Prediction of the progression of undifferentiated arthritis to rheumatoid arthritis using DNA methylation profiling

Calle-Fabregat, C. de la; Niemantsverdriet, E.; Canete, J.D.; Li, T.L.; Helm-van Mil, A.H.M. van der; Rodriguez-Ubreva, J.; Ballestar, E.

Note: To cite this publication please use the final published version (if applicable).

AMERICAN COLLEGE
of RHEUMATOLOGY
Empowering Rheumatology Professionals

Prediction of the Progression of Undifferentiated Arthritis to Rheumatoid Arthritis Using DNA Methylation Profiling

Carlos de la Calle-Fabregat,[1] [iD] Ellis Niemantsverdriet,[2] [iD] Juan D. Cañete,[3] [iD] Tianlu Li,[1] Annette H. M. van der Helm-van Mil,[2] Javier Rodríguez-Ubreva,[1] and Esteban Ballestar[1] [iD]

Objective. The term "undifferentiated arthritis (UA)" is used to refer to all cases of arthritis that do not fit a specific diagnosis. A significant percentage of UA patients progress to rheumatoid arthritis (RA), others to a different definite rheumatic disease, and the rest undergo spontaneous remission. Therapeutic intervention in patients with UA can delay or halt disease progression and its long-term consequences. It is therefore of inherent interest to identify those UA patients with a high probability of progressing to RA who would benefit from early appropriate therapy. This study was undertaken to investigate whether alterations in the DNA methylation profiles of immune cells may provide information on the genetically or environmentally determined status of patients and potentially discriminate between disease subtypes.

Methods. We performed DNA methylation profiling of a UA patient cohort, in which progression to RA occurred for a significant proportion of the patients.

Results. We found differential DNA methylation in UA patients compared to healthy controls. Most importantly, our analysis identified a DNA methylation signature characteristic of those UA cases that differentiated to RA. We demonstrated that the methylome of peripheral mononuclear cells can be used to anticipate the evolution of UA to RA, and that this methylome is associated with a number of inflammatory pathways and transcription factors. Finally, we designed a machine learning strategy for DNA methylation-based classification that predicts the differentiation of UA toward RA.

Conclusion. Our findings indicate that DNA methylation profiling provides a good predictor of UA-to-RA progression to anticipate targeted treatments and improve clinical management.

INTRODUCTION

Undifferentiated arthritis (UA) is a form of early arthritis that involves joint inflammation that cannot be classified as any definite rheumatic disorder (1). Eventually, ~30% of patients with UA will develop rheumatoid arthritis (RA) or other differentiated forms of arthritis, whereas 45–55% of patients will achieve spontaneous remission (1). UA represents a unique window of opportunity to intervene during the course of the disease before more severe manifestations become established.

The ability to provide early indicators for the treatment of UA patients at high risk of developing RA is of utmost relevance for decision making regarding whether and when to start treatment with disease-modifying antirheumatic drugs (DMARDs), which usually hamper RA progression but are not recommended for UA patients who achieve eventual remission (2). To that end, prediction rules have been proven to be crucial tools to provide guidance to clinicians by estimating patient outcome probabilities. In fact, a prediction rule for UA patients, based strictly on patient clinical data, has been developed previously (3). This model accurately

estimates the risk of developing RA in >75% of patients with UA. However, rules based on clinical data, although easy to implement in the clinical setting, usually fail to identify a detailed biologic basis for individual phenotypic presentations of the disease, and usually do not succeed in all predictions. In this regard, approaches including "-omics" data may provide compelling alternatives or complementary tools for both improving prediction accuracy and allowing an in-depth characterization of the molecular alterations in patients (4).

Epigenetic alterations are associated with both genetic and environmentally driven determinants, which can in turn characterize pathogenic phenotypes. Specifically, DNA methylation and histone modifications, which are altered in multiple pathologic contexts, are proposed to be both a causal factor (5) and a consequence of disease (6), as well as an intermediary for genetic susceptibility (7). In all cases, the exhaustive study of these alterations using high-throughput technologies allows a detailed description and identification of novel molecular pathways that undergo alterations in a pathogenic context. DNA methylation is one of the most stable and easily comparable epigenetic modifications, and thus stands out as an ideal candidate for biomarker discovery (8).

In the present study, we characterized the DNA methylome of patients with UA in comparison to healthy controls. We also analyzed the data using different patient classification criteria, which proved to have a pivotal effect on DNA methylation profiles. In addition, we obtained the DNA methylation profiles of UA patients with known diverging future phenotypes. Moreover, we analyzed the profiles of patients with definite RA and compared them to those of patients with UA. The identification and interpretation of these data stand out as a valuable resource to delve into the molecular alterations in UA patients. Finally, we propose the use of DNA methylation data as a candidate biomarker with the ability to improve clinical prediction rules by integrating molecular insights and clinical knowledge for the prediction of patient outcomes.

PATIENTS AND METHODS

Patient cohort. A total of 72 samples from patients with UA and 8 samples from patients with RA were obtained from the Leiden Early Arthritis Clinic (EAC) cohort, which has been described previously (9). Thirteen healthy donor samples were also obtained. Patient samples were collected at the first visit (baseline). Patients had not received prior treatment with DMARDs (including glucocorticoids and antimalarial agents) and were diagnosed according to the American College of Rheumatology (ACR) 1987 criteria for RA (10). Within the group of UA patients, 39 had developed RA, while 33 remained classified as having UA, 1 year after baseline. The study was approved by the medical ethical testing committee (METC) Leiden Den Haag Delft, with cohort METC number P11.210, and the board of the Bellvitge Hospital Ethical Committee (PR275/17). The demographic and clinical characteristics of the patients and healthy donors are summarized in Supplementary Tables 1 and 2, available on the *Arthritis & Rheumatology* website at http://onlinelibrary.wiley.com/doi/10.1002/art.41885/abstract. Data on additional patient samples from the Leiden EAC cohort (with other disease subtypes) are also included in Supplementary Table 1.

DNA methylation profiling, bioinformatics analysis, and machine learning methods. Infinium HumanMethylation450K BeadChip arrays (Illumina) were used for DNA methylation analysis in the discovery cohort. By the time of the analysis of the validation cohort, 450K microarrays had been commercially discontinued; therefore, Infinium HumanMethylation EPIC
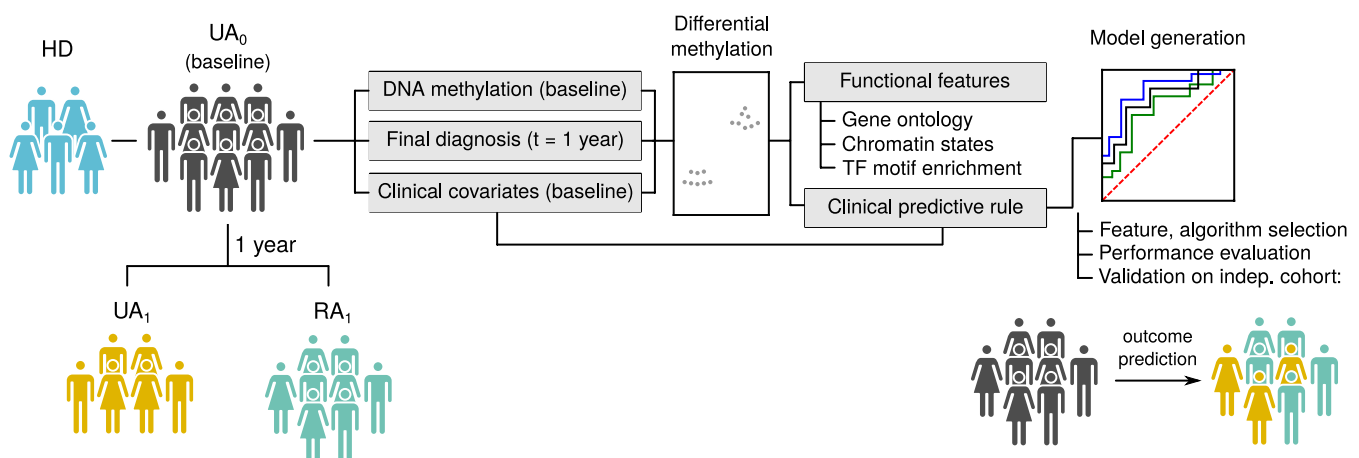
Figure 1. Flow chart of the study design, showing the conceptual and analytical workflow. Samples were obtained from healthy donors (HDs) and patients with undifferentiated arthritis at baseline (UA_0). One year after the initial visit, patients with UA at baseline were classified as continuing to have UA (UA_1) or as having developed rheumatoid arthritis (RA_1). DNA methylation profiles and clinical covariates were used to generate models to predict RA diagnosis. TF = transcription factor. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/art.41885/abstract.

BeadChip arrays (Illumina) were used instead. Methylation array data have been deposited in the National Center for Biotechnology Information GEO database and are accessible through GEO series accession number GSE175364. Details on the bead array analysis, downstream bioinformatics methods, machine learning methods, and representation are provided in full in the Supplementary Methods, available on the *Arthritis & Rheumatology* website at http://onlinelibrary.wiley.com/doi/10.1002/art.41885/abstract.

RESULTS

Altered DNA methylome in inflammation-related genes in peripheral blood mononuclear cells (PBMCs) from patients with UA. First, we determined the DNA methylation profile of PBMCs obtained from patients in the Leiden Early Arthritis Clinic (EAC) cohort (as described in Patients and Methods). Patient demographic and clinical characteristics, including age, sex, anti–cyclic citrullinated peptide (anti-CCP) antibody status, rheumatoid factor (RF) status, and Disease Activity Score (DAS) (11) were also obtained (Supplementary Table 1). Samples from a total of 64 UA patients at the first visit (baseline), referred to as UA_0, and 13 healthy donors were analyzed. An illustrated flow chart of the study design is depicted in Figure 1. After data correction for age and sample balancing by sex and cell type proportions (Supplementary Methods and Supplementary Figure 1A, available on the *Arthritis & Rheumatology* website at http://onlinelibrary.wiley.com/doi/10.1002/art.41885/abstract), a comparison of UA_0 and healthy donor DNA methylation profiles revealed 321 hypermethylated and 3,029 hypomethylated CpG sites (Figure 2A). Of note, the differentially methylated positions (DMPs) identified did not vary significantly with regard to the different microarray chips (slide) or position within the chip (array; see Supplementary Figure 1D and the Supplementary Methods).

Gene ontology (GO) analysis of the DMPs revealed enrichment for multiple categories related to inflammatory response, immune cell activation, vitamin metabolism, and cytokine and chemokine signaling pathways in both the hypermethylated and hypomethylated clusters (Figure 2B). Among those, interleukin-1 (IL-1), IL-6, IL-12, IL-10, tumor necrosis factor (TNF), macrophage colony-stimulating factor, and chemokine CXCL2 signaling pathways were shown to be enriched within the affected regions. The hypomethylated regions were specifically enriched in categories related to antimicrobial response and type I interferon (IFN) production. Detailed examples of the methylation of CpG sites proximal to genes contained in the GO categories, showing B values in the healthy donor and UA_0 groups, are depicted in Figure 2C. Two examples of differentially methylated regions are shown in Supplementary Figure 1B. These genes were selected due to their previously reported direct involvement in rheumatic diseases and their underlying molecular pathways. For instance, we found differences between healthy donors and patients with UA at baseline

(UA_0) in CpG sites located in cytokine and chemokine genes, such as *CXCR5*, *IL10*, *IL1R1*, and *IRAK2*; TNF signaling pathway genes, such as *LTA*, *TNFSF10*, and *TRAF4*; type I IFN–activated transcription factor IRF8; and others.

Analysis of transcription factor binding motifs revealed enrichment of motifs belonging to the RUNX transcription factor family in the hypermethylated cluster. Within the hypomethylated cluster, motifs of transcription factors in the basic leucine zipper and ETS families were predominantly enriched (Figure 2D).

Additionally, we performed a differentially variable position (DVP) analysis, which revealed a greater heterogeneity of DNA methylation within the UA_0 group (Figure 2E). Those DVPs exhibited <2% overlap with the previously identified DMPs, further suggesting the presence of intrinsic variance within the UA_0 group (Supplementary Figure 1C). Figure 2F depicts examples of 2 DVPs proximal to genes related to the previously identified GO categories. These data suggest the existence of an underlying epigenetic heterogeneity among UA patients, which might play a role in the diverse clinical presentation of the disease.

To ascertain the relationship between DNA methylation and genomic functional features, we calculated enrichment of the identified DMPs in 15 distinct chromatin states, defined by combinations of epigenetic modifications in PBMCs (12) (Figure 2G). DMPs in the hypermethylated cluster were enriched in regions containing gene coding sequences and transcription start sites (TSS), while hypomethylated DMPs were enriched in actively transcribed regions. Both clusters displayed an enrichment in enhancer regulatory regions, consistent with previously published studies focusing on the analysis of dynamic DNA methylation (13).

The methylome of UA patients anticipating subsequent evolution of the disease. The higher variability of methylation profiles among UA_0 samples compared to healthy individuals is consistent with the clinical heterogeneity in the UA_0 group. UA_0 was composed of 2 subgroups, one of patients who underwent subsequent differentiation to RA 1 year after the initial visit (designated as RA_1) and one of patients who remained classified as having UA 1 year after the initial visit (designated as UA_1) (Figure 1). In fact, slight clinical dissimilarities were found between these 2 groups (Supplementary Table 1). For instance, RA_1 patients had a higher frequency of seropositivity for rheumatoid factor (RF) (Figure 3A). The DAS (14) and some of the parameters included in its calculation, of note, the erythrocyte sedimentation rate and the number of swollen joints, were also higher in the RA_1 group (Figure 3B). However, technically, such differences do not allow the identification of those patients as having definite RA in the clinical setting. Therefore, we aimed to identify DNA methylation alterations that might help predict a future diagnosis in a prospective manner. In our analysis, we included the clinical features that were significantly different between the 2 groups (RF and DAS) as covariates, in order to identify methylation changes that were not due to the effect of those differences.
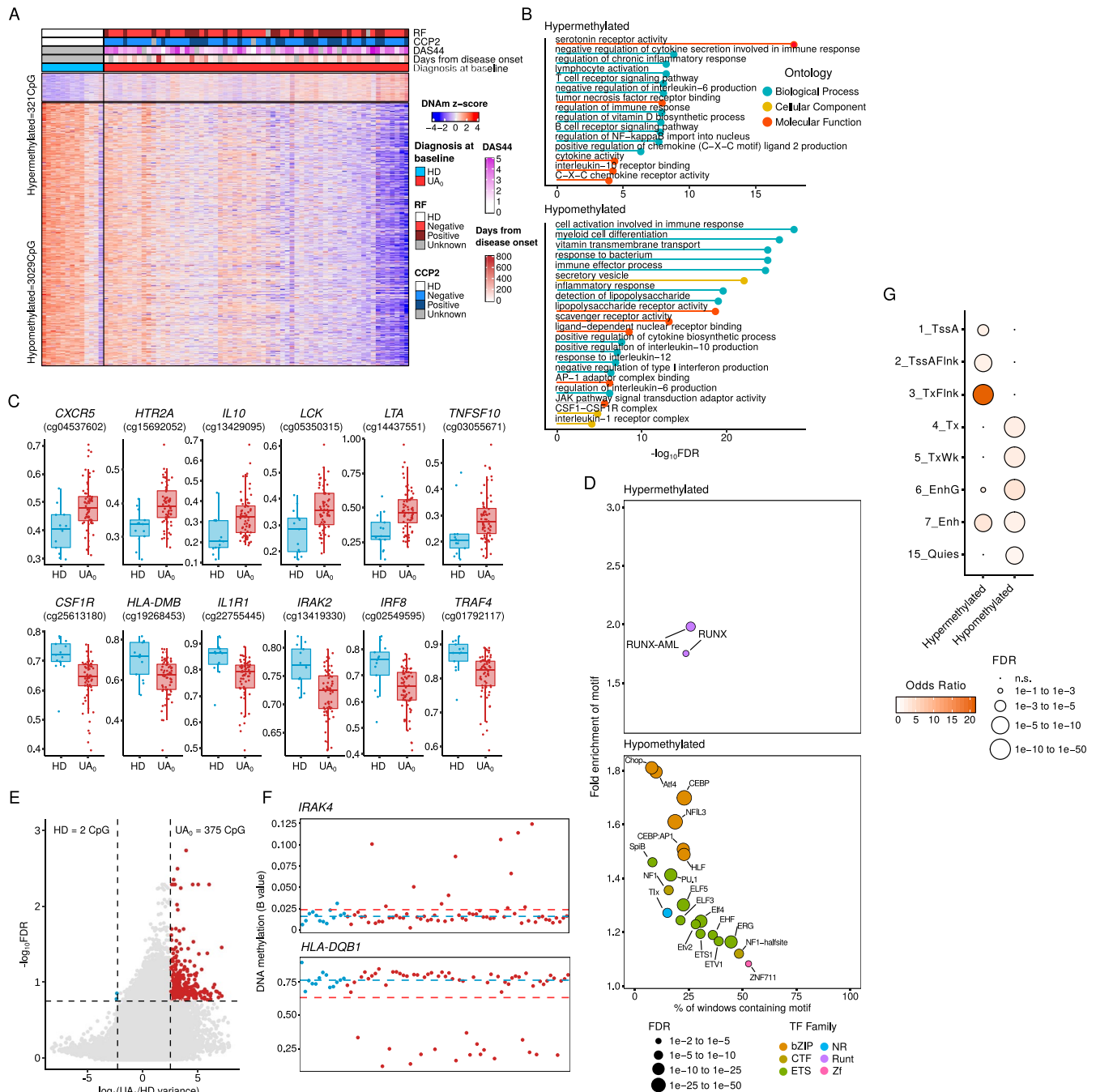
Figure 2. Characterization of the DNA methylation (DNAm) profiles of patients with undifferentiated arthritis at baseline (UA_0) compared to healthy donors (HDs). **A**, Heatmap showing differentially methylated positions (DMPs) between UA_0 and healthy donors (false discovery rate [FDR] < 0.05). Blue indicates lower methylation and red indicates higher methylation. RF = rheumatoid factor; CCP2 = anti–cyclic citrullinated peptide antibody; DAS44 = Disease Activity Score (44 joints assessed). **B**, Significant gene ontology (GO) categories in each cluster, selected by Genomic Regions Enrichment of Annotations Tool analysis of the DMPs identified. AP-1 = activator protein 1; CSF1 = colony-stimulating factor 1; CSF1R = CSF1 receptor. **C**, B values for selected significant CpG sites in the GO categories shown in **B**. Data are shown as box plots. Each box represents the 25th to 75th percentiles. Lines inside the boxes represent the median. Lines outside the boxes represent the 25th percentile minus 1.5 times the interquartile range (IQR) and the 75th percentile plus 1.5 times the IQR. Circles represent individual subjects. **D**, Significantly enriched motifs in DMPs from both clusters, analyzed by HOMER. TF = transcription factor; bZIP = basic leucine zipper; CTF = CCAAT box-binding transcription factor; NR = nuclear receptor; Zf = zinc finger domain. **E**, Variability plot depicting $\log_2$ ratio of variance (var_{UA0}:var_{HD}) for individual CpG sites by $\log_{10}$ FDR of the mean comparison t-test. Significant differentially variable position (DVPs) for both groups identified by the *iEVORA* package are shown in color. **F**, Two examples of DVPs, showing DNA methylation in individual healthy donors (blue) and patients with UA at baseline (red). Broken lines show the mean. **G**, Chromatin functional state enrichment in each cluster, based on public peripheral blood mononuclear cell data from the Roadmap Epigenomics Project (http://www.roadmapepigenomics.org/). TssA = active transcription start site; TssAFlnk = flanking active TSS; TxFlnk = transcript at gene 5′ and 3′; Tx = strong transcription; TxWk = weak transcription; EnhG = genic enhancers; Quies = quiescent; NS = not significant.
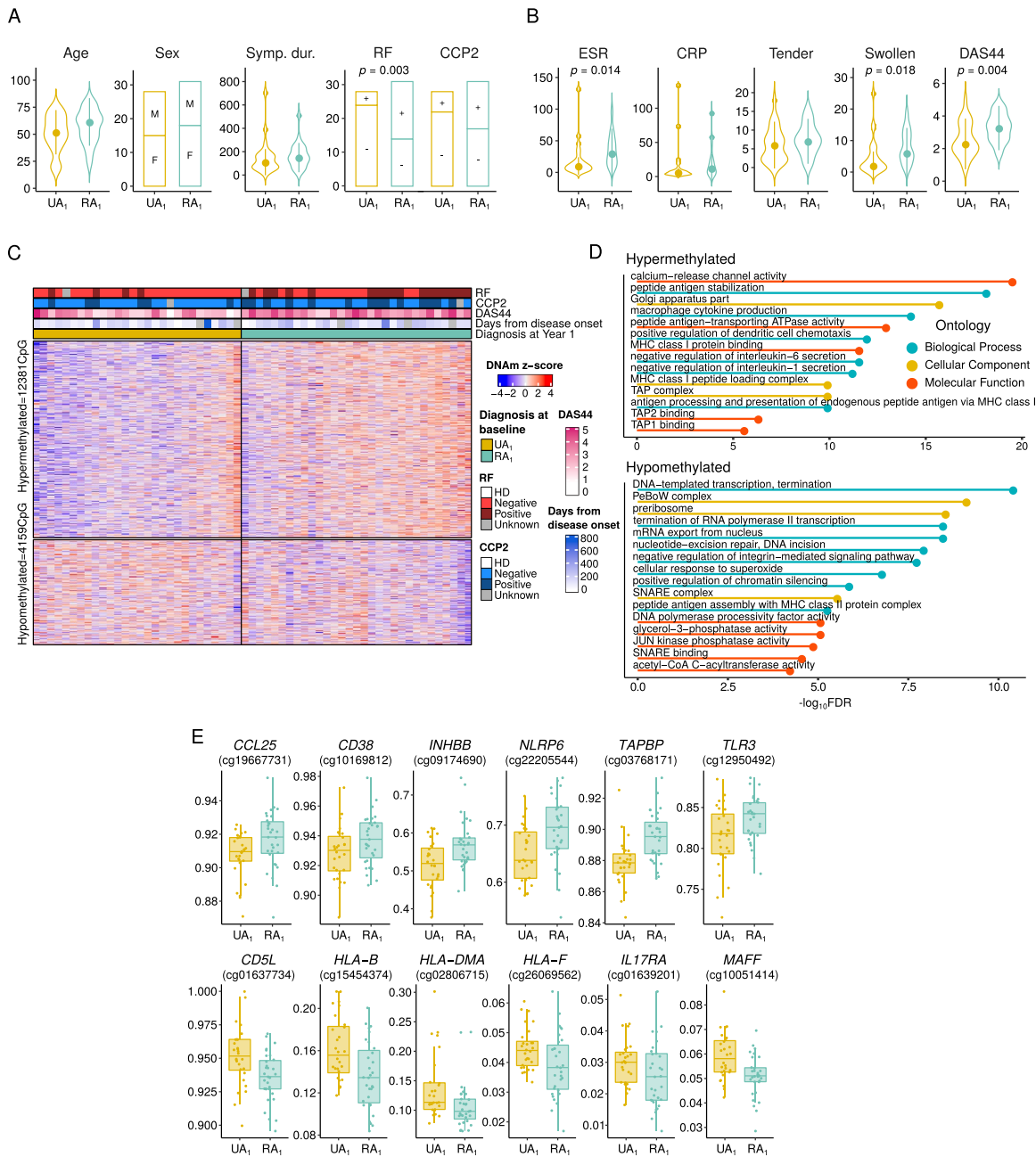
Figure 3. Characterization of the DNA methylation profiles of UA patients with diverging outcomes after 1 year. **A** and **B**, Demographic and clinical characteristics (**A**) and clinical variables included in the DAS (**B**) in patients with UA 1 year after baseline (UA$_1$) and patients with RA 1 year after baseline (RA$_1$). Violin plots show density curves; circles and vertical lines show the mean ± SD. Bars show the absolute number of patients. Significance was determined by Wilcoxon's test for numeric variables and by chi-square test for categorical variables. Symp. dur. = symptom duration in days; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein. **C**, Heatmap showing DMPs between UA$_1$ and RA$_1$ ($P < 0.05$). Blue indicates lower methylation and red indicates higher methylation. **D**, Significant GO categories in each cluster, selected by Genomic Regions Enrichment of Annotations Tool analysis of the DMPs identified. MHC = major histocompatibility complex; TAP = transporter associated with antigen processing; SNARE = soluble N-ethylmaleimide–sensitive factor attachment protein receptor. **E**, B values for selected significant CpG sites in the GO categories shown in **D**. Data are shown as box plots. Each box represents the 25th to 75th percentiles. Lines inside the boxes represent the median. Lines outside the boxes represent the 25th percentile minus 1.5 times the IQR and the 75th percentile plus 1.5 times the IQR. Circles represent individual subjects. See Figure 2 for other definitions. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/art.41885/abstract.

After verification of the similarity of cell type composition (Supplementary Figure 2A, available on the *Arthritis & Rheumatology* website at http://onlinelibrary.wiley.com/doi/10.1002/art.41885/abstract), the comparison of the DNA methylation profiles determined from baseline samples in relation to the clinical groups defined 1 year later (UA$_1$ versus RA$_1$) led to the identification of 12,381 hypermethylated CpGs and 4,159 hypomethylated CpGs (Figure 3C). These DMPs did not vary significantly with regard to slide or array (Supplementary Figure 2D).

DMPs in the hypermethylated cluster were enriched in GO categories related to antigen presentation through major histocompatibility complex (MHC) class I, as well as inflammatory cytokine signaling (such as IL-1 and IL-6). GO categories in the hypomethylated cluster were mainly related to basic cellular processes such as gene transcription, translation, and metabolism, and an additional category related to antigenic presentation through MHC class II (Figure 3D). CpG sites proximal to genes contained in the aforementioned GO categories and related to inflammatory cytokines and chemokine pathways, such as *CCL25*, *CD5L*, and *IL17RA,* were selected, and their B values in the UA$_1$ and RA$_1$ groups are depicted in Figure 3E.

Analysis of transcription factor binding motifs in the hypermethylated cluster revealed an enrichment of motifs belonging to the basic helix-loop-helix and zinc finger domain (Zf) families. The hypomethylated cluster showed enrichment of transcription factors from the ETS and Zf families (Supplementary Figure 2B). Chromatin state enrichment for the hypomethylated cluster revealed an enrichment of regions located in active and poised TSS or their flanking regions. The hypermethylated cluster showed enrichment in actively transcribed regions, enhancers, and repressed chromatin (Supplementary Figure 2C). None of the chromatin states were commonly enriched in the 2 clusters, suggesting the involvement of distinct pathways underlying the identified alterations.

Additionally, samples from patients with UA at baseline that had differentiated into other arthritis subtypes 1 year after the initial visit (psoriatic arthritis, spondyloarthritis, osteoarthritis, or reactive arthritis), labeled "other subtypes" (see Supplementary Table 1), were compared to UA$_1$ and RA$_1$. Due to the sparsity of samples of each of the other subtypes (n = 2 patients per group), we decided to identify DMPs from the UA$_1$ versus RA$_1$ comparison, and to represent the DNA methylation values of the additional samples in an unsupervised manner. In a principal components analysis (PCA), the distribution of the samples from the other subtypes group largely overlapped with the distribution of the RA$_1$ samples (Supplementary Figure 2E). This tendency was corroborated by inspecting the mean methylation value of the DMPs. The mean value in the other subtypes group appeared closer to that of RA$_1$ than to that of UA$_1$, in both the hypomethylated and hypermethylated CpGs (Supplementary Figure 2F).

Of note, this tendency was not reproduced individually by all of the samples, as shown by a pairwise mean comparison between the UA$_1$ and RA$_1$ groups (Supplementary Figure 2G). However, after comparing UA$_1$ and RA$_1$ to each of the samples in the group of other subtypes, we found that significant differences in the mean values occurred more frequently between the UA$_1$ group and the other subtypes group (6 of 8 in the hypomethylated cluster and 8 of 8 in the hypermethylated cluster) than between the RA$_1$ group and the other subtypes group (2 of 8 in the hypomethylated cluster and 6 of 8 in the hypermethylated cluster). Although these results need to be further confirmed, this tendency suggests the existence of an altered signature shared by patients with differentiated arthritis. Taken together, these results indicate for the first time the existence of a pre-established epigenetic signature in UA patients whose disease will evolve to RA.

Improvement of patient classification by incorporation of DNA methylation data into the clinical parameters–based model. Given our findings of DNA methylation differences between UA patient groups that had divergent diagnoses 1 year after baseline, we investigated the possibility of using DNA methylation data to obtain predictive markers of disease progression. To this end, we applied machine learning approaches to build a classification system based on DNA methylation data alone or DNA methylation data in combination with clinical data. The pipeline of the methodology included a random split of the original data into "training" and "test" sets, followed by a selection of predictor CpG sites, and a cross-validation for the internal evaluation of the model (Figure 4A). Models developed and evaluated through this procedure were constructed using logistic regression, random forest, and support vector machine (SVM) algorithms. Aiming at obtaining a relatively simple classifier, we generated models with increasing numbers of CpG sites as predictors (from 1 to 50 CpG sites). In parallel, patient clinical data (RF and DAS) were included as explanatory variables. These variables, which showed significant differences among groups, have also been included in previous studies describing classification rules that were based strictly on 9 clinical parameters (3).

The comparison of the accuracies of all models (see Supplementary Methods) showed the highest precision for SVM-generated models with RF and DAS covariates included (SVM+RF,DAS models) (Figure 4B). The top 10 most frequent CpGs (after performing 100-fold cross-validation) in the SVM+RF,-DAS model are shown in Figure 4C. Given that SVM+RF,DAS models discriminate relatively well with >10 CpGs, we selected 2 examples of models, representing a complex classifier, with 40 CpGs, and a simpler classifier, with 25 CpGs, that might potentially be implemented in the clinical setting. Finally, these models were applied to an external validation cohort (n = 8), recruited independently (Supplementary Table 2, available on the *Arthritis & Rheumatology* website at http://onlinelibrary.wiley.com/doi/10.1002/art.41885/abstract). The predicted outcome was then compared to the observed outcome 1 year after baseline for every patient (Figure 4D). For benchmarking purposes, the previously described clinical classification score (3) (named "composite score") was also used alone or along with DNA methylation in the analysis, in parallel.

Within the validation cohort, the prediction accuracy of the composite score alone was 75%, while the simplified model, which included only 2 variables (DAS and RF), showed an accuracy of 62.5% (Figure 4D). The simplified model with 25 CpGs increased the accuracy (area under the curve [AUC] 0.875) of the prediction by the clinical covariates alone (AUC 0.625). The simplified model with 40 CpGs accurately predicted the class of all 8 patients (AUC 1) (Figure 4D and Supplementary Figure 3A, available on the *Arthritis & Rheumatology* website at http://online
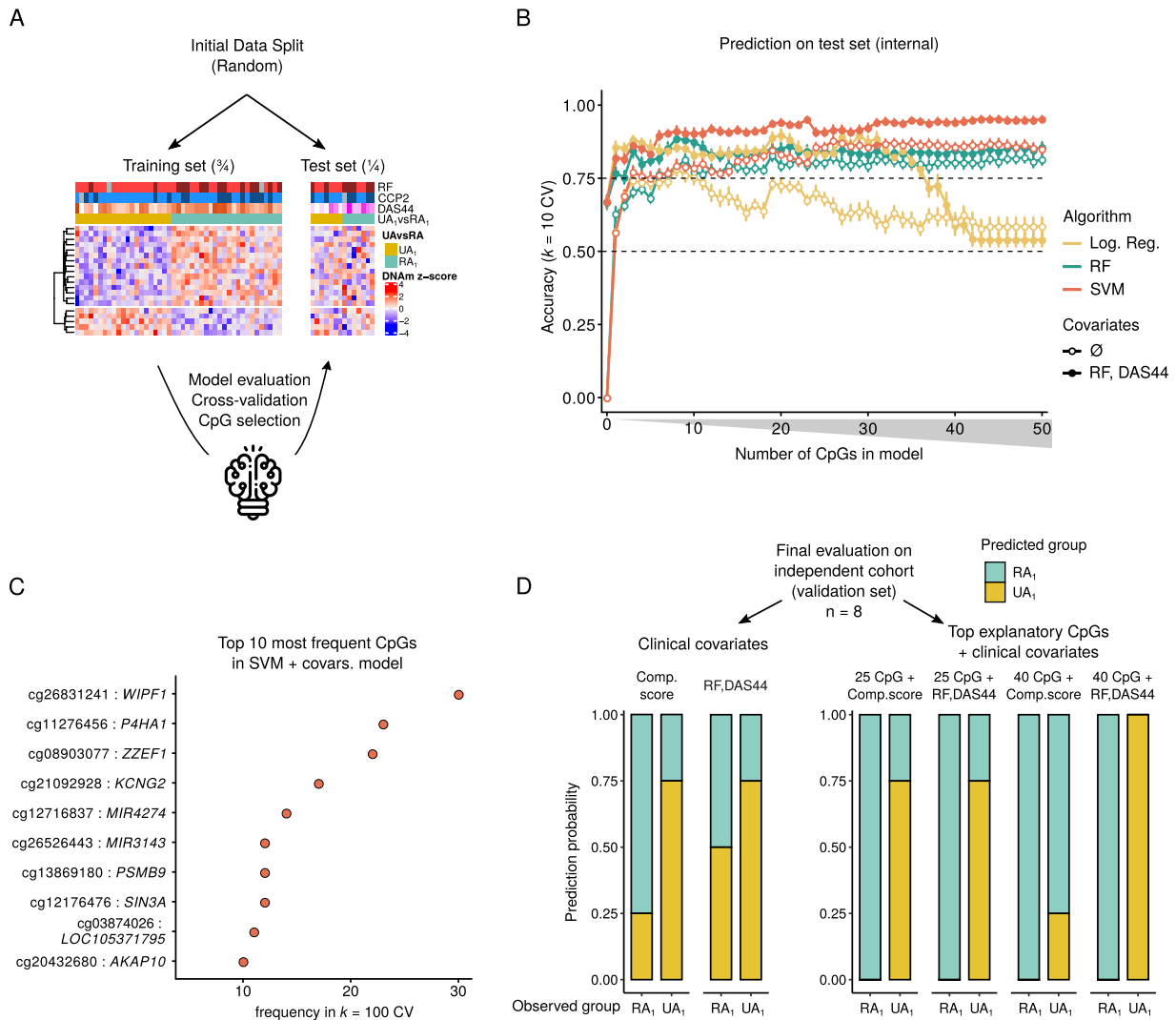
Figure 4. Development of a DNA methylation-based prediction rule by machine learning. Models were constructed to predict whether patients with UA at baseline would have RA 1 year after the initial visit (RA_1) or UA 1 year after the initial visit (UA_1). **A**, Schematic representation of the machine learning methodology, including splitting of data into training and test sets, feature (CpG) selection, evaluation of the model parameters, and cross-validation. **B**, Accuracy of the models developed using logistic regression (Log. Reg.), random forest (RF), and support vector machine (SVM) algorithms. Models included varying numbers of most frequent CpGs as explanatory variables (1–50 CpGs). Values are the mean ± SEM from 10 independent cross-validations. **C**, Top 10 most frequent CpGs after 100-fold cross-validation in the SVM model with covariates (SVM + covars). CV = cross-validation. **D**, Classification results of selected models in an independent validation cohort. Left, Prediction models based on clinical covariates only. Right, Prediction models based on the combination of DNA methylation data plus clinical covariates. Comp. score = composite score (see Figure 2 for other definitions). Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/art.41885/abstract.

library.wiley.com/doi/10.1002/art.41885/abstract). In fact, simplified models with >25 CpGs predicted future diagnosis with an average accuracy of >75%, and the addition of the CpG methylation data improved the predictive ability of clinical parameters in the majority of the models (Supplementary Figure 3B). Although the prediction accuracy of the composite score was higher than that of DAS+RF alone, after the addition of DNA methylation data, the accuracy of the simplified models was higher when compared to the composite score models, in the majority of the cases. In fact, in models with >30 CpGs, the accuracy of the models that included the composite score as a covariate dropped to random

classifier levels, ~50% accuracy (Figure 4D and Supplementary Figure 3B). Taken together, these results highlight the potential of adding DNA methylation as a diagnostic predictive biomarker.

Comparison of UA and definite RA profiles, revealing progression of RA_1 status to RA status. To further characterize the UA_0 subgroups, we compared the DNA methylation profiles of UA_0 with those of patients with terminally differentiated RA (diagnosed as having RA at baseline), labeled RA_0 (Supplementary Table 2). The RA_0 group displayed the most distinct methylation profile, as shown by the greatest differences
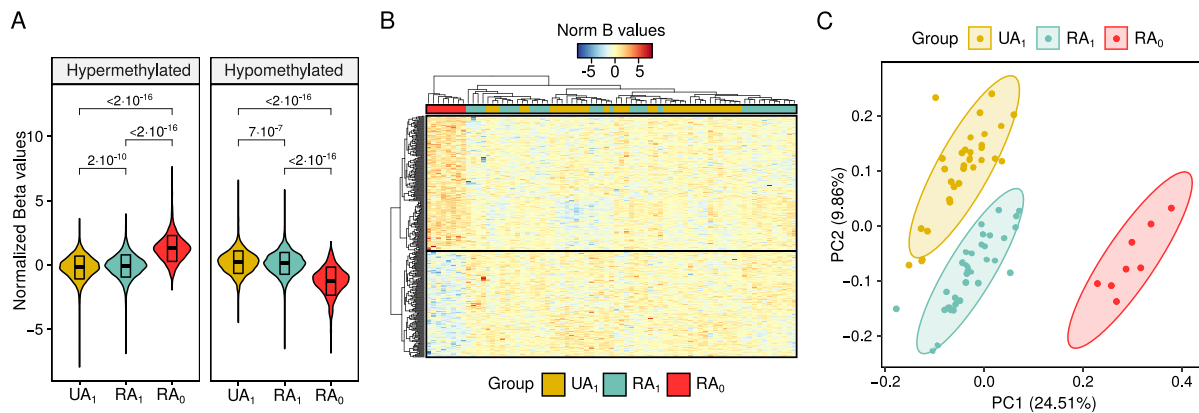
Figure 5. Comparison of the methylome profiles of patients with undifferentiated arthritis (UA) at baseline who continued to be classified as having UA 1 year after baseline (UA_1), patients with UA at baseline that progressed to RA 1 year after baseline (RA_1), and patients with RA at baseline (RA_0). **A**, Violin plots showing normalized B value distributions of differentially methylated positions (DMPs) between UA_1, RA_1 and RA_0 profiles. Boxes show the 25th to 75th percentiles. Lines inside the boxes represent the median. Values above the violin plots are P values. The microarray model (450k or EPIC) was included as a covariate in the *limma* model. **B**, Heatmap of the identified DMPs. Columns (samples) were clustered by a complete-linkage clustering algorithm. **C**, Principal components analysis of the DMPs. Ellipses show the 95% confidence interval for the distribution of each group. Circles represent individual patients. PC1 = principal component 1. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/art.41885/abstract.

in mean DNA methylation when compared to both UA_1 and RA_1 (Figure 5A). We then performed unsupervised clustering of the significant DMPs among the 3 groups (RA_0, UA_1, and RA_1). We observed that all UA_1 and RA_1 samples (both subgroups of UA_0) aggregated together in the same cluster, while all RA_0 samples clustered independently (Figure 5B). Overall methylation of the identified DMPs showed significant differences among the 3 groups. In addition, these regions appeared to experience a progressive dynamic from UA_1 to RA_1 to RA_0, both for the hypermethylated and the hypomethylated clusters (Figure 5A). This tendency was further reinforced after reducing the dimensionality of the DMP data by PCA, where RA_1 patients lie in between RA_0 and UA_1, which appeared as the most extreme groups when projected in principal component 1 (PC1) and PC2 (Figure 5C).

Additionally, unsupervised K-means clustering of the DMP data revealed a total association of RA_0 in an isolated cluster (cluster 3), while UA_1 and RA_1, which were largely associated with independent clusters (clusters 1 and 2), showed a certain degree of interspersing (Supplementary Figure 4A, available on the *Arthritis & Rheumatology* website at http://onlinelibrary.wiley.com/doi/10.1002/art.41885/abstract). When samples were assigned to clusters, 5 (15.6%) of 32 UA_1 samples belonged to cluster 2 ("RA_1 cluster"), while 7 (20%) of 35 RA_1 samples belonged to cluster 1 ("UA_1 cluster"). All (100%) of the 8 RA_0 samples belonged to cluster 3 (Supplementary Figure 4B). These data reinforce the notion of a pre-existing RA-like epigenetic profile underlying UA in some patients, which reveals a progression of the disease in these patients to RA.

DISCUSSION

Our analysis of the DNA methylomes of UA patients showed a distinct signature in comparison with healthy individuals, as well

as specific differences between patients whose disease subsequently evolved to RA and those whose disease remained undifferentiated. These observations prompted us to design a machine learning–based method, which improved previous classification systems (3), to predict outcomes in UA patients in our cohort. The finding that UA patients who will develop RA have a more similar DNA methylation signature to patients with well-established RA supports the notion of pre-existing epigenetic signatures that might be used to anticipate patient outcomes and, therefore, improve therapeutic decisions.

Our results show for the first time that UA patients display epigenetic alterations when compared to healthy individuals. Those alterations, which occur in regions that are functionally associated with inflammatory pathways, are common to those previously observed in other inflammatory diseases. In particular, enriched functional categories of inflammation, immune cell activation, and cytokine signaling have also been found in RA (6,15,16), SLE (17,18), asthma (19), and inflammatory bowel disease (20) in comparable studies, supporting the idea that UA shares epigenetic similarities with other inflammatory diseases and thus can be molecularly considered as such. Furthermore, the identification of vast DNA methylation differences at the TNF locus, as well as alterations at several CpGs within the HLA class II region (both at the DMP and DVP level) confirms that UA is an arthritide like RA, with which it shares clinical characteristics (7,21–23). Of note, the results of this particular analysis may be partially limited due to the exclusion of sex chromosomes and age-associated CpGs, in which UA-associated alterations could also occur.

Furthermore, we identified DNA methylation differences among UA patients based on their prospective status, namely, the diagnosis 1 year after the first visit (evolution to RA or persistent UA). After correcting for clinical features among the 2 groups, we

identified DNA methylation differences mainly localized in regions near genes related to inflammation and antigen presentation (24–27). For instance, HLA genes have been recursively linked to autoimmunity, showing both association with genetic susceptibility and epigenetic alterations in several studies (7,28,29). HLA–B, HLA–DMA, and HLA–F have further been linked to autoimmunity (30–32). Other genes, such as CD38, have been shown to be up-regulated at the protein level in UA patients, and CD38 has been proposed as a therapeutic target in UA and early arthritis (33). These observations suggest that those patients possess early biologic alterations before undergoing diverging clinical outcomes.

Identification of disease onset in the clinical setting is often preceded by the presence of unapparent molecular triggers, as previously described for RA treatment response (34) and flare outbreaks (35). Those determinants appear early in the disease course and cannot easily be detected by clinicians through noninvasive means. However, their sustained presence and effect at several levels may contribute to a specific pathologic phenotype. We believe that this study underpins the potential of using epigenetic modifications as a molecular sensor for those early disease determinants in UA, in order to improve the classification criteria for UA and prevent damage caused by sustained inflammation. However, future longitudinal studies that include data from follow-up visits would provide further insights into the mechanisms by which UA patients diverge, and their underlying epigenetic dynamics. Also, further evidence is needed to confirm whether the observed phenomenon is common to differentiated subtypes of arthritis other than RA, as suggested by the preliminary data included in this study.

Autoimmune arthritides are characterized by a high level of heterogeneity in terms of patient prognosis, joint damage, and response to treatment, for which mechanistic causality remains largely unknown (36). In this sense, the use of high-throughput technologies has enabled the development of computational methods for the processing of patient -omic data in search of novel and more precise conclusions (37–39). For instance, the implementation of machine learning algorithms in high-dimensional data analysis has previously been used to improve stratification of patients (40–42) or to predict disease activity (43,44) in RA and SLE. In the present study, we used DNA methylation in addition to clinical data on UA patients by applying machine learning approaches, fine-tuning the prediction performance of previously existing classifiers (3) in an independent validation cohort. Nevertheless, although the use of data obtained from PBMCs might limit the identification of alterations in specific cell subtypes, it simplifies the generation of data in clinical practice, avoiding the need for cell sorting. Our conclusions highlight the convenience of using both clinical and basic research data in conjunction for a complete and robust patient prognostic and therapeutic assessment. The results obtained herein are presented as a proof of concept to be further confirmed in independent studies with larger sample sizes. We hope our methodology can also be applied to other disease contexts.

The comparison of the methylation profiles of all of the UA patients included in our cohort (regardless of their prospective status at year 1, i.e., UA or RA) versus those initially classified as having RA, showed that patients with UA and those with RA displayed differential methylation profiles, further supporting the idea that these 2 groups actually belong to distinct disease entities from a molecular/epigenetic perspective. Upon exploration by unsupervised analyses, the 2 UA subgroups, UA_1 and RA_1, showed a higher resemblance to each other than either did to RA_0, suggesting that despite the existing differences among them, the 2 UA groups (UA_1 and RA_1) still behaved as an entity when compared to a differentiated group. Interestingly, UA_1 and RA_0 had the most extreme distributions, while RA_1 displayed an intermediate distribution. In all, these data suggest the pre-existence of a molecular/epigenetic signature in UA patients that develop RA in the future.

Many efforts have been devoted to promptly abort the inflammatory process and the progression of the disease to a more severe form, facilitating a rapid halt of the dysregulated inflammatory process and avoiding inflammation-associated tissue damage. Indeed, a delayed treatment of these patients is commonly associated with a worse global response to treatment, joint destruction, and impaired quality of life. In this context, our results regarding epigenetic signatures associated with distinctive UA evolution suggest that, in addition to specific clinical parameters, molecular features such as DNA methylation should be considered to be integrated into the clinic with the aim of a better classification of these patients.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Ballestar had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study conception and design. de la Calle-Fabregat, van der Helm-van Mil, Ballestar.
Acquisition of data. de la Calle-Fabregat, Niemantsverdriet, van der Helm-van Mil, Rodríguez-Ubreva, Ballestar.
Analysis and interpretation of data. de la Calle-Fabregat, Cañete, Li, Rodríguez-Ubreva, Ballestar.

REFERENCES

1. Van Aken J, van Dongen H, le Cessie S, Allaart CF, Breedveld FC, Huizinga TW. Comparison of long term outcome of patients with rheumatoid arthritis presenting with undifferentiated arthritis or with rheumatoid arthritis: an observational cohort study. Ann Rheum Dis 2006;65:20–5.

2. Van Dongen H, van Aken J, Lard LR, Visser K, Ronday HK, Hulsmans HM, et al. Efficacy of methotrexate treatment in patients with probable rheumatoid arthritis: a double-blind, randomized, placebo-controlled trial. Arthritis Rheum 2007;56:1424–32.

3. Van Der Helm-van Mil AH, le Cessie S, van Dongen H, Breedveld FC, Toes RE, Huizinga TW. A prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: how to guide individual treatment decisions. Arthritis Rheum 2007;56:433–40.

4. De Maturana EL, Alonso L, Alarcón P, Martín-Antoniano IA, Pineda S, Piorno L, et al. Challenges in the integration of omics and non-omics data [review]. Genes (Basel) 2019;10:238.

5. Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene [letter]. Nature 1999;402:187–91.

6. Rodríguez-Ubreva J, de la Calle-Fabregat C, Li T, Ciudad L, Ballestar ML, Català-Moll F, et al. Inflammatory cytokines shape a changing DNA methylome in monocytes mirroring disease activity in rheumatoid arthritis. Ann Rheum Dis 2019;78:1505–16.

7. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 2013;31:142–7.

8. Ballestar E, Sawalha AH, Lu Q. Clinical value of DNA methylation markers in autoimmune rheumatic diseases [review]. Nat Rev Rheumatol 2020;16:514–24.

9. De Rooy DP, van der Linden MP, Knevel R, Huizinga TW, van der Hel-van Mil AH. Predicting arthritis outcomes—what can be learned from the Leiden Early Arthritis Clinic? Rheumatology (Oxford) 2011;50:93–100.

10. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.

11. Van der Heijde DM, van 't Hof MA, van Riel PL, Theunisse LM, Lubberts EW, van Leeuwen MA, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Ann Rheum Dis 1990;49:916–20.

12. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization [letter]. Nat Methods 2012;9:215–6.

13. Luo C, Hajkova P, Ecker JR. Dynamic DNA methylation: In the right place at the right time [review]. Science 2018;361:1336–40.

14. Ranganath VK, Yoon J, Khanna D, Park GS, Furst DE, Elashoff DA, et al. Comparison of composite measures of disease activity in an early seropositive rheumatoid arthritis cohort. Ann Rheum Dis 2007;66:1633–40.

15. Zhu H, Wu LF, Mo XB, Lu X, Tang H, Zhu XW, et al. Rheumatoid arthritis-associated DNA methylation sites in peripheral blood mononuclear cells. Ann Rheum Dis 2019;78:36–42.

16. Mok A, Rhead B, Holingue C, Shao X, Quach HL, Quach D, et al. Hypomethylation of CYP2E1 and DUSP22 promoters associated with disease activity and erosive disease among rheumatoid arthritis patients. Arthritis Rheumatol 2018;70:528–36.

17. Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Ignacio Martin-Subero J, Rodriguez-Ubreva J, et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. Genome Res 2010;20:170–9.

18. Lanata CM, Paranjpe I, Nititham J, Taylor KE, Gianfrancesco M, Paranjpe M, et al. A phenotypic and genomics approach in a multi-ethnic cohort to subtype systemic lupus erythematosus. Nat Commun 2019;10:3902.

19. Yang IV, Pedersen BS, Liu A, O'Connor GT, Teach SJ, Kattan M, et al. DNA methylation and childhood asthma in the inner city. J Allergy Clin Immunol 2015;136:69–80.

20. McDermott E, Ryan EJ, Tosetto M, Gibson D, Burrage J, Keegan D, et al. DNA methylation profiling in inflammatory bowel disease provides new insights into disease pathogenesis. J Crohn's Colitis 2016;10:77–86.

21. Van Steenbergen HW, Luijk R, Shoemaker R, Heijmans BT, Huizinga TW, van der Helm-van Mil AH. Differential methylation within the major histocompatibility complex region in rheumatoid arthritis: a replication study. Rheumatology (Oxford) 2014;53:2317–8.

22. Anaparti V, Agarwal P, Smolik I, Mookherjee N, El-Gabalawy H. Whole blood targeted bisulfite sequencing and differential methylation in the C6ORF10 gene of patients with rheumatoid arthritis. J Rheumatol 2020;47:1614–23.

23. Pitaksalee R, Burska AN, Ajaib S, Rogers J, Parmar R, Mydlova K, et al. Differential CpG DNA methylation in peripheral naïve CD4+ T-cells in early rheumatoid arthritis patients. Clin Epigenetics 2020;12:54.

24. Yokoyama W, Kohsaka H, Kaneko K, Walters M, Takayasu A, Fukuda S, et al. Abrogation of CC chemokine receptor 9 ameliorates collagen-induced arthritis of mice. Arthritis Res Ther 2014;16:445.

25. Wu X, Li M, Chen T, Zhong H, Lai X. Apoptosis inhibitor of macrophage/CD5L is associated with disease activity in rheumatoid arthritis. Clin Exp Rheumatol 2021;39:58–65.

26. Wang C, Yosef N, Gaublomme J, Wu C, Lee Y, Clish CB, et al. CD5L/AIM regulates lipid biosynthesis and restrains Th17 cell pathogenicity. Cell 2015;163:1413–27.

27. Van den Berg WB, Miossec P. IL-17 as a future therapeutic target for rheumatoid arthritis [review]. Nat Rev Rheumatol 2009;5:549–53.

28. Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease [review]. Nat Rev Immunol 2018;18:325–39.

29. Kular L, Liu Y, Ruhrmann S, Zheleznyakova G, Marabita F, Gomez-Cabrero D, et al. DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. Nat Commun 2018;9:2397.

30. Bowness P. HLA B27 in health and disease: a double-edged sword? [review]. Rheumatology (Oxford) 2002;41:857–68.

31. Morel J, Roch-Bras F, Molinari N, Sany J, Eliaou JF, Combe B. HLA-DMA*0103 and HLA-DMB*0104 alleles as novel prognostic factors in rheumatoid arthritis. Ann Rheum Dis 2004;63:1581–6.

32. Afroz S, Giddaluru J, Vishwakarma S, Naz S, Khan AA, Khan N. A comprehensive gene expression meta-analysis identifies novel immune signatures in rheumatoid arthritis patients. Front Immunol 2017;8:74.

33. Cole S, Walsh A, Yin X, Wechalekar MD, Smith MD, Proudman SM, et al. Integrative analysis reveals CD38 as a therapeutic target for plasma cell-rich pre-disease and established rheumatoid arthritis and systemic lupus erythematosus. Arthritis Res Ther 2018;20:85.

34. Lewis MJ, Barnes MR, Blighe K, Goldmann K, Rana S, Hackney JA, et al. Molecular portraits of early rheumatoid arthritis identify clinical and treatment response phenotypes. Cell Rep 2019;28:2455–70.

35. Orange DE, Yao V, Sawicka K, Fak J, Frank MO, Parveen S, et al. RNA identification of PRIME cells predicting rheumatoid arthritis flares. N Engl J Med 2020;383:218–28.

36. Pitzalis C, Kelly S, Humby F. New learnings on the pathophysiology of RA from synovial biopsies [review]. Curr Opin Rheumatol 2013;25:334–44.

37. Donlin LT, Park SH, Giannopoulou E, Ivovic A, Park-Min KH, Siegel RM, et al. Insights into rheumatic diseases from next-generation sequencing [review]. Nat Rev Rheumatol 2019;15:327–39.

38. Lewis MJ, Barnes MR. RNA sequencing and machine learning as molecular scalpels [review]. Nat Rev Rheumatol 2018;14:388–90.

39. Stafford IS, Kellermann M, Mossotto E, Beattie RM, MacArthur BD, Ennis S. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases [review]. NPJ Digit Med 2020;3:30.

40. Orange DE, Agius P, DiCarlo EF, Robine N, Geiger H, Szymonifka J, et al. Identification of three rheumatoid arthritis disease subtypes by

machine learning integration of synovial histologic features and RNA sequencing data. Arthritis Rheumatol 2018;70:690–701.

41. Figgett WA, Monaghan K, Ng M, Alhamdoosh M, Maraskovsky E, Wilson NJ, et al. Machine learning applied to whole-blood RNA-sequencing data uncovers distinct subsets of patients with systemic lupus erythematosus. Clin Transl Immunol 2019;8:e01093.

42. Robinson GA, Peng J, Dönnes P, Coelewij L, Naja M, Radziszewska A, et al. Disease-associated and patient-specific immune cell signatures in juvenile-onset systemic lupus erythematosus: patient stratification using a machine-learning approach. Lancet Rheumatol 2020;2:e485–96.

43. Lin C, Karlson EW, Canhao H, Miller TA, Dligach D, Chen PJ, et al. Automatic prediction of rheumatoid arthritis disease activity from the electronic medical records. PLoS One 2013;8:e69932.

44. Kegerreis B, Catalina MD, Bachali P, Geraci NS, Labonte AC, Zeng C, et al. Machine learning approaches to predict lupus disease activity from gene expression data. Sci Rep 2019;9:9617.

Clinical Images: Cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy syndrome, a central nervous system vasculitis mimic
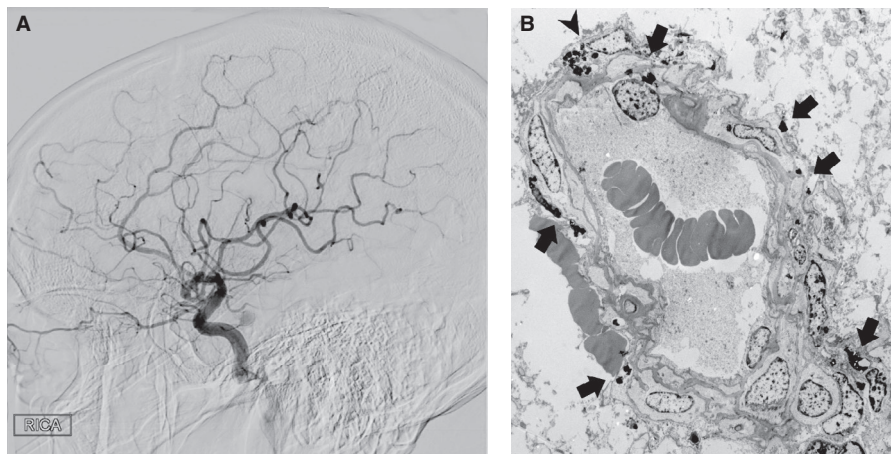
The patient, a 37-year-old man with diabetes mellitus and hypertension, presented with severe headache. Over a 1-month period, magnetic resonance imaging showed acute strokes in the right paramedian pons, left thalamus/globus pallidus/subinsula, and right corona radiata. Evaluation for primary angiitis of the central nervous system (PACNS) included lumbar puncture revealing 7 white blood cells, as well as normal protein and glucose levels. Cerebral arteriography demonstrated diffuse small vessel beading in the anterior and middle territories of the cerebral artery bilaterally (**A**). The patient was started on pulse-dose therapy with methylprednisolone empirically for PACNS. Brain biopsy did not show the expected finding of vasculitis on light microscopy, and transmission electron microscopy revealed the presence of granular osmophilic material (GOM) around smooth muscle cells in cerebral white matter arterioles. An abundance of lipofuscin in association with GOM (as indicated by **arrows** in **B** with **arrowhead** denoting a scavenger cell containing GOM and lipofuscin) suggested the presence of a degenerative process. In contrast, in patients with PACNS, brain biopsies show transmural lymphocytic vasculitis with fibrinoid necrosis (1). Given the young age of the patient, genetic testing was performed to assess for possible hereditary syndromes. The test identified a Notch homolog 3 mutation, which confirmed the diagnosis of cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). CADASIL is the most common genetic cause of ischemic stroke, often presenting with early-onset strokes, migraines, and white matter lesions (2). Mutations in the Notch-3 gene, which encodes a transmembrane receptor expressed in arterial smooth muscle cells, result in an arteriopathy that can mimic CNS vasculitis with hypointense lesions at the cortico–subcortical junction and white matter hyperintensities in the anterior temporal lobes (1,2). The ultrastructural findings in this case could explain the beaded appearance of arterioles on arteriography, mimicking vasculitis. This case demonstrates that findings suggestive of PACNS on arteriography often lack specificity, and brain biopsy and genetic testing can be critical tools to secure the right diagnosis. Familiarity with this rare vasculitis mimic can ensure early diagnosis and avoid unnecessary immunosuppression.

1. Byram K, Hajj-Ali RA, Calabrese L. CNS vasculitis: an approach to differential diagnosis and management [review]. Curr Rheumatol Rep 2018;20:37.

2. Chabriat H, Joutel A, Dichgans M, Tournier-Lasserve E, Npusser MG. Cadasil [review]. Lancet Neurol 2009;8:643–53.

Mithu Maheswaranathan, MD [iD]
Anne F. Buckley, MD, PhD
Andrew B. Cutler, MD
Lisa Criscione-Schreiber, MD, Med
Ankoor Shah, MD
Duke University School of Medicine
Durham, NC