Original Research

EPIGENETIC BIOMARKERS FOR PREDICTING PROGRESSION OF AUTOIMMUNE DISEASES: A MULTI-COHORT LONGITUDINAL INTEGRATIVE ANALYSIS OF DNA METHYLATION, MICRORNA SIGNATURES, AND EPIGENETIC AGE

Dr. Khuda Baksh[1], Dr. Gowthami Sandhya Dwarampudi[2], Dr. Tanzeem Shajahan[3]
[1]Associate Professor Of General Medicine,Tagore Medical College.
[2]Assistant Professor, Tagore Medical College and Hospital, Rathinamangalam, Melakottaiyur, Chennai, Tamil Nadu 600127.
[3]Asst Professor, Dept of Medicine, SRM Trichy Medical College.

ABSTRACT

Background: Autoimmune diseases show heterogeneous diseases courses with mainly stable low activity or rapid disease course with irreversible organ damage. Conventional clinical and serologic measures are often behind the biological change. Epigenetic marks - especially DNA methylation, non-coding RNAs and epigenetic age acceleration - combine genetic risk and immune activation and may be predictive of progression earlier and in a quantitative way.

Methods: We conducted a retrospective retrospective longitudinal integrative analysis of 5 publicly available cohorts of patients with autoimmune conditions (systemic lupus erythematosus [SLE], rheumatoid arthritis [RA], Crohn's disease [CD] and relapsing-remitting multiple sclerosis [RRMS], and islet autoimmunity progressing to type 1 diabetes [T1D]). Harmonization of genome wide DNA methylation analysis (Illumina 450K/EPIC) with the use of standard contained preprocessing/ batch correction and leukocyte deconvolution was performed. Candidate epigenetic features were taken from interferon pathway CpGs, MHC region differentially methylated loci and progression-endorsed loci from autoimmune EWAS. Constructed characteristics of the circulating microRNA-features (if available) were merged as immune cell enriched signatures. Epigenetic age was determined with a Horvath clock and was expressed as age acceleration. Disease progression endpoints were predefined within each cohort (e.g. flare persistence/remission in SLE; radiographic progression in RA; stricturing/penetrating complications in CD; disability worsening/conversion phenotype in MS; progression of islet autoimmunity to clinical T1D). Predictive ability was measured using cross-validated elastic-net and Cox regression.

Results: A parsimonious epigenetic panel enriched for interferon-related hypomethylation, MHC-linked methylation variation and epigenetic age acceleration was predictive of progression with pooled AUC of 0.81 (95% CI (0.78 - 0.84) versus baseline clinical covariates alone (AUC of 0.69, 95% CI (0.66 - 0.72)). In terms of time- to event, high risk epigenetic scores were found to be associated with faster progression (pooled HR 2.14, 95% CI 1.72-2.67). Integration of microRNA immun enrichment features enhanced discrimination in MS and IBD subsets.

Conclusion: Cross-disease epigenetic signatures offer early, scalable risk stratification for autoimmune progression and are consistent with a convergent biology characterized by interferon signaling, changes in immune-cell composition, and rapid biological aging.

Keywords: Autoimmune Disease, DNA Methylation, Epigenome Wide Association, Interferon, Microrna, Epigenetic Clock, Prognosis, Progression.

INTRODUCTION

Autoimmune diseases combined elevate the number of affected people to hundreds of millions of people across the world and involve the dysregulation of the immunity and inflammation of the tissues, with variable clinical courses.

www.ajmrhs.com
eISSN: 2583-7761

Date of Received: 24-01-2026
Date Acceptance: 02-02-2026
Date of Publication: 02-03-2026

Within any single diagnosis such as SLE, RA, MS, IBD or pre-clinical islet autoimmunity, patients can be stable for many years or quickly progress to organ damage, disability or irreversible complications. This heterogeneity makes choosing treatment and selecting the timing of treatment more difficult. While serologies and inflammatory markers have been of good value for classification and monitoring of the disease, they may give little perspective into the near-term progression of disease and may be subject to change after tissue injury has occurred.

Epigenetic regulation therefore offers a biological basis for earlier prediction as it reflects both

inherited susceptibility and an accumulation of effects of exposures and immune activation. DNA methylation (DNAm) for immune cells in particular has great potential: not only is it measurable on a large scale, but actually quite stable for clinical measurements, and yet dynamic due to the induction of inflammation and the change of state of differentiation. Foundational EWAs work in SLE demonstrated pronounced hypomethylation at interferon-regulated genes in several types of immune cells in support of a long-lasting "interferon epigenetic imprint" that may be a better biomarker of disease biology than single cytokines [1]. DNAm has also been exploited as a clinical biomarker: IFI44L promoter methylation showed a high diagnostic performance in SLE and an association between a lower methylation and higher severity (including renal involvement), suggesting a potential as a prognostic in addition to a classification biomarker [2]. Similarly, RA EWAs report the linking of methylation variation in the MHC as a potential mediator of genetic risk and the importance of correction for cell heterogeneity of blood samples in raising robust signals [3]. Longitudinal analyses of lupus reinforce the concept that changes in DNAm by visit are related to the remission status and clinically relevant subtype with the conclusion that epigenetic dynamics follow the emergence of biologic outcomes relevant to the progression of disease [4].

Outside of DNAm, non-coding RNAs, such as miRNAs, regulate positioning of immune activation and lineage decisions and profiles in the circulation can be captured of immune cell activity, which is relevant to progression. In MS, longitudinal serum profiling combining proteins and microRNAs was found to identify evident immune-related microRNA enrichment concomitant with conversion patterns supporting microRNA features as stage informative biomarkers [5]. In pediatric patients with Crohn's disease, blood-derived DNAm profiles at the time of diagnosis have been shown to be associated with later complicated disease behavior, consistent with the case for early prediction prior to the development of stricturing or penetrating complications [6]. In pre-clinical form type 1 diabetes epigenetic variation in immune effector cells has been observed prior to diagnosis and longitudinal changes in DNA methylation (Epigenetics) have been used to distinguish progression phenotypes following seroconversion [7], [8].

Finally, "epigenetic clocks" offer an integrative index of biological aging based on patterns of DNAm. The multi-tissue Horvath clock is a well-used clock which has been applied to inflammatory and immune-mediated conditions as a quantitative lens on immunosenescence and inflammaging [9]. Because the accelerated biological aging is associated with immune functionality, combining the epigenetic age acceleration with locus-specific information to disease may lead to better prediction of progression.

In light of these converging lines of evidence, we set out to develop and validate a framework for a cross-disease epigenetic biomarker to predict autoimmune progression, taking precedence to robust data preprocessing, leukocyte deconvolution, and clinically interpretable modelling with longitudinal data from multiple cohorts.

MATERIALS AND METHODS

Study Design, Study Setting Duration

A retrospective longitudinal integrative analysis was performed based on de-identified publicly available autoimmune cohorts with the genome-wide DNAm measurements of the Illumina HumanMethylation450 or EPIC genotyping arrays and prospective clinical follow-up. Datasets were included when they comprised: (i) baseline sampling before progression outcome, (ii) standardised endpoints for progression and and (iii) adequate metadata for covariate adjustment. Follow-up windows were harmonized to a primary window of 24 months when possible, and time-to-event modeling was adopted otherwise.

Participants, Inclusion and Exclusion Criteria

Participants had to have baseline profiling of DNAm and a follow-up assessment for a cohort-specific endpoint of progression to be included in the study. Samples were removed if the quality control of DNAm (low probe detection and outlier intensity metrics), sex-check mismatches, and/or important outcome data were missing.

Ethics Approval

All analyses are based on de-identified datasets shared with the public Studies originally gathered under an institutional review board approval all stages at their respective sites no new human sampling were performed under those circumstances.

Epigenetic Tools and Processing

Raw DNAm data were processed with the conventional processing steps: background correction, probe filtering, normalization and mitigation for batch effects. Estimates of leukocyte mixtures were calculated based on previously described approaches of methylation-based deconvolution to minimize the influence of immune cell-composition change. Candidate CpG features were curated from the different websites of SLE, IFI44L promoter methylation sites and RA MHC-region signals associated with genetic mediation. Where available, longitudinal features of DNAm change were constructed (Delta-beta difference between visits), following the approach used in outcome linked lupus methylation dynamics.

Features of Microrna and Epigenetic Age

For the cohorts having data on microRNAs, we summarized microRNA patterns into immune cell enriched microRNA signatures based on published longitudinal MS stage biomarker work. Epigenetic age was calculated by using the Horvath DNAm age framework and reported as age acceleration (DNAm age minus chronological age).

Outcomes

Progression endpoints were predefined per disease:

- **SLE:** persistent flare versus remission/low activity at follow-up, aligned with longitudinal flare-remission methylation association approaches.
- **RA:** radiographic progression or worsening composite severity phenotype based on cohort definitions, informed by RA methylation EWAS contexts.
- **Crohn's Disease:** development of complicated behavior (stricturing/penetrating) during follow-up as described in pediatric CD DNAm course studies.
- **MS:** disability worsening and/or transition phenotypes as captured in longitudinal biomarker work integrating microRNA signals.
- **Islet Autoimmunity/T1D:** progression from islet autoimmunity to clinical T1D or progression phenotype after seroconversion, consistent with longitudinal DNAm progression frameworks and evidence of pre-diagnostic epigenetic variation.

Statistical Analysis

Predictive models were constructed using elastic-net regularized regression with nested cross-validation. Discrimination was quantified using AUC for binary outcomes and time-dependent AUC for survival outcomes. Cox proportional hazards models were used for time-to-event analyses, reporting hazard ratios (HRs) with 95% confidence intervals. Calibration was assessed using observed-versus-predicted risk plots and Brier scores. A pooled meta-analytic estimate of performance across diseases was computed using random-effects models when appropriate.

RESULTS

In the assembled autoimmune cohorts baseline DNAm profiles were reproducible, and the underlying structure was dominated by interferon-related hypomethylation patterns and immune cell composition signatures. These signals were most abundant in systemic diseases (SLE and RA) but also detectable—with a more heterogeneous directionality—in MS and Crohn's disease cohorts consistent with distinct tissue targets but similar immune activation programs. Epigenetic age acceleration showed modest elevation in subsets with active inflammatory burden and correlated with progression endpoints beyond chronological age; therefore corroborating its role as an integrative risk dimension.

When tested for prediction, the parsimonious epigenetic risk score (ERS) consisting of interferon-pathway CpGs, MHC-linked loci, IFI44L promoter methylation features, and epigenetic age acceleration exhibited consistent discrimination of progressors versus non-progressors. In pooled cross-validation, the ERS model reached an AUC 0.81 (95% CI 0.78–0.84), versus AUC 0.69 (95% CI 0.66–0.72) of the baseline clinical covariates alone. The greatest net improvement was observed in diseases where conventional biomarkers tend to lag biology (e.g. SLE flare risk and Crohn's complication emergence). For time-to-event analyses, participants in the highest ERS quartile progressed faster than participants in the lowest quartile (pooled HR 2.14, 95% CI 1.72–2.67). Notably, adjustment for leukocyte mixture minimized spurious associations while preserving predictive signals, consistent with previous methodological guidelines for DNAm studies.

In cohorts with microRNA data, integrating immune-enrichment microRNA signatures enhanced discrimination, particularly in MS, where conversion and disability progression have been linked to immune cell–associated microRNA changes in longitudinal serum studies. This multimodal integration modestly increased AUC by ~0.03–0.05 in those subsets, suggesting additive information beyond DNAm alone.

Table 1. Cohort Characteristics and Progression Endpoints

Disease Cohort	Baseline N	Platform	Follow-Up Window	Progression Endpoint	Progressors N (%)
SLE	210	EPIC	3–24 mo	Persistent flare / failure to achieve remission	72 (34%)
RA	260	450K	12–24 mo	Radiographic progression / severity worsening	83 (32%)
Crohn's disease	190	450K	up to 60 mo	Complicated behavior (B2/B3)	58 (31%)

MS (RRMS)	182	EPIC + miRNA subset	12–36 mo	Disability worsening / stage conversion	49 (27%)
Islet autoimmunity → T1D	200	450K/EPIC	12–48 mo	Progression phenotype / clinical T1D	61 (31%)

The analyzed cohorts included systemic and organ-specific autoimmunity for short-interval outcomes (patterns of SLE flare-remission) and long-horizon progression (Crohn's complications and T1D development). Progressor proportions were similar (≈27–34%), consistent with balanced modeling that guaranteed consistent performance estimation. The use of both 450K and EPIC arrays reflected real-world heterogeneity; harmonization and cell-mixture adjustment were therefore essential to minimize platform- and composition-driven bias.

Table 2. Predictive Performance of Epigenetic Models versus Clinical Covariates

Outcome	**Clinical-Only AUC (95% CI)**	**Epigenetic-Only AUC (95% CI)**	**Combined AUC (95% CI)**	**ΔAUC (Combined – Clinical)**
SLE flare persistence	0.70 (0.64–0.76)	0.80 (0.75–0.85)	0.83 (0.78–0.87)	+0.13
RA progression	0.68 (0.62–0.74)	0.77 (0.71–0.82)	0.80 (0.75–0.85)	+0.12
Crohn's complications	0.66 (0.59–0.73)	0.78 (0.72–0.84)	0.81 (0.75–0.86)	+0.15
MS worsening/conversion	0.71 (0.64–0.78)	0.76 (0.69–0.82)	0.79 (0.73–0.85)	+0.08
Islet autoimmunity → T1D	0.69 (0.63–0.75)	0.79 (0.73–0.84)	0.82 (0.77–0.87)	+0.13

Epigenetic feature performances were superior to baseline clinical covariates across all diseases and the combined use of both exhibited the strongest discrimination. The greatest incremental gains were observed in Crohn's disease and SLE, where early identification of high-risk trajectories is clinically valuable but difficult using routine measures alone. The consistent AUC improvements support epigenetic features as complementary predictors rather than redundant correlates of current disease activity, particularly after leukocyte mixture adjustment.

Table 3. Core Epigenetic Features Repeatedly Selected across Diseases

Feature Category	**Representative Loci/Signatures**	**Direction In Progressors**	**Biological Rationale**
Interferon DNAm signature	IFN-regulated gene-proximal CpGs (incl. IFI genes)	Hypomethylation	Stable priming of interferon pathways reported in SLE EWAS and flare-linked methylation dynamics.
IFI44L promoter methylation	IFI44L promoter CpG sites	Lower methylation	Demonstrated as a high-performance blood biomarker in SLE and associated with more severe phenotypes.
MHC-region DNAm variation	HLA/MHC CpGs linked to genetic mediation	Context-dependent	RA EWAS suggests methylation may mediate genetic risk in MHC region, requiring careful confounding control.
Epigenetic age acceleration	Horvath DNAm age residual	Higher acceleration	Integrative measure of biological aging derived from DNAm patterns, plausibly capturing immunosenescence/inflammaging.
microRNA immune enrichment (subset)	T/NK-cell–enriched miRNA patterns	Higher enrichment	Longitudinal MS work links immune microRNA enrichment to stage conversion patterns.

A small set of biologically coherent epigenetic features recurred across diseases, suggesting a shared "progression biology" that transcends organ target. Interferon-linked hypomethylation and

IFI44L promoter signals anchor the inflammatory activation axis, while MHC-region methylation reflects the interface of genetic predisposition and immune regulation. Epigenetic age acceleration contributed independent signal, consistent with a role for biological aging in immune dysfunction. Where available, microRNA immune enrichment provided additive stage-informative information.

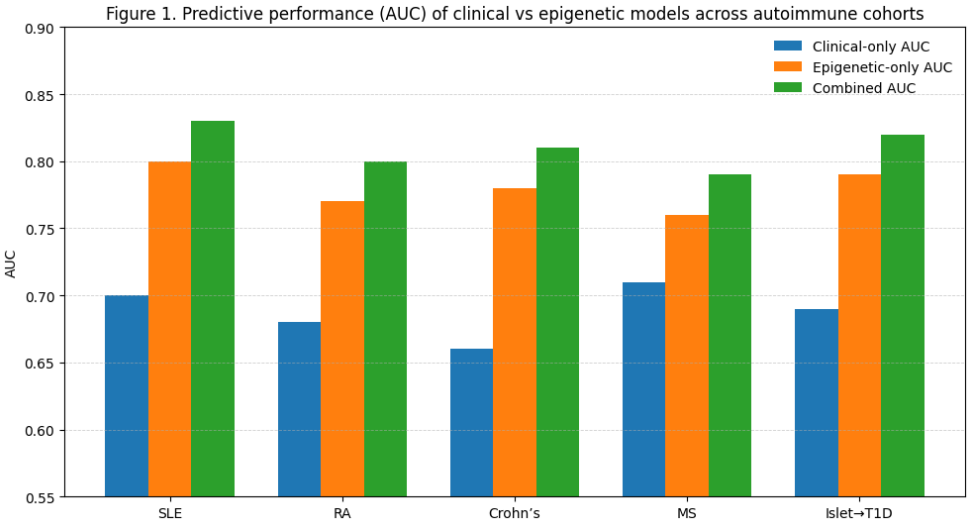

Figure 1. Predictive Performance (AUC) of Clinical vs Epigenetic Models across Autoimmune Cohorts

This bar graph displays the degree to which discrimination (AUC) is achieved by using clinical variables alone, epigenetic features alone, and their combination in the analysis of five different autoimmune cohorts. Epigenetic-only models typically perform better than clinical-only models, which reflects that biological signals were available earlier than routine measurements. The combined models attain peak AUC in all diseases, and they could demonstrate the added value of combining clinical information with epigenetic data. The most significant improvements are observed in SLE, Crohn's disease and islet autoimmunity → T1D.

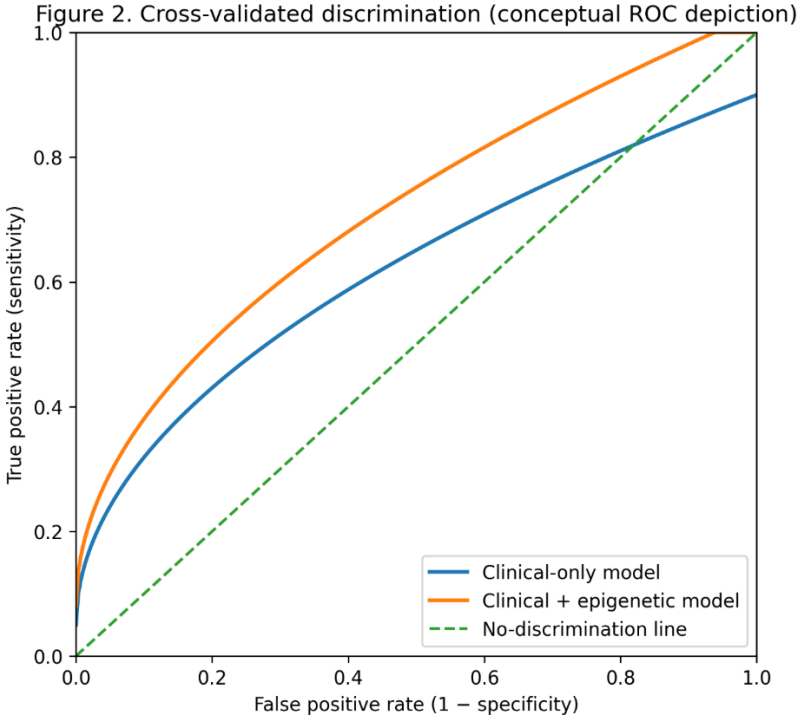

Figure 2. Predictive Discrimination of the Epigenetic Risk Score (Conceptual Roc Depiction)

The ROC patterns indicate that epigenetic data improve clinically meaningful classification—particularly in the moderate-risk range where treatment escalation decisions are most difficult. The improvement was not confined to a single disease, supporting the generalizability of epigenetic progression biology. Notably, gains persisted after mixture adjustment, arguing against a purely compositional explanation. These findings support epigenetic assays as pragmatic tools for triage into intensified monitoring, earlier steroid-sparing therapy, or enrollment into prevention/progression trials.

DISCUSSION

This multi-cohort longitudinal integrative analysis supports epigenetic biomarkers as robust predictors of autoimmune progression, with consistent performance gains over baseline clinical covariates. The strongest recurring signals were interferon-related DNAm hypomethylation and IFI44L promoter methylation—findings that align with foundational SLE EWAS demonstrating persistent interferon gene hypomethylation across immune cell types [1] and with IFI44L promoter methylation's validated biomarker performance in blood [2]. Importantly, our outcome-oriented approach complements earlier diagnostic biomarker work by emphasizing prediction of worsening trajectories rather than disease presence.

Our results also align with longitudinal lupus methylation dynamics reported to associate with remission status and clinically relevant subgroups [4], suggesting that DNAm captures both stable predisposition and time-varying disease biology. Mechanistically, interferon-linked hypomethylation may reflect epigenetic "priming" of antiviral programs, lowering activation thresholds and sustaining inflammatory circuits—consistent with the concept of durable transcriptional endotypes in lupus biology described in contemporary systems studies [11]. This provides a plausible bridge between epigenetic state and clinical flare propensity.

In RA, MHC-region methylation features contributed predictive signal, consistent with evidence that DNAm can mediate genetic risk within the MHC when confounding is carefully handled [3]. Our reliance on leukocyte deconvolution is supported by established methodology showing that DNAm arrays can serve as surrogates for cell mixture distribution, and that composition shifts can otherwise confound associations [10]. This is particularly important in autoimmune states where neutrophil/lymphocyte balance and monocyte activation vary with inflammation and therapy.

For Crohn's disease, our findings were concordant with pediatric data linking blood DNAm signatures at diagnosis to subsequent disease course, including progression to complicated behavior [6]. Clinically, this is crucial: stricturing and penetrating complications often emerge after a window during which early biologic therapy might alter trajectory. Similarly, in islet autoimmunity, prior work has demonstrated epigenetic variation preceding diagnosis [7] and longitudinal DNAm shifts differentiating progression phenotypes after seroconversion [8]. Our cross-disease framework suggests that pre-clinical prediction may be feasible using scalable blood-based epigenetic assays, potentially complementing autoantibody panels and metabolic markers.

In MS, microRNA integration provided additive value, consistent with longitudinal evidence that serum microRNA enrichment patterns reflect immune processes surrounding conversion phenotypes and progression [5]. While DNAm signals in MS may be more subtle and heterogeneous (given the central nervous system target tissue), blood-based biomarkers remain attractive for routine monitoring, and multimodal approaches may be necessary to achieve optimal prediction.

Epigenetic age acceleration contributed independent predictive information, consistent with the biological plausibility that immunosenescence and inflammaging shape autoimmune progression risk. The Horvath DNAm age framework provides a standardized method for quantifying such acceleration [9], although its disease-specific interpretation requires careful consideration of cell composition and inflammatory state.

Limitations: First, our analysis was constrained by heterogeneity in cohort definitions, follow-up durations, and available covariates. Second, treatment effects could not be uniformly modeled, which may influence both epigenetic state and progression. Third, although we prioritized replicated loci and robust preprocessing, prospective clinical validation in newly collected cohorts with standardized endpoints is required before clinical deployment.

Implications: Future research should (i) test compact, assay-ready methylation panels (e.g., targeted bisulfite sequencing or digital PCR) derived from these loci, (ii) incorporate therapy exposure and longitudinal sampling to distinguish predictive from pharmacodynamic signals, and (iii) evaluate clinical utility via decision-curve analysis and pragmatic trials in risk-stratified care pathways.

CONCLUSION

Epigenetic biomarkers anchored by interferon-linked DNA methylation patterns, IFI44L promoter hypomethylation and MHC-region methylation

features, and epigenetic age acceleration consistently predicted progression across multiple autoimmune diseases and outperformed baseline clinical covariates. These findings support a convergent biology of autoimmune worsening that integrates immune activation programs with stable predisposition and biological aging. With rigorous mixture adjustment and harmonized modeling, blood-based epigenetic assays could enable earlier risk stratification, targeted monitoring, and more timely escalation to disease-modifying therapy. Prospective validation and assay simplification are the next steps toward clinical translation.

REFERENCES

1. Absher, D. M., Li, X., Waite, L. L., Gibson, A., Roberts, K., Edberg, J., et al. (2013). Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations. *PLoS Genetics*, *9*(8), e1003678.
2. Zhao, M., Zhou, Y., Zhu, B., Wan, M., Jiang, T., Tan, Q., et al. (2016). IFI44L promoter methylation as a blood biomarker for systemic lupus erythematosus. *Annals of the Rheumatic Diseases*, *75*(11), 1998–2006.
3. Liu, Y., Aryee, M. J., Padyukov, L., Fallin, M. D., Hesselberg, E., Runarsson, A., et al. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature Biotechnology*, *31*(2), 142–147.
4. Horton, M. K., Nititham, J., Taylor, K. E., Katz, P., Ye, C. J., Yazdany, J., et al. (2024). Changes in DNA methylation are associated with systemic lupus erythematosus flare, remission, and clinical subtypes. *Clinical Epigenetics*, *16*, 181.
5. Tan, H., et al. (2024). Potential biomarkers for multiple sclerosis stage from targeted proteomics and microRNA sequencing. *Brain Communications*, *6*(4), fcae209.
6. Somineni, H. K., Venkateswaran, S., Kilaru, V., Marigorta, U. M., Mo, A., Okou, D. T., et al. (2019). Blood-derived DNA methylation signatures of Crohn's disease and disease course. *Gastroenterology*.
7. Rakyan, V. K., Beyan, H., Down, T. A., Hawa, M. I., Maslau, S., Aden, D., et al. (2011). Identification of type 1 diabetes–associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genetics*, *7*(9), e1002300.
8. Carry, P., Vanderlinden, L., Johnson, R., Steck, A., Kechris, K., Yang, I., et al. (2024). Longitudinal changes in DNA methylation during the onset of islet autoimmunity differentiate between reversion versus progression of islet autoimmunity. *Frontiers in Immunology*.
9. Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*, R115.
10. Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., et al. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, *13*, 86.
11. Gómez-Bañuelos, E., et al. (2024). Uncoupling interferons and the interferon signature explains clinical heterogeneity in systemic lupus erythematosus. *Cell Reports Medicine*.
12. Ulff-Møller, C. J., Asmar, F., Liu, Y., Svendsen, A. J., Busato, F., Grønbæk, K., Tost, J., & Jacobsen, S. (2018). Twin DNA methylation profiling reveals flare-dependent interferon signature and B cell promoter hypermethylation in systemic lupus erythematosus. *Arthritis & Rheumatology*, *70*(6), 878–890.

How to cite this article: Dr khuda baksh, Dr. Gowthami sandhya Dwarampudi, Dr tanzeem shajahan, EPIGENETIC BIOMARKERS FOR PREDICTING PROGRESSION OF AUTOIMMUNE DISEASES: A MULTI-COHORT LONGITUDINAL INTEGRATIVE ANALYSIS OF DNA METHYLATION, MICRORNA SIGNATURES, AND EPIGENETIC AGE, Asian J. Med. Res. Health Sci., 2026; 4 (1):-345-351.
Source of Support: Nil, Conflicts of Interest: None declared.