneous gene-level replication, with no gene satisfying the complete predefined replication criterion. GNLY remained a priority transcriptomic candidate, GZMK retained directional evidence, and KLRB1 provided the strongest positional genetic support, although its FDR was 0.0670. Collectively, these findings support the nine-gene panel as a research-stage candidate biomarker signature rather than a clinically validated diagnostic test. Larger prospective cohorts, selection-independent model development, complete executable model preservation, orthogonal molecular validation, and rigorous clinical evaluation are required before clinical translation.

References:

[1] D. van der Heijde *et al.*, "Goodbye to the term 'ankylosing spondylitis', hello 'axial spondyloarthritis': time to embrace the ASAS-defined nomenclature," *Ann. Rheum. Dis.*, vol. 83, no. 5, pp. 547–549, Dec. 2023, doi: 10.1136/ard-2023-225185.

[2] V. Navarro-Compán *et al.*, "ASAS consensus definition of early axial spondyloarthritis," *Ann. Rheum. Dis.*, vol. 83, no. 9, pp. 1093–1099, Aug. 2024, doi:

10.1136/ard-2023-224232.

[3] J. D. Taurog, A. Chhabra, and R. A. Colbert, "Ankylosing Spondylitis and Axial Spondyloarthritis," *N. Engl. J. Med.*, vol. 374, no. 26, pp. 2563–2574, Jun. 2016, doi: 10.1056/NEJMRA1406182;SUBPAGE:STRING:METEREDCONTENT.

[4] J. Sieper and D. Poddubnyy, "Axial spondyloarthritis," *Lancet*, vol. 390, no. 10089, pp. 73–84, Jul. 2017, doi: 10.1016/S0140-6736(16)31591-4.

[5] M. A. Brown, T. Kenna, and B. P. Wordsworth, "Genetics of ankylosing spondylitis--insights into pathogenesis," *Nat. Rev. Rheumatol.*, vol. 12, no. 2, pp. 81–91, Feb. 2016, doi: 10.1038/NRRHEUM.2015.133.

[6] K. Du, C. Y. Zhang, A. Li, Q. H. Zuo, R. Guo, and S. M. Li, "Unraveling new therapeutic targets in ankylosing spondylitis: Multi-omics Mendelian randomization on immune cells, metabolites, and inflammation proteins," *Medicine (Baltimore).*, vol. 104, no. 16, p. e42177, Apr. 2025, doi: 10.1097/MD.0000000000042177.

[7] V. Ranganathan, E. Gracey, M. A. Brown, R. D. Inman, and N. Haroon, "Pathogenesis of ankylosing spondylitis - recent advances and future directions," *Nat. Rev. Rheumatol.*, vol. 13, no. 6, pp. 359–367, Jun. 2017, doi: 10.1038/NRRHEUM.2017.56.

[8] F. Navid, L. Chen, P. Bowness, and R. A. Colbert, "HLA-B27 and spondyloarthritis: at the crossroads of innate and adaptive immunity," *Nat. Rev. Rheumatol.*, vol. 21, no. 2, pp. 77–87, Feb. 2025, doi: 10.1038/S41584-024-01189-3;SUBJMETA.

[9] D. M. Evans *et al.*, "Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility," *Nat. Genet.*, vol. 43, no. 8, pp. 761–767, Aug. 2011, doi: 10.1038/NG.873.

[10] P. R. Burton *et al.*, "Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants," *Nat. Genet.*, vol. 39, no. 11, pp. 1329–1337, Nov. 2007, doi: 10.1038/NG.2007.17.

[11] E. Gracey *et al.*, "Sexual Dimorphism in the Th17 Signature of Ankylosing Spondylitis," *Arthritis Rheumatol. (Hoboken, N.J.)*, vol. 68, no. 3, pp. 679–689, Mar. 2016, doi: 10.1002/ART.39464.

[12] L. Dai, L. Xia, G. Su, Y. Gao, Q. Jiang, and P. Yang, "Identifying prioritization of therapeutic targets for ankylosing spondylitis: a multi-omics Mendelian randomization study," *J. Transl. Med.*, vol. 22, no. 1, Dec. 2024, doi: 10.1186/S12967-024-05925-X.

[13] Y. Chen, Y. Xu, S. Cao, Q. Lv, Y. Ye, and J. Gu, "Oxidative Phosphorylation Pathway in Ankylosing Spondylitis: Multi-Omics Analysis and Machine Learning," *Int. J. Rheum. Dis.*, vol. 28, no. 5, May 2025, doi: 10.1111/1756-185X.70175.

[14] X. Xu *et al.*, "Characterization of immune features and discovery of potential biomarkers for ankylosing spondylitis using deep plasma proteomics," *J. Adv. Res.*, vol. 81, pp. 223–234, Mar. 2026, doi: 10.1016/J.JARE.2025.05.052.

[15] L. Cuesta-López *et al.*, "Transcriptomic and proteomic analysis stratifies patients with axial spondyloarthritis based on disease activity, structural damage and radiographic progression," *RMD open*, vol. 12, no. 1, Jan. 2026, doi: 10.1136/RMDOPEN-2025-006136.

[16] M. Brown *et al.*, "Multi-omic insights from a multi-ancestry genome-wide meta-analysis of ankylosing spondylitis reveal novel pathways of disease susceptibility," *Res. Sq.*, Jul. 2025, doi: 10.21203/RS.3.RS-6917334/V1.

[17] GEO Accession viewer, "Sexual Dimorphism in the Th17 Signature of Ankylosing," 2011. Accessed: Oct. 17, 2026. [Online]. Available: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73754.

[18] Y. Zheng *et al.*, "Identification of immune related cells and crucial genes in the

peripheral blood of ankylosing spondylitis by integrated bioinformatics analysis," *PeerJ*, vol. 9, Sep. 2021, doi: 10.7717/PEERJ.12125.

[19] Y. Han *et al.*, "Identification of diagnostic mRNA biomarkers in whole blood for ankylosing spondylitis using WGCNA and machine learning feature selection," *Front. Immunol.*, vol. 13, Sep. 2022, doi: 10.3389/FIMMU.2022.956027.

[20] GEO Accession viewer, "Gene Expression in Inflammatory Diseases," 2011. Accessed: Oct. 17, 2026. [Online]. Available: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18781

[21] J. Wen, L. Wan, and X. Dong, "Novel peripheral blood diagnostic biomarkers screened by machine learning algorithms in ankylosing spondylitis," *Front. Genet.*, vol. 13, p. 1032010, Nov. 2022, doi: 10.3389/FGENE.2022.1032010/TEXT.

[22] H. Liu and Y. Peng, "Identification of a Novel Gene Expression Signature Associated with Amino Acid Metabolism (AAM) in Ankylosing Spondylitis (AS)," *Int. J. Gen. Med.*, vol. 17, pp. 597–609, Feb. 2024, doi: 10.2147/IJGM.S435650.

[23] "Differentially Expressed Genes in Males and Females with Ankylosing Spondylitis (AS): Transcriptomics of IL-17-producing Cells in Patients with AS - ACR Meeting Abstracts." Accessed: Oct. 10, 2026. [Online]. Available: https://acrabstracts.org/abstract/differentially-expressed-genes-in-males-and-females-with-ankylosing-spondylitis-as-transcriptomics-of-il-17-producing-cells-in-patients-with-as/

[24] D. Li *et al.*, "Identification of potential biomarkers for ankylosing spondylitis based on bioinformatics analysis," *BMC Musculoskelet. Disord.*, vol. 24, no. 1, p. 413, Dec. 2023, doi: 10.1186/S12891-023-06550-3.

[25] X. Cao *et al.*, "Integrated long-read transcriptomic profiling of peripheral blood from ankylosing spondylitis patients identifies regulatory shifts and core genes associated with programmed cell death," *Front. Immunol.*, vol. 17, 2026, doi: 10.3389/FIMMU.2026.1640271.

[26] G. S. Collins *et al.*, "TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods," *BMJ*, vol. 385, Apr. 2024, doi: 10.1136/BMJ-2023-078378.

[27] K. G. M. Moons *et al.*, "PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods," *BMJ*, vol. 388, 2025, doi: 10.1136/BMJ-2024-082505.

[28] Y. Gao, X. Li, F. Luo, R. Chen, and X. Zhang, "Integrative mass spectrometry-driven multi-omics and single cell technologies in ankylosing spondylitis: insights into pathogenesis, biomarker discovery, and precision medicine," *J. Transl. Autoimmun.*, vol. 11, Dec. 2025, doi: 10.1016/j.jtauto.2025.100319.

[29] X. Dan *et al.*, "Multi-Omics Integration Identifies a Six-Gene Diagnostic Signature for Ankylosing Spondylitis via Metabolic–Immune Crosstalk," *Int. J. Mol. Sci.*, vol. 27, no. 9, p. 3860, May 2026, doi: 10.3390/IJMS27093860/S1.

[30] M. E. Ritchie *et al.*, "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Res.*, vol. 43, no. 7, pp. e47–e47, Apr. 2015, doi: 10.1093/NAR/GKV007.

Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.