



OPEN DNA methylome biomarkers of rheumatoid arthritis-associated interstitial lung disease reflecting lung fibrosis pathways, an exploratory case–control study

Bartosz Kaczmarczyk^{1,23}, Carlos de la Calle-Fabregat^{2,23}, Adrian Conde¹, Ana Catarina Duarte³, Natalia Mena-Vazquez⁴, Antonio Fernandez-Nebro⁴, Ana Triguero-Martinez⁵, Santos Castañeda⁵, Raquel Dos-Santos Sobrin¹, Antonio Mera-Varela^{1,22}, Chary Lopez-Pedrerá⁶, Alejandro Escudero-Contreras⁶, Paloma Vela-Casasempere⁷, Maria Molina⁸, Javier Narvaez⁹, Miriam Retuerto-Guerrero¹⁰, Jose L. Pablos¹⁰, Juan C. Sarmiento-Monroy¹¹, Raimon Sanmarti¹¹, Luis Gomez-Carrera¹², Gema Bonilla¹³, Sara Remuzgo-Martinez¹⁴, Miguel Angel Gonzalez-Gay^{15,16}, Virginia Leiro-Fernandez¹⁷, Nair Perez-Gomez¹⁸, Cristina Vadillo-Font¹⁹, Lydia Abasolo¹⁹, Ivette Casafont-Sole²⁰, Lourdes Mateo-Soria²⁰, Ana Cristina Castillo-Gonzalez²¹, Carlos Marras²¹, Eva Perez-Pampin^{1,22}, Esteban Ballestar^{2,23}, Antonio Gonzalez^{1,23}✉ & On behalf of the MARILD network*

Rheumatoid Arthritis-associated Interstitial Lung Disease (RA-ILD) significantly reduces life quality and survival, necessitating improvements in its understanding and clinical management. We addressed these goals using DNA methylation analysis, which has not been done in RA-ILD samples, by comparing 32 RA patients with ILD diagnosed less than one year before (cases) and 32 matched RA patients without ILD (controls). This analysis identified 6679 differentially methylated positions (DMPs) with $\Delta\beta \geq 2\%$ and $FDR < 0.05$, and 576 differentially methylated regions in RA-ILD. Some DMPs were near mucin, collagen, and telomere maintenance genes. Also, the most notably enriched gene set (up to $p_{adj} = 1.9 \times 10^{-38}$) included genes overexpressed in fibrosis by monocytes and alveolar macrophages. Other significantly enriched gene sets, known to be dysregulated in fibrosis, included the mitotic spindle and the Rho GTPases. Additionally, analysis of transcription factor binding sites around DMPs showed unique enrichment near the liver X receptor element (LXRE), which is associated with fibrosis in multiple tissues. These results were consistent and unaffected by stricter significance thresholds. They indicated that differential DNA methylation may serve as blood biomarkers for RA-ILD including some related to lung fibrosis pathways.

Keywords Rheumatoid arthritis, Interstitial lung disease, Biomarkers, Lung fibrosis, DNA methylation, Epigenetics

¹Experimental and Observational Rheumatology and Rheumatology Unit, Instituto Investigación Sanitaria-Hospital Clínico Universitario de Santiago, Santiago de Compostela, Spain. ²Epigenetics and Immune Disease Group, Josep Carreras Leukaemia Research Institute (IJC), Barcelona, Spain. ³Rheumatology Department, Unidade Local de Saúde de Almada-Seixal - Hospital Garcia de Orta, Almada, Portugal. ⁴Department of Rheumatology, University Regional Hospital of Malaga (HRUM). Institute for Biomedical Research in Malaga (IBIMA), Malaga University, Málaga, Spain. ⁵Rheumatology Department, Hospital Universitario de La Princesa, Instituto de Investigación Sanitaria la Princesa (IIS-Princesa), Madrid, Spain. ⁶Rheumatology Service, Maimonides Institute of Biomedical Research of Cordoba (IMIBIC), Reina Sofia University Hospital, University of Cordoba, Cordoba, Spain. ⁷Rheumatology Department, Hospital General Universitario Dr Balmis, ISABIAL, Alicante, Spain. ⁸Pneumology Department, Hospital Universitario Belvitge, Barcelona, Spain. ⁹Rheumatology Department, Hospital Universitario Belvitge, Barcelona, Spain. ¹⁰Rheumatology Department, Hospital 12 de Octubre and Universidad Complutense de Madrid, Madrid, Spain.

Spain. ¹¹Rheumatology Department, Hospital Clinic and IDIBAPS, Barcelona, Spain. ¹²Pneumology Department, Instituto de Investigación Hospital Universitario La Paz (IDIPAZ), Madrid, Spain. ¹³Rheumatology Department, Instituto de Investigación Hospital Universitario La Paz (IDIPAZ), Madrid, Spain. ¹⁴Rheumatology Department, Hospital Universitario Marques de Valdecilla, Santander, Spain. ¹⁵Department of Medicine and Psychiatry, University of Cantabria, Santander, Spain. ¹⁶Rheumatology Division, IIS-Fundación Jiménez Díaz, Madrid, Spain. ¹⁷Pneumology Department, NeumoVigo I+i Research Group, Complejo Hospitalario Universitario de Vigo, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO. CIBERES. ISCIII, Vigo, Spain. ¹⁸Rheumatology Unit, Hospital Alvaro Cunqueiro, Vigo, Spain. ¹⁹Rheumatology Department, Hospital Clínico San Carlos – Instituto Investigación Sanitaria San Carlos (IdISSC), Madrid, Spain. ²⁰Rheumatology Department, Hospital Universitari Germans Trias i Pujol, Badalona, Spain. ²¹Rheumatology Unit, Hospital Universitario Virgen de la Arrixaca, Murcia, Spain. ²²Present address: Department of Medicine. Faculty of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain. ²³Bartosz Kaczmarczyk, Carlos de la Calle-Fabregat, Esteban Ballestar and Antonio Gonzalez contributed equally to this work. *A list of authors and their affiliations appears at the end of the paper. ✉email: agmartinezp@ser.es

Approximately 10% of the patients with RA are affected by clinically relevant RA-associated Interstitial Lung Disease (RA-ILD)¹. This fraction of patients has become a focus of interest for the identification of biomarkers since ILD can drastically reduce the quality of life and survival. Notwithstanding, the pathogenesis of RA-ILD is still incompletely understood. Current knowledge derives mostly from idiopathic pulmonary fibrosis (IPF) research given that the two diseases share important features². The recognized risk factors are advanced age, male sex, smoking, and a genetic polymorphism in the *MUC5B* gene, which also are risk factors for IPF; whereas the other two risk factors, inflammatory activity and presence of autoantibodies are RA-ILD exclusive^{1,2}. Additionally, the pathophysiology of the two diseases seems to involve alveolar epithelial cell (AEC) dysfunction and aberrant wound healing resulting in fibroblast activation, myofibroblast differentiation, and lung fibrosis^{2,3}. Some of the pathogenic elements are reflected in biomarkers. Examples are the *MUC5B* polymorphism, which is associated with the overproduction of mucus and impaired mucociliary clearance; genetically determined telomere shortening, that contributes to AECs senescence^{4,5}; the serum biomarkers Krebs von den Lungen-6 (KL-6) and surfactant protein D (SP-D), that reflect AECs dysfunction, and metalloproteinase-7 (MMP-7), that is associated with increased matrix turnover^{2,3}. However, this knowledge does not satisfy the clinical needs because the known risk factors and biomarkers have proven insufficient for RA-ILD prediction, early diagnosis, and management^{1,2}. A promising approach to address these unmet needs is DNA methylation analysis.

This type of analysis is promising because DNA methylation includes a dynamic and reversible component that reflects genetic, environmental, and endogenous factors. In recent years, the assessment of DNA methylation and other epigenetic layers has contributed to the molecular characterization of multiple diseases^{6–9}. The changes in methylation can be dramatic, as in cancer, or more subtle as observed in a wide range of complex diseases, including RA and other autoimmune and inflammatory diseases^{6,8–10}. Most often the small (<10%) methylation changes lack a biological interpretation^{6,8,9}. This “subtle change” paradigm contrasts with the large differences in methylation associated with switching genes on or off observed during development, imprinting, and oncogenesis^{6–9}. However, the subtle changes at specific positions (DMPs) and regions (DMRs) can be useful biomarkers reflecting a range of disease features, including onset, activity, and prognosis^{6,8–19}. These associations are frequently discovered through epigenome-wide association studies (EWAS), which evaluate the relationship of disease phenotypes with modest DNA methylation changes at numerous CpG sites^{6,8,9}. However, no such study has ever been done for RA-ILD. The published RA EWAS have already shown the potential of this technology. It has revealed fibroblast-like synovocyte (FLS) DMPs that are specific to RA^{11,12}, differentiate the RA phases¹³, and the joint from which the FLS originate¹¹. Besides FLS, synovial and blood monocytes have also proven informative on the progression of early arthritis over time, in the work of the de la Calle-Fabregat and colleagues¹⁴, and of RA disease activity, in the Rodriguez-Ubreva and colleagues study¹⁵. Additional EWAS on whole blood DNA have identified a set of biomarkers able to predict the evolution from undifferentiated arthritis to RA¹⁶, and cell-type-specific biomarkers predicting the response of RA patients to treatment with methotrexate¹⁷. Other studies on whole blood DNA identified an interferon-inducible gene network involved in the RA pathogenesis¹⁸ and evidence suggesting that DNA methylation mediates the HLA genetic risk in patients presenting anti-cyclic citrullinated peptide (CCP) antibodies¹⁹. Notably, many of these studies were successful with total sample sizes ranging from 35 to 72^{11–18}.

The abundance of blood DNA methylation findings in RA has inspired the current exploratory study considering that methylation changes in the blood could identify biomarkers related to molecular processes involved in RA-ILD. Specifically, we hypothesized that an EWAS would be an efficient strategy to obtain new RA-ILD biomarkers reflecting underlying pathogenic disturbances. In addition, we limited our analysis to patients with an early ILD diagnosis and performed a matched case-control study to increase our chances of genuine results. Finally, we accounted for multiple tests by applying the false discovery rate (FDR) method and verified the robustness of thresholds and the concordance of gene set enrichments to ensure the validity of our findings.

Results

RA-ILD-associated DMPs and DMRs are predominantly hypermethylated

Firstly, we matched patients with RA-ILD and RA^{controls} for known potential confounders (Suppl. Fig. 1). The two groups showed an identical frequency in sex (23 women), smoking (14 ever smokers), anti-CCP status (29 seropositive), and an identical median age (64.7 years). Additionally, they were similar in other demographic and clinical aspects: ethnicity, level of studies, rheumatoid factor (RF) positive or negative status, *MUC5B* promoter genotype, disease activity assessed with the Disease Activity Score 28 joints (DAS28), age at RA diagnosis, time since RA diagnosis, and drugs used at the time of sample extraction (Suppl. Table 2). Subsequently, we obtained

DNA methylation data and identified 814,043 CpGs that passed our QC pipeline, thus rendering them available for analysis.

The differential methylation analysis revealed that 6679 CpGs qualified as DMPs ($\Delta\beta \geq 2\%$ and $\text{FDR} < 0.05$) between RA-ILD and RA_{controls}, most of them (91.5%) hypermethylated in the RA-ILD samples (Fig. 1A and Suppl. Table 4). The largest differences were $\Delta\beta = 0.18$ among the hypermethylated DMPs and $\Delta\beta = -0.10$ among the hypomethylated DMPs. However, the $\Delta\beta$ of most DMPs was smaller, as is characteristic of complex diseases except cancer (median $\Delta\beta$ of hypermethylated DMPs = 0.039, IQR = 0.028 to 0.059; and median $\Delta\beta$ of hypomethylated DMPs = -0.030, IQR = -0.024 to -0.037). The DMPs were annotated to numerous genes, many of which may be associated with RA-ILD pathogenesis. Some examples include DMPs present in mucins (*MUC6*, *MUC13*, *MUC15*, *MUC22*), collagen (*COL9A3*, *COL2A1*, *COL3A1*, *COL6A3*, and others), telomere maintenance genes (*TERT*, *PARN*, *TERF1*, *POT1*, *WRAP53*), and immune response-related genes (*HLA-DRA*, *VCAM1*, *IL15*, *IL17*, *IL20*, *IL13RA*). Several genes from the above groups that contain multiple DMPs with the same direction of change are presented in Fig. 1B. Many DMPs mapped to the same region (either gene body, gene promoter, or CGI), a circumstance that was reflected in the 576 DMRs with ≥ 5 leading CpGs, mean $\Delta\beta \geq 0.02$ and $\text{FDR} \leq 0.05$ we identified (Fig. 1A, inner circos plot, and Suppl. Table 5). They were predominantly hypermethylated in RA-ILD (551 of the 576), and some DMRs, especially in the gene bodies, included more than 20 CpGs with differential methylation (representative examples in Fig. 1C). In more detail, the median number of differential CpGs was 12 (IQR = 8–18) among a median of 21 CpGs (IQR = 14–36) annotated to the gene bodies reflecting that many DMRs include more than half of the annotated CpGs (median = 55.6%, IQR = 46.6–66.7%).

We estimated the frequencies of 12 immune cell types using deconvolution methods²⁰. These frequencies are necessary to assess whether differences in blood cell subpopulations could affect the identification of DMPs. This possibility was excluded because all the subpopulations showed similar frequencies in the RA-ILD and RA_{controls} (Suppl. Fig. 2 and Suppl. Table 2). Nevertheless, we investigated the possibility of blood subpopulation-specific DMPs with two algorithms (Suppl. Table 6A). The CellDMC algorithm identified a DMP that was hypermethylated in lymphocytes and hypomethylated in monocytes of the RA-ILD patients. The TCA algorithm found three other DMPs as specifically hypermethylated in the granulocytes of RA-ILD patients. Therefore, we considered negligible the impact of the blood cell subpopulations in our DMPs. Additionally, other potential sources of confounding were excluded because the DMPs were not significantly associated with any patient features, including RA disease activity and the time since ILD diagnosis (Suppl. Table 6B and Suppl. Fig. 3).

DMPs and DMRs highlight functions related to lung fibrosis

We used two types of enrichment analysis, overrepresentation analysis (ORA) and functional class scoring (FCS). We performed ORA of the differential results (DMPs on uncorrected and consistent data, and DMRs) considering their genome context (all, promoter, gene body or CGI) and direction of change (all, hypermethylated and hypomethylated) across several gene set libraries (Fig. 2 and Suppl. Table 7). The most notable enrichment was observed with overexpressed genes in the “late fibrosis 5” cluster from a lung fibrosis mouse model²¹. The enrichment was highly significant in multiple analyses: DMPs (up to $p_{\text{adj}} = 1.9 \times 10^{-38}$) and DMRs (up to $p_{\text{adj}} = 1.1 \times 10^{-29}$); all sequences and gene bodies; and all methylation differences and hypermethylated changes. Many genes (594 with DMPs and 144 with DMR) overlapped with the “late fibrosis 5” cluster, which contains overexpressed genes by blood monocytes 19 days after the induction of lung fibrosis with bleomycin²¹. Our DMPs were also significantly overrepresented in another cluster from the lung fibrosis model (up to $p_{\text{adj}} = 7.8 \times 10^{-4}$)²¹. However, this “infiltrating monocytes 2” cluster was not significantly enriched under stricter thresholds (Fig. Suppl. 4 and Suppl. Table 8).

No other ORA result was comparable to the “late fibrosis 5” cluster. The next enriched gene sets showed p_{adj} values in the 10^{-10} level. They included the mitotic spindle gene set and the Rho GTPase pathways. Many of the mitotic spindle genes showed DMPs (up to 75 of the 199 in the MSigDB_Hallmark gene set) and DMRs (18 of the 199) in our RA-ILD patients (Suppl. Table 7). The mitotic spindle mediates chromatid separation to daughter cells during cell division and is involved in IPF as revealed by four susceptibility loci^{22–25}. The Rho GTPases appeared in our ORA enrichment analysis as multiple enriched pathways and gene sets at different levels of significance (Fig. 2 and Suppl. Table 7). The Rho GTPases are a critical component in the organization of the actin cytoskeleton and there is ample evidence of their participation in fibrosis^{26,27}. Other Rho GTPases-related gene sets were overrepresented at the level of p_{adj} of 10^{-5} . Also, the Guanyl-Nucleotide Exchange Factor Activity is highly interconnected with Rho GTPases signaling. Therefore, all these gene sets could be interpreted as part of the larger Rho GTPase pathway. The mitotic spindle and Rho GTPase pathways were also significantly overrepresented under stricter thresholds (Suppl. Fig. 4 and Suppl. Table 8). It is worth noting that the enriched gene sets observed at lower levels of significance did not remain with the stricter thresholds (Suppl. Fig. 4 and Suppl. Table 8).

We then performed FCS as a further validation because differences in the number of CpGs per gene introduce biases in ORA²⁸. FCS corrects for this source of bias but is only available for the GO and Reactome libraries among the four we examined with ORA. The FCS analysis provided detailed results with many significantly enriched gene sets (Suppl. Tables 9–11). Given the abundance of results, we grouped similar gene sets and pathways using overlapping genes (Fig. 3 and Suppl. Fig. 5) and GO terms semantic similarity (Suppl. Fig. 6). The mitotic spindle and Rho GTPases gene sets were significantly enriched, validating the ORA results. The mitotic spindle appeared as part of the largest and most prominent cluster. This cluster contained many mitosis-related Reactome pathways and GO gene sets (Fig. 3 and Suppl. Fig. 5, respectively). They included components of the mitotic spindle, like the kinetochore and separation of sister chromatids, but other aspects of mitosis were also represented, like the cell cycle checkpoints, and processes that overlap with the mitosis nuclear changes, like chromatin organization. The Rho GTPase pathways were also significantly enriched in the FCS analysis (Fig. 3),

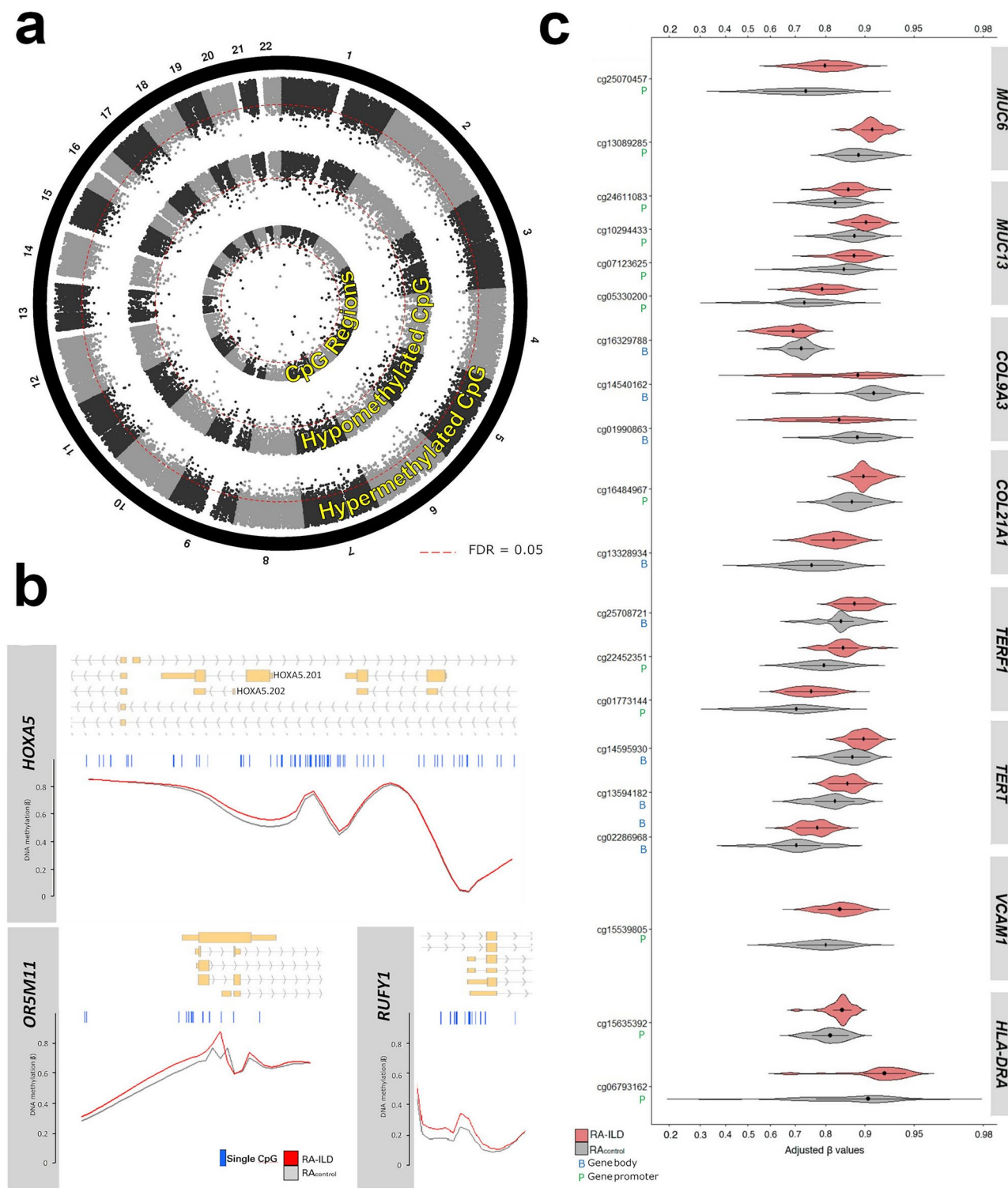


Fig. 1. Differentially methylated features (DMPs and DMRs). **(A)** Circos plot representing the features as dots in their schematic position on chromosomes (alternating gray and black colors and numbered in the outermost circle) with the corresponding $-\log(p\text{-value})$ in the y-axis toward the circle center (those over the red scattered lines are below FDR 0.05). The outer features are hypermethylated CpG, the middle circle represents hypomethylated CpG and the inner circle shows the DMR. **(B)** Violin plots of DNA methylation β values corresponding to DMP in selected genes. The x-axis shows the β values adjusted for sex, age, smoking, and anti-CCP in the logit scale. The right y-axis shows the CpG code and its localization in the promoter (P) or gene body (B) of the selected gene (indicated in the left y-axis). Red violin plots correspond to RA-ILD and gray plots to RA_{controls}, respectively. **(C)** Line plots of DNA methylation β values (y-axis) in selected DMR. The blue bars over the lines represent CpG included in the EPIC array. The upper part shows the gene-specific tracks corresponding to different splice isoforms.

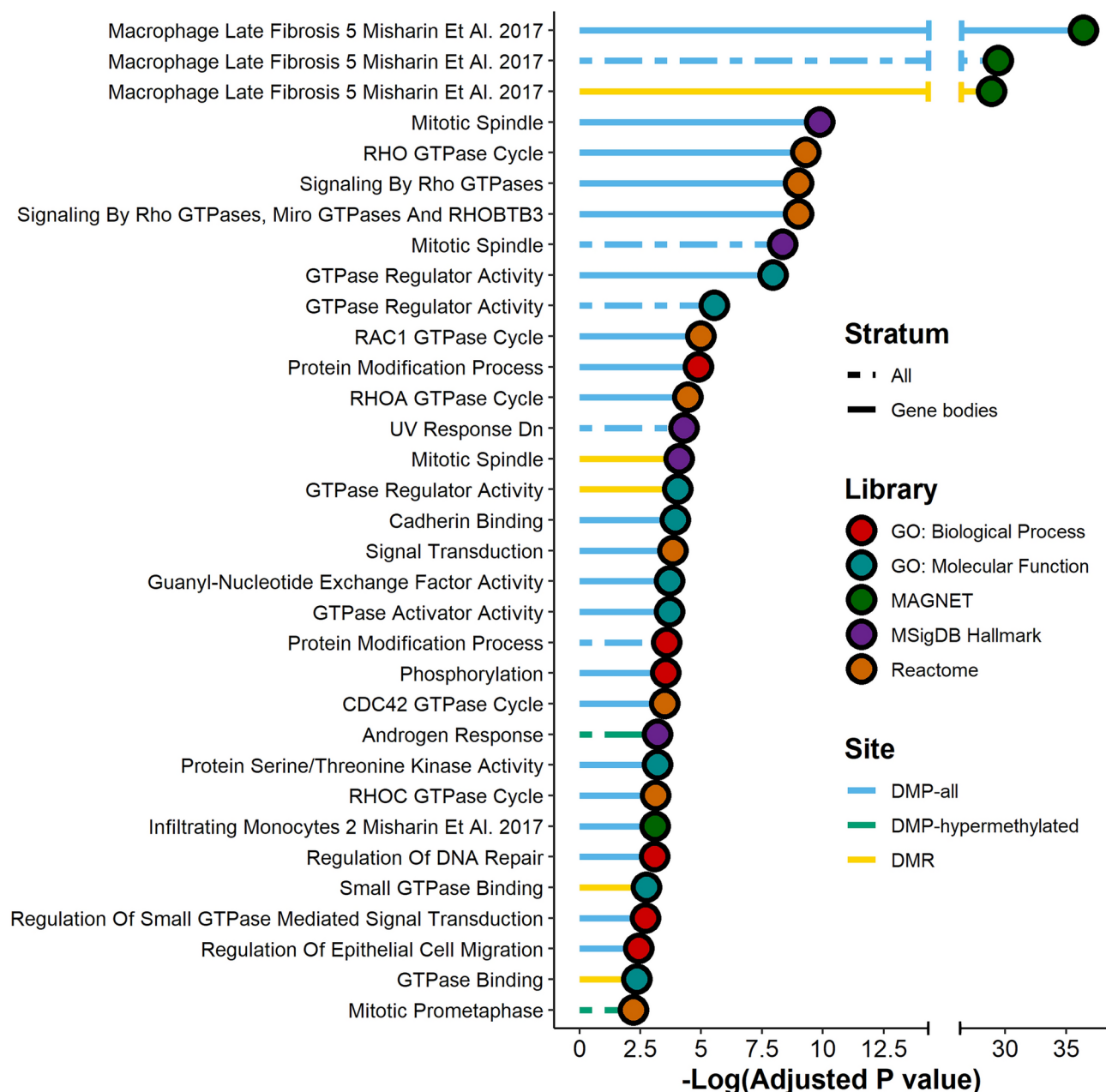


Fig. 2. Enriched gene sets and pathways in differentially methylated features (DMPs and DMRs). This exploratory ORA was done with *Enrichr*. We show the most informative gene sets in increasing order of p values. The legend indicates with the circle color the library where the enriched gene set was found (none in the GO: Cellular Compartment library), with the line type the genome stratum (none in the promoters and CGI), and with the line color the DMP or DMR feature and the type of methylation change (none with the hypomethylated features). The complete results of the analysis are provided in Suppl. Table 7.

together with multiple pathways and gene sets related to intracellular vesicle traffic and location (Fig. 3 and Suppl. Fig. 5), which are linked to Rho GTPase regulation. In addition, the FCS analysis identified other gene sets that were not in the ORA results (Fig. 3).

Significant enrichment of liver X receptor element (LXRE) among the TFBS

We searched for enriched TFBS in 500 and 200 bp windows around the DMPs. The search included subanalyses focused on promoters and enhancers (Fig. 4 and Suppl. Table 12). We used the most restricted analyses to increase the specificity of the enrichment, but the reduction of sequences decreased the sensitivity of the tests. Despite this difficulty, three motifs showed consistent enrichment across the four subanalyses, LXRE, GLIS3, and GSC (Fig. 4). The most striking was the liver X response element (LXRE) motif that binds liver X receptor α (LXR α) and β (LXR β) and have an established role in fibrosis^{29–31}. It was highly significantly enriched (all subanalyses

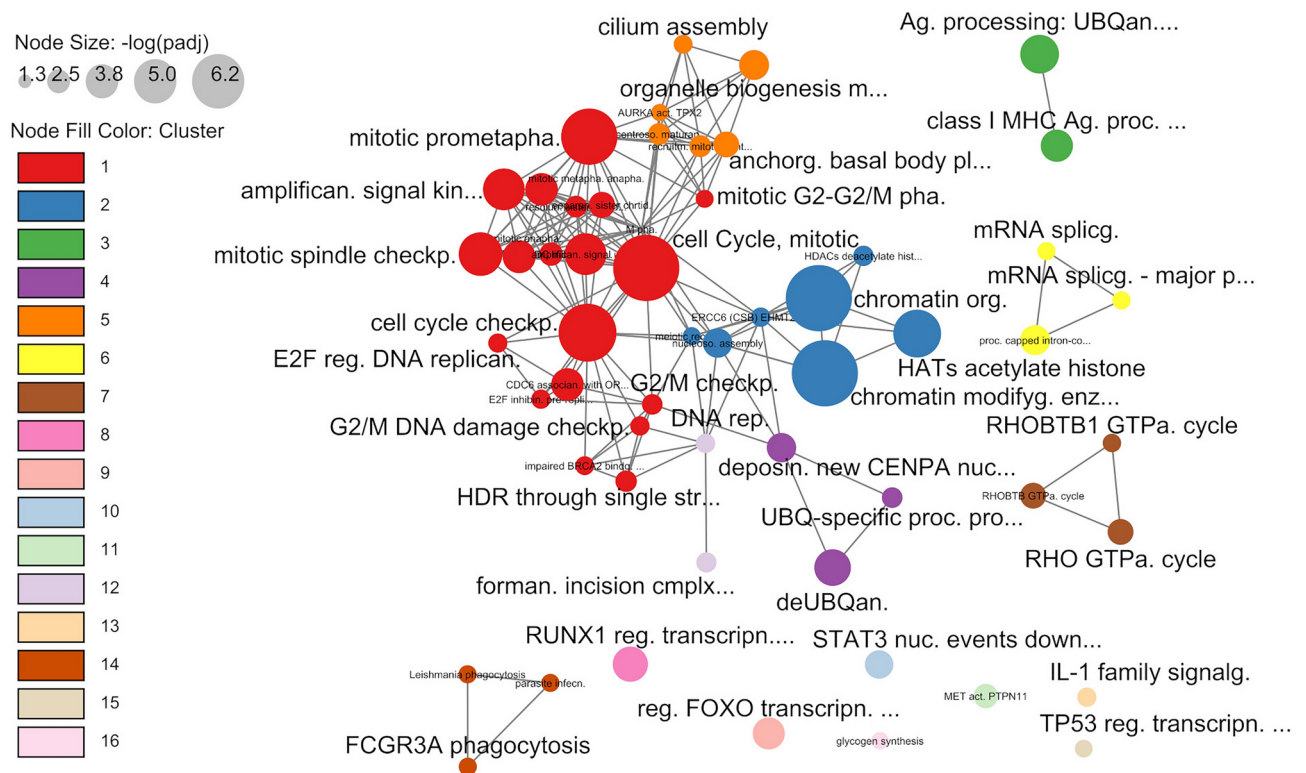


Fig. 3. Network representation of the enriched Reactome pathways in DMPs mapping to gene bodies. Each pathway and gene set corresponds to a node with a size proportional to the enrichment significance ($-\log(p_{adj})$) and color for the cluster. The edges represent shared genes between gene sets. We shortened the labels (detailed correspondence in Suppl. Table 9).

$p \leq 10^{-8}$), and its relative abundance increased with the focus on the small windows and the promoters (fold enrichment from 1.99 in the 500 bp window on all sequences to 3.65 in the 200 bp window on promoters). Perhaps in relation to this finding, we observed hypermethylated DMPs in the gene bodies of *ABCA1*, *ABCG1*, and *ABCG5*, which are important targets of LXRE regulation of cholesterol in multiple tissues^{32,33}. However, we did not find DMPs in the genes coding for the LXR β and LXR α proteins and did not detect changes in any of these five genes in RNAseq analysis of the blood cells.

Discussion

Our exploratory study has uncovered blood DNA methylation biomarkers for RA-ILD. These included a considerable number of mostly hypermethylated DMPs and DMRs. These changes significantly overlapped with genes overexpressed by monocytes in a mouse model of lung fibrosis²¹. Additionally, DMPs and DMRs indicated dysfunction in processes of broad relevance, such as Rho GTPase regulation and mitosis. The DMPs were also specifically enriched near the LXRE motif. Each of these findings can be interpreted to reflect processes linked to fibrosis in past studies but taking care to avoid inferring causal relationships.

These results have been possible because DNA methylation reflects cellular processes with high sensitivity^{6,8}. Illustrative examples are the DMPs distinguishing synovial fibroblasts from different joints^{11,34} and the age-sensitive CpG sites that are included in the epigenetic clocks³⁵. We think that patient selection was another component contributing to the sensitivity of our study. It involved the recruiting of recently diagnosed RA-ILD patients to reduce the heterogeneity of disease evolutions; and the use of matched RA_{controls} to avoid confounding. The abundance of DMPs and the lack of association of the DMPs with demographic and clinical features other than ILD indicate these patient selection criteria have been successful. We think this experience could serve to stimulate similar studies in other CTD-ILD where a need for prediction and early diagnosis biomarkers also exists.

The most significant enriched gene set corresponds to a cluster of overexpressed genes in a lung fibrosis model²¹. These genes were overexpressed in the blood monocytes, and less markedly in monocyte-derived alveolar macrophages and other lung macrophages. The genes were described as associated with epigenetic regulation late in fibrosis but without any detailed characterization except for the absence of fibrosis biosynthetic pathways²¹. The absence of pathways like the TGF β , Wnt- β catenin, and PDGF pathways, was also a feature of our findings. Therefore, the enrichment in these genes indicates that many of our DMPs are associated with lung fibrosis irrespective of the trigger, bleomycin in the experimental model, and RA in our study. This interpretation is consistent with the generality of many fibrosis mechanisms that are shared across tissues and diseases^{3,36}.

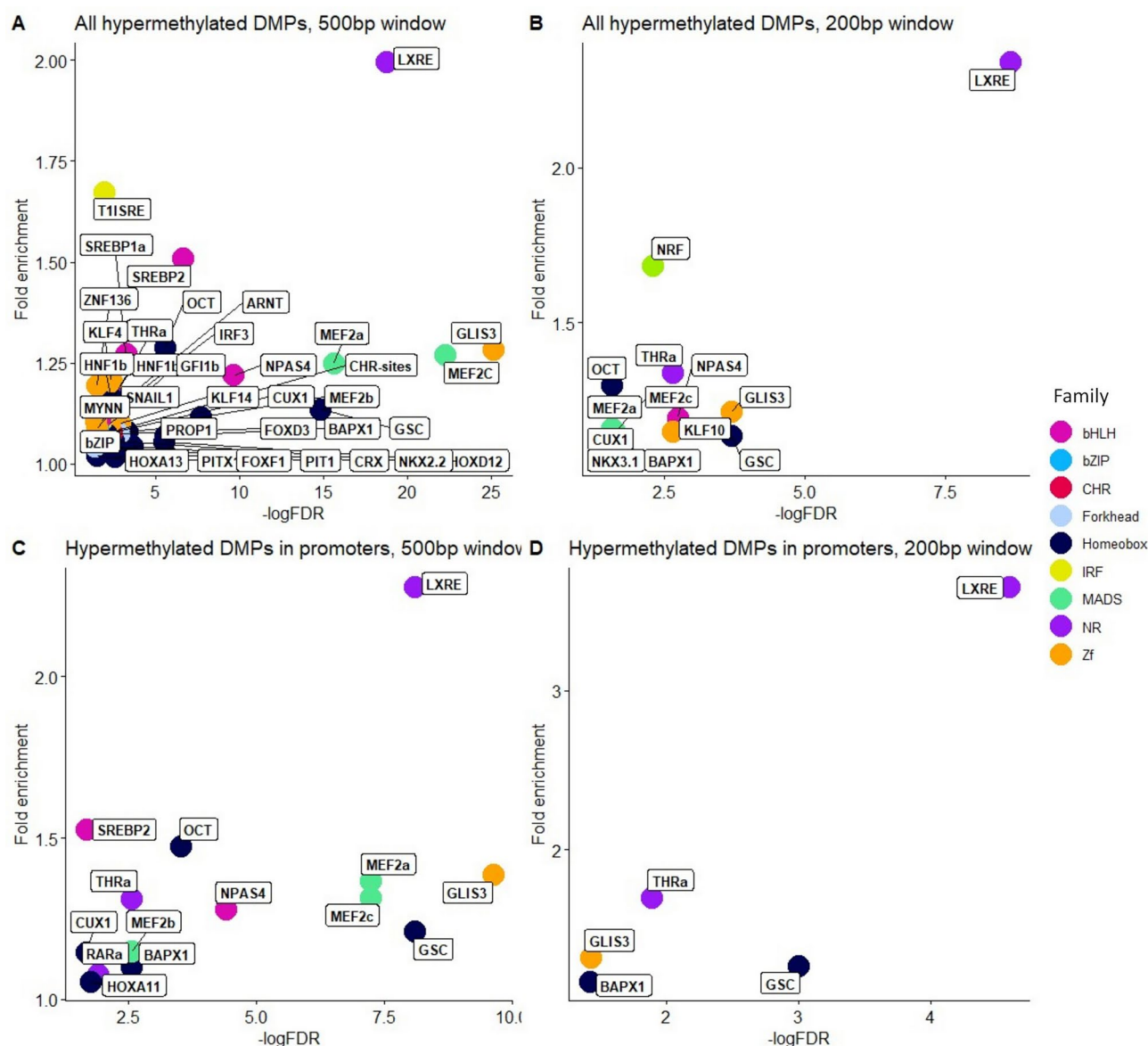


Fig. 4. Representation of the transcription factor binding sites (TFBS) enriched around the DMPs. We searched enriched motifs in the 500 bp (A,B) and 200 bp (C,D) centered on the DMPs and considering either all the DMPs (A,C) or only the DMPs annotated to gene promoters (B,D). The x-axis shows the $-\log(p)$ of enrichment, whereas the y-axis represents the fold enrichment of the motifs.

The other enriched gene sets also evidence the relationship of our findings with lung fibrosis. The processes that contribute to nuclear division in the M phase of mitosis and are related to the mitotic spindle are necessary for all dividing cells. However, they have a singular relevance for lung fibrosis because aberrant activation of the TP53 pathway contributes to the AEC senescence and abnormal phenotype in the fibrotic lung^{37,38}. Also, common genetic variants in *TP53* and *CDKN1A*, a cell cycle regulation gene, increase susceptibility to IPF³. Aside from the general involvement of mitosis, the particular role of the mitotic spindle is supported by four IPF susceptibility loci that reveal the special relevance of this process: *KNL1* and *SPDL1*, two components of the kinetochore, *MAD1L1*, a member of the mitotic spindle assembly complex, and *KIF15*, a kinesin involved in spindle separation^{22–25}. All these factors likely contribute to the dysregulated wound repair response and compromised regenerative capacity of lung epithelial cells³. Also, the enriched Rho GTPase pathways are related to aberrant wound healing and lung fibrosis^{26,27}. These pathways have a particular impact on the regulation of the actin cytoskeleton, which is essential for fibroblast migration and lung contraction, but they contribute to the pathogenesis of lung fibrosis in other ways too. For example, the Rho GTPases are involved in fibroblast activation and extracellular matrix production, mediation of profibrotic signals, and activation of downstream effectors, most notably ROCK, to initiate profibrotic cellular responses^{26,27,39,40}. Some of these pathways are targets of the antifibrotic drugs nintedanib and pirfenidone^{26,39}.

Furthermore, the regulation of the actin cytoskeleton by Rho GTPases connects to mitosis and intracellular traffic and location. The interaction between Rho GTPases, mitosis, and the mitotic spindle is multifaceted and critical for the correct sequential order of mitotic stages⁴¹. For example, CDC42 promotes spindle assembly during early mitosis, while RhoA plays a crucial role during cytokinesis⁴¹. Rho GTPases are involved in various mitosis processes: cell cortex stiffening, mitotic spindle formation, attachment of the spindle microtubules to the kinetochore, formation of the division plane, the contractile ring, membrane ingression, and abscission⁴¹.

However, the enrichment of DMPs around LXRE does not have clear connections with mitosis or the Rho GTPases implying that its role is independent of the already considered pathways. In effect, the LXR ligands and LXRE activation are mainly involved in cholesterol regulation, although they also have well-known roles in the control of inflammation and fibrosis^{29–31}. The mechanisms by which LXR prevents fibrosis depend on the tissue and cell type, with different targets in infiltrating macrophages, fibroblasts, and alveolar epithelial cells in the lung^{29–31}. However, all tissues share the activation of LXR in response to the overproduction of oxysterols. These oxidation cholesterol metabolites are increased in RA as part of the associated lipid changes and cardiovascular risk⁴². Therefore, the enrichment around LXRE suggests a link between cholesterol metabolism and RA-ILD that will require future investigation.

Despite the interesting results, we acknowledge some limitations in our exploratory study. Perhaps, the most fundamental weakness is the retrospective design. However, this limitation is difficult to avoid due to the rare and unpredictable onset of RA-ILD¹. Additionally, the nature of the study limits the potential utility of the biomarkers to RA-ILD prediction or early diagnosis, but not to prognosis or differential diagnosis with other ILD. A more general limitation of the EWAS is the complex relationship between differences in DNA methylation and gene expression^{6–9}. This complexity is evidenced by the small number of DNA methylation quantitative trait loci showing a detectable causal relationship with the associated phenotypes⁴³. In addition, the nature of the relationships is unclear as the DNA methylation changes can be mediators, modifiers, or consequences of the phenotypes^{6–9}. Notable examples of the latter are the DMPs associated with obesity that are the consequence of adiposity, rather than its cause⁴⁴ and the hypermethylation of CGI promoters that are a byproduct of the rapid cellular proliferation of malignant myeloid cells⁴⁵. As such, the DMPs cannot be assumed to have an important effect on gene expression and should only be regarded as biomarkers of underlying processes until more evidence is obtained. Another shortcoming comes from the limited understanding of many DMPs, especially if they are located in genes with unknown functions or do not associate with genes. This incomplete knowledge is exemplified by Reactome, one of the largest pathway libraries but only contains information on about half of the human protein-coding genes (<https://reactome.org/about/statistics>). Therefore, we could have missed important biological processes for RA-ILD that are still poorly covered in the gene set and pathway libraries.

In conclusion, our exploratory study shows that biomarkers associated with RA-ILD can be obtained by studying blood DNA methylation. Many of these biomarkers point to molecular or cellular processes relevant to lung fibrosis. Further research is needed to replicate the findings and determine if the identified DNA methylation biomarkers have clinical value for prediction or early diagnosis.

Methods

Patient recruitment and ethical requirements

We studied patients with RA according to the 2010 ACR/EULAR classification criteria⁴⁶ recruited in collaboration with 13 hospitals from Spain and Portugal (Suppl. Table 1). The recruitment process followed a pre-specified protocol to ensure criteria homogeneity. Accordingly, all patients were older than 18 years at inclusion, had undergone high-resolution computed tomography (HRCT) at their recruitment hospitals, and had provided informed consent to participate. Other rheumatic and respiratory diseases, as well as professional exposure to pro-fibrotic substances, were considered exclusion causes. The criteria for inclusion as RA-ILD comprise the presence of an interstitial pneumonia pattern in HRCT that is directly attributable to RA while excluding any other potential causes. Among the recruited patients, we selected RA-ILD patients diagnosed with ILD less than a year before sampling. In turn, the controls (RA_{controls}) were patients with RA showing no ILD signs in HRCT and matched to the patients with RA-ILD. The matching variables were sex, smoking status (never or ever), anti-cyclic citrullinated peptide antibodies (anti-CCP, also named ACPA) status, and age, which are known RA-ILD risk factors¹. These variables were employed to make cases and controls identical in the qualitative factors and the closest possible in age. This was achieved with the coarsened exact matching technique incorporated in the *MatchIt* R package⁴⁷.

The study was approved by the ethics committee of Santiago-Lugo (code 2019/332) and by the ethics committee of each participating center (Suppl. Table 1). Besides, all research protocols followed current legislation (Spanish Laws 14/2007 of Biomedical Research and 3/2018 on the Protection of Personal Data and Guarantee of Digital Rights) and the ethical guidelines of the Declaration of Helsinki and the Belmont Report.

Analysis of differential DNA methylation

We kept blood collected in EDTA tubes at -80°C until we extracted DNA using the NucleoSpin® Blood L kit (Macherey–Nagel). Subsequently, we checked DNA in a NanoDrop One (ThermoScientific) and performed bisulfite DNA conversion using the EZ DNA Methylation kit (Zymo Research) and 300 ng of DNA. The samples were hybridized in Infinium Methylation EPIC v1.0 BeadChip (Illumina) arrays with cases and controls equally distributed to prevent biases caused by discrepancies between arrays. Consequently, we obtained the methylation of over 850,000 CpG on the BeadArray Reader (Illumina) as fluorescence signals. The preprocessing and analysis of these results were done with the R application *ShinyÉPICO*⁴⁸ which implements *minfi*^{49,50}, *limma*⁵¹, and *mCSEA*⁵². The workflow involved the transformation of raw signals into β and M values (logit transformed β values), exclusion of CpGs with detection p-values >0.1 , overlapping with SNPs according to the Illumina annotation (MAF >0), or in the X and Y chromosomes, and normalization. In more detail, normalization was

done with *Noob*, which consists of within-array background correction and dye-bias normalization. No batch effect correction was required as the data were generated in a single batch. However, we tested the possibility of batch effects associated with the slides within the single chip using the ComBat function of the *sva* R package. We ran ComBat with the default option involving parametric empirical Bayesian adjustments for mean and variance. Three different settings were assayed, including only the batch variable (-batch), batch and the group (RA-ILD or RA-control) variable (-group), and with the addition of the four covariates used for patient matching (-covariates). The three settings led to DMPs that were inconsistent overall but were consistent in the subgroup of DMPs that overlapped with those identified in the uncorrected data (Suppl. Table 3). Therefore, we used the uncorrected data for analysis except for labeling the DMPs regarding the ComBat corrections (Suppl. Table 4) and performing an ORA enrichment analysis using only the overlapping DMPs (Suppl. Table 8). Regarding the contrast analysis, we used M values for model construction, DMP, and DMR detection. The model incorporated the covariates utilized for group matching (sex, age, smoking status, and anti-CCP status). An empirical Bayes-moderated t-test (*limma*)⁵¹ with a False Discovery Rate (FDR) ≤ 0.05 and an absolute β difference ≥ 0.02 was used to identify DMPs⁴⁸. The three options, ArrayWeights, Trend, and Robust, were selected to enhance the accuracy and reproducibility of the findings⁵¹. We checked the number of beads per probe: 12,246 probes had less than 3 beads in 3 or more samples, however, none were present in the detected DMPs. In turn, the DMRs were detected with the *mCSEA* method that employs set enrichment analysis in predefined regions and separately in promoters, gene bodies, and CpG islands⁵². We only report DMR with FDR ≤ 0.05 , a mean $\Delta\beta$ of the leading-edge CpGs ≥ 0.02 , and at least five leading-edge CpGs, although *mCSEA* default settings do not filter by $\Delta\beta$ and accept two leading-edge CpGs. We needed these restrictive thresholds to avoid DMRs without DMPs. Besides, we used other stricter thresholds to verify the results' robustness. In these verifications, DMPs required an absolute $\Delta\beta \geq 0.05$ (FDR as above), and DMRs needed at least ten leading-edge CpGs (FDR and mean $\Delta\beta$ as above).

Cell type deconvolution of DNA methylation

We estimated the frequency of blood cell subpopulations from CpG methylation data with the FlowSorted.Blood.EPIC package⁵³ using the reference matrix for 12 blood cell subpopulations²⁰. This package implements a modified version of the constrained projection/quadratic programming algorithm that is used to estimate the frequency of T regulatory cells, natural killer cells, neutrophils, monocytes, eosinophils, CD8⁺ naïve cells, CD8⁺ memory cells, CD4⁺ naïve cells, CD4⁺ memory cells, B naïve cells, B memory cells, and basophils in each patient. We then compared the raw and centered log-ratio transformed frequencies between RA-ILD and RA_{controls} using a Mann–Whitney U test per each cell type. The raw frequencies are common in the bibliography, but log-ratio transformations are required for correct compositional analysis⁵⁴. An additional analysis examined whether the DMPs could be cell-type-specific using two methods: the CellDMC function of the EpiDISH R package⁵⁵ and a tensor analysis tool, TCA⁵⁶. We restricted this analysis to the three major blood cell lineages (lymphocytes, granulocytes, and monocytes) because of the low percentages represented by more detailed subpopulations.

Gene set enrichment analysis

We employed overrepresentation analysis (ORA) for DMRs and DMPs. Additionally, we used functional class scoring (FCS) for all the ranked CpGs. ORA was done with default settings on *Enrichr*, a tool that performs a uniform analysis across a wide range of gene set libraries⁵⁷. The analysis was done separately for DMPs and DMRs and stratified by CpG islands (CGI), gene promoters, and gene bodies; and according to their hyper- or hypo-methylated change. All these factors reflect diverse DNA methylation processes. We report only the statistically significant gene sets (FDR ≤ 0.05) from the most informative libraries. Additionally, we used FCS with the *methyglm* tool of *methyGSA*²⁸. This tool uses logistic regression to adjust for the different number of CpGs annotated to each gene. The gene sets were taken from Gene Ontology via org.Hs.eg.db annotation (Bioconductor v3.16, last accession on December 9th 2022) and Reactome via reactome.db annotation (Bioconductor v3.16, last accession on December 9th, 2022). These gene sets reflect the complexity of biology showing a hierarchical and overlapping structure. Therefore, we restricted the size of gene sets to the 5 to 500 genes range and grouped the outcomes based on semantic similarity and shared genes. For semantic similarity, we used the binary cut method (cutoff at 0.7) implemented in *simplifyEnrichment*⁵⁸. This tool provides a heatmap of the enriched GO gene sets, labeled with words appearing frequently in the GO definitions. In addition, we used *GOMCL*⁵⁹ for the grouping based on shared genes (with overlap coefficients and a clustering threshold of 0.5). This tool identifies networks within the FCS results using the Markov Clustering algorithm. We added three gene sets larger than 500 genes to act as parents in cluster formation, but they are not shown. Finally, Cytoscape 3.10 was used to represent the networks (<https://cytoscape.org/>). Our focus was on the gene sets most significantly enriched, concordant across the diverse enrichment analyses, and including a larger number of DMRs or DMPs.

Enrichment of transcription factor binding sites

We investigated the enrichment of transcription factor binding sites (TFBS) around DMPs with HOMER (version homer2, v4.11)⁶⁰ using the “known motifs”, which are derived from ChIP-Seq experiments. We verified the robustness of our results by doing the analysis six times with an increasingly strict selection of the windows around the DMPs. In this way, we considered two window sizes around DMPs (500 and 200 bp) and an unrestricted analysis for “any sequence” or restricted to DMPs in promoters or enhancers. The promoters were taken from the Illumina annotation, whereas the enhancers were taken from FANTOM5 (<https://fantom.gsc.riken.jp/5/>). However, we do not present the enhancer analysis because the sequences were too few to be meaningful. The background sequences for comparison were all QC-filtered and preprocessed regions of the same size around the remaining CpGs in the EPIC array. Only results with FDR ≤ 0.05 are reported.

Data availability

The DNA methylation dataset supporting the conclusions of this article is deposited in the Gene Expression Omnibus database under accession number GSE275597 and is freely available at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE275597>. The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 18 October 2024; Accepted: 22 April 2025

Published online: 29 April 2025

References

- Koduri, G. & Solomon, J. J. Identification, monitoring, and management of rheumatoid arthritis-associated interstitial lung disease. *Arthritis Rheumatol.* **75**(12), 2067–2077 (2023).
- Matson S, Lee J, Eickelberg O: Two sides of the same coin? A review of the similarities and differences between idiopathic pulmonary fibrosis and rheumatoid arthritis-associated interstitial lung disease. *Eur Respir J* 2021, 57(5).
- Moss, B. J., Ryter, S. W. & Rosas, I. O. Pathogenic mechanisms underlying idiopathic pulmonary fibrosis. *Annu. Rev. Pathol.* **17**, 515–546 (2022).
- Juge, P. A. et al. Shared genetic predisposition in rheumatoid arthritis-interstitial lung disease and familial pulmonary fibrosis. *Eur. Respir. J.* **49**(5), 1602314 (2017).
- Juge, P. A. et al. MUC5B promoter variant and rheumatoid arthritis with interstitial lung disease. *N. Engl. J. Med.* **379**(23), 2209–2219 (2018).
- Ehrlich, M. DNA hypermethylation in disease: Mechanisms and clinical relevance. *Epigenetics* **14**(12), 1141–1163 (2019).
- Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* **20**(10), 590–607 (2019).
- Jin, Z. & Liu, Y. DNA methylation in human diseases. *Genes Dis.* **5**(1), 1–8 (2018).
- Leenen, F. A., Muller, C. P. & Turner, J. D. DNA methylation: Conducting the orchestra from exposure to phenotype?. *Clin Epigenet.* **8**(1), 92 (2016).
- Ospelt, C., Gay, S. & Klein, K. Epigenetics in the pathogenesis of RA. *Semin. Immunopathol.* **39**(4), 409–419 (2017).
- Ai, R. et al. Joint-specific DNA methylation and transcriptome signatures in rheumatoid arthritis identify distinct pathogenic processes. *Nat. Commun.* **7**(1), 11849 (2016).
- Nakano, K., Whitaker, J. W., Boyle, D. L., Wang, W. & Firestein, G. S. DNA methylome signature in rheumatoid arthritis. *Ann. Rheum. Dis.* **72**(1), 110–117 (2013).
- Karouzakis, E. et al. Analysis of early changes in DNA methylation in synovial fibroblasts of RA patients before diagnosis. *Sci. Rep.* **8**(1), 7370 (2018).
- de la Calle-Fabregat, C. et al. The synovial and blood monocyte DNA methylomes mirror prognosis, evolution, and treatment in early arthritis. *JCI Insight* **7**(9), e158783 (2022).
- Rodriguez-Ubreva, J. et al. Inflammatory cytokines shape a changing DNA methylome in monocytes mirroring disease activity in rheumatoid arthritis. *Ann. Rheum. Dis.* **78**(11), 1505–1516 (2019).
- de la Calle-Fabregat, C. et al. Prediction of the progression of undifferentiated arthritis to rheumatoid arthritis using DNA methylation profiling. *Arthritis Rheumatol.* **73**(12), 2229–2239 (2021).
- Adams, C. et al. Identification of cell-specific differential DNA methylation associated with methotrexate treatment response in rheumatoid arthritis. *Arthritis Rheumatol.* **75**(7), 1088–1097 (2023).
- Zhu, H. et al. Rheumatoid arthritis-associated DNA methylation sites in peripheral blood mononuclear cells. *Ann. Rheum. Dis.* **78**(1), 36–42 (2019).
- Liu, Y. et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* **31**(2), 142–147 (2013).
- Salas, L. A. et al. Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. *Nat. Commun.* **13**(1), 761 (2022).
- Misharin, A. V. et al. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J. Exp. Med.* **214**(8), 2387–2404 (2017).
- Allen, R. J. et al. Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **201**(5), 564–574 (2020).
- Allen, R. J. et al. Genome-wide association study across five cohorts identifies five novel loci associated with idiopathic pulmonary fibrosis. *Thorax* **77**(8), 829–833 (2022).
- Peljo, A. L. et al. Idiopathic pulmonary fibrosis is associated with common genetic variants and limited rare variants. *Am. J. Respir. Crit. Care Med.* **207**(9), 1194–1202 (2023).
- Zhang, D. et al. Rare and common variants in KIF15 contribute to genetic risk of idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **206**(1), 56–69 (2022).
- Monaghan-Benson, E., Wittchen, E. S., Doerschuk, C. M. & Burridge, K. A Rnd3/p190RhoGAP pathway regulates RhoA activity in idiopathic pulmonary fibrosis fibroblasts. *Mol. Biol. Cell* **29**(18), 2165–2175 (2018).
- Tsou, P. S., Haak, A. J., Khanna, D. & Neubig, R. R. Cellular mechanisms of tissue fibrosis. 8. Current and future drug targets in fibrosis: Focus on Rho GTPase-regulated gene transcription. *Am. J. Physiol. Cell Physiol.* **307**(1), C2–13 (2014).
- Ren, X. & Kuan, P. F. methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing. *Bioinformatics* **35**(11), 1958–1959 (2019).
- Beyer, C. et al. Activation of liver X receptors inhibits experimental fibrosis by interfering with interleukin-6 release from macrophages. *Ann. Rheum. Dis.* **74**(6), 1317–1324 (2015).
- Hernandez-Hernandez, I. et al. Endogenous LXR signaling controls pulmonary surfactant homeostasis and prevents lung inflammation. *Cell. Mol. Life Sci.* **81**(1), 287 (2024).
- Shichino, S. et al. Transcriptome network analysis identifies protective role of the LXR/SREBP-1c axis in murine pulmonary fibrosis. *JCI Insight* **4**(1), e122163 (2019).
- Ito, A. et al. LXRs link metabolism to inflammation through Abca1-dependent regulation of membrane composition and TLR signaling. *Elife* **4**, e08009 (2015).
- Zhang, R., Wuersch, E., Yong, V. W. & Xue, M. LXR agonism for CNS diseases: Promises and challenges. *J. Neuroinflamm.* **21**(1), 97 (2024).
- Frank-Bertoncelj, M. et al. Epigenetically-driven anatomical diversity of synovial fibroblasts guides joint-specific fibroblast functions. *Nat. Commun.* **8**, 14852 (2017).
- Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**(6), 371–384 (2018).
- Rockey, D. C., Bell, P. D. & Hill, J. A. Fibrosis—a common pathway to organ injury and failure. *N. Engl. J. Med.* **372**(12), 1138–1149 (2015).

37. Xu, Y. et al. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* **1**(20), e90558 (2016).
38. DePianto, D. J. et al. Molecular mapping of interstitial lung disease reveals a phenotypically distinct senescent basal epithelial cell population. *JCI Insight* **6**(8), e143626 (2021).
39. Nakamura, Y. et al. A protective effect of pirfenidone in lung fibroblast-endothelial cell network via inhibition of Rho-Kinase activity. *Biomedicine* **11**(8), 2259 (2023).
40. Knipe, R. S. et al. The Rho kinase isoforms ROCK1 and ROCK2 each contribute to the development of experimental pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **58**(4), 471–481 (2018).
41. Chircop, M. Rho GTPases as regulators of mitosis and cytokinesis in mammalian cells. *Small GTPases* **5**, e29770 (2014).
42. Bilotta, M. T., Petillo, S., Santoni, A. & Cipitelli, M. Liver X receptors: Regulators of cholesterol metabolism, inflammation, autoimmunity, and cancer. *Front. Immunol.* **11**, 584303 (2020).
43. Min, J. L. et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat. Genet.* **53**(9), 1311–1321 (2021).
44. Wahl, S. et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541**(7635), 81–86 (2017).
45. Spencer, D. H. et al. CpG Island hypermethylation mediated by DNMT3A is a consequence of AML progression. *Cell* **168**(5), 801–816 e813 (2017).
46. Aletaha, D. et al. 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum.* **62**(9), 2569–2581 (2010).
47. Ho DE, Imai K, King G, Stuart EA: MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software* 2011, 42(8).
48. Morante-Palacios, O. Interactive DNA methylation array analysis with ShinyEPICo. *Methods Mol Biol* **2624**, 7–18 (2023).
49. Aryee, M. J. et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**(10), 1363–1369 (2014).
50. Fortin, J. P., Triche, T. J. Jr. & Hansen, K. D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**(4), 558–560 (2017).
51. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**(7), e47 (2015).
52. Martorell-Marugan, J., Gonzalez-Rumayor, V. & Carmona-Saez, P. mCSEA: Detecting subtle differentially methylated regions. *Bioinformatics* **35**(18), 3257–3262 (2019).
53. Salas, L. A. et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol.* **19**(1), 64 (2018).
54. Greenacre, M. Compositional data analysis. *Annu. Rev. Stat. Appl.* **8**(1), 271–299 (2021).
55. Zheng, S. C., Breeze, C. E., Beck, S. & Teschendorff, A. E. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat. Methods* **15**(12), 1059–1066 (2018).
56. Rahmani, E. et al. Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology. *Nat. Commun.* **10**(1), 3417 (2019).
57. Kuleshov, M. V. et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**(W1), W90–97 (2016).
58. Gu, Z. & Hubschmann, D. simplifyEnrichment: A bioconductor package for clustering and visualizing functional enrichment results. *Genom. Proteom. Bioinform.* **21**(1), 190–202 (2023).
59. Wang, G., Oh, D. H. & Dassanayake, M. GOMCL: A toolkit to cluster, evaluate, and extract non-redundant associations of Gene Ontology-based functions. *BMC Bioinform.* **21**(1), 139 (2020).
60. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**(4), 576–589 (2010).

Author contributions

BK, CC-F, EB and AG contributed to the conception of the work, ACD, NM-V, AF-N, AM-V, AE-C, PV-C, JN, JLP, RS, MGB, MAG-G, LA, LM-S, CM, EP-P and AG contributed to the design of the MARILD network; BK, CC-F, ACD, NM-V, AF-N, AT-M, SC, RD-SS, AM-V, CL-P, AE-C, PV-C, MM, JN, MR-G, JLP, JCS-M, RS, LG-C, MGB, SR-M, MAG-G, VL-F, NP-G, CV-F, LA, IC-S, LM-S, ACC-G, CM, and EP-P contributed to the acquisition of data; BK, CC-F, AC, EB and AG contributed to the analysis of the results; all authors contributed the interpretation of data and have approved the submitted version. In addition, all authors have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Funding

Funding for this study was provided by the Instituto de Salud Carlos III (ISCIII, Spain) through grants PI23/00841, PI20/01268 and RD21/0002/0003, co-funded by the European Union by the ERDF “A way for making Europe” and Next Generation EU/PRTR, respectively.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

The study was approved by the ethics committee of Santiago-Lugo (code 2019/332) and by the ethics committee of each participating center (Suppl. Table 1).

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-99755-6>.

Correspondence and requests for materials should be addressed to A.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

On behalf of the MARILD network

Bartosz Kaczmarczyk¹, Adrian Conde¹, Yolanda Lopez-Golan¹, Angeles Villaronga-Torres¹, Raquel Dos-Santos Sobrin¹, Antonio Mera-Varela^{1,22}, Eva Perez-Pampin^{1,22}, Antonio Gonzalez¹, Ana Catarina Duarte³, Natalia Mena-Vazquez⁴, Antonio Fernandez-Nebro⁴, Ana Triguero-Martinez⁵, Santos Castañeda⁵, Chary Lopez-Pedrer⁶, Alejandro Escudero-Contreras⁶, Rocio Caño⁷, Silvia Gomez-Sabater⁷, Paloma Vela-Casasempere⁷, Maria Molina⁸, Javier Narvaez⁹, Miriam Retuerto-Guerrero¹⁰, María Martin-Lopez¹⁰, Jose L. Pablos¹⁰, Juan C. Sarmiento-Monroy¹¹, Raúl Castellanos¹¹, Raimon Sanmarti¹¹, Luis Gomez-Carrera¹², Ma Gema Bonilla¹³, Sara Remuzgo-Martinez¹⁴, Raquel Lopez-Mejias¹⁴, Miguel Angel Gonzalez-Gay^{15,16}, Virginia Leiro-Fernandez¹⁷, Nair Perez-Gomez¹⁸, Cristina Vadillo-Font¹⁹, Arkaitz Mucientes-Ruiz¹⁹, Lydia Abasolo¹⁹, Ivette Casafont-Sole²⁰, Susana Holgado²⁰, Lourdes Mateo-Soria²⁰, Ana Cristina Castillo-Gonzalez²¹ & Carlos Marras²¹