Pers Jacques-Olivier (Orcid ID: 0000-0001-7287-5541)

Machine learning identifies a common signature for anti-SSA/Ro60 antibody expression across autoimmune diseases

Nathan Foulquier[1] (PhD), Christelle Le Dantec[1] (PhD), Eleonore Bettacchioli[1] (PharmD), Christophe Jamin[1,2] (PhD), PRECISESADS Clinical Consortium[3], PRECISESADS Flow Cytometry Consortium[3], Marta E. Alarcón-Riquelme[4] (MD, PhD), and Jacques-Olivier Pers[1,2,*] (DDS, PhD)

[1]LBAI, UMR1227, Univ Brest, Inserm, Labex IGO, Brest, France.

[2]CHU de Brest, Brest, France.

[3]List of members of the PRECISESADS consortia and their affiliations appear at the end of the paper.

[4]Department of Medical Genomics, Center for Genomics and Oncological Research (GENYO), Granada, Spain.

*Correspondence: Pers Jacques-Olivier. Univ Brest, Inserm, LBAI, UMR1227, Brest, France. Email: pers@univ-brest.fr

Word count: 4179

Keywords: Autoimmune diseases, anti-SSA/Ro60 antibodies, machine learning, autoantibodies, DNA methylation, GWAS, RNA-Seq, omics.

Key messages:

1. While anti-Ro60 autoantibodies are among the most frequently detected extractable nuclear antigen autoantibodies, a signature common to all patients expressing anti-Ro60 autoantibodies has not yet been established.
2. Machine learning identifies features from RNA-Seq, GWAS and DNA methylation datasets that discriminate anti-Ro60$^+$ patients regardless of the autoimmune disease, remain stable over time and are not influenced by treatment.
3. Three genes (*ATP10A*, *PARP14* and *MX1*) are overexpressed, hypomethylated and mutated in anti-Ro60$^+$ patients and are remarkably associated with the IFN signature regardless of the autoimmune disease.
4. Targeting Ro60-associated RNAs and Ro60-specific autoantibodies will reduce interferon signature in systemic autoimmune diseases.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/art.42243

Abstract

Objectives: Anti-Ro autoantibodies are among the most frequently detected extractable nuclear antigen autoantibodies, mainly associated with primary Sjögren's syndrome (pSS), systemic lupus erythematosus (SLE) and undifferentiated connective tissue disease (UCTD). Is there a common signature to all patients expressing anti-Ro60 autoantibodies regardless of their disease phenotype?

Methods: Using high-throughput multi-omics data collected within the cross-sectional cohort from the PRECISESADS IMI project (genetic, epigenomic, transcriptomic, combined with flow cytometric data, multiplexed cytokines, classical serology and clinical data), we assessed by machine learning the integrated molecular profiling of 520 anti-Ro60-positive (anti-Ro60$^+$) compared to 511 anti-Ro60-negative (anti-Ro60$^-$) patients with pSS, SLE and UCTD, and 279 healthy controls (HCs).

Results: The selected features for RNA-Seq, DNA methylation and GWAS data allowed a clear separation between anti-Ro60$^+$ and anti-Ro60$^-$ patients. The different features selected by machine learning from the anti-Ro60$^+$ patients constitute specific signatures when compared to anti-Ro60$^-$ patients and HCs. Remarkably, the transcript *z*-score of three genes (*ATP10A*, *MX1* and *PARP14*), presenting an overexpression associated with a hypomethylation and genetic variation, and independently identified by the Boruta algorithm, was clearly higher in anti-Ro60$^+$ patients compared to anti-Ro60$^-$ patients in all the diseases. We demonstrate that these signatures, enriched in interferon stimulated genes, were also found in anti-Ro60$^+$ patients with rheumatoid arthritis and systemic sclerosis and remained stable over time and not influenced by treatment.

Conclusion: Anti-Ro60$^+$ patients present a specific inflammatory signature regardless of their disease suggesting that a dual therapeutic approach targeting both Ro-associated RNAs and anti-Ro60 autoantibodies should be considered.

1. Introduction

Anti-Ro autoantibodies are among the most frequently detected extractable nuclear antigen (ENA) autoantibodies and have mainly been associated with primary Sjögren's syndrome (pSS). These autoantibodies are also frequently observed in systemic lupus erythematosus (SLE) and undifferentiated connective tissue disease (UCTD) (1,2). Additionally, anti-Ro autoantibodies have been reported in other autoimmune diseases such as systemic sclerosis (SSc), mixed connective tissue diseases (MCTD), rheumatoid arthritis (RA) and myositis (3).

Anti-Ro autoantibodies comprise reactivity against two autoantigens (Ro52 and Ro60) encoded by separate genes and found in distinct cellular compartments (4). Ro52 is a type I and type II interferon (IFN)-inducible protein (5,6) and is a negative regulator for proinflammatory cytokine production (7). Ro60 antigen binds to ~100 nt noncoding RNAs called hY-RNA (8) and acts as a quality checkpoint for RNA misfolding with molecular chaperones for defective RNA (9).

Variation in clinical manifestations or outcome based on presence or absence of anti-Ro autoantibodies has been highlighted in previous studies. Thus, SLE subjects with anti-Ro60 antibodies have an increased prevalence of skin disease, photosensitivity, and nephritis, along with elevated expression of IFN-inducible genes in immune cells and tissues (10). In pSS, patients with both anti-Ro60 and Ro52 antibodies were distinguished by a higher prevalence of markers of B cell hyperactivity and glandular inflammation (11). Those patients also had earlier disease onset and presented more systemic extraglandular manifestations such as leukopenia, hypergammaglobulinaemia and major salivary gland swelling (12). Recently, two subgroups of pSS patients were defined based on HLA association, Ro60/SSB antibodies and clinical manifestations. The Ro60/SSB antibody-positive subgroup was younger at disease onset and diagnosis and more frequently presented anaemia, leukopenia, hypergammaglobulinaemia, purpura, major salivary gland swelling, lymphadenopathy and lymphoma. These results confirmed an overall more severe disease phenotype compared with patients negative for both anti-Ro60 and anti-SSB antibodies (13). Solo anti-Ro60 reactivity correlated strongly with oral ulcers, a characteristic manifestation of SLE, while the combination of anti-Ro60 and anti-Ro52 was significantly more prevalent in patients demonstrating interstitial kidney disease and sicca symptoms (14).

Due to the presence of anti-Ro60 antibodies in different autoimmune diseases and the reported clinical manifestations which characterize this expression, a question remains. Is there a signature common to all patients expressing anti-Ro60 autoantibodies that would allow us to consider a suitable therapy regardless of their disease phenotype?

With algorithms derived from machine learning, the present study was undertaken to establish a precise signature of anti-Ro60-positive (anti-Ro60^{+}) patients in the diseases where this autoantibody is the most

frequently observed (pSS, SLE and UCTD) using high-throughput multi-omics data collected within the PRECISESADS IMI JU project (genetic, epigenomic, transcriptomic, combined with flow cytometric data, multiplexed cytokines, as well as classical serology and clinical data). Here we report on the integrated molecular profiling of 520 anti-Ro60^{+} patients compared to 511 anti-Ro60-negative (anti-Ro60^{-}) patients and 279 healthy controls (HCs). We then observe whether this signature was also present in the 41/725 anti-Ro60^{+} patients with other autoimmune diseases such as MCTD, RA and SSc all from the same PRECISESADS cohort.

2. Methods

2.1. Patient population

The present study was conducted in 1755 patients (367 pSS, 508 SLE, 156 UCTD, 307 RA, 327 SSc, and 90 MCTD) and 279 HCs included in the European multi-centre cross-sectional study of the PRECISESADS IMI consortium (15). The classification criteria were: for RA on the 2010 ACR/EULAR classification criteria (16), for SLE on the 1997 update of 1982 ACR criteria (17), for SSc on ACR/EULAR 2013 classification criteria (18), for pSS on AECG pSS classification criteria (19) with at least the presence of anti-Ro and/or a positive focus on a minor salivary gland biopsy, for MCTD on Alarcon-Segovia criteria (20), and for UCTD on patients with clinical features of systemic autoimmune diseases (SADs) not fulfilling any of the above or any other SADs criteria for at least 2 years with the presence of unspecific antibodies, antinuclear antibodies (ANA) $\geq$ 1:160. Patients fulfilling 3 out of 4 SLE classification criteria and patients with early systemic sclerosis (21) were not classified as UCTD. Recruitment was performed between December 2014 and October 2017 involving 19 institutions in 9 countries (Austria, Belgium, France, Germany, Hungary, Italy, Portugal, Spain and Switzerland). The PRECISESADS study adhered to the standards set by International Conference on Harmonization and Good Clinical Practice (ICH-GCP), and to the ethical principles that have their origin in the declaration of Helsinki (2013). Each patient signed an informed consent prior to study inclusion. The Ethical Review Boards of the 19 participating institutions approved the protocol of the study. The protection of the confidentiality of records that could identify the included subjects is ensured as defined by the EU Directive 2001/20/EC and the applicable national and international requirements relating to data protection in each participating country. The study is registered in ClinicalTrials.com with numbers NCT02890121 (cross-sectional cohort) and NCT02890134 (inception cohort). The anti-Ro60^{+} signature identified by machine learning was validated using the transcriptome of 106 patients recruited in the PRECISESADS inception study (NCT02890134), followed up and sampled at the time of recruitment and at 6 and/or 14 months. Of note, patients in the inception cohort were diagnosed as above within less than a year since diagnosis and had not had high doses of immuno-suppressants, cyclophosphamide or Belimumab at least 3 months prior to recruitment. For time points at 6 and 14 months, patients could have any standard of care therapy indicated by their physician. HCs were individuals without chronic medication who do not suffering from any inflammatory autoimmune,

allergic or infectious condition, and without a history of autoimmune disease, particularly thyroid disease or other diseases that may modify cellular profiles in blood.

2.2. Determination of autoantibodies, antinuclear antibodies, free light chains and complement fractions

All autoantibodies were determined in a single center (Brest) for all samples between March 2016 and June 2019. Anti-ENA (comprising Sm, U1-RNP, Scl70, Ro52, Ro60 and SSB) and specific autoantibodies anti-Ro52 and anti-Ro60, anti-CCP2, IgG and IgM anti-β2GPI, IgG and IgM anti-cardiolipin, anti-dsDNA, and anti-centromere autoantibodies were determined using the chemiluminescent immunoanalyser IDS-ISYS (Immunodiagnostic, Boldon, UK). Rheumatoid factor (RF) was determined regardless of the isotypes by turbidimetry with the SPA plus (The Binding Site), as well as C3 and C4 complement fractions and κ and λ free light chains. All patients and HCs were tested. For more technical details on sample and data collection, please refer to Supplementary methods section 2. Autoantibodies and RF distribution have been described by concentration level (Negative/Low/Medium/Elevated/High) and a Fisher's exact test was applied to compare the proportion and the concentration across the anti-Ro60^{+} and anti-Ro60^{-} patients in each disease. Complements C3 and C4 and circulating free light chains have been described in continued concentration expressed in g/L and mg/L respectively and a Kruskal–Wallis test was applied to compare the concentration level across the anti-Ro60^{+} and anti-Ro60^{-} patients in each disease. Positive samples for anti-Ro60 autoantibodies were also classified according to their degree of positivity. Positive samples with concentrations between 10 and 640 arbitrary unit (AU)/ml were considered as anti-Ro60^{low} whereas samples with concentration > 640 AU/ml were considered as anti-Ro60^{high} patients.

ANA detection was performed by an in-house technique on HEp-2 cells (ATCC strain: CCL23). Each sample was systematically tested on 5 successive dilutions (1:80; 1:160; 1:320; 1:640; 1:1280) and the threshold of positivity was set at 1:160, according to international recommendations (22). Information on current or past presence of hypergammaglobulinemia was collected in each center at the time of inclusion and defined as the presence within 12 weeks of serum IgG > upper limit of normal and/or gammaglobulin > 20%.

2.3. Clinical data.

Clinical data obtained from 520 anti-Ro60^{+} (306 pSS, 175 SLE and 39 UCTD) patients, 511 anti-Ro60^{-} (61 pSS, 333 SLE and 117 UCTD) patients and 279 HCs, were collected using an electronic case report form (eCRF). Clinical data included patient's age, sex, ethnicity, disease duration, the physician global assessment of disease activity (PGA), SLEDAI for SLE, ESSDAI for pSS, and current use of treatments.

2.4. Other Available data.

High-dimensional omics genotype, RNA-seq, DNA methylation and proportions of relevant cell types using flow cytometry custom marker panels were analysed from whole blood samples. Additional information, such as cytokines, chemokines and inflammatory mediator expression levels were obtained from serum samples. All these parameters are detailed in Supplementary methods. Repartition of patients with a full dataset per omic type and across disease is available in Supplementary Table 1.

2.7. Dimension reduction

Our strategy of dimensionality reduction was driven by artificial intelligence approaches involving machine learning. Patients were first split according to their disease (pSS, SLE or UCTD). We then separately considered each of the datasets describing these patients (RNA-Seq, DNA methylation, GWAS and flow cytometry associated with cytokine expression). For each of these datasets, we performed a Boruta (23) analysis to discriminate anti-Ro60^{+} and anti-Ro60^{-} patients in order to extract, within each dataset, features that significantly contributed to distinguish the two groups. The Boruta algorithm creates an extended dataset by adding copies of each feature in the original dataset. Values of the duplicated features are then shuffled and the resulting features are called "shadow features". The random permutation of the modality within these features leads to the removal of any pre-existing correlation with the target variable, in our case, anti-Ro60 positivity. Once shadow features were crafted, a random forest classifier was run on the whole dataset and z-scores were computed for all features (real and shadow). Shadow features were then sorted according to their z-score and the maximum score was kept in memory as a threshold. The algorithm assigned a hit to each real feature that had a z-score above this threshold. Finally, Boruta marked the features which had a z-score significantly lower than the shadow with maximum z-score as "unimportant" and removed them from the dataset, before removing all shadow features and returning a clean dataset.

We ran the Boruta algorithm on 300 iterations with a max depth set to 5. Extracted features were run on a linear discriminant analysis (LDA), only used to visually assess the separation between anti-Ro60^{+}, anti-Ro60^{-} patients and HCs. No classification metrics were computed with LDA.

3. Results

3.1. Anti-Ro60-positive patients present specific biological and clinical features

The characteristics of the 279 anti-Ro60^{-} HCs and the 520 anti-Ro60^{+} patients (306 pSS, 175 SLE and 39 UCTD) compared to the 511 anti-Ro60^{-} patients (61 pSS, 333 SLE and 117 UCTD) patients are presented in Table 1. Regarding the antibody profile, compared to anti-Ro60^{-}, anti-Ro60^{+} patients from the three diseases had significantly increased levels of ANA, kappa and lambda free light chains, rheumatoid factor, anti-Ro52 and anti-SSB antibodies (Supplementary Table 2 and Figure 1). Anti-Ro52 and anti-SSB autoantibodies were also significantly increased in anti-Ro60^{high} patients compared to anti-Ro60^{low} patients. Past and/or present hypergammaglobulinemia was increased in anti-Ro60^{+} patients

regardless of the disease (Supplementary Figure 1A). No difference in disease activity score (ESSDAI, SLEDAI, PGA) was observed between anti-Ro60$^-$ and anti-Ro60$^+$ patients (Supplementary Figure 1B). However, in pSS, anti-Ro60$^+$ patients had a lower ESSPRI, and the higher is the anti-Ro60 scale, the lower are the ESSPRI and its components (dryness, fatigue and pain) (Supplementary Figure 1C).

3.2. Machine learning identifies specific signature common to anti-Ro60$^+$ patients in the different omics datasets

We used the Boruta algorithm (23) on all datasets to extract features that significantly contributed to the prediction of patients who were anti-Ro60$^+$ according to the different omics (RNA-Seq, DNA methylation, GWAS and cytokine expression associated with cell subset distribution). A total of 923 features were selected from RNA-Seq variables, 64 from DNA methylation, 5749 from GWAS (Supplementary Tables 3, 4 and 5 respectively) and 8 from the association of cytokine expression levels and cell subset distribution. An LDA for each omics is presented in Figure 2. We have considered the combined analysis of pSS, SLE and UCTD patients within the framework of the Boruta approach. Features were selected on disease dataset to capture a maximum of discriminating information. We then considered the union of Boruta results for pSS, SLE and UCTD to constitute the final signature. Of note, the selected features for RNA-Seq and GWAS data allowed a clear separation between anti-Ro60$^+$ and anti-Ro60$^-$ patients. Remarkably, even if data from HCs were not used for feature selection, their integration into the different LDA, based on the features identified as discriminating between anti-Ro60$^+$ and anti-Ro60$^-$ patients using the Boruta algorithm resulted in a separation from the patients. These results demonstrate that the different features selected by machine learning from the anti-Ro60$^+$ patients constitute specific signatures when compared to anti-Ro60$^-$ patients and HCs.

3.3. Characterisation and pathway analysis of the transcriptomic signature found in anti-Ro60$^+$ patients

The 923 transcripts identified by machine learning to discriminate anti-Ro60$^+$ patients (Figure 2A) were analysed with Reactome (24). The 25 most relevant pathways are presented in Supplementary Table 6. Anti-Ro60$^+$ patients were enriched in genes involved in IFN signaling (type I and II), cytokine signaling, activation of C3 and C5, antiviral mechanism by IFN-stimulated genes and IL-10 signaling.

To go further in the IFN signature, we analysed IFN-annotated modules previously described as strongly up-regulated in SLE (25,26). The different type I and type II IFN *z*-scores were increased in anti-Ro60$^+$ patients regardless of the disease (Figure 3).

3.4. Characterisation and pathway analysis of the DNA methylation signature found in anti-Ro60$^+$ patients

The 37 genes associated with the 64 CpG identified by machine learning to discriminate anti-Ro60$^+$ patients (Figure 2B) were analysed with Reactome. Interestingly, the most relevant pathways were the same as those previously found in the transcriptome analysis such as IFN signaling (Type I and II), cytokine signaling in the immune system and antiviral mechanism by IFN-stimulated genes (Supplementary Table 7).

Among these 37 differentially methylated genes, 33 were also found by the Boruta algorithm in the RNA-Seq analysis. Interaction network of these 33 common genes, determined by STRING analysis with a confidence cut-off of 0.4, revealed and confirmed the common IFN signature (Supplementary Figure 2) (27). Of note, all transcripts were overexpressed in anti-Ro60$^+$ patients regardless of the disease, and a global hypomethylation of CpG was observed for all but one gene (*ISG15*). For one gene (*IFITM1*), up to eight hypomethylated CpGs were assessed (Supplementary Figure 3).

3.5. Genome-wide association study analysis of anti-Ro60$^+$ patients

Our machine learning approach identified 5749 SNPs able to differentiate anti-Ro60$^+$ from anti-Ro60$^-$ patients (Figure 2C). Interestingly, three of these SNPs were located on genes previously selected by the algorithm from the previous RNA-Seq and DNA methylation analyses (Figure 4A). The three corresponding genes were *ATP10A*, *MX1,* and *PARP14*. Remarkably, the transcript *z*-score of these three genes was clearly higher in anti-Ro60$^+$ patients compared to anti-Ro60$^-$ patients, but also in anti-Ro60^{high} *versus* anti-Ro60^{low} patients, when all diseases were merged (Figure 4B). The same was true in all the diseases and constituted a clear signature (Figure 4C). Given the strong association of anti-Ro60 with anti-Ro52/TRIM21 antibodies, we considered that the positivity for anti-Ro52/TRIM21 could define the signature. We have then divided pSS, SLE, and UCTD patients into three groups (anti-Ro60$^-$/anti-Ro52$^-$, anti-Ro60$^+$/anti-Ro52$^-$ and anti-Ro60$^+$/anti-Ro52$^+$) and assessed the identified anti-Ro60 signature defined by the transcript *z*-score of the three genes (*ATP10A*, *MX1*, and *PARP14*). The *z*-score was higher in anti-Ro60$^+$/Ro52$^-$ patients compared to anti-Ro60$^-$/anti-Ro52$^-$ patients in pSS and SLE. In contrast, the *z*-score was only higher in anti-Ro60$^+$/anti-Ro52$^+$ pSS patients compared to anti-Ro60$^+$/anti-Ro52$^-$ pSS patients and no significant difference was observed in SLE and UCTD (Supplementary Figure 4A).

We also considered that anti-Ro60 positivity could just be a marker of B cell reactivity given that the majority of anti-Ro60$^-$ patients are ENA negative. We then assessed the transcript *z*-score of the three genes (*ATP10A*, *MX1*, and *PARP14*) in five groups of patients regardless of their disease: patients without any autoantibody, anti-Ro60$^-$/anti-Ro52$^+$ patients negative for any of the autoantibodies analyzed (anti-RNP, anti-Sm, anti-SSB, anti-Scl70, anti-CCP, anti-dsDNA, anti-MPO, anti-PR3, anti-CentB), anti-Ro60$^-$/anti-Ro52$^+$ patients positive for any other autoantibodies, anti-Ro60$^-$/anti-Ro52$^-$ patients positive for any of the autoantibodies, and anti-Ro60$^+$ patients. The z-score was clearly higher

in anti-Ro60$^+$ patients compared to all the other groups (Supplementary Figure 4B). All these data confirm that the determined signature is specific to anti-Ro60$^+$ patients and is not just a marker of B cell activation or due to the presence of any other autoantibody.

3.6. Characterisation of the flow cytometry signature and cytokine expression found in anti-Ro60$^+$ patients

Machine learning identified six parameters among flow cytometry and two parameters among cytokine expression (assessed by Luminex-based quantitative assay) data able to differentiate anti-Ro60$^+$ from anti-Ro60$^-$ patients (Figure 2D). The robustness of the six flow cytometry features was poor and only associated to one disease (Supplementary Figure 5). Interestingly, however, cytokine expression in serum showed an increase of IFNγ-induced protein (CXCL10/IP-10) and a down regulation of IL-1 RII, the decoy receptor for cytokine belonging to the IL-1 family, in anti-Ro60$^+$ patients regardless of the disease (Supplementary Figure 5).

3.7. The established common signature of anti-Ro60$^+$ patients with pSS, SLE and UCTD is also common to patients with RA, SSc or MCTD expressing anti-Ro60 antibodies and is stable over time

To confirm the robustness of the identified signature, we observed whether this signature was also present in an independent cohort of 106 newly diagnosed patients with pSS, SLE or UCTD, from the inception cohort of the PRECISESADS study, which provided an additional validation set to test the generalization of our signature on patients that were not used for the feature selection process. At inclusion (first time point), we had 46 anti-Ro60$^+$ and 60 anti-Ro60$^-$ patients. Again, using the 923 transcripts of the RNA-Seq signature, the LDA showed a clear separation between anti-Ro60$^+$ and anti-Ro60$^-$ patients regardless of the disease (Supplementary Figure 6A). Furthermore, the *z*-score for the 33 genes identified by the Boruta algorithm and common to RNA-Seq and DNA methylation data were significantly increased in anti-Ro60$^+$patients (Supplementary Figure 6B). This was also true with the *z*-score for the three genes (*ATP10A*, *MX1* and *PARP14*) previously selected by the algorithm from the RNA-Seq, GWAS and DNA methylation analyses (Supplementary Figure 6C).

Finally, the signature's robustness was also assessed in 724 patients with other autoimmune diseases such as MCTD, RA and SSc all from the PRECISESADS cross-sectional cohort. A clear separation between anti-Ro60$^+$ (n=40) and anti-Ro60$^-$ (n=684) patients using the representation space generated by the LDA is shown in Figure 5A. In all diseases but MCTD, anti-Ro60$^+$ patients had significantly increased *z*-scores for the 33 common genes (Figure 5B) and the three genes constituting the clear signature (Figure 5C). We can therefore conclude that anti-Ro60$^+$ patients have a specific signature regardless of their disease.

Finally, we have assessed the transcript *z*-score of the three genes (*ATP10A*, *MX1*, and *PARP14*) in the inception cohort restricted to 86 patients (pSS, SLE, UCTD, RA, SSc and MCTD) followed up and

sampled at three time points (recruitment, at 6 and/or 14 months). For time points at 6 and 14 months, patients could have any standard of care therapy indicated by their physician (Supplementary Table 8). Anti-Ro60$^+$ patients remained positive and anti-Ro60$^-$ patients remained negative over time (data not shown). We confirm that the *z*-score remained stable in anti-Ro60$^+$ and anti-Ro60$^-$ patients over time (Figure 5D). Overall, the identified signature for anti-Ro60$^+$ patients does not depend on treatment and is stable over time.

4. Discussion

Our study demonstrates that anti-Ro60$^+$ patients have a specific signature regardless of their disease. Anti-Ro60$^+$ compared to anti-Ro60$^-$ patients presented the same clinical and biological characteristics as those previously described in the literature such as hypergammaglobulinemia (12,13) and association with other autoantibodies (Anti-Ro52, anti-SSB and RF) (14). Anti-Ro60 positivity was reported to be higher in the low symptom burden subgroups of pSS patients (28) in accordance with our observation that anti-Ro60$^+$ patients had a lower ESSPRI.

Our study has some limitations. First, it could be argued that this is a cross-sectional study and it is assumed that single samples (cells and sera) were collected at an arbitrary time during the disease course of the different autoimmune diseases. However, we demonstrated in our inception cohort with a follow-up of 14 months, that the identified signature for anti-Ro60$^+$ patients was stable over time and was not influenced by treatment. Second, virtually all subjects were Caucasian and although common variants are expected to be evolutionarily old and shared across ethnicities, some risk loci show considerable ethnic differences in frequency and/or effect size.

The novelty of our study was to use machine learning to conduct a robust signature extraction specific to anti-Ro60$^+$ patients through dimensionality reduction approaches, using high-throughput multi-omics data. Assessment of the signature's robustness occurred in three steps. First, we used discriminant features extracted from the different omics data sets to perform LDA. The new representation spaces crafted from the selected features through this analysis allowed a sufficient separation of anti-Ro60$^+$ and anti-Ro60$^-$ patients in each of the three diseases studied (SLE, pSS and UCTD). Second, we considered the intersection of the selected features between RNA-Seq and DNA methylation and the intersection of the selected features between RNA-Seq, DNA methylation and GWAS, narrowing down the original selection to two signatures composed of 33 and three genes, respectively. Both *z*-scores, generated by the 33 genes and the three genes, were significantly different between anti-Ro60$^+$ and anti-Ro60$^-$ for SLE, pSS, and UCTD patients. Third, because we used a machine learning approach to extract the features, we assessed the possibility of overfitting by testing the validity of these signatures in patients who were not used in the training process of the algorithm. Generalization of the signature was then evaluated by computing RNA-Seq features from another cohort of pSS, SLE and UCTD patients and

from RA, SSc and MCTD patients. The LDA consistently showed a clear separation among anti-Ro60$^+$ and anti-Ro60$^-$ patients. Again, *z*-scores were significantly different between anti-Ro60$^+$ and anti-Ro60$^-$ for RA and SSc patients but not for MCTD patients. Consequently, the discriminating properties of the representation space obtained through the computation of LDA, the statistical tests of distributions and the generalization to other diseases constituted strong indicators of the signature's robustness.

Reactome pathway analysis of the 33 genes differentially expressed and methylated showed a link between anti-Ro60 antibodies and IFN signature, cytokines secretions and IRF7 which were associated with TLR signaling. The striking association between anti-Ro60 autoantibodies and inflammation in autoimmune diseases led to the hypothesis that the RNA-binding properties of Ro60 produce aberrant Toll-like receptor (TLR) signaling (29). Alu retroelements activate TLR7 and TLR8 as oligoribonucleotides and associate with Ro60 in cell lines (30); consequently, inflammatory and IFN signatures associated with anti-Ro60 autoantibodies could be due to the RNA-binding properties of Ro60.

Remarkably, the transcript *z*-score of three genes (*ATP10A*, *MX1* and *PARP14*) were clearly higher in anti-Ro60$^+$ patients compared to anti-Ro60$^-$ patients in all the diseases and constituted a clear signature. The first gene, *ATP10A*, encoded one of the five P4 ATPase which requires interaction with TMEM30A (Transmembrane Protein 30A) to exit from the endoplasmic reticulum to the plasma membrane. *ATP10A* was recently linked to autoimmunity, as a study demonstrated that meQTLs regulated the methylation of the *ATP10A* gene in blood from pSS patients (31). Since this enzyme transports mainly two aminophospholipids: phosphatidylserine and phosphatidylethanolamine, which may be the target of minor autoantibodies in anti-phospholipid syndrome (APS) (32), it is legitimate to speculate on a link between the presence of antiphospholipid antibodies and the increase in *ATP10A* transcript. We re-ran the analysis excluding patients positive for the major autoantibodies found in the APS (i.e. anti-B2GPI and anti-CL IgG and IgM) to eliminate potential patients with secondary APS, and the signature found persists (p=4,2e-10, data not shown). Moreover, to our knowledge, no association with APS has yet been described in the literature. Thus, the signature carried by *ATP10A* seems to be specific to anti-Ro60$^+$ patients. Another GWAS study on cytokine responses found that genetic variants of *ATP10A* were associated with IFNα production (33). The second gene, *PARP14* (Polyadenosine diphosphate ribose polymerase 14), encoded for a member of PARP family proteins which contain macrodomain binding proteins influencing many biological processes (34). PARP14 suppressed proinflammatory IFN-STAT1 signaling and activated the anti-inflammatory IL4-STAT6 pathway in primary human macrophages (35). PARP14 also enhanced histone activation to promote transcription of type 1 IFN genes such as *IFNb1* after LPS stimulation in RAW264.7 cells (36). Interestingly, *PARP-14* was identified as one of the five genes that can distinguish pSS from controls (37). The third gene, *MX1*, encoded the MX dynamic like GTPase 1 or MXA which participates in the cellular antiviral response by antagonizing the replication process of several different RNA or DNA viruses. *MX1* gene expression

is induced by IFN via Jak1/Tyk2 followed by activation of STAT1/STAT2 pathway (38). Furthermore, MX1 protein levels were recently reported as surrogate for the IFN-I gene scores in SLE (39). Consequently, these three overexpressed, hypomethylated and mutated genes in anti-Ro60^{+} patients are remarkably associated with the IFN signature regardless of the autoimmune disease.

To control IFN signature in anti-Ro60^{+} patients with autoimmune diseases, a key challenge would be to break the continual turnover of Ro60-specific clones that seems to drive lifelong Ro60 humoral autoimmunity (40). This may entail a dual approach targeting both Ro60-associated RNAs (including Alu transcripts and Y RNAs) and Ro60-specific autoantibody clonotypes as elegantly suggested by Reed and Gordon (29).

Funding: The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under the Grant Agreement Number 115565 (PRECISESADS project), resources of which are composed of financial contribution from the European Union's Seventh Framework Program (FP7/2007–2013) and EFPIA companies' in-kind contribution.

Acknowledgments: LBAI was supported by the Agence Nationale de la Recherche under the "Investissement d'Avenir" program with the Reference ANR-11-LABX-0016-001 (Labex IGO). The authors would like to particularly express their gratitude to the patients, nurses, technicians and many others who helped directly or indirectly in the consecution of this study. They are grateful to the Institut Français de Bioinformatique (ANR-11-INBS-0013), the Roscoff Bioinformatics platform ABiMS (http://abims.sb-roscoff.fr) for providing computing and storage resources and the Hypérion platform at LBAI (Brest, France) for flow cytometry facilities.

1. Aggarwal A. Role of autoantibody testing. Best Pract Res Clin Rheumatol. 2014 Dec 1;28(6):907–20.

2. Radin M, Rubini E, Cecchi I, Foddai SG, Barinotti A, Rossi D, et al. Disease evolution in a long-term follow-up of 104 undifferentiated connective tissue disease patients. Clin Exp Rheumatol. 2021 Jun 26;

3. Robbins A, Hentzien M, Toquet S, Didier K, Servettaz A, Pham B-N, et al. Diagnostic Utility of Separate Anti-Ro60 and Anti-Ro52/TRIM21 Antibody Detection in Autoimmune Diseases. Front Immunol. 2019;10:444.

4. Schulte-Pelkum J, Fritzler M, Mahler M. Latest update on the Ro/SS-A autoantibody system. Autoimmun Rev. 2009 Jun;8(7):632–7.

5. Strandberg L, Ambrosi A, Espinosa A, Ottosson L, Eloranta M-L, Zhou W, et al. Interferon-alpha induces up-regulation and nuclear translocation of the Ro52 autoantigen as detected by a panel of novel Ro52-specific monoclonal antibodies. J Clin Immunol. 2008 May;28(3):220–31.

6. Rhodes DA, Ihrke G, Reinicke AT, Malcherek G, Towey M, Isenberg DA, et al. The 52 000 MW Ro/SS-A autoantigen in Sjögren's syndrome/systemic lupus erythematosus (Ro52) is an interferon-gamma inducible tripartite motif protein associated with membrane proximal structures. Immunology. 2002 Jun;106(2):246–56.

7. Yoshimi R, Chang T-H, Wang H, Atsumi T, Morse HC, Ozato K. Gene disruption study reveals a nonredundant role for TRIM21/Ro52 in NF-kappaB-dependent cytokine expression in fibroblasts. J Immunol Baltim Md 1950. 2009 Jun 15;182(12):7527–38.

8. Lerner MR, Boyle JA, Hardin JA, Steitz JA. Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. Science. 1981 Jan 23;211(4480):400–2.

9. Belisova A, Semrad K, Mayer O, Kocian G, Waigmann E, Schroeder R, et al. RNA chaperone activity of protein components of human Ro RNPs. RNA N Y N. 2005 Jul;11(7):1084–94.

10. Kirou KA, Lee C, George S, Louca K, Peterson MGE, Crow MK. Activation of the interferon-alpha pathway identifies a subgroup of systemic lupus erythematosus patients with distinct serologic features and active disease. Arthritis Rheum. 2005 May;52(5):1491–503.

11. Armağan B, Robinson SA, Bazoberry A, Perin J, Grader-Beck T, Akpek EK, et al. Antibodies to both Ro52 and Ro60 may identify Sjögren's syndrome patients best suited for clinical trials of disease-modifying therapies. Arthritis Care Res. 2021 Mar 20;

12. Brito-Zerón P, Acar-Denizli N, Ng W-F, Zeher M, Rasmussen A, Mandl T, et al. How immunological profile drives clinical phenotype of primary Sjögren's syndrome at diagnosis: analysis of 10,500 patients (Sjögren Big Data Project). Clin Exp Rheumatol. 2018 Jun;36 Suppl 112(3):102–12.

13. Thorlacius GE, Hultin-Rosenberg L, Sandling JK, Bianchi M, Imgenberg-Kreuz J, Pucholt P, et al. Genetic and clinical basis for two distinct subtypes of primary Sjögren's syndrome. Rheumatol Oxf Engl. 2021 Feb 1;60(2):837–48.

14. Zampeli E, Mavrommati M, Moutsopoulos HM, Skopouli FN. Anti-Ro52 and/or anti-Ro60 immune reactivity: autoantibody and disease associations. Clin Exp Rheumatol. 2020 Aug;38 Suppl 126(4):134–41.

15. Barturen G, Babaei S, Català-Moll F, Martínez-Bueno M, Makowska Z, Martorell-Marugán J, et al. Integrative Analysis Reveals a Molecular Stratification of Systemic Autoimmune Diseases. Arthritis Rheumatol Hoboken NJ. 2021 Jun;73(6):1073–85.

16. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum. 2010 Sep;62(9):2569–81.

17. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1997 Sep;40(9):1725.

18. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. Ann Rheum Dis. 2013 Nov;72(11):1747–55.

19. Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, et al. Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. Ann Rheum Dis. 2002 Jun;61(6):554–8.

20. Alarcón-Segovia D, Cardiel MH. Comparison between 3 diagnostic criteria for mixed connective tissue disease. Study of 593 patients. J Rheumatol. 1989 Mar;16(3):328–34.

21. LeRoy EC, Medsger TA. Criteria for the classification of early systemic sclerosis. J Rheumatol. 2001 Jul;28(7):1573–6.

22. Sack U, Conrad K, Csernok E, Frank I, Hiepe F, Krieger T, et al. Autoantibody detection using indirect immunofluorescence on HEp-2 cells. Ann N Y Acad Sci. 2009 Sep;1173:166–73.

23. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. J Stat Softw. 2010 Sep 16;36:1–13.

24. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 2018 Jan 4;46(D1):D649–55.

25. Kirou KA, Lee C, George S, Louca K, Papagiannis IG, Peterson MGE, et al. Coordinate overexpression of interferon-alpha-induced genes in systemic lupus erythematosus. Arthritis Rheum. 2004 Dec;50(12):3958–67.

26. Chiche L, Jourde-Chiche N, Whalen E, Presnell S, Gersuk V, Dang K, et al. Modular transcriptional repertoire analyses of adults with systemic lupus erythematosus reveal distinct type I and type II interferon signatures. Arthritis Rheumatol Hoboken NJ. 2014 Jun;66(6):1583–95.

27. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013 Jan;41(Database issue):D808–15.

28. Tarn JR, Howard-Tripp N, Lendrem DW, Mariette X, Saraux A, Devauchelle-Pensec V, et al. Symptom-based stratification of patients with primary Sjögren's syndrome: multi-dimensional characterisation of international observational cohorts and reanalyses of randomised clinical trials. Lancet Rheumatol. 2019 Sep 25;1(2):e85–94.

29. Reed JH, Gordon TP. Autoimmunity: Ro60-associated RNA takes its toll on disease pathogenesis. Nat Rev Rheumatol. 2016 Mar;12(3):136–8.

30. Hung T, Pratt GA, Sundararaman B, Townsend MJ, Chaivorapol C, Bhangale T, et al. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. Science. 2015 Oct 23;350(6259):455–9.

31. Teruel M, Barturen G, Martínez-Bueno M, Castellini-Pérez O, Barroso-Gil M, Povedano E, et al. Integrative epigenomics in Sjögren´s syndrome reveals novel pathways and a strong interaction between the HLA, autoantibodies and the interferon signature. Sci Rep. 2021 Dec 2;11(1):23292.

32. Alessandri C, Conti F, Pendolino M, Mancini R, Valesini G. New autoantigens in the antiphospholipid syndrome. Autoimmun Rev. 2011 Aug;10(10):609–16.

33. Kennedy RB, Ovsyannikova IG, Pankratz VS, Haralambieva IH, Vierkant RA, Jacobson RM, et al. Genome-wide genetic associations with IFNγ response to smallpox vaccine. Hum Genet. 2012 Sep;131(9):1433–51.

34. Fehr AR, Singh SA, Kerr CM, Mukai S, Higashi H, Aikawa M. The impact of PARPs and ADP-ribosylation on inflammation and host-pathogen interactions. Genes Dev. 2020 Mar 1;34(5–6):341–59.

35. Iwata H, Goettsch C, Sharma A, Ricchiuto P, Goh WWB, Halu A, et al. PARP9 and PARP14 cross-regulate macrophage activation via STAT1 ADP-ribosylation. Nat Commun. 2016 Oct 31;7:12849.

36. Caprara G, Prosperini E, Piccolo V, Sigismondo G, Melacarne A, Cuomo A, et al. PARP14 Controls the Nuclear Accumulation of a Subset of Type I IFN-Inducible Proteins. J Immunol Baltim Md 1950. 2018 Apr 1;200(7):2439–54.

37. Yao Q, Song Z, Wang B, Qin Q, Zhang J-A. Identifying Key Genes and Functionally Enriched Pathways in Sjögren's Syndrome by Weighted Gene Co-Expression Network Analysis. Front Genet. 2019;10:1142.

38. Haller O, Staeheli P, Kochs G. Protective role of interferon-induced Mx GTPases against influenza viruses. Rev Sci Tech Int Off Epizoot. 2009 Apr;28(1):219–31.

39. Chasset F, Ribi C, Trendelenburg M, Huynh-Do U, Roux-Lombard P, Courvoisier DS, et al. Identification of highly active systemic lupus erythematosus by combined type I interferon and neutrophil gene scores vs classical serologic markers. Rheumatol Oxf Engl. 2020 Nov 1;59(11):3468–78.

40. Lindop R, Arentz G, Bastian I, Whyte AF, Thurgood LA, Chataway TK, et al. Long-term Ro60 humoral autoimmunity in primary Sjögren's syndrome is maintained by rapid clonal turnover. Clin Immunol Orlando Fla. 2013 Jul;148(1):27–34.

Figure legends

Figure 1: Serological distributions in pSS, SLE and UCTD patients. The presence of anti-Ro52, anti-Ro60, anti-SSB antibodies, rheumatoid factor (RF) and circulating free light chains (cFLc) were measured in the serum of 520 anti-Ro60^{+} (306 pSS, 175 SLE, 39 UCTD) and 511 anti-Ro60^{-} (61 pSS, 333 SLE, 117 UCTD) patients in the same center, using an automated chemiluminescent immunoanalyzer (IDS-iSYS). Turbidimetry was used for the detection of RF and cFLc (Kappa and Lambda). Anti-Ro60^{+} patients were divided in two groups: anti-Ro60^{low} patients (samples with concentrations between 10 and 640 AU/ml), and anti-Ro60^{high} patients (samples with concentrations > 640 AU/ml). Statistical significance was determined by the two-tailed pairwise Wilcoxon-rank sum test. Plots show median, with error bars indicating ± interquartile range. (pSS: primary Sjögren's syndrome, SLE: systemic lupus erythematosus, UCTD: undifferentiated connective tissue disease).

Figure 2: Machine learning identifies specific signatures common to anti-Ro60^{+} patients in the different omics datasets. Linear discriminant analysis (LDA) representation of the systemic lupus erythematosus (SLE), undifferentiated connective tissue disease (UCTD) and primary Sjögren's syndrome (pSS) patients using features selected by the Boruta algorithm. (**A**) A total of 923 features were selected from RNA-Seq data. (**B**) A total of 64 features were selected from methylation data. (**C**) A total of 5749 features were selected from GWAS data. (**D**) A total of eight features were selected among flow cytometry distribution and cytokine expression data.

Figure 3: Ro60^{+} patients show a higher IFN signature regardless of the disease. IFN *z*-score analyses were performed for 411 anti-Ro60^{+} (249 pSS, 136 SLE and 26 UCTD) patients compared to 392 anti-Ro60^{-} (46 pSS, 267 SLE and 79 UCTD) patients and 254 HCs. The genes (*IFI44*,*IFI44L*,*IFIT1* and *MX1*) of the M1.2 module are induced by IFNα, while genes from both M1.2 and M3.4 (*ZBP1*, *IFIH1*, *EIF2AK2*, *PARP9* and *GBP4*) are upregulated by IFNβ. The genes (*PSMB9*, *NCOA7*, *TAP1*, *ISG20* and *SP140*) from the M5.12 module are poorly induced by IFNα and IFNβ alone while they are upregulated by IFNγ. Moreover, transcripts belonging to M3.4 and M5.12 were only fully induced by a combination of Type I and Type II IFNs (26). Other modules identified genes preferentially induced by IFNα (*IFIT1*, *IFI44* and *EIF2AK2*) or IFNγ (*IRF1*, *GBP1* and *SERPING1*) (25). Two-tailed pairwise Wilcoxon-rank sum test results are shown. Plots show median, with error bars indicating ± interquartile range. (pSS: primary Sjögren's syndrome, SLE: systemic lupus erythematosus, UCTD: undifferentiated connective tissue disease, HCs: healthy controls).

Figure 4: Three genes common to RNA-Seq, DNA methylation and GWAS analysis characterize anti-Ro60^{+} patients. (**A**) Venn diagram showing the gene overlaps according to the different omics data analyses conducted by machine learning (RNA-Seq, DNA methylation and SNPs) to discriminate anti-Ro60^{+} from anti-Ro60^{-} patients. (**B**) *ATP10*/*MX1*/*PARP14* *z*-score analyses were performed for 803

patients and 254 HCs according to anti-Ro60 expression. (**C**) *ATP10*/*MX1*/*PARP14* *z*-score analyses were performed for 295 pSS, 403 SLE and 105 UCTD patients and 254 HCs. Two-tailed pairwise Wilcoxon-rank sum test results are shown. Plots show median, with error bars indicating ± interquartile range. (pSS: primary Sjögren's syndrome, SLE: systemic lupus erythematosus, UCTD: undifferentiated connective tissue disease, HCs: healthy controls).

Figure 5: The established signature of anti-Ro60$^+$ patients is common to patients with rheumatoid arthritis (RA), systemic sclerosis (SSc) or mixed connective tissue disease (MCTD) expressing anti-Ro60 antibodies. (**A**) Linear discriminant analysis (LDA) representation of the patients, using the 923 features selected by the Boruta algorithm from RNA-Seq data in SLE, UCTD and pSS patients, discriminates anti-Ro60$^+$ and anti-Ro60$^-$ patients in RA, SSc and MCTD. (**B**) *z*-score analyses of the 33 genes, identified by the Boruta algorithm and common to RNA-Seq and methylome data, were performed for 295 pSS, 403 SLE, 105 UCTD, 307 RA, 327 SSc and 90 MCTD patients and 254 HCs. (**C**) *ATP10*/*MX1*/*PARP14* *z*-score analyses were performed for 295 pSS, 403 SLE, 105 UCTD, 307 RA, 327 SSc and 90 MCTD patients and 254 HCs. Two-tailed pairwise Wilcoxon-rank sum test results are shown. (**D**) *ATP10*/*MX1*/*PARP14* *z*-score analyses were performed in 86 patients from the inception cohort followed up and sampled at the time of recruitment (M0) and at 6 (M6) and/or 14 months (M14). Patients were distributed as anti-Ro60$^+$ (n=29) and anti-Ro60$^-$ (n=57) regardless of the disease (pSS, SLE, UCTD, RA, SSc and MCTD). Pairwise t-tests are shown. Plots show individual values and median, with error bars indicating ± interquartile range.

PRECISESADS Clinical Consortium

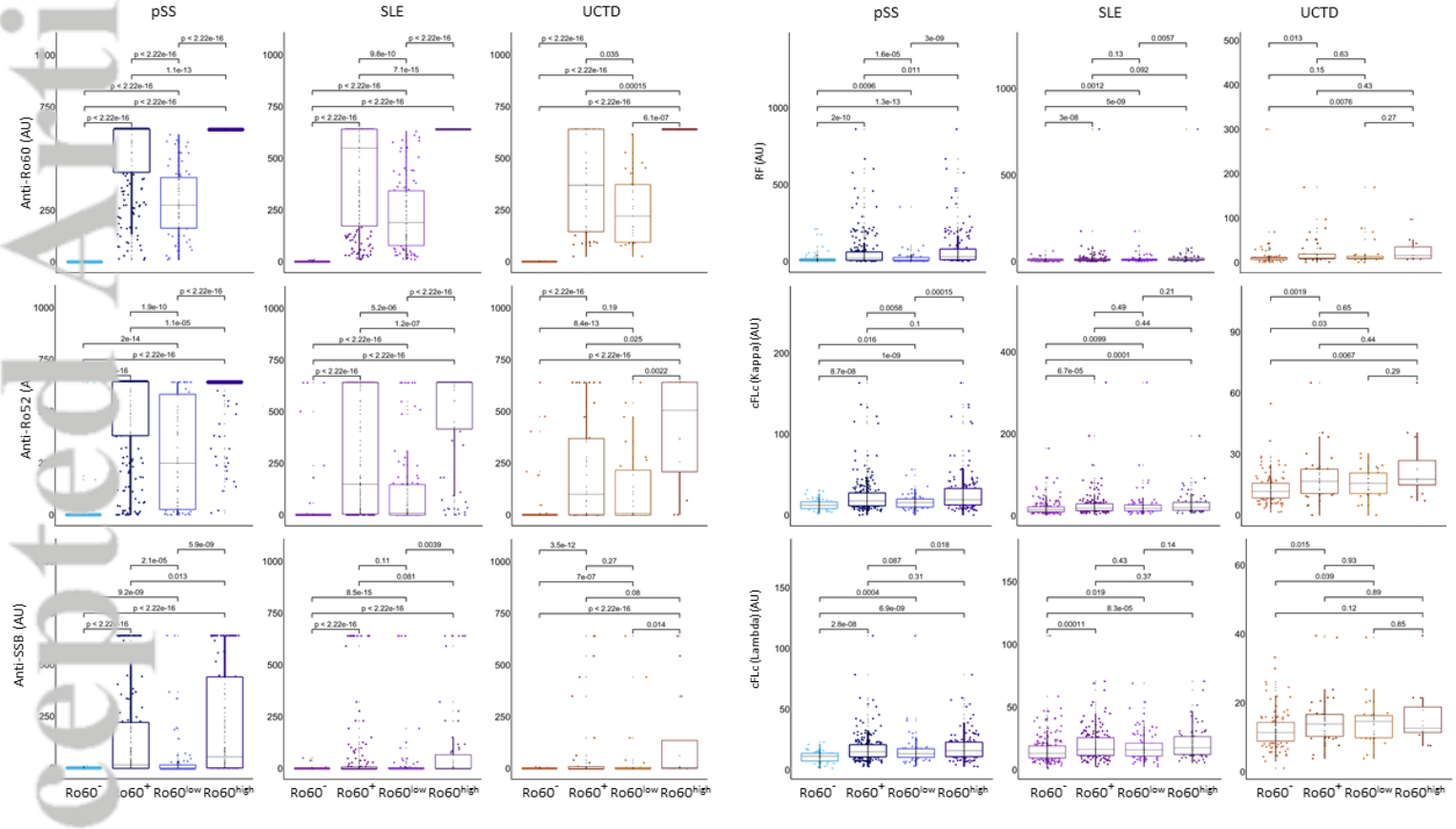
Lorenzo Beretta[1], Barbara Vigone[1], Jacques-Olivier Pers[2,3], Alain Saraux[2,3], Valérie Devauchelle-Pensec[2,3], Divi Cornec[2,3], Sandrine Jousse-Joulin[2,3], Bernard Lauwerys[4], Julie Ducreux[4], Anne-Lise Maudoux[4], Carlos Vasconcelos[5], Ana Tavares[5], Esmeralda Neves[25], Raquel Faria[5], Mariana Brandão[5], Ana Campar[5], António Marinho[5], Fátima Farinha[5], Isabel Almeida[5], Miguel Ángel González-Gay[6], Ricardo Blanco Alonso[6], Alfonso Corrales Martínez[6], Ricard Cervera[7], Ignasi Rodríguez-Pintó[7], Gerard Espinosa[7], Rik Lories[8], Ellen De Langhe[8], Nicolas Hunzelmann[9], Doreen Belz[9], Torsten Witte[10], Niklas Baerlecken[10], Georg Stummvoll[11], Michael Zauner[11], Michaela Lehner[11], Eduardo Collantes[12], Rafaela Ortega-Castro[12], Mª Angeles Aguirre-Zamorano[12], Alejandro Escudero-Contreras[12], Mª Carmen Castro-Villegas[12], Yolanda Jiménez Gómez[12], Norberto Ortego[13], María Concepción Fernández Roldán[13], Enrique Raya[14], Inmaculada Jiménez Moleón[14], Enrique de Ramon[15], Isabel Díaz Quintero[15], Pier Luigi Meroni[16], Maria Gerosa[16], Tommaso Schioppo[16], Carolina Artusi[16], Carlo Chizzolini[17], Aleksandra Dufour[17], Donatienne Wynar[17], Laszló Kovács[18], Attila Balog[18], Magdolna Deák[18], Márta Bocskai[18], Sonja Dulic[18], Gabriella Kádár[18], Falk Hiepe[19], Velia Gerl[19], Silvia Thiel[19], Manuel Rodriguez Maresca[20], Antonio López-Berrio[20], Rocío Aguilar-Quesada[20], Héctor Navarro-Linares[20], Yiannis Ioannou[21], Chris Chamberlain[21], Jacqueline Marovac[21], Marta Alarcón Riquelme[23], Tania Gomes Anjos[23].

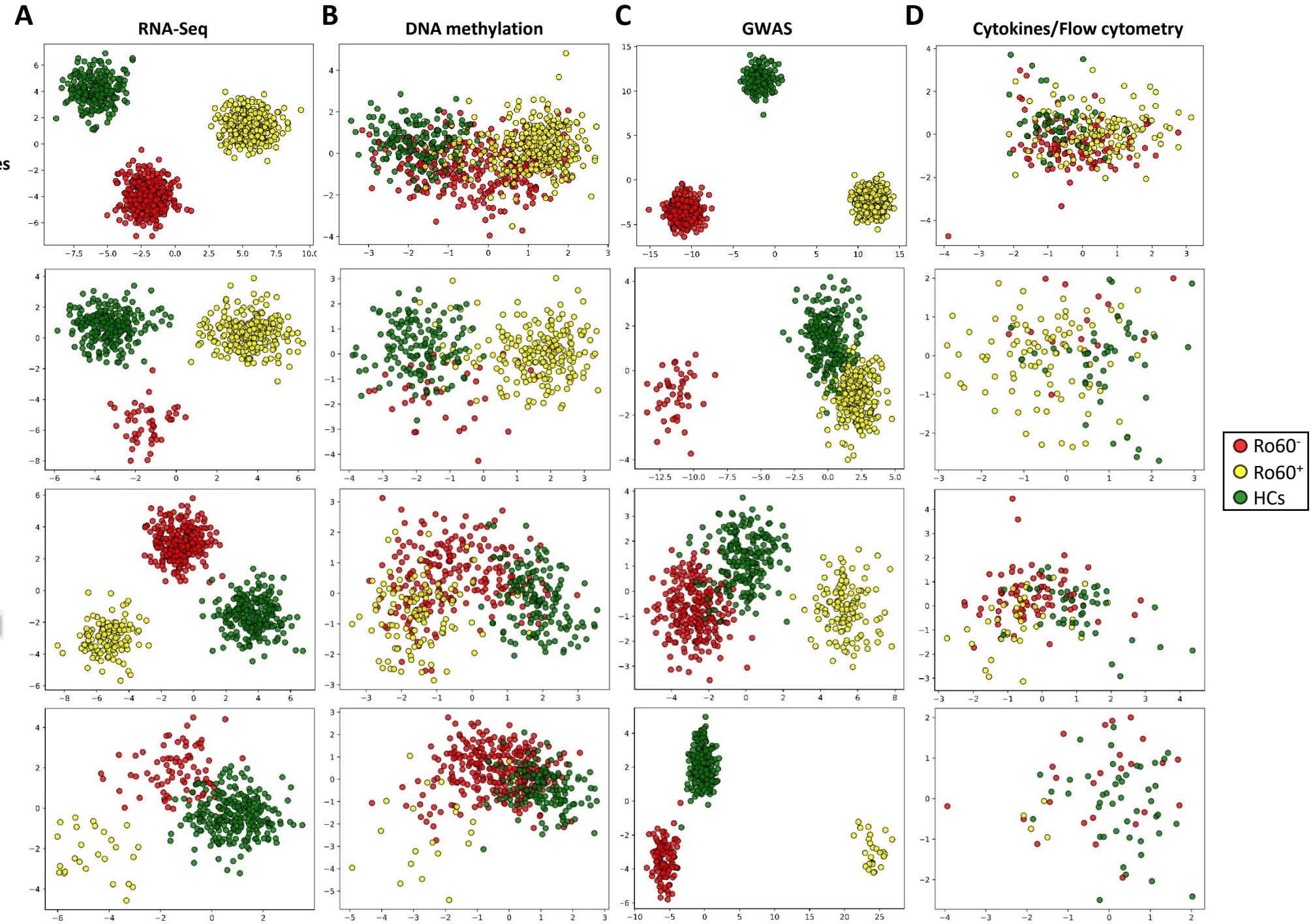
PRECISESADS Flow Cytometry Consortium

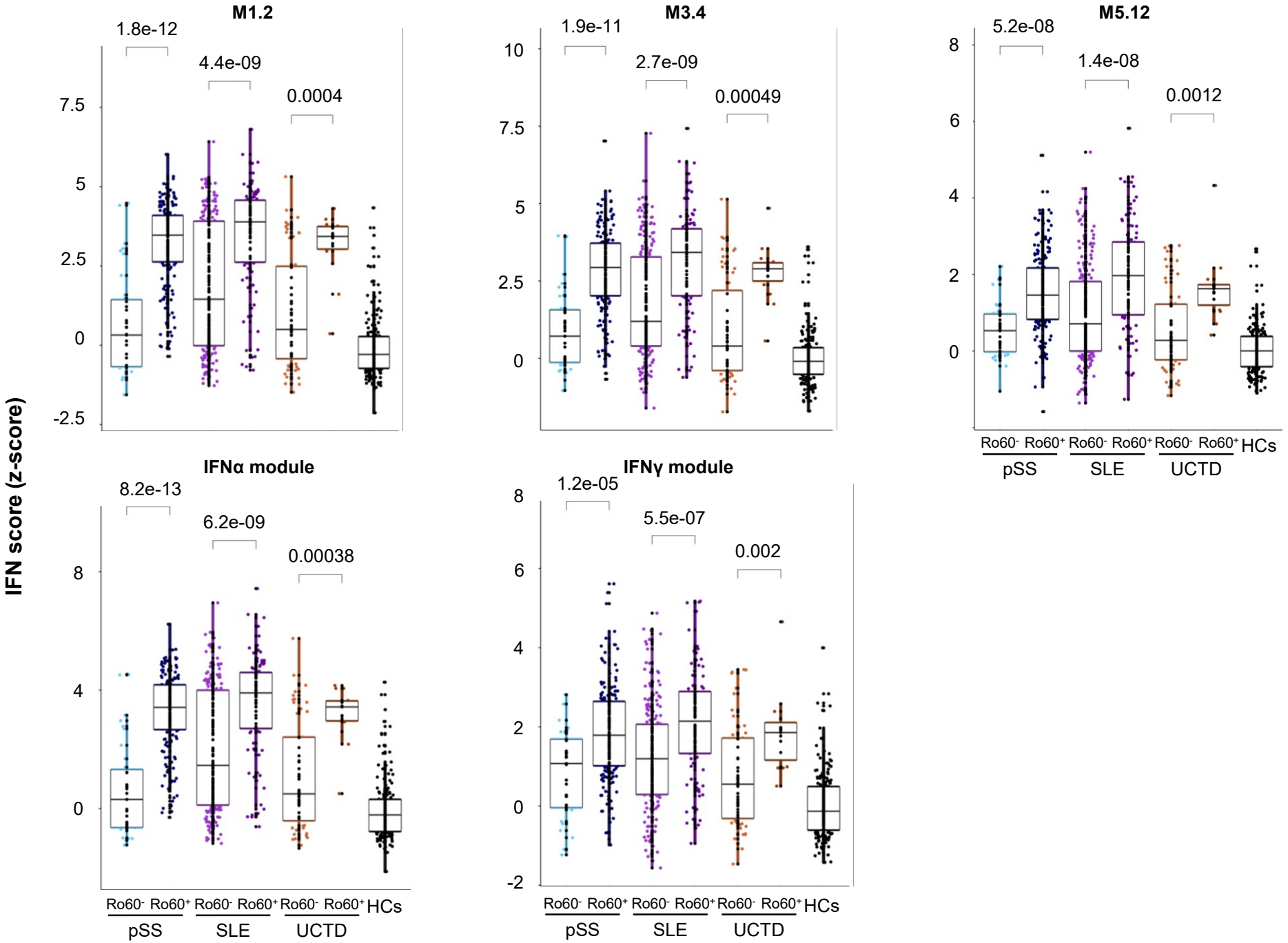
Christophe Jamin[2,3], Concepción Marañón[23], Lucas Le Lann[2], Quentin Simon[2], Bénédicte Rouvière[2,3], Nieves Varela[23], Brian Muchmore[23], Aleksandra Dufour[17], Montserrat Alvarez[17], Carlo Chizzolini[17], Jonathan Cremer[8], Ellen De Langhe[8], Nuria Barbarroja[12], Chary Lopez-Pedrera[5], Velia Gerl[19], Laleh Khodadadi[19], Qingyu Cheng[19], Anne Buttgereit[24], Zuzanna Makowska[24], Aurélie De Groof[4], Julie Ducreux[4], Elena Trombetta[1], Tianlu Li[22], Damiana Alvarez-Errico[22], Torsten Witte[10], Katja Kniesch[10], Nancy Azevedo[25], Esmeralda Neves[25], Maria Hernandez-Fuentes[21], Pierre-Emmanuel Jouve[26] and Jacques-Olivier Pers[2,3]

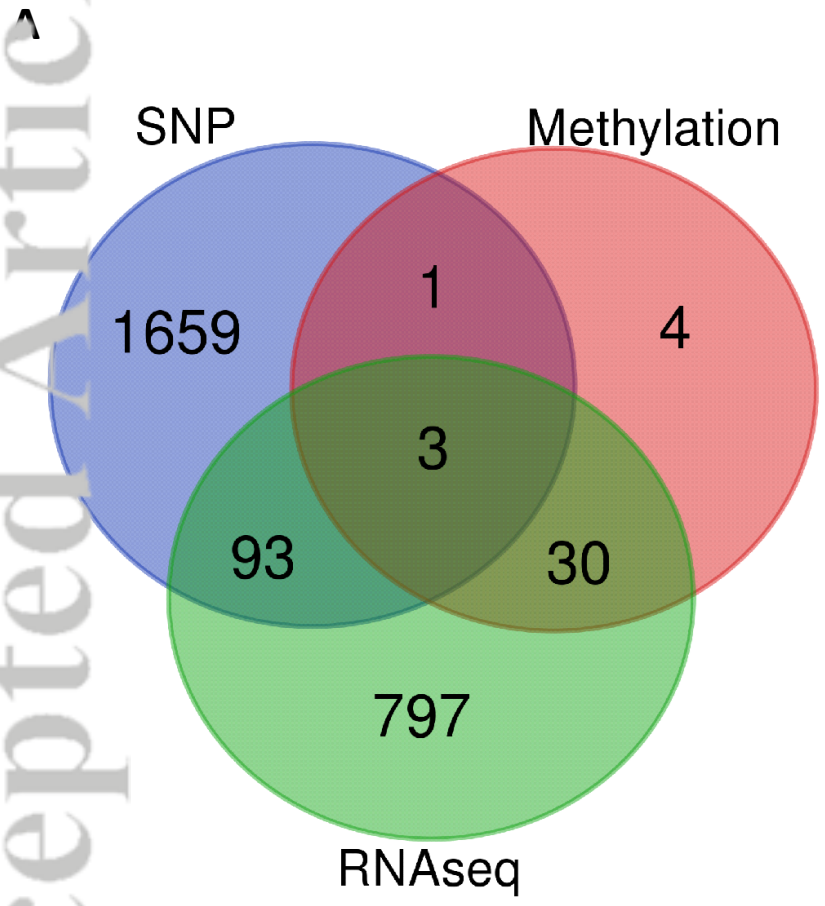
[1]Scleroderma Unit, Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca'Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy. [2]LBAI, UMR1227, Univ Brest, Inserm, Labex IGO, Brest, France. [3]CHU de Brest, Brest, France. [4]Pôle de pathologies rhumatismales systémiques et inflammatoires, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Brussels, Belgium. [5]Centro Hospitalar do Porto, Portugal. [6]Hospital Universitario Marqués de Valdecilla, IDIVAL, Universidad de Cantabria, Santander, Spain. [7]Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Catalonia, Spain. [8]Skeletal Biology and Engineering Research Center, KU Leuven and Division of Rheumatology, UZ Leuven, Belgium. [9]Klinikum der Universitaet zu Koeln, Cologne, Germany. [10]Klinik für Immunologie und Rheumatologie, Medical University Hannover, Hannover, Germany. [11]Medical University Vienna, Vienna, Austria. [12]Reina Sofia Hospital, Maimonides Institute for Research in Biomedicine of Cordoba (IMIBIC), University of Cordoba, Cordoba, Spain. [13]Complejo hospitalario Universitario de Granada (Hospital Universitario San Cecilio), Spain. [14]Complejo hospitalario Universitario de Granada (Hospital Virgen de las Nieves), Spain. [15]Hospital Regional Universitario de Málaga, Spain. [16]Università degli studi di Milano, Milan, Italy. [17]Immunology & Allergy, University Hospital and School of Medicine, Geneva, Switzerland. [18]University of Szeged, Szeged, Hungary. [19]Charite, Berlin, Germany. [20]Andalusian Public Health System Biobank, Granada, Spain. [21]UCB Pharma, Slough, United Kingdom (PRECISESADS Project office). [22]Chromatin and Disease Group, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain. [23]Department of Medical Genomics,

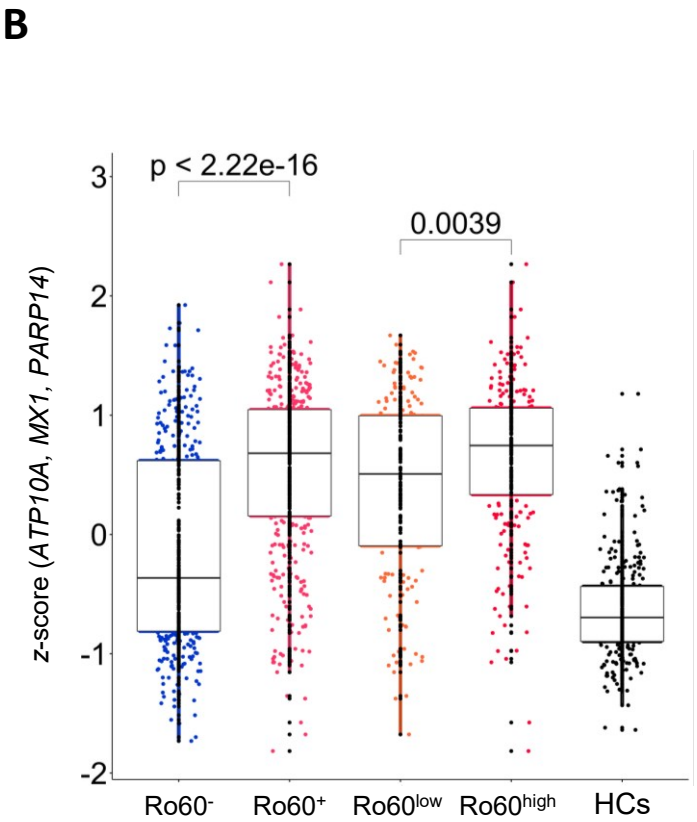
Center for Genomics and Oncological Research (GENYO), Granada, Spain. [24]Pharmaceuticals Division, Bayer Pharma Aktiengesellschaft, Berlin, Germany. [25]Serviço de Imunologia EX-CICAP, Centro Hospitalar e Universitário do Porto, Porto, Portugal. [26]AltraBio SAS, Lyon, France.

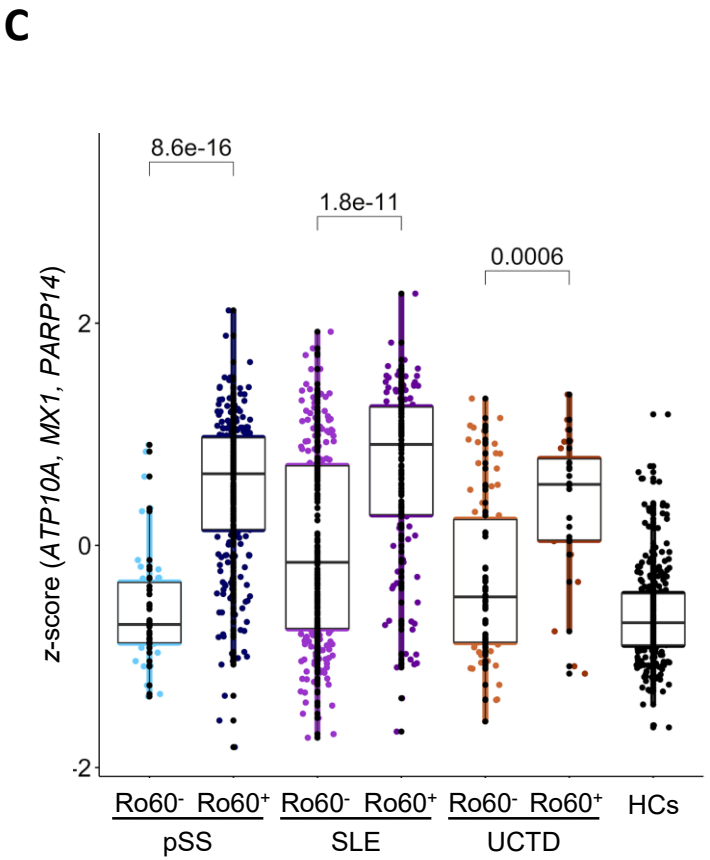


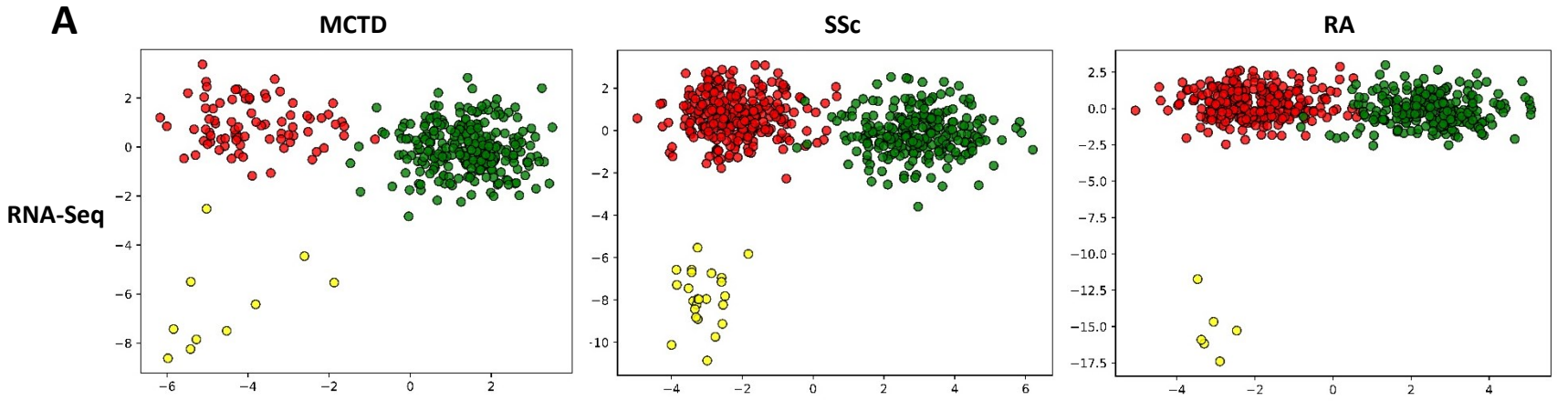


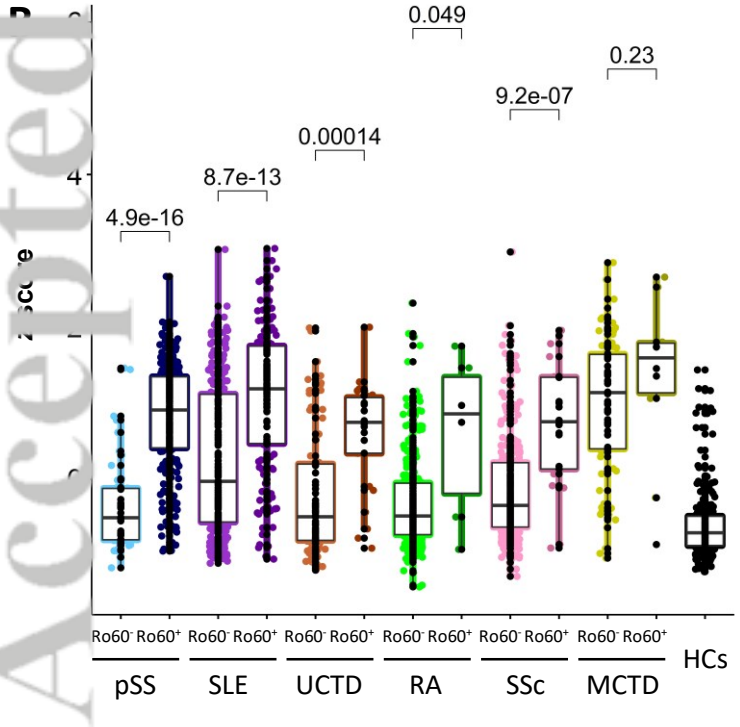


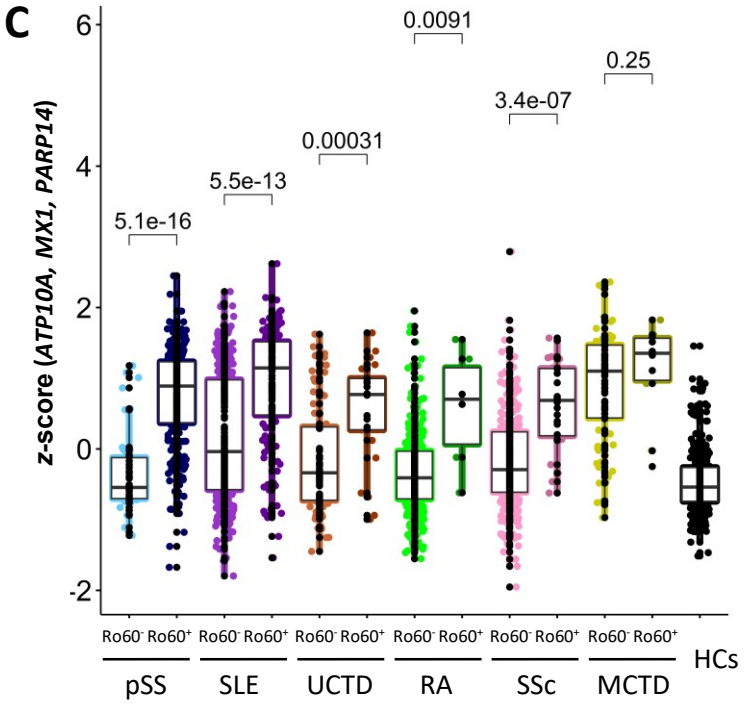












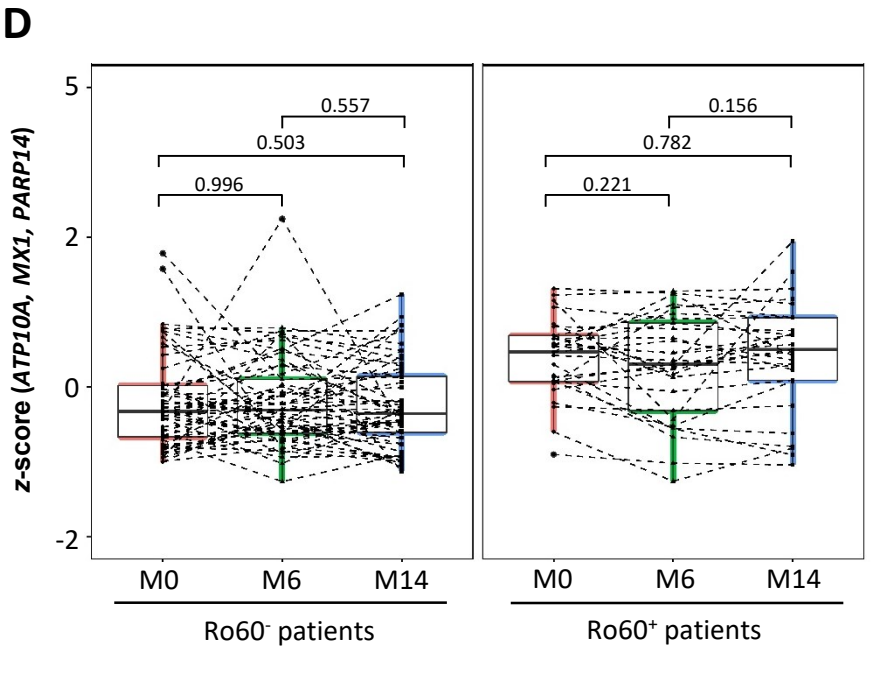


Table 1: Characteristics of the healthy controls (HCs), primary Sjögren's syndrome (pSS), systemic lupus erythematosus (SLE) and undifferentiated connective tissue disease (UCTD) patients according to anti-Ro60 expression (anti-Ro60$^+$ vs anti-Ro60$^-$).

HCs			pSS			
Characteristics	**N**	**N = 279**	**Characteristics**	**N**	**anti-Ro60$^-$, N = 61**	**anti-Ro60$^+$, N = 306**	**p-value[1]**
Age	279		Age	367			0.3
Mean (SD)		52 (9)	Mean (SD)		59 (13)	57 (13)	
MD		0	MD		0	0	
Gender	279		Gender	367			0.5
Female		262 (94%)	Female		57 / 61 (93%)	293 / 306 (96%)	
Male		17 (6.1%)	Male		4 / 61 (6.6%)	13 / 306 (4.2%)	
MD		0	MD		0	0	
Ethnicity	279		Ethnicity	367			>0.9
American Indian/Alaska native		0 (0%)	American Indian/Alaska native		0 / 61 (0%)	0 / 306 (0%)	
Asian		2 (0.7%)	Asian		0 / 61 (0%)	2 / 306 (0.7%)	
Black/African American		0 (0%)	Black/African American		0 / 61 (0%)	1 / 306 (0.3%)	

Caucasian/White		277 (99%)
Native Hawaiian/ Other Pacific Islander		0 (0%)
Other		0 (0%)
MD		0
Obesity	279	16 (5.8%)
MD		1
Cigarette smoking	279	38 (14%)
MD		10

MD: Missing data

Caucasian/White		61 / 61 (100%)	299 / 306 (98%)	
Native Hawaiian/ Other Pacific Islander		0 / 61 (0%)	0 / 306 (0%)	
Other		0 / 61 (0%)	4 / 306 (1.3%)	
MD		0	0	
Obesity	354	7 / 58 (12%)	37 / 296 (12%)	>0.9
MD		3	10	
Cigarette smoking	348	7 / 57 (12%)	29 / 291 (10.0%)	0.6
MD		4	15	
Disease duration	365			0.2
Mean (SD)		9 (8)	10 (8)	
MD		1	1	
Steroids	367			0.3
No		45 / 61 (74%)	244 / 306 (80%)	
Yes		16 / 61 (26%)	62 / 306 (20%)	

	MD	0	0	
Antimalarials	367			0.7
	No	37 / 61 (61%)	194 / 306 (63%)	
	Yes	24 / 61 (39%)	112 / 306 (37%)	
	MD	0	0	
Immunosuppressants	367			0.067
	No	47 / 61 (77%)	264 / 306 (86%)	
	Yes	14 / 61 (23%)	42 / 306 (14%)	
	MD	0	0	
Biologicals	39			>0.9
	No	10 / 10 (100%)	29 / 29 (100%)	
	Yes	0 / 10 (0%)	0 / 29 (0%)	
	MD	51	277	
PGA	339			0.007

	Mean (SD)	30 (18)	24 (19)	
	MD	2	26	
ESSDAI	240	4 (6)	5 (5)	0.11
	MD	15	112	
ESSPRI	183	5.59 (2.29)	4.71 (2.33)	0.029
	MD	17	167	

1 Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test, MD: Missing data, PGA: Physician global assessment

	SLE			
Characteristics	**N**	**anti-Ro60$^-$, N = 333**	**anti-Ro60$^+$, N = 175**	**p-value[1]**
Age	508			0.6
Mean (SD)		46 (14)	45 (13)	
MD		0	0	
Gender	508			0.3

	UCTD			
Characteristics	**N**	**anti-Ro60$^-$, N = 117**	**anti-Ro60$^+$, N = 39**	**p-value[1]**
Age	156			0.6
Mean (SD)		47 (12)	46 (12)	
MD		0	0	
Gender	156			>0.9

Female		302 / 333 (91%)	163 / 175 (93%)	
Male		31 / 333 (9.3%)	12 / 175 (6.9%)	
MD		0	0	
Ethnicity	508			0.010
American Indian/Alaska native		0 / 333 (0%)	0 / 175 (0%)	
Asian		3 / 333 (0.9%)	1 / 175 (0.6%)	
Black/African American		3 / 333 (0.9%)	9 / 175 (5.1%)	
Caucasian/White		318 / 333 (95%)	162 / 175 (93%)	
Native Hawaiian/ Other Pacific Islander		0 / 333 (0%)	1 / 175 (0.6%)	
Other		9 / 333 (2.7%)	2 / 175 (1.1%)	
MD		0	0	
Obesity	491	23 / 321 (7.2%)	15 / 170 (8.8%)	0.5
MD		12	5	
Cigarette smoking	476	60 / 312 (19%)	30 / 164 (18%)	0.8

Female		108 / 117 (92%)	36 / 39 (92%)	
Male		9 / 117 (7.7%)	3 / 39 (7.7%)	
MD		0	0	
Ethnicity	156			0.6
American Indian/Alaska native		1 / 117 (0.9%)	0 / 39 (0%)	
Asian		0 / 117 (0%)	1 / 39 (2.6%)	
Black/African American		1 / 117 (0.9%)	0 / 39 (0%)	
Caucasian/White		113 / 117 (97%)	38 / 39 (97%)	
Native Hawaiian/ Other Pacific Islander		0 / 117 (0%)	0 / 39 (0%)	
Other		2 / 117 (1.7%)	0 / 39 (0%)	
MD		0	0	
Obesity	154	11 / 116 (9.5%)	5 / 38 (13%)	0.5
MD		1	1	
Cigarette smoking	154	17 / 115 (15%)	6 / 39 (15%)	>0.9

	MD	21	11	
Disease duration	508			0.079
	Mean (SD)	14 (10)	12 (9)	
	MD	0	0	
Steroids	508			0.037
	No	179 / 333 (54%)	77 / 175 (44%)	
	Yes	154 / 333 (46%)	98 / 175 (56%)	
	MD	0	0	
Antimalarials	508			0.066
	No	112 / 333 (34%)	45 / 175 (26%)	
	Yes	221 / 333 (66%)	130 / 175 (74%)	
	MD	0	0	
Immunosuppressants	508			0.12
	No	234 / 333 (70%)	111 / 175 (63%)	

	MD	2	0	
Disease duration	155			>0.9
	Mean (SD)	6 (6)	7 (8)	
	MD	1	0	
Steroids	156			<0.001
	No	78 / 117 (67%)	38 / 39 (97%)	
	Yes	39 / 117 (33%)	1 / 39 (2.6%)	
	MD	0	0	
Antimalarials	156			0.5
	No	62 / 117 (53%)	23 / 39 (59%)	
	Yes	55 / 117 (47%)	16 / 39 (41%)	
	MD	0	0	
Immunosuppressants	156			0.045
	No	100 / 117 (85%)	38 / 39 (97%)	

	Yes	99 / 333 (30%)	64 / 175 (37%)	
	MD	0	0	
Biologicals	44			>0.9
	No	27 / 27 (100%)	17 / 17 (100%)	
	Yes	0 / 27 (0%)	0 / 17 (0%)	
	MD	306	158	
PGA	477			0.067
	Mean (SD)	19 (18)	21 (17)	
	MD	16	15	
SLEDAI	253	4(6)	5(5)	0.2
	MD	159	96	

1 Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test, MD: Missing data, PGA: Physician global assessment

	Yes	17 / 117 (15%)	1 / 39 (2.6%)	
	MD	0	0	
Biologicals	34			0.6
	No	23 / 27 (85%)	7 / 7 (100%)	
	Yes	4 / 27 (15%)	0 / 7 (0%)	
	MD	90	32	
PGA	149			0.12
	Mean (SD)	26 (21)	18 (15)	
	MD	5	2	

1 Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test, MD: Missing data, PGA: Physician global assessment