Exploring the potential of lipid mediator profiling in understanding the development and progression of inflammatory diseases

Blood pro-resolving mediators as biomarkers to predict the response of DMARDs treatment in rheumatoid arthritis patients

---

9 Month Progression Report

Esteban Alberto Gomez Cifuentes

Primary supervisor: Dr Jesmond Dalli

Secondary supervisor: Dr Conrad Bessant

William Harvey Research Institute

Barts and The London School of Medicine and Dentistry

Queen Mary University of London

Table of Contents:

Chapter 1: Introduction

Rheumatoid arthritis and treatments

Rheumatoid arthritis (RA) is a chronic autoimmune inflammatory disease characterized by a continuous inflammation of the joints, leading to the progressive destruction of bone and cartilage that results in structural damage, functional disability and a reduction of quality of life [1].
RA is known to affect primarily the synovium [2]. The progression of the disease can be separated into three stages: the non-specific inflammation, the amplification in the synovium and the chronic inflammation [3].

The first stage is characterized by the presence of antibodies, such as rheumatoid factor (RF), anti-perinuclear factor (APF) and anti-cyclic citrullinated peptide (anti-CPP), produced by genetic and enviromental, such as smoking, factors not yet fully understood [3,4]. The accumulation of these antibodies leads to the second stage of the disease. This is characterised by a specific inflammatory reaction in the synovium caused by the formation of new blood vessels that allows the presence of synovial fluid in the joints and the infiltration of different immune cells that promotes inflammation. $CD4^+$ and $CD8^+$ T cells, B cells, and, especially, macrophages and synovial fibroblasts will accumulate to promote synovial hyperplasia, known as pannus [5].

Under normal circumstances the balance between pro and anti-inflammatory soluble mediators, including cytokines, protects the joints and cartilage from the inflammation. However, in RA, pannus formation leads to chronic inflammation with the release of proinflammatory cytokines, such as tumour necrosis factor (TNF), interleukin (IL)-1, IL-6 and granulocyte-macrophage colony-stimulating factor (GM-CSF). These cytokines will promote the production of metalloproteinases, prostaglandins and nitric oxide that contributes to the invasion of the pannus into the cartilage and bone, leading to matrix degradation and joint destruction [5,6].

Recent studies have demonstrated that based on the synovial histological and molecular features, RA can be separated into three categories or pathotypes. These are the Lympho-myeloid phatotype, characterized by the infiltration of T cells, B cells and myeloid cells; Diffuse-myeloid characterized by a predominance of myeloid lineage and a low number of B cells; and Pauci-immune-fibroid characterized by a poor number of immune cells but a large presence of connective tissues. Of note, different pathotypes are linked with distinct disease progression rates, specific symptoms and response to different treatments [7].

Although there is not a cure for RA, different treatments have been used to control the disease. Disease-modifying antirheumatic drugs (DMARDs) are the front line in the

treatment of RA and encompass a series of drugs that slow down the progression of the disease by targeting the processes thought to underlie this disorder [8,9]. DMARDs can be synthetic or biological. The therapies based on biologic agents, such as TNF, IL-1 and RA synovial fibroblast proliferation inhibitors, have demonstrated efficacy against the disease and relief in symptoms [5,6,9]; however, synthetic DMARDs are still preferred, with low-dose methotrexate (MTX) as the most common drug used [8].

Despite MTX being the main drug for the treatment of RA, its precise mechanism of action is still not fully understood. MTX was originally developed as an antimetabolite drug, interfering with the folate pathway, responsible of the purine/pyrimidine synthesis (and therefore the DNA and RNA synthesis), through the inhibition of the dihydrofolate reductase (DHFR) [10,11]. It was proposed that its action was associated with the inhibition of proliferation of cells involved in inflammation; however, low-dose of MTX administered with folic acid did not change the efficacy of the treatment in RA patients, which indicated that the inhibition of purine/pyrimidine synthesis is not the main mechanism of MTX in RA [10,11]. Other proposed mechanisms include the regulation of CD39 in peripheral blood regulatory T-cells, enhanced adenosine release, inhibition of transmethylation reactions required for diverse cellular functions, nitric oxide synthase uncoupling and alteration of the cytokine profile [10–12].

Currently, MTX in combination with other DMARDs, such as hydroxychloroquine and sulfasalazine, is the first regime used against RA since these combinations can improve the efficacy of the treatment. Nonetheless, the presence of unwanted side effects, such as reduction of T cells counts, bone marrow toxicity and liver function abnormalities [8]; the high percentage of patients who do not respond to these treatments and the fact that most of the patients rarely go into full remission complicate disease control [13].

Even though the advances in RA treatments in the last 20 years have changed the way the disease is managed; RA still represents an economic burden for both patients and health services, affecting around 1% of the worldwide population [14]. It is then necessary, besides the development of new treatments, to discover biomarkers that can help to predict whether or not a patient will respond to the currently available therapies.

Specialised pro-resolving mediators in RA

It’s now well known that RA may arise from a decreased ability of the host immune response to control the inflammation and avoid the chronicity of the disease [15]. In the past it was thought that resolution of the inflammation was a passive process originated by the dissipation of pro-inflammatory molecules; however, the role of a series of mediators, known as specialized pro-resolving mediators (SPM), in the termination of inflammation has refuted that theory [16]. The first group of discovered

SPM, the lipoxins, were originally identified in extracts of human leukocytes incubations exposed to arachidonic acid metabolites using liquid chromatography tandem mass spectrometry (LC-MS/MS) based lipid mediator lipidomics [17]. Since then, other families of SPM has been discovered including resolvins, protectins and maresins [16].

SPMs are produced by immune cells via the enzymatic conversion of essentials fatty acids such as docosahexaenoic acid (DHA), n-3 docosapentaenoic acid (n-3 DPA), eicosapentaenoic acid (EPA) and arachidonic acid (AA). The SPM pathways start with the insertion of oxygen into the fatty acids backbone, a reaction primarily carried out by lipoxygenases (ALOX), cyclooxygenase-2 (COX-2) and cytochrome P450 [16,18]. Each of the fatty acids gives rise to structurally distinct SPM families.

DHA by the action of ALOX15 and ALOX5 produces Resolvins (Rv) D series, protectins (PD) and Protectin conjugates in tissue regeneration (PCTR); while ALOX12 converts this fatty acid to Maresins (MaR) and Maresin conjugates in tissue regeneration (MCTR). When COX-2 is acetylated in the presence of aspirin, Aspirin-triggered resolvins (17R-RvDs) and protectins (17R-PDs) are produced [19]. Resolvins counter regulate proinflammatory mediator actions and promote resolution via macrophages uptake of debris and clearing microbes. Protectins display anti-inflammatory, neuroprotective and antiapoptotic actions; while maresins promote macrophage phagocytosis and play a role in pain reduction and tissue regeneration [20].

n-3 DPA is precursor to structurally distinct pro-resolving mediators, namely: 13-series resolvins (RvTs), biosynthesized via the action of COX-2 and ALOX5. RvTs down-regulate the production of proinflammatory molecules and regulate leukocyte trafficking and bacterial clearance [20,21]. $RvD_{n\text{-}3\ DPA}$ are produced through the enzymatic action of ALOX15 and ALOX5, whereas the biosynthesis of $PD_{n\text{-}3\ DPA}$ is initiated by ALOX15. Whereas, ALOX12 is the initiating enzyme in the $MaR_{n\text{-}3\ DPA}$ pathway. n-3 DPA derived $RvD_{n\text{-}3\ DPA}$, $PD_{n\text{-}3\ DPA}$ and $MaR_{n\text{-}3\ DPA}$ have demonstrated to reduce neutrophil infiltration in peritonitis models. These mediators also counterregulate the production of inflammatory cytokines, enhancing macrophage phagocytosis and also reducing neutrophil chemotaxis and adhesion to endothelial cells [20–22].

E-series resolvins (RvEs) are produced from EPA by the action of COX-2/CYP450 and ALOX15 (RvE3) or ALOX5 (RvE1 and RvE2). RvEs stop neutrophil infiltration, suppress oxidative stress in periodontitis models and stimulate the phagocytosis of apoptotic cells, debris and bacteria [16,18].

AA, in addition of producing proinflammatory mediators such as prostaglandins (PG), leukotrienes (LK) and thromboxane (TX), is the substrate for Lipoxins (LX), produced by either COX-2 or ALOX15, and ALOX5. LX inhibit chemotaxis and the production of proinflammatory cytokines by neutrophils, monocytes and macrophages. LX also help

in the migration of neutrophils during inflammation and the clearing of these cells during the resolution step [23].

In the inflammation process, as the case of RA, proinflammatory lipid mediators such as PGE_2, LTB_4 and PGD_2 play a role mediating pain, heat and swelling, while SPM help to control the ongoing inflammation. It has been observed that an increase of SPM in plasma reduces joint inflammation and promotes joint protection in RA patients. An increase of RvE2 in the synovial fluid has been associated with decreased of joint pain, while during ongoing joint inflammation there is a downregulation of several SPM including RvDs and aspirin-triggered RvDs [15,24]. Increase of RvTs production was associated with a decrease of disease activity in the joint and reduction of leukocyte migration and activation [21]; while the presence of PGE_2 in chronic inflammation mediates the production of LXA_4 and 15R-LXB_4, essentials to restore the homeostasis after inflammation in RA [25]. The evidence of resolution activity of SPM in RA, suggest that these mediators can be considered as possible new therapies, and, in the same note, can be used as biomarkers of the disease.

Bioinformatic strategies to identify biomarkers

The identification of biomarkers by a conventional approach, meaning identifying a candidate and running tests to validate it, requires a great amount of biological evidence, is time-consuming, expensive and risky. To date, new techniques to generate and collect large amounts of data, such as DNA/RNA sequencing and mass spectrometry, have been used in biology research, including drug discovery and biomarker development, for the analysis of a high number of molecules, improving the chances of identifying markers of interest [26]. Specifically, mass spectrometry, used for either proteomics or metabolomics, can provide a precise picture of biological processes from cell-free biological fluids; for example, lipid mediator profiles can be obtained from serum, urine and synovial fluid.

The value of the data generated by mass spectrometry depends on the strategies used to analyse it. The comparison of samples with different conditions is mandatory for the identification of putative biomarkers. Machine learning methodologies, algorithms to build mathematical models based on patterns and interferences that then can make decisions [27], have been applied to metabolomics data to diagnose illness and, in a more advanced application, to classify and predict outcomes of different diseases, and therefore for the identification of biomarkers associated with them. These biomarkers can be used to determine the efficacy of pre-existing treatments [27–30].

Biological and medical data are complex, containing multiple layers of information that are usually incomplete, heterogeneous, noisy and present unwanted variability (e.g, technician and batch effect) [28]. Machine learning, different than univariate association analyses, can address some of the issues associated with biological data while also

provides strategies to consider unknown associations between the different variables present in the samples [31].

Furthermore, the combination of machine learning with other bioinformatic strategies, such as pathway analysis, exploring how different responses affects the production of SPM, and the identification of genomic variants (single nucleotide polymorphisms; SNPs) associated with RA and the dysregulation of resolution of inflammation, can generate a snapshot of the dynamic of the lipid mediators in RA and other inflammatory diseases.

Therefore, this study aims to demonstrate if the plasma SPM concentrations are predictive of DMARD treatment in RA patients. For this purpose I will use plasma from deeply characterized early-arthritis patients and machine learning methodologies In addition, we expect to establish novel tools to interrogate the lipid mediator biosynthetic pathways and explore the utility of combining lipid mediator profiles. I will also utilize DNA/RNA sequencing in order to provide new insights into molecular mechanisms that may be linked with differential SPM regulation in DMARD responders and non-responders.

Objectives

I will use a multidisciplinary approach to investigate the following aims:

Aim 1: Using lipid mediator profiles to identify protective mechanisms engaged by therapeutics in patients with inflammatory conditions.

- *Lipid mediator profiling using LC-MS/MS in deeply phenotyped cohorts of individuals with RA*: we will use samples of individuals with RA that were treated with either MTX (~100 patients) or anti-TNF-α (~100 patients; obtained from PEAC and MATURA cohorts). Lipid mediator profiling will be obtained using a well established LC-MS/MS platform to assess whether specific SPM may predict the outcome of treatment in individuals with RA.

- *Machine learning to identify mediators/mediator families that are regulated by different therapeutics:* in order to gain insight into the regulation of specific SPM families by distinct therapeutics, we will use machine learning to interrogate lipid mediator profiles described before. Here we will build classification models with optimal predictive capability and then we will investigate these to determine which mediators/mediator families are most important in the prediction.

- *Machine learning models evaluation and improvement:* independent cohorts of RA patients will be used to improve the models designed in previous steps, as

well to evaluate their prediction performance. This with the aim of building strong predictors that can be used in the medical field.

Aim 2: Construct lipid mediator biosynthetic networks to gain insights into the etiopathology of inflammatory diseases and responses to treatment.

- *Establish a novel tool to interrogate lipid mediator biosynthetic networks:* biosynthetic networks for each of the lipid mediator metabolomes will be constructed having in consideration the different components of each of the lipid mediator pathways including the substrate, precursor and further metabolite concentrations for each of the mediator families as well as the biosynthetic enzymes.

- *Biosynthetic networks to identify pathways that become dysregulated in inflammatory diseases:* We will next use these biosynthetic networks to assess which mediator pathways become dysregulated in RA patients with distinct disease phatotypes.

- *Biosynthetic networks to identify factors regulating lipid mediator biosynthetic pathways:* The biosynthetic pathways developed in the previous steps will be used to identify potential regulators of the biosynthetic pathways found to be differentially expressed in the distinct patient groups. These will be validated by measuring the expression of molecules known to regulate the specific biosynthetic enzymes found to be differentially regulated.

Aim 3: Identify genomic variants associated with the development of inflammatory disorders and explore their link with dysregulated lipid mediator profiles.

- *Data and samples collection:* Genomic, phenotypic and clinical data from patients with RA will be obtained using health resources such as UK Biobank (~500,000 individuals), East London Genes & Health (~100,000 individuals) and The 100,000 Genomes Project.

- *Identification of associations between polymorphisms and the response to different treatments in RA patients*: we want to identify SNP associated to the presence of RA and run correlation analysis between the genotypes and the different responses to treatmente with the aim of trying to understand the role of the genetic variance in the progression of the disease. We plan to run quality control over the identified SNPs in the data and create logistic regression models to find these associations. False discovery rate and other relevant artefacts when working in Genome-wide association studies will be considered if required.

- *Identify the possible correlation between polymorphisms of interest and SPM dysregulation:* once identified relevant polymorphisms that can be associated with a dysregulation of SPM, we plan of requesting samples from those patients to evaluate how genetic variants can have an effect in the lipid mediator profile and how this association is related with progression of inflammatory diseases.

Chapter 2: Methods

Materials

Liquid chromatography (LC)-grade solvents were purchased from Fisher Scientific (Pittsburgh, PA, USA); Poroshell 120 EC-C18 column (100 mm x 4.6 mm x 2.7 µm) was obtained from Agilent (Cheshire, UK); C18 SPE columns were from Biotage (Uppsala, SE); synthetic standards for LC-tandem mass spectrometry (MS-MS) quantitation and deuterated (d) internal standards (d8-5S-HETE, d5-RvD2, d5-LXA_4, d4-PGE_2, d4-LTB_4, d5-LTE_4, d5-LTD_4 and d5-LTC_4) were purchased from Cambridge Bioscience (Cambridge, UK) or provided by Charles N. Serhan (Harvard Medical School, Boston, Massachusetts, USA.).

Pathobiology of Early Arthritis Cohort

Baseline and 6 months post-treatment plasma samples were obtained from the Pathobiology of Early Arthritis Cohort (PEAC). The PEAC cohort study was approved by the King's College Hospital Research Ethics Committee (REC 05/Q0703/198). Following written informed consent, peripheral blood samples and synovial tissue were obtained from patients recruited at Barts Health NHS Trust into the Pathobiology of Early Arthritis Cohort (PEAC, http://www.peac-mrc.mds.qmul.ac.uk) undergoing ultrasound (US)-guided synovial biopsy of the most inflamed joint (knee, wrist or small joints of hands or feet) [7]. All patients were disease-modifying anti rheumatic drugs (DMARDs) and steroid-naïve, had symptoms duration less than 12 months and fulfilled the ACR/EULAR 2010 classification criteria for RA. RA individuals were categorised into three pathotypes based on histological classification of synovial tissue: Lympho-myeloid, Diffuse-Myeloid and pauci-immune Fibroid. Patients were treated with DMARDs. Response status after 6 months of mixed DMARD therapy was determined by EULAR response criteria based on DAS28-ESR.

Lipid mediator profiling

Protein precipitation:

Plasma was obtained from peripheral blood following centrifugation at 2500 rpm for 10 min at room temperature. Methanol (MeOH) containing each deuterated internal standard (IS; 500 pg each) was added to samples for lipid mediator profiling. The proportion MeOH+IS/sample was 4:1 (vol/vol); in this case, to 0.5 mL of samples was added 2 mL of MeOH+IS. 100% samples, MeOH+IS that did not contain any sample and did not get extracted, were prepared to allow the estimation of recovery.

The samples were kept on ice at least 30 minutes before protein precipitation by centrifugation (3000 rpm, 4°C, 10 minutes). The supernatant was collected, excess of MeOH was evaporated using a gentle stream of nitrogen gas and a TurboVap LV (Biotage) system at 37°C until a final volume of 1 mL.

Lipid mediator extraction:

Samples were extracted using ExtraHera (Biotage) automated solid-phase extraction with ISOLUTE C18 columns (500 mg, 3 mL; Biotage). C18 columns were placed in the ExtraHere while elution tubes were added to the collection plates (A for methyl formate fractions and B for MeOH fractions). The extractions consisted in the conditioning of the C18 columns with MeOH, follows by the loading of the samples onto the columns with pH 3.5 water. A further washing step with pH 7 water and hexane was done to neutralize the acid in the C18 columns and to elute hydrophobic molecules. This was done four times. Lipid mediators were then eluted in collection tubes with the addition of methyl formate (MF) followed by the addition of MeOH.

The MF fraction contains the Rv, MaR, PD, PG, LX, LT and TX lipid mediator families, while the MeOH fraction contains the PCTR, MCTR and cys-LT lipid mediator families.

Solvent drying:

After extraction, samples were transferred into 15 mL falcon tubes and placed in the TurboVap LV. The 100% samples were also transferred into the TurboVap LV at this stage. Samples were drying as follow: with a constant flux of nitrogen, the walls of the falcon tube were washed when the solvent was dried, twice with MF and once with MeOH. 40 µL solution containing MeOH and water in 1:1 (vol/vol) was added after drying and vortex was applied immediately after. Samples were then centrifuged at 3000 rpm for 5 min and the supernatant was transferred to a glass injection vial. The insert was placed into Eppendorf tube and then centrifuged at 9900 rpm for 10 seconds. The insert was observed to identify the presence of a pellet. In case there was one, the supernatant was transferred to another glass insert; otherwise, the insert was transferred directly to an amber vial.

Liquid chromatography tandem mass spectrometry (LC-MS/MS):

Samples were injected on a Shimadzu LC-20AD HPLC and a Shimadzu SIL-20AC autoinjector, paired with a QTrap 5500 or QTrap 6500 plus (Sciex). An Agilent Poroshell 120 EC-C18 column (100 mm x 4.6 mm x 2.7 µm) was kept at 50°C and mediators eluted using a mobile phase consisting of MeOH/water/acetic acid of 20:80:0.01 (vol/vol/vol) that was ramped to 50:50:0.01 (vol/vol/vol) over 0.5 min and then to 80:20:0.01 (vol/vol/vol) from 2 min to 11 min, maintained till 14.5 min and then rapidly ramped to 98:2:0.01 (vol/vol/vol) for the next 0.1 min. This was subsequently maintained at 98:2:0.01 (vol/vol/vol) for 5.4 min, and the flow rate was maintained at

0.5 ml/min. QTrap 5500 or QTrap 6500 plus were operated using a multiple reaction monitoring (MRM; Appendix 1-3 for MRM settings in detail), which monitors the mass-to-charge ratio (m/z) of Q1 (parent ion) and a Q3 (characteristic daughter ion) for each lipid mediator. This generates a chromatogram with a peak at a retention time (RT) that reflects the hydrophilicity and the strength of the interaction with the C18 column of each mediator. For the correct identification of the RT, a mix of known lipid mediator standards was also profiled to check any shift in the individual RTs. The LC-MS/MS methodology also generated a full MS/MS scan when the ion count in the MRM transition is above a pre-set threshold. The MS/MS spectrum is a representation of the m/z fragments produced when the parent ion is fragmented in the collision chamber of the mass spectrometer [21].

The criteria used to corroborate the identity of each lipid mediator were the following: the presence of a peak with a minimum area of 2000 counts, matching RT to synthetic or authentic standards with maximum drift between the expected RT and the observed RT of 0.05 seconds, a peak with a minimum of 5 data points and matching of at least 6 diagnostic ions to that of the reference standard, with a minimum of one backbone fragment being identified.

Calculation of lipid mediators concentration:

Calibration curves were obtained for each lipid mediator using synthetic and authentic lipid mediator mixtures at 0.78, 1.56, 3.12, 6.25, 12.5, 25, 50, 100, and 200 pg that gave linear calibration curves with an r^2 values of 0.98–0.99. To determine these concentrations, the Beer-Lambert law was applied to convert ultraviolet (UV) absorbance to concentration:

$$A = \varepsilon . l . c$$

where *A* is UV absorbance, ε is absorbance coefficient, *l* is the length of solution light passes through and *c* is the concentration of the solution.

When a specific mass is injected into the LC-MS/MS, a specific area under the curve of the MRM peak will be generated. This area was used to construct the standard curve for each lipid mediator.

To calculate the loss percentage of lipid mediators during the extraction step, a comparison between the IS concentrations in the samples and the concentrations of those in the 100% sample were made. The percentage of recovery can be calculated since it's known the amount of IS in the 100% and the samples before the solid-phase extraction:

$$Recovery\ (\%) = (pg\ IS\ in\ sample / pg\ IS\ in\ 100\%\ sample) * 100$$

Knowing the percentage of recovery, we calculated the concentrations correcting the loss of lipid mediators. The percentage of recovery of the IS that best represents the properties of the lipid mediator in question was used.

Data

The data used for the machine learning models consisted of the lipid mediator profiles (54 lipid mediators) of patients with RA who responded (n = 30) or did not (n = 24) to the treatment with DMARDs for a PEAC-derived patient cohort. The lipid mediator profile included DHA-derived resolvins (RvD1, RvD2, RvD3, RvD4, RvD5, RvD6, 17-RvD1 and 17-RvD3), protectins (PD1, 17-PD1, 10S,17S-diHDAH and 22-OH-PD1), PCTRs (PCTR1, PCTR2 and PCTR3), maresins (MaR1, MaR2, 7S,14S-diHDHA, 4,14-diHDHA, 14-oxo-MaR1 and 22-OH-MaR1), MCTRs (MCTR1, MCTR2 and MCTR3), 13-series resolvins (RvT1, RvT2, RvT3 and RvT4), n-3 DPA-derived resolvins ($RvD1_{n\text{-}3\ DPA}$, $RvD2_{n\text{-}3\ DPA}$ and $RvD5_{n\text{-}3\ DPA}$), n-3-DPA derived protectins ($PD1_{n\text{-}3\ DPA}$ and 10S, 17S-diHDPA), n-3 DPA-derived maresins ($MaR1_{n\text{-}3\ DPA}$), E-series resolvins (RvE1, RvE2 and RvE3), leukotrienes (LXA_4, LXB_4, 5S,15S-diHETE, 20-OH-LTB_4, 20-COOH-LTB_4, 6-trans-LTB_4 and 12-epi-6-trans-LTB_4), cysteinyl leukotrienes (LTC_4, LTD_4 and LTE_4), prostaglandins (PGD_2, PGE_2, PGF_{2a}) and thromboxane (TXB_2).

A Clinical Score model was obtained using the following available clinical parameters: disease duration, erythrocyte sedimentation rate (ESR), rheumatoid factor (RF titre), tiredness visual analogue scale (VAS), pain VAS, patient global health VAS, physician global assessment VAS, swollen joints number, disease activity score-28 (DAS28) and 12 max US Synovial Thickness and US Power Doppler scores. Age, sex and clinical parameters not mentioned before were not considered for this first approximation of creating a model able to classify the response of RA patients to DMARD treatment.

Machine learning models

Data were preprocessed and analysed using R Software (version 3.5.1; https://www.r-project.org/) and RStudio environment (version 1.1.456; https://www.rstudio.com/).

From the exploratory analysis, two samples were removed for showing outlier concentrations of TXB_2, which is likely due to coagulation during sample collection, and an additional sample was removed due to lack of clinical records. Although no normalization was required since all the lipid mediator concentrations were calculated based on the same amount of standard, the concentrations were scaled by subtracting the mean and dividing by the standard deviation of each feature. Scaling has proved to be an important step during data preparation in machine learning since it helps to normalize the magnitudes of all the features in a common range, which avoids misleading results due to the overrepresentation of features with high absolute values [32]. In this case, for example, prostaglandin, leukotrienes and thromboxane levels are regularly higher in plasma when compared with other lipid mediators, and this can

result in machine learning models wrongfully based only on proinflammatory lipid mediators.

Two supervised machine learning methodologies were used to create the classifier models: Support Vector Machine (SVM) [33] and random forest [34]. SVM separates groups by organizing the samples in two spaces divided by a hyperplane in a way that the distances between the samples in the same group are not too wide and the distance between the groups is as large as possible [35]. The algorithm increases systematically the dimensionality of the data using a kernel function until it finds the optimal hyperplane that serves as a threshold between the groups [36]. The kernel function used here was a nonlinear kernels radial basis function, meaning that new samples will be classified based on how close they are to the samples used to create the SVM model [36,37].

The SVM models were created using the R Package "classyfire" (https://cran.r-project.org/src/contrib/Archive/classyfire/), that uses bootstrapping (new datasets are created from the original data by randomly sampling with replacement) as the validation strategy. In order to identify the best model, we created models testing different times of the resampling and a different number of ensembles (fusion of the individual classifiers created during the bootstrapping step), with 70 bootstrap iterations and 70 individual classifiers in each ensemble that gave stable models for all groups tested. Furthermore, we also used the inbuilt automatic optimization step that includes minimization of the bootstrapping error to improve and validate the models [37].

Random forest operates by getting the consensus of weak decision tree classifiers [35]. The decision trees are created using the features as vertices and classes as leaves; each tree is designed using a different set of randomly chosen features [38]. It's useful that the algorithm does not use all the available features to create each tree because this allows identification of the stronger predictors in the data set and permits further analysis to improve the prediction models [39].

In the present studies, we applied the random forest methodology using the R package "randomForest" (https://cran.r-project.org/package=randomForest), which also uses bootstrapping as the validation method. We first created a small loop to define the optimal number of variables randomly sampled as candidates at each split (the *mtry* value); with the *mtry* value that gives the best classification performance for each model, we tested a number of *ntree* (number of decision trees used to create the consensus classifier tree) to find the value that gave us the most stable models. Here we found that a *ntree* of 10,000 gave us stable models for all the lipid mediator groups. Increasing the number of *ntree* beyond this value did not markedly improve the outcomes.

Statistical analysis

We performed all statistical analyses and data derivation using R (https://www.r-project.org/), SIMCA 14.1 (Umetrics, Umea, Sweden) and MetaboAnalyst 4.0 [39,40]. Results represented in the lipid mediator concentration table are displayed as mean ± sem.

Sample sizes for each experiment were determined on the variability observed in prior experiments. Exploratory analysis were performed using Orthogonal partial least squares-discriminant analysis (OPLS-DA) by MetaboAnalyst 4.0 and SIMCA 14.1 software after mean centering and unit variance scaling of lipid mediator concentrations. OPLS-DA is based on a linear multivariate model that identifies variables that contribute to the class separation of observations (e.g. treatment response) on the basis of their variables (lipid mediator concentrations). During classification, observations were projected onto their respective class model. The plot illustrates the clusters among the observations based on the variable importance in projection (VIP) score, which identify the contribution of each mediator in the observed separation between groups. These analyses were made with the purpose to confirm if lipid mediator profiles were able to differentiate between DMARD responders and non responder and identify possible outliers.

Data and software availability

R scripts used for the machine learning methodologies, together with the data and expected results (including tables and figures), are found in the GitHub repository: https://github.com/eagomezc/2020_Biomarkers_identification_ML_and_RA

Chapter 3: Results

Plasma SPM concentrations can predict the response to DMARD treatment

To determine whether peripheral blood SPM concentrations are able to predict the response to DMARD treatment in patients with RA, we measured the plasma lipid mediator concentrations in a matched, deeply phenotyped early RA patients cohort before treatment initiation (Appendix to see patients information). Plasma lipid mediators were identified using LC-MS/MS strategy and criteria including matching of retention time of lipid mediators in liquid chromatography when compared against standards and the identification of at least 6 diagnostic ions in the tandem mass spectrum (Appendix 5 to see the criteria in detail). In RA patient plasma we identified lipid mediators from all four major essential fatty acid metabolomes: DHA, n-3 DPA, EPA and AA (Appendix 6). We next used OPLS-DA, a multivariate analysis that generates a regression model based on continuous variables to discriminate between two conditions [40], to define the difference between DMARD-responders and DMARD-non-responders patients based on the lipid mediator profile. OPLS-DA was able to identify two clusters representing each of the patient groups (Figure 1A, B).

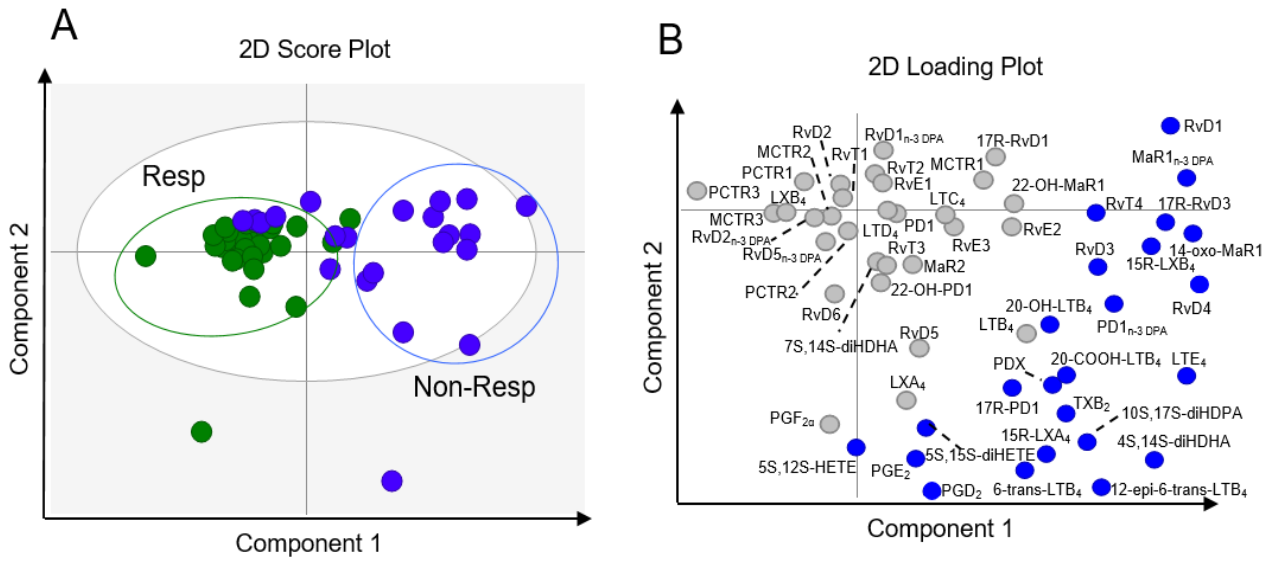


Figure 1. Plasma SPM can differentiate between DMARD-responders and DMARD-non-responders in RA patients. Plasma was collected from RA patients prior to the initiation of treatment with DMARDs and lipid mediator concentrations established using LC-MS/MS based lipid mediator profiling (see methods for details). OPLS-DA of peripheral blood lipid mediator concentrations for DMARD-responders (Resp) and DMARD-non-Responders (Non-Resp). (**A**) 2-dimensional score plot with the grey circle representing the 95% confidence regions. (**B**) 2-dimensional loading plots. Lipid mediators with VIP score greater than 1 are highlighted in blue and upregulated in Non-Resp. Results are representative of n = 30 Resp and n = 22 Non-Resp.

We then applied machine learning methodologies to create predictive models of the responsiveness to DMARD treatment in RA patients using plasma lipid mediator profiles. We focused on two machine learning methodologies, SVM and random forest, to build models based on either all the lipid mediators, SPM from each of the fatty acid metabolomes or the clinical score. First, we established the optimal parameters to get the most stable models of each methodology by testing different *mtry* and *ntree* values in the random forest models and ensemble iteration values in the SVM models. The *mtry* value, through a small loop in the algorithm, was chosen automatically for each model, while a *ntree* value of 10,000 and ensemble interaction value of 70 were proved to be a stable configuration for random forest and SVM, respectively, and gave us the most accurate models (Figure 2A, B).

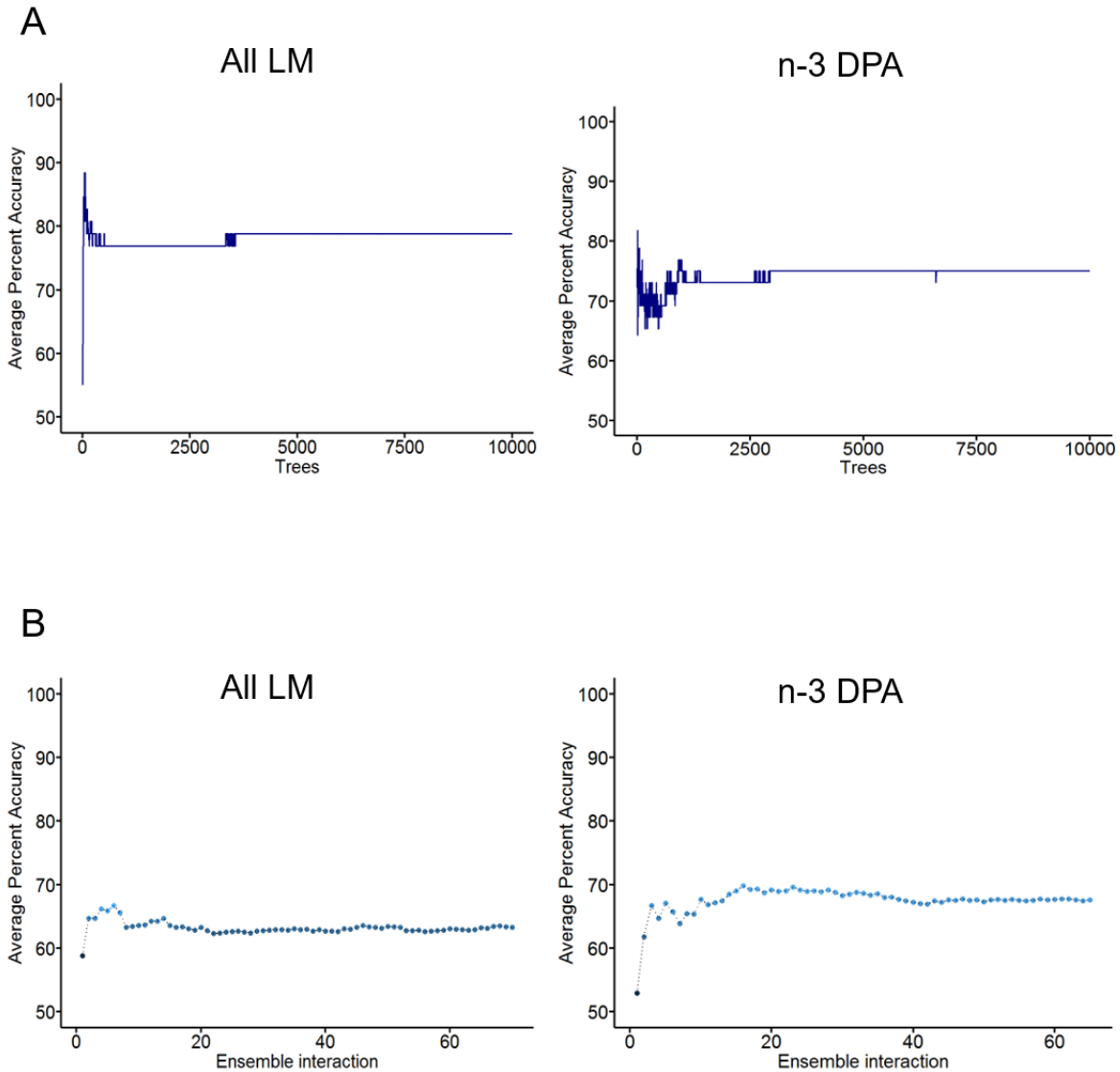


Figure 2. Establishing the optimal parameters for random forest and SVM models. Representative results obtained for average percentage accuracy for (A) random forest models and (B) SVM models by increasing the number of (A) decision trees (*ntree* value) or (B) ensemble interactions. All LM = all the lipid mediators model.

Overall, the random forest models were more accurate in predicting the response of DMARD treatment in RA patients using baseline plasma SPM concentrations (Figure 3 and Table 1). Specifically, we found that cumulative concentrations of the DHA (that includes the D-series resolvins, protectins and maresins) and n-3 DPA (that includes 13-series resolvins, D-series resolvins, protectins and maresins) metabolomes were the most accurate at predicting whether a patient would respond to treatment or not. For the random forest models, the accuracy for the DHA metabolome at predicting outcome was of ~81% (~62% using SVM) and that of the n-3 DPA metabolome was of ~69% (~61% using SVM). These values were higher than those obtained using a combination of clinical parameters, including DAS28-ESR and RF titre. The accuracy score for the model with all the lipid mediator was of ~73% using random forest and ~61% using SVM.

Since the best predictors are the ones with the fewest number of features, we further evaluated the ability of the DHA metabolome-based model (best score both random forest and SVM) to accurately categorize patients using the resulting confusion matrix of the model. Here we found that the model based on concentrations of DHA derived mediators was able to correctly classify ~90% of responders in the appropriate category (Figure 4).

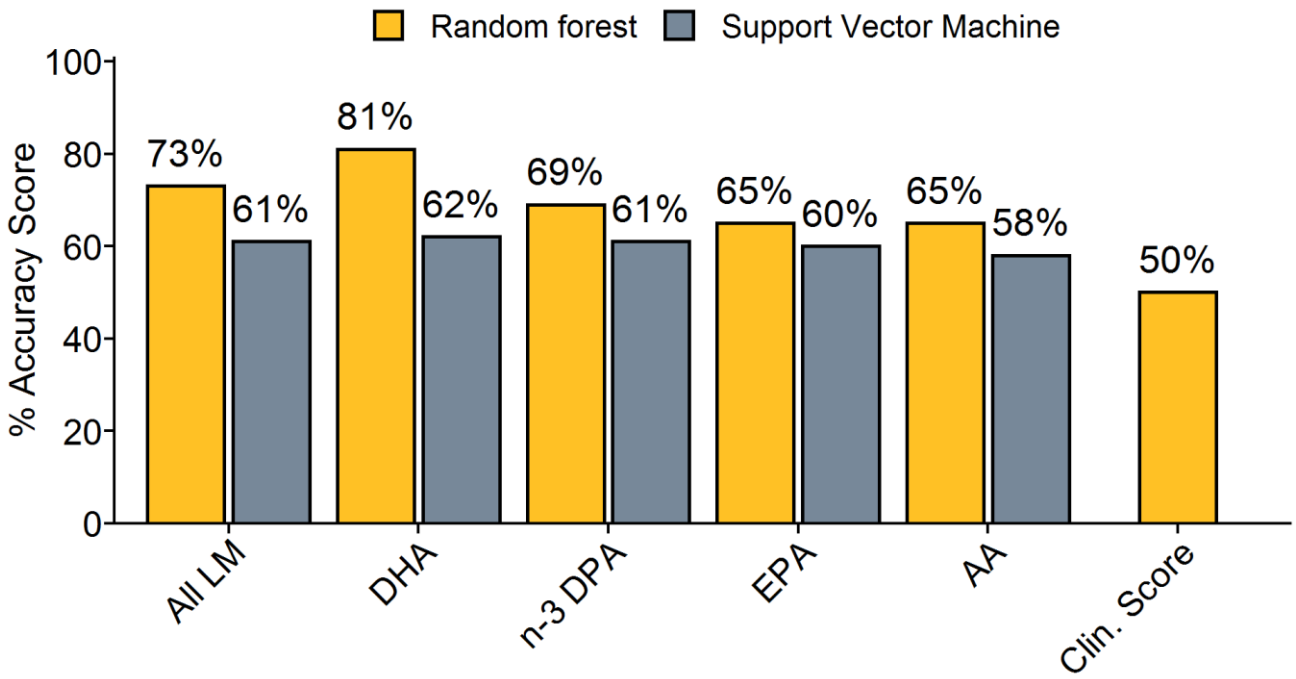


Figure 3. Baseline plasma lipid mediator profiles are predictive of treatment responsiveness in DMARD naive RA patients. Peripheral blood was collected in RA patients prior to DMARD treatment initiation and lipid mediator profiles were established using LC-MS/MS. % accuracy score of prediction models based on the combination of all lipid mediators identified and quantified (ALL LM) or individual fatty acid metabolomes as indicated. Clin. Score = clinical score (see methods for parameters included). All the models were created using the random forest ("randomForest" package from R) or the support vector machine ("classyfire" package from R) methodology. Clin. Score model was only created using random forest because of missing values.

Table 1: Summary of prediction models created using the support vector machine and random forest methodologies.

Machine Learning methodology	Samples	Model	Number of variables	% Accuracy Score	Model validation					
					Sensitivity	Specificity	TP	FP	TN	FN
randomForest (RF)	All samples	Four metabolomes	54	73	0.87	0.55	87	45	55	13
randomForest (RF)	All samples	DHA metabolome	23	81	0.9	0.68	90	32	68	10
randomForest (RF)	All samples	n-3 DPA metabolome	10	69	0.83	0.5	83	50	50	17
randomForest (RF)	All samples	EPA metabolome	3	65	0.8	0.45	80	55	45	20
randomForest (RF)	All samples	AA metabolome	18	65	0.77	0.5	77	50	50	23
randomForest (RF)	All samples	Clin. Score	11	50	0.6	0.36	60	64	36	40
Classyfire (SVM)	All samples	Four metabolomes	54	61	0.63	0.54	63	46	54	37
Classyfire (SVM)	All samples	DHA metabolome	23	62	0.65	0.56	65	44	56	35
Classyfire (SVM)	All samples	n-3 DPA metabolome	10	61	0.63	0.54	63	46	54	37
Classyfire (SVM)	All samples	EPA metabolome	3	60	0.62	0.52	62	48	52	38
Classyfire (SVM)	All samples	AA metabolome	18	58	0.6	0.48	60	52	48	40

TP = True Positives, FP = False Positives, TN = True Negatives, FN = False Negatives.

Future work

Taken together, these results indicate that baseline peripheral blood lipid mediators are linked with DMARD treatment outcome and can predict the responsiveness of these treatments in RA patients with more accuracy than the clinical scores. Further samples are required to validate the robustness of our models against independent datasets and make our methodological approach relevant; besides that, new analysis, looking forward to identify the most relevant lipid mediators in our models, are needed for improving the prediction ability of the models and their usefulness to classify the DMARD treatment response in different RA pathotypes.

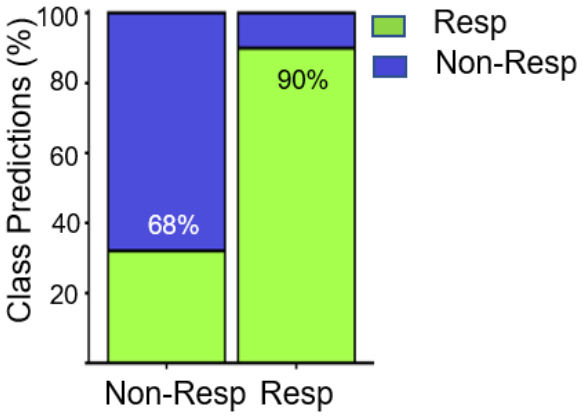


Figure 4. The DHA metabolome-based model can classify correctly DMARD-responder patients. Peripheral blood was collected in RA patients prior to DMARD treatment initiation and lipid mediator profiles were established using LC-MS/MS. Classification predictions for each class (sensitivity and specificity) of the DHA model. Green indicates the samples that were predicted as Resp while blue indicates those patients predicted Non-Resp. Percentages indicate true positives (Resp class) and true negatives (Non-Resp class). The model was created using the random forest ("randomForest" package from R) methodology.

Chapter 4: Method *critique*: Machine learning

Introduction of machine learning

Machine learning is an application of artificial intelligence that allows a system to learn and improve from data exposition without the need to be programmed every step of the process. In simple words, it's an algorithm created to build mathematical models than based on patterns and correlations can make decisions.

The way how machine learning works can be described as a flow of knowledge: the model is built by learning from sample data, known as training data, and then applies the knowledge to make predictions or classify data. This process can be repeated a *n* number of times so the model can teach itself and improve its ability to make decisions [27].

Most of the machine learning strategies can be divided into three categories: supervised learning, in which the training data comes with labels and the system knows from the beginning the distribution of the samples in the groups, and the aim of the model is to be able to predict in which group new samples belong to; unsupervised learning, in which no classification is given with the training data and the goal of the algorithm is to identify the patterns and cluster of the dataset; and reinforcement learning, in which the model interacts constantly with a changing environment and needs to learn from experience what is the best set of actions to take to fulfil a specific goal [27,41].

The use of machine learning has expanded in recent decades and there are now multiple applications of machine learning in different industries. Supervised machine learning is commonly used in banking to predict risk associated with investments and loans, and pattern identification is very useful in image and speech recognition. Unsupervised machine learning strategies have been widely used in marketing with the aim of finding associations between customers trends and product placement; while reinforcement learning represents a milestone in the video game and robotic industries [42]. Other applications include financial market, computer networks, data quality, economics, etc.

In medical science, where complex biological data is used, both supervised and unsupervised machine learning are applied to answer specific questions related to how different omics, such as proteomics, metabolomics, transcriptomics and genomics, can be associated with different conditions or outcomes. Particularly, a common case of study is to understand how relevant patient information can explain the development of disease; here, supervised machine learning is used to identify patterns in the patient data and predict if other patients would develop the same disease or not. Since numerous machine learning applications in biology and medical

diagnose involve tasks that can be set up as supervised, the present analysis will be focused on supervised machine learning methodologies, the steps for good practice of these strategies and how to overcome their limitations.

Machine learning methodologies

Three steps have to be followed for good practice in machine learning application: training, validation and evaluation.

The *training step* consists in the presentation of the data to the machine learning methodology. The training data is formed by a set of samples containing the same features (measurable explanatory variables) that can be continuous (e.g. protein concentrations, expression level, etc.), categorical (e.g. range of age, type of injury, level of pain, etc.) or binary (e.g. presence or absence of symptoms, male or female, etc.). As mentioned before, in supervised machine learning the samples in the training data come with a label that established the different groups in the dataset; these groups are the expected output and the response variable that needs to be explained by the explanatory variables. In other words, the machine learning model uses the selected feature to determined if a sample belongs to a group or another.

During the training step, the training data is separated in two subsets: the training dataset, used to train the machine learning model on how to predict or classify; and the testing dataset, usually a third of the original samples so it conserves the balance of sample groups as the initial training data. The testing dataset does not participate in the training process as a way of reducing the risk of overfitting (error causes when a model is too closely fit of a limited set of data points) [32,37].

The training dataset is then used to feed a machine learning methodology, which is going to design a mathematical predictive model. A basic representation of the model looks like:

classes ~ features

where *classes* represents the different groups in which the data can be classified and *features* are the different variables used to classify the samples.

Different machine learning methodologies have been developed and researched in the past years. Every method has its pros and cons, and select the best system will depend on the training data and the goal to achieve.

Among the most used methodologies when analysing data obtained from mass spectrometry (the main type of data used in this project) are Bayesian classifiers, artificial neural networks (ANN), support vector machine (SVM) and random forest [29].

A Bayesian classifier is a statistical approach based on applying Bayes theorem (the probability of having a first condition based on a second one can help to predict the second condition based on having the first one) [43]. This simple approach has the advantage to be fast and easy to implement; however, its use is very limited since assumes that all the features are independent variables which rarely applies to biological data.

ANN are formed by a collection of units (neurons) that are linked through a pattern of connections. There are three types of units: inputs, outputs and hidden units. Specific connections are activated according to the input unit and each neuron is a variant of a linear classifier. The information travels through the network to generate an output or prediction [41]. ANN can perform as well than other machine learning methodologies; however, the results are usually very complex to interpret and the training time is much longer.

SVM method, briefly, tries to establish a threshold between groups creating a hyperplane to separate the data. The method increases systematically the dimension of the data until it finds the best hyperplane to discriminate the samples. SVM, when implemented correctly, serves to identify outliers and allows misclassifications (which is useful since biological data usually don't involve intuitive separable data) reducing the overfitting of the model. On the other hand, the training time is longer than other machine learning methodologies, SVM method only works to solve binary problems and it's very sensitive to parameters manipulation.

Random forest follows the same theory of decision trees, which makes decisions based on examples in a tree structure, where each decision in every step reduces the possible outcomes. Random forest creates thousand of different trees variating the number of samples and features used in every training and, in the end, designs a consensus decision tree. This method works well with large datasets with a high number of features since the data is divided at every step, allowing the identification of the most important features. On the downside, it's not very sensitive to outliers and requires a large number of features to generate reliable results.

The best way to overcome some issues while using biological samples in machine learning applications is to run some exploratory analysis in the data before to create the machine learning models. Multivariate analysis, such as principal component analysis (PCA) and heatmap data visualization (cluster analysis), can give a hint of corrupt, inconsistent or outlier values that can decrease the performance of the models. These analyses also can help to identify the differences in magnitude of the features and indicate if the data needs to be scale and what scaling method can be used; usually scaling by the mean and standard deviation of each feature avoids misleading results due to features with big units [32].

After preparing the training dataset and creating the model, the next step is the *validation step*. The validation step answer two questions: how good is my model to predict? and can it be improved? Two strategies are the most used to answer those questions.

The first one is cross validation, which is the procedure by which the training data is separated differently in the training and test datasets a *n* number of times with the purpose of finding the best training dataset that creates the best predictive model. For example, using 10-fold cross validation, the data will split ten times, producing ten training and ten test datasets, and the system will highlight the model with the best ability to predict. The main downside of this strategy is the overfitting as a result of presenting the same data all the time resulting in an overestimation of the model performance [37,44].

A solution to this issue is using the bootstrapping technique, the currently preferred validation strategy, which is a resampling technique that creates new training datasets allowing repeat samples by ignoring other enties in the data. Doing this a *n* number of times gives a good idea of the sampling distribution in the data without having to use the same samples all the time. This reduces the overfitting of the machine learning model and allows a robust evaluation of its performance [37,45].

Finally, the last step in machine learning is the *evaluation step*, which basically consists of evaluating the model using a new dataset, one that has not been used in the training step. This is perhaps the most important step. Using a new cohort to evaluate the machine learning model can elucidate how bias is and how good is to classify new data. Different statistical scores and tests can be used to assess the performance of the model: with the confusion matrix (true negatives, true positives, false negatives and false positives) is possible to calculate sensitivity and specificity of the model, while ROC curves, precision-recall curves and Matthews correlation coefficient give statistical estimates of the accuracy of the model (the last two used when there is an imbalanced dataset).

Final Remarks

Machine learning helps to understand complex biological data without missing unknown interactions between the features; however, the training data needs to be strong enough (large number of samples and features) to create reliable models. A preprocessing step derived from exploratory analysis can be helpful to make the data stronger taking out outlier values and scaling the features. Choosing the best machine learning methodology is always fundamental and it will depend on the available data and the biological question that it needs to be answered. It's important to consider the disadvantages of each methodology so measures can be taken. Finally, the evaluation step is the most important step during machine learning application if the purpose of

the analysis is to create a model able to predict between classes and gives an insight of the biological processes behind the prediction logic.

References

1. Sayah, A. & English, J. C. Rheumatoid arthritis: A review of the cutaneous manifestations. *Journal of the American Academy of Dermatology*. **53**, 191–209 (2005).
2. van de Sande, M. G. H. *et al.* Different stages of rheumatoid arthritis: features of the synovium in the preclinical phase. *Ann. Rheum. Dis.* **70**, 772–777 (2011).
3. Holmdahl, R., Malmström, V. & Burkhardt, H. Autoimmune priming, tissue attack and chronic inflammation - the three stages of rheumatoid arthritis. *Eur. J. Immunol.* **44**, 1593–1599 (2014).
4. Avouac, J., Gossec, L. & Dougados, M. Diagnostic and predictive value of anti-cyclic citrullinated protein antibodies in rheumatoid arthritis: a systematic literature review. *Ann. Rheum. Dis.* **65**, 845–851 (2006).
5. Neumann, E., Lefèvre, S., Zimmermann, B., Gay, S. & Müller-Ladner, U. Rheumatoid arthritis progression mediated by activated synovial fibroblasts. *Trends Mol. Med.* **16**, 458–468 (2010).
6. Firestein, G. S. Evolving concepts of rheumatoid arthritis. *Nature* **423**, 356–361 (2003).
7. Humby, F., Lewis, M., Townsend, M. J. & Pitzalis, C. Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. *Ann. Rheum. Dis.* **78**, 761-772 (2019).
8. Whittle, S. L. & Hughes, R. A. Folate supplementation and methotrexate treatment in rheumatoid arthritis: a review. *Rheumatology* **43**, 267–271 (2004).
9. Gartlehner, G., Hansen, R. A., Jonas, B. L., Thieda, P. & Lohr, K. N. The comparative efficacy and safety of biologics for the treatment of rheumatoid

arthritis: a systematic review and metaanalysis. *J. Rheumatol.* **33**, 2398–2408 (2006).

10. Cronstein, B. N. & Aune, T. M. Methotrexate and its mechanisms of action in inflammatory arthritis. *Nat. Rev. Rheumatol.* **16**, 145-154 (2020).
11. Friedman, B. & Cronstein, B. Methotrexate mechanism in treatment of rheumatoid arthritis. *Joint Bone Spine*. **86,** 301–307 (2019).
12. Peres, R. S. *et al.* Low expression of CD39 on regulatory T cells as a biomarker for resistance to methotrexate therapy in rheumatoid arthritis. *Proceedings of the National Academy of Sciences*. **112,** 2509–2514 (2015).
13. Aletaha, D. & Smolen, J. S. Diagnosis and Management of Rheumatoid Arthritis. *JAMA.* **320,** 1360 (2018).
14. Wasserman, A. Rheumatoid Arthritis: Common Questions About Diagnosis and Management. *Am. Fam. Physician* **97**, 455–462 (2018).
15. Perretti, M., Cooper, D., Dalli, J. & Norling, L. V. Immune resolution mechanisms in inflammatory arthritis. *Nature Reviews Rheumatology*. **13**, 87–99 (2017).
16. Serhan, C. N., Chiang, N. & Dalli, J. New pro-resolving n-3 mediators bridge resolution of infectious inflammation to tissue regeneration. *Mol. Aspects Med.* **64**, 1–17 (2018).
17. Serhan, C. N., Hamberg, M. & Samuelsson, B. Lipoxins: novel series of biologically active compounds formed from arachidonic acid in human leukocytes. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 5335–5339 (1984).
18. Leuti, A., Maccarrone, M. & Chiurchiù, V. Proresolving Lipid Mediators: Endogenous Modulators of Oxidative Stress. *Oxid. Med. Cell. Longev.* **2019**, 1-12 (2019).
19. Sok, M. C. P., Tria, M. C., Olingy, C. E., San Emeterio, C. L. & Botchwey, E. A.

Aspirin-Triggered Resolvin D1-modified materials promote the accumulation of pro-regenerative immune cell subsets and enhance vascular remodeling. *Acta Biomater.* **53**, 109–122 (2017).

20. Serhan, C. N. & Levy, B. D. Resolvins in inflammation: emergence of the pro-resolving superfamily of mediators. *J. Clin. Invest.* **128**, 2657–2669 (2018).
21. Walker, M. E., Souza, P. R., Colas, R. A. & Dalli, J. 13-Series resolvins mediate the leukocyte-platelet actions of atorvastatin and pravastatin in inflammatory arthritis. *FASEB J.* **31**, 3636–3648 (2017).
22. Dalli, J., Colas, R. A. & Serhan, C. N. Novel n-3 immunoresolvents: structures and actions. *Sci. Rep.* **3**, 1940 (2013).
23. Sharma-walia, N. & Chandrasekharan, J. Lipoxins: nature's way to resolve inflammation. *Journal of Inflammation Research.* **30**, 181-192 (2015).
24. Arnardottir, H. H. *et al.* Resolvin D3 Is Dysregulated in Arthritis and Reduces Arthritic Inflammation. *J. Immunol.* **197**, 2362–2368 (2016).
25. Chan, M. M.-Y. & Moore, A. R. Resolution of inflammation in murine autoimmune arthritis is disrupted by cyclooxygenase-2 inhibition and restored by prostaglandin E2-mediated lipoxin A4 production. *J. Immunol.* **184**, 6418–6426 (2010).
26. Mamoshina, P. *et al.* Machine Learning on Human Muscle Transcriptomic Data for Biomarker Discovery and Tissue-Specific Drug Target Identification. *Front. Genet.* **9**, 242 (2018).
27. Bishop, C. M. *Pattern Recognition and Machine Learning*. (Springer, 2016).
28. Zitnik, M. *et al.* Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. *Inf. Fusion* **50**, 71–91 (2019).
29. Swan, A. L., Mobasheri, A., Allaway, D., Liddell, S. & Bacardit, J. Application of machine learning to proteomics data: classification and biomarker identification in

postgenomics biology. *OMICS* **17**, 595–610 (2013).

30. Williams, F. M. K. Biomarkers: in combination they may do better. *Arthritis Research & Therapy*. **11,** 130 (2009).
31. Valletta, J. J. & Recker, M. Identification of immune signatures predictive of clinical protection from malaria. *PLoS Comput. Biol.* **13**, e1005812 (2017).
32. Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Min.* **10**, (2017).
33. Bennett, K. P. & Campbell, C. Support vector machines. *ACM SIGKDD Explorations Newsletter*. **2**, 1–13 (2000).
34. Breiman, L. *Classification and Regression Trees*. (Routledge, 2017).
35. Han, T., Jiang, D., Zhao, Q., Wang, L. & Yin, K. Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery. *Transactions of the Institute of Measurement and Control*. **40**, 2681–2693 (2018).
36. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning: with Applications in R*. (Springer Science & Business Media, 2013).
37. Eleni Anthippi Chatzimichali, C. B. Novel application of heuristic optimisation enables the creation and thorough evaluation of robust support vector machine ensembles for machine learning applications. *Metabolomics* **12**, (2016).
38. Lötsch, J. *et al.* Machine-learning based lipid mediator serum concentration patterns allow identification of multiple sclerosis patients with high accuracy. *Scientific Reports*. **8** (2018).
39. Huang, J. Z. An Introduction to Statistical Learning: With Applications in R By Gareth James, Trevor Hastie, Robert Tibshirani, Daniela Witten. *Journal of Agricultural, Biological, and Environmental Statistics*. **19**, 556–557 (2014).

40. Chong, J., Yamamoto, M. & Xia, J. MetaboAnalystR 2.0: From Raw Spectra to Biological Insights. *Metabolites*. **9** 57 (2019).
41. Maglogiannis, I. G. *Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*. (IOS Press, 2007).
42. Alpaydin, E. *Introduction to Machine Learning*. (MIT Press, 2020).
43. Casella, G. & Berger, R. L. Statistical Inference. *Biometrics*. **49** 320 (1993).
44. Varma, S. & Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* **7**, 91 (2006).
45. Tsamardinos, I., Greasidou, E. & Borboudakis, G. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. *Mach. Learn.* **107**, 1895–1922 (2018).

Appendix

Appendix 1: Mass Spectrometer settings for Multiple Reaction Monitoring and Enhance product Ion for AB Sciex 5500 Q TRAP and AB Sciex 6500+ Q TRAP.

AB Sciex 5500 Q TRAP		
Parameter	**MRM**	**EPI**
Curtain Gas	35	35
Collision Gas	Medium	Medium
Ion Spray Voltage	-3500	-3500
Temperature	550 °C	550 °C
Ion Source Gas 1	30	30
Ion Source Gas 2	75	75
AB Sciex 6500+ Q TRAP		
Parameter	**MRM**	**EPI**
Curtain Gas	30	30
Collision Gas	Medium	Medium
Ion Spray Voltage	-4500	-4500
Temperature	440 °C	440 °C
Ion Source Gas 1	45	45
Ion Source Gas 2	70	70

Appendix 2: Mass Spectrometer settings for Multiple Reaction Monitoring and Enhance product Ion for AB Sciex 5500 Q TRAP by lipid mediator.

AB Sciex 5500 Q TRAP					
	Transition	**DP**	**EP**	**CE**	**CXP**
DHA Bioactive Metabolome					
RvD1	*215*	-70	-10	-26.8	-19
	233	-70	-10	-20.5	-19
RvD2	*141*	-64	-10	-22.5	-14
	175	-60	-10	-30	-14
RvD3	*147*	-90	-10	-26	-13
	137	-90	-10	-26	-13
RvD4	*101*	-70	-10	-23	-11
	225	-70	-10	-27.1	-10
RvD5	*199*	-76	-10	-22.5	-17
	261	-79	-10	-19.3	-17
RvD6	*101*	-80	-10	-23	-14
	159	-80	-10	-23	-14
17R-RvD1	*215*	-70	-10	-26.8	-19
	233	-70	-10	-20.5	-19
17R-RvD3	*147*	-90	-10	-26	-13
	137	-90	-10	-26	-13
PD1	*153*	-80	-10	-21.5	-13
	181	-80	-10	-21	-15
17R-PD1	*153*	-80	-10	-21.5	-13
	181	-80	-10	-21	-15
PDx	*181*	-80	-10	-21	-15
	153	-80	-10	-21.5	-13
22-OH-PD1	*153*	-70	-10	-24.5	-12
PCTR1	*231*	80	9	28	15
PCTR2	*231*	80	5	23.5	13
PCTR3	*231*	80	5	23.5	13
MaR1	*221*	-75	-10	-20	-18
	250	-75	-10	-21	-12
MaR2	*191*	-80	-10	-20	-16
22-OH-MaR1	*221*	-70	-10	-24	-12
22-COOH-MaR1	*221*	-70	-10	-24	-12
14-oxo-MaR1	*248*	-80	-10	-20	-11
7S,14S-diHDHA	*221*	-75	-10	-20	-18
	250	-75	-10	-20	-18
4,14-diHDHA	*101*	-80	-10	-23	-14
	159	-80	-10	-23	-14
MCTR1	*191*	80	9	28	15
	227	80	9	28	15
MCTR2	*191*	80	5	23.5	13
MCTR3	*191*	80	5	23.5	13
n-3 DPA bioactive Metabolole					
RvT1	*193*	-70	-10	-26.8	-19
	211	-70	-10	-26.8	-19
RvT2	*197*	-64	-10	-22.5	-14
	255	-64	-10	-22.5	-14
RvT3	*197*	-60	-10	-30	-14
	173	-60	-10	-30	-14
RvT4	*211*	-70	-10	-23	-12

	193	-70	-10	-23	-12
$RvD1_{n\text{-}3\ DPA}$	*143*	-70	-10	-20.2	-13
	215	-70	-10	-26.8	-13
$RvD2_{n\text{-}3DPA}$	*261*	-80	-10	-24	-13
	233	-80	-10	-27	-13
$RvD5_{n\text{-}3DPA}$	*199*	-80	-10	-22.5	-13
	263	-80	-10	-19.3	-13
$PD1_{n\text{-}3\ DPA}$	*155*	-75	-10	-30	-10
	183	-75	-10	-24	-11
7S,14S-diHDPA	*223*	-75	-10	-24.2	-18
	249	-75	-10	-22	-15
$MaR1_{n\text{-}3\ DPA}$	*233*	-75	-10	-24.2	-18
	249	-75	-10	-22	-15
EPA bioative Metabolome					
RvE1	*195*	-90	-10	-23.5	-16
	161	-75	-10	-23.5	-13.5
RvE2	*199*	-80	-10	-25.5	-18
	159	-80	-10	-25	-18
RvE3	*201*	-75	-10	-22	-16
	251	-75	-10	-24.2	-20
AA bioative Metabolome					
LXA_4	*115*	-80	-10	-19.5	-11
	217	-80	-10	-27.2	-18
LXB_4	*221*	-75	-10	-22.5	-15
	115	-80	-10	-19.5	-11
5S,15S-diHETE	*235*	-80	-10	-22	-13
	115	-80	-10	-22	-13
15-epi-LXA_4	*115*	-80	-10	-19.5	-11
	217	-80	-10	-27.2	-18
15-epi-LXB_4	*221*	-75	-10	-22.5	-15
	115	-80	-10	-19.5	-11
LTB_4	*195*	-90	-10	-23	-16
5S,12S-diHETE	*195*	-90	-10	-23	-16
6-trans-LTB_4	*195*	-90	-10	-23	-16
6-trans-12-epi LTB_4	*195*	-90	-10	-23	-16
20-OH-LTB_4	*195*	-100	-10	-25	-16
20-COOH-LTB_4	*195*	-80	-10	-19	-15
LTC_4	*189*	80	9	28	15
LTD_4	*189*	80	5	23.5	13
LTE_4	*189*	80	5	23.5	13
PGD_2	*189*	-70	-10	-27.5	-13
PGE_2	*189*	-70	-10	-26	-16
PGF_{2a}	*193*	-90	-10	-34.5	-16
TxB_2	*169*	-70	-10	-25	-15

DP=Declustering Potential, EP=Entrance Potential, CE=Collision Energy, CXP=Collision Cell Exist Potential.

Appendix 3: Mass Spectrometer settings for Multiple Reaction Monitoring and Enhance product Ion for AB Sciex 6500+ Q TRAP by lipid mediator.

AB Sciex 6500+ Q TRAP					
	Transition	**DP**	**EP**	**CE**	**CXP**
DHA Bioactive Metabolome					
RvD1	*215*	-33	-10	-25.8	-10
	233	-33	-10	-20.5	-10
RvD2	*141*	-33	-10	-21.5	-10
	215	-33	-10	-25.8	-10
RvD3	*147*	-33	-10	-26	-10
	137	-33	-10	-26	-10
RvD4	*101*	-33	-10	-23	-10
	225	-33	-10	-27.1	-10
RvD5	*199*	-33	-10	-22.5	-10
	141	-33	-10	-18.3	-10
RvD6	*101*	-33	-10	-23	-10
	159	-33	-10	-23	-10
17R-RvD1	*215*	-33	-10	-25.8	-10
	233	-33	-10	-20.5	-10
17R-RvD3	*147*	-33	-10	-26	-10
	137	-33	-10	-26	-10
PD1	*153*	-33	-10	-21.5	-10
	137	-33	-10	-21	-10
17R-PD1	*153*	-33	-10	-21.5	-10
	137	-33	-10	-21	-10
PDx	*153*	-33	-10	-21.5	-10
	137	-33	-10	-21	-10
22-OH-PD1	*153*	-33	-10	-24.5	-10
	137	-33	-10	-24.5	-10
PCTR1	*231*	40	9	28	15
PCTR2	*231*	40	5	23.5	13
PCTR3	*231*	40	5	23.5	13
MaR1	*221*	-33	-10	-19	-10
	177	-33	-10	-20	-10
MaR2	*191*	-33	-10	-19	-10
	221	-33	-10	-20	-10
22-OH-MaR1	*221*	-33	-10	-24.5	-10
	177	-33	-10	-24.5	-10
22-COOH-MaR1	*221*	-33	-10	-24.5	-10
	177	-33	-10	-24.5	-10
14-oxo-MaR1	*248*	-33	-10	-20	-10
7S,14S-diHDHA	*221*	-33	-10	-19	-10
	177	-33	-10	-20	-10
4,14-diHDHA	*101*	-33	-10	-23	-10
	159	-33	-10	-23	-10
MCTR1	*191*	40	9	28	15
	227	40	9	28	15
MCTR2	*191*	40	5	23.5	13
MCTR3	*191*	40	5	23.5	13
n-3 DPA bioactive Metabolole					
RvT1	*193*	-33	-10	-26.8	-10
	211	-33	-10	-26.8	-10
RvT2	*197*	-33	-10	-22.5	-10
	255	-33	-10	-22.5	-10
RvT3	*197*	-33	-10	-30	-10
	173	-33	-10	-30	-10
RvT4	*211*	-33	-10	-23	-10
	193	-33	-10	-23	-10

$RvD1_{n-3\ DPA}$	*143*	-33	-10	-20.2	-10
	215	-33	-10	-26.8	-10
$RvD2_{n-3DPA}$	*261*	-33	-10	-24	-10
	233	-33	-10	-27	-10
$RvD5_{n-3DPA}$	*199*	-33	-10	-22.5	-10
	263	-33	-10	-19.3	-10
$PD1_{n-3\ DPA}$	*155*	-33	-10	-30	-10
	183	-33	-10	-24	-10
7S,14S-diHDPA	*223*	-33	-10	-24.2	-10
	205	-33	-10	-25.8	-10
$MaR1_{n-3\ DPA}$	*223*	-33	-10	-24.2	-10
	205	-33	-10	-25.8	-10
EPA bioative Metabolome					
RvE1	*195*	-33	-10	-23.5	-10
	161	-33	-10	-23.5	-10
RvE2	*199*	-33	-10	-25.5	-10
	159	-33	-10	-25	-10
RvE3	*201*	-33	-10	-22	-10
	275	-33	-10	-22	-10
AA bioative Metabolome					
LXA_4	*115*	-33	-10	-19.5	-10
	217	-33	-10	-27.2	-10
LXB_4	*221*	-33	-10	-21.5	-10
	115	-33	-10	-19.2	-10
5S,15S-diHETE	*235*	-33	-10	-22	-10
	115	-33	-10	-22	-10
15-epi-LXA_4	*115*	-33	-10	-19.5	-10
	217	-33	-10	-27.2	-10
15-epi-LXB_4	*221*	-33	-10	-21.5	-10
	115	-33	-10	-19.2	-10
LTB_4	*195*	-33	-10	-23	-10
5S,12S-diHETE	*195*	-33	-10	-23	-10
6-trans-LTB_4	*195*	-33	-10	-23	-10
6-trans-12-epi LTB_4	*195*	-33	-10	-23	-10
20-OH-LTB_4	*195*	-33	-10	-25	-10
20-COOH-LTB_4	*195*	-33	-10	-19	-10
LTC_4	*189*	40	9	28	15
LTD_4	*189*	40	5	23.5	13
LTE_4	*189*	40	5	23.5	13
PGD_2	*189*	-33	-10	-27.5	-10
PGE_2	*189*	-33	-10	-26	-10
	175	-33	-10	-26	-10
PGF_{2a}	*193*	-33	-10	-34.5	-10
TxB_2	*169*	-33	-10	-25	-10

DP=Declustering Potential, EP=Entrance Potential, CE=Collision Energy, CXP=Collision Cell Exist Potential.

Appendix 4: RA patients demographics and clinical information.

	DMARD - Responders (n=30)	DMARD- Non-Responders (n=24)
Pathotype (n)	Lymphoid (10), Fibroid (10), Myeloid (10)	Lymphoid (9), Fibroid (8), Myeloid (7)
Ethnicity (n)	Caucasian (19), Black (2), Indian (1), Caribbean (1), Asian (3), Bangladeshi (2), Black African (1)	Asian (1), Bangladeshi (1), Black (7), British (1), Caribbean (3), Caucasian (9), Filipino (1)
Gender (n)	Female (16), Male (14)	Female (21), Male(3)
Age at Recruitment – years	51 (±21)	56 (±9.5)
Onset	5.6 (±3.4)	6.0 (±3.42)
Currently smoking (%)	5 (16.7)	10 (43.5)
Co-Morbs Baseline (n)	Acne (1), Anaemia (1), Vitamin D deficiency (3), Hypercholesterolemia (2), Asthma (3), Hypertension (9), Osteoarthritis (2), Hypothyroidism (3), Gout (1), Shingles (1), Graves' disease (1), Kidney disease (1), TIA (1), Cardiovascular disease (1), Type 1 diabetes (1), Reactive iritis (1), Raised cholesterol (1), Hypothyroidism (3), Glaucoma (1), Thalassaemia (1), Poor vision (1), Heart surgery (1), Rubello in utero (1), Scleritis (1), COPD (1), Enlarged prostate (1), Sinusitis (1), Peptic ulcer (1), Lower back pain (1), Acute MI (1), Ischaemic heart disease (1), Ca bladder (1)	Hypercholesterolemia (2), Asthma (5), Hypertension (12), Osteoarthritis (2), Hypothyroidism (1), COPD (1), Ischaemic heart disease (2), Osteoporosis (1), Sickle cell trait (1), IBS (2), Cervical spondylitis (1), Psoriasis (1), Fatty liver (1), Renal impairment (1), Coronary artery disease (1), Angina (1), Dyslipidaemia (1), Menorrhagia (1), Multiple sclerosis (1), Rectal incontinence (1), Detrusor instability (1), Meniscus tear/knee (1), Depression (1), Spina Bifida occulta (1), Axonal neuropathy (1), BCC (1), Gastritis (1), Heart valve repaired (1), Hysterectomy (1), foot & shoulder surgery (1), Endometriosis (1), Hay fever (1), Hysterectomy (1)
Concomitant Med. Baseline (n)	Tetracycline (1), Ferrous sulphate (1), Co-codamol (6), NSAIDs (3), Calcium vitamin D (2), Inhaler (1), Candesartan (1), Etoricoxib (1), Furosemide (1), Ramipril (2), Levothyroxine (1), Dihydrocodeine (1), Lisinopril (1), Lansoprazole (1), Bisoprolol (1), Tamsulosin (1), Carbimazole (1), Aspirin (2), Adcal (1), Simvastatin (5), Ibuprofen (4), Codeine (1), Losartan (1), Citalopram (1), Naproxen (5), Metformin (1), Lantus (1), Lacri-lube (1), Atenolol (1), Irbesartan (1), Amlodipine (2), Timolol (1), GTN spray (1), Frusemide (1), Celecoxib (1), Omeprazole (2), Budromide (1), Salbutamol (3), Diclofenac (1), Finasteride (1), Thyroxine (1), Talisartan (1)	Co-codamol (3), NSAIDs (3), Calcium vitamin D (3), Inhaler (1), Candesartan (2), Furosemide (1), Ramipril (3), Dihydrocodeine (1), Allopurinol (1), Lansoprazole (4), Aspirin (4), Simvastatin (5), Ibuprofen (2), Losartan (1), Ca-antagonist (1), Atenolol (2), Amlodipine (2), Alendronate (1), GTN spray (1), Pregabalin (2), Frusemide (1), Omeprazole (1), Salbutamol (1), Diclofenac (4), Thyroxine (2), Isosorbide (1), Adizem (1), Clopidogrel (2), Atorvastatin (1), Ivabradine (1), Co-tenidone (1), Nicorandil (1), Paracetamol (4), Tramadol (5), Amitriptyline (5), Piroxicam (1), Quinine (1), Bendroflumethiazide (3), Budesonide (2), Formoterol (1), Cetirizine (1), Arthrotec (1), Insulin (1), Doxazosin (2), Seretide (2), Spiriva (1), Arcoxia (1), Metformin (2), Glimepiride (1), Fluoxetine (1), Anti-hyper (1), Lyrica (1), CaD3 (1), Peppermint (1), Tiotropium (1), Detrusitol (1), Docusate (1), Voltron (1), HRT, (1), Aminophylline (1), Xolair (1), Ventolin (1), Oramorph (1), Temazepam (1)
Concomitant Med. 6mth (n)	Ferrous sulphate (1), MTX (28), Folic acid (8), ASA (21), Co-codamol (6), HCQ (8), Prednisolone (6), NSAIDs (3), Calcium vitamin D (3), Azathioprine (1), Salbutamol (3), Inhaler (1), Candesartan (1), Etoricoxib (1), Furosemide (1), Ramipril (2), Levothyroxine (1), Dihydrocodeine (1), Lisinopril (1), Allopurinol (1), Lansoprazole (1), Bisoprolol (1), Tamsulosin (1), Carbimazole (1), Aspirin (2), Adcal (1), Simvastatin (5), Ibuprofen (4), Codeine (1), Losartan (1), Citalopram (1), Naproxen (5), Metformin (1), Lantus (1), Lacri-lube (1), Atenolol (1), Irbesartan (1), Amlodipine (3),Timolol (1), Frusemide (1), Celecoxib (1), Omeprazole (2), Budromide (1), Diclofenac (1), Thyroxine (1), GTN spray (1), Finasteride (1)	MTX (18), Folic acid (6), ASA (9), Co-codamol (2), HCQ (11), Prednisolone (5), NSAIDs (3), Calcium vitamin D (3), Salbutamol (1), Inhaler (1), Candesartan (1), Furosemide (1), Ramipril (3), Dihydrocodeine (1), Allopurinol (1), Lansoprazole (5), Aspirin (4), Simvastatin (5), Ibuprofen (2), Losartan (1), Atenolol (2), Amlodipine (2), Frusemide (1), Omeprazole (1), Diclofenac (4), Thyroxine (2), Atorvastatin (1), GTN spray (1), Ca-antagonist (1), Alendronate (1), Pregabalin (2), Co-tenidone (1), Leflunomide (2), Paracetamol (5), Tramadol (5), Amitriptyline (4), Piroxicam (1), Quinine (1), Spironolactone (1), Bendroflumethiazide (2), Clopidogrel (2), Sulfasalazine (2), Budesonide (2), Formoterol (1), Cetirizine (1), Insulin (1), Doxazosin (2), Isosorbide (1), Adizem (1), Ivabradine (1), Nicorandil (1), Seretide (2), Spiriva (1), Arcoxia (1), Metformin (2), Glimepiride (1), Fluoxetine (1), Anti-hyper (1), Lyrica (1), Peppermint (1), Tiotropium (1), CaD4 (1), Detrusitol (1), Docusate (1), Voltarol (1), Aminophylline (1), Xolair (1), Ventolin (1), Co-Amoxiclav (1), Calcichew (1)
DMARD Treatment (n)	MTX/ASA/HCQ (2), MTX/ASA (21), MTX/HCQ (3), ASA/HCQ (1), MTX (1)	MTX/ASA/HCQ (2), MTX/ASA (11), MTX/HCQ (6), MTX (1), ASA (1), HCQ (2)
Recent Steroid Therapy Baseline (n)	No (25), Yes (5)	No (19), Yes (3)
Steroid Treatment Baseline (n)	Depo-Medro (5)	Pred. (3)
Recent Steroid Therapy 6mth (n)	No (13), Yes (15)	No (8), Yes (13)
Steroid Treatment 6mth (n)	Pred (15)	Depo-Medro (1), Pred (11)
ESR Baseline (mm/hr)	28 (±23)	41 (±28)
CRP Baseline (mg/l)	14 (±17)	24 (±37)
CCP Baseline (UI/L)	210 (±201)	249 (±255)

RF Baseline (UI/L)	87 (±120)	93 (±150)
ESR 6mth (mm/hr)	7 (±7)	30 (±14)
CRP 6mth (mg/l)	7 (±5)	15 (±18)
Tiredness VAS Baseline (1-100)	34 (±29)	48 (±26)
Pain VAS Baseline (1-100)	50 (±29)	62 (±23)
Pt. VAS Global Health baseline (1-100)	67 (±24)	72 (±19)
Physician VAS Global Assess. Baseline (1-100)	58 (±)21	65 (±20)
Tiredness VAS 6mth (1-100)	28 (±27)	64 (±17)
Pain VAS 6mth (1-100)	12 (±14)	66 (±22)
Pt. VAS Global Health 6mth (1-100)	17 (±20)	71 (±24)
Physician VAS Global Assess. 6mth (1-100)	10 (±12)	63 (±22)
Tender Joints Baseline (Number/28)	12 (±7)	13 (±9)
Swollen Joints Baseline (Number/28)	7 (±5)	8 (±5)
Tender Joints 6mth (Number/28)	1 (±1)	17 (±8)
Swollen Joints 6mth (Number/28)	1 (±1)	8 (±4)
HAQ Baseline (Max 38)	1.41 (±0.68)	1.83 (±0.60)
HAQ 6mth (Max 38)	0.57 (±0.73)	1.79 (±0.51)
DAS28 Baseline (0-10)	5.58 (±0.98)	5.81 (±1.08)
DAS28 6mth (0-10)	1.86 (±0.65)	6.17 (±0.98)
Delta DAS28 (Difference between 2 time points)	-3.72 (±1.13)	0.36 (±0.98)
US Synovial Thickness (12max) Baseline (number/36)	18 (±7)	15 (±9)
US Power Doppler (12max) Baseline (number/36)	7 (±6)	6 (±8)
US Synovial Thickness (Biopsied Joint) Baseline (number/3)	2 (±0.5)	3 (±1)
US Power Doppler (Biopsied Joint) Baseline (number/3)	1 (±1)	2 (±1)
US Synovial Thickness (12max) 6mths (number/36)	6 (±4)	7 (±8)
US Power Doppler (12max) 6mths (number/36)	2 (±3)	3 (±5)
Radiographic Erosion (n)	No (23), Yes (4)	No (20), Yes (2)

Erythrocyte sedimentation rate = ESR, C-reactive protein = CRP, Anti-cyclic citrullinated peptide = CCP, rheumatoid factor =RF, VAS = visual analogue scale, HAQ = Health Assessment Questionnaire, DAS28 = disease activity score-28, US = ultrasound, MTX = methotrexate, ASA = aspirin, HCQ = hydroxychloroquine, LEF = Leflunomide.

Appendix 5: Chromatography and MS-MS spectral criteria employed in the identification of lipid mediators.

	Standard RT	Reference diagnostic ions	Identified RT	S/N ratio	Data points	Identified diagnostic ions
DHA Bioactive Metabolome						
RvD1	11.1	**375**, 357, 331, 313, 295, **277, 233**, 215, **171, 113**	11.1	12	7	**375**, 357, 331, 295, **277**, 215, **171, 113**
RvD2	10.6	**375**, 357, 331, 313, 295, 287, **247, 233**, 215, 203 **141, 113**	10.6	8	9	**375**, 357, 331, 313, 295, 287, **247, 233**, 215, 203, **113**
RvD3	10.7	**375**, 357, 313, 295, 259, 215, **181, 165**, 147, 137, **101**	10.7	6	9	**375**, 357, 313, 295, 259, 215, **165**, 137
RvD4	12.1	**375**, 357, 339, 331, 313, 295, **277**, 255, 225, 215, 113, **101**	12.1	9	11	**375**, 357, 339, 331, 313, 295, **277**, 225, 215, **101**
RvD5	13.2	**359**, 341, 323, 315, 297, 279, 245, 243, 227, **217**, 199, **141, 113**	13.2	8	6	**359**, 341, 315, 297, 279, 227, 199, **141, 113**
RvD6	13.8	**359**, 341, 315, 297, **289**, 279, 245, 243, 227, 217, 199, **101**	13.8	11	4	**359**, 341, 315, **289**, 279, 243, 227, 217, 199, **101**
17R-RvD1	11.3	**375**, 357, 331, 313, 295, 243, 215, **171**, 135, 123	11.3	6	6	**375**, 357, 331, 313, 295, 243, 215, **171**, 123
17R-RvD3	10.6	**375**, 357, 339, 313, 295, 259, 191, **165**, 147, 137, **101**	10.6	5	8	**375**, 357, 339, 313, 295, 259, 191, **165**, 147, 137
PD1	13.3	**359**, 341, 323, 315, 297, 243, 217, **206**, 199, **181**, 159, **153**	13.3	5	6	**359**, 341, 323, 315, 297, 243, 217, 199, **153**
17R-PD1	12.5	**359**, 341, 323, 315, 297, 279, **261**, 217, **206, 181**, 177, 153	12.5	5	5	**359**, 341, 323, 315, 297, 279, 217, **206, 181**, 153
PDx	13.2	**359**, 341, 323, 315, 297, 279, 243, 217, **206, 181,** 159, **153**, 119	13.2	4	5	**359**, 341, 315, 297, 279, 243, 217, 159, **153**
22-OH-PD1	9.8	**375**, 357, 331, 313, 295, **289, 261**, 243, 217, 199, **181**, **153**, 137	9.8	31	11	**375**, 357, 331, 313, 295, **289**, 243, 199, **153**
PCTR1	9.5	**650**, 632, 503, **343, 308**, 264, 227	9.5	7	6	**650**, 632, **343**, **308**, 264, 227
PCTR2	8.5	**521**, 503, **343**, 325, **231**, 213, **179**, 161	8.5	6	6	**521**, 503, **343**, 325, **231**, 213, **179**, 161
PCTR3	10	**464**, 446, **343**, 325, **231**, 213, 187, **121**	10	12	6	**464**, 446, **343**, **231**, 213, 187, **121**
MaR1	13.4	**359**, 341, 323, 315, 297, 279, **221, 141, 113**	13.4	6	5	**359**, 341, 323, 315, 297, 279, **221, 141, 113**
MaR2	14.1	**359**, 341, 323, 315, 297, **249, 221, 191, 167**, 149	14.1	30	7	**359**, 341, 323, 315, 297, **221, 191, 167,** 149
22-OH-MaR1	9.5	**375**, 357, 339, 331, 313, 295, **262**, 244, **221**, 177, **153**	9.5	6	9	**375**, 357, 339, 331, 313, 295, 244, **221, 153**
14-oxo-MaR1	13.3	**357**, 339, 321, 313, 295, **248, 221, 141, 113**	13.3	23	5	**357**, 339, 321, 313, 295, **248, 141, 113**
7S,14S-diHDHA	13.5	**359**, 341, 323, 315, 297, 279, **221, 141, 113**	13.5	23	11	**359**, 341, 323, 315, 297, 279, **221, 141, 113**
4,14-diHDHA	14.1	**359**, 341, 323, 315, 297, 279, **257**, 239, **221**, 203, 159, **109**	14.1	11	8	**359**, 341, 323, 315, 297, 279, **221**, 159
MCTR1	9.4	**650**, 632, **418, 308, 235, 109**	9.4	17	6	**650**, 632**, 418, 308, 235, 109**
MCTR2	8.5	**521**, 503, **343**, 325, **191, 179**, 161	8.5	5	5	**521**, 503, **343**, 325, **191, 179**, 161
MCTR3	10	**464**, 446, **343**, 325, **235, 205, 191**, 187, 147	10	13	7	**464**, 446, **343**, 325, 235, **205, 191**, 147
n-3 DPA bioactive Metabolole						
RvT1	10.4	**377**, 359, 333, **319**, 315, 297, **233**, 193, **143**	10.4	13	6	**377**, 359, 333, **319**, 315, 297, 193, **143**
RvT2	11.3	**377**, 359, 341, 333, 315, 297, **263, 233, 197, 143**	11.3	6	5	**377**, 359, 341, 333, 315, 297, **263, 233**
RvT3	11.7	**377**, 359, 341, 333, 315, 297, **255, 233**, 215, **173,143**	11.7	8	9	**377**, 359, 333, 315, 297, **233**, 215, **173, 143**
RvT4	13.7	**361**, 343, 325, 317, 299, 233, 221, **217, 211**, 193, **143**	13.7	7	6	**361**, 343, 325, 317, 299, 233, 221, **143**
$RvD1_{n\text{-}3\ DPA}$	11.4	**377**, 359, 341, 333, 315, 297, 289, **279**, 261, 235, **233**, 125	11.4	11	9	**377**, 359, 341, 333, 315, 297, 289, **279**, 261, 245, **233**
$RvD2_{n\text{-}3DPA}$	11	**377**, 359, 341, 333, 315, 297, 261, **233**, 215, **143**, 125	11	7	11	**377**, 359, 341, 333, 315, 297, 233, 143, 125

$RvD5_{n\text{-}3DPA}$	13.6	**361**, 343, 325, 317, 299, 281, **263, 217,** 201, 199, **143**	13.6	12	6	**361**, 343, 325, 317, 299, 281, **263, 217**, 201, **143**
$PD1_{n\text{-}3\ DPA}$	13.7	**361**, 343, 325, 317, 299, 281, **263**, 219, **183**	13.7	12	5	**361**, 343, 325, 317, 299, 281, **263**, 219, **183**
10S,17S-diHPDA	13.6	**361**, 343, 325, 317, 299, 281, **263**, 245, 219, **183, 179, 155**	13.6	5	10	**361**, 343, 325, 317, 299, 281, **263**, 245, 219, **183, 179**, **155**
$MaR1_{n\text{-}3\ DPA}$	13.7	**361**, 343, 317, 299, 281, **223**, 205, 179, 161, **143, 115**	13.7	5	11	**361**, 343, 317, 299, 281, **223**, 205, 161 , **115**
EPA bioative Metabolome						
RvE1	8.2	**349**, 331, 305, 287, 273, 269, 255, 205, **195**, 177, 161, 151	8.2	8	7	**349**, 331, 305, 287, 273, 269, 205, **195**, 177
RvE2	12	**333**, 315, 297, 289, **275**, 271, 257, 253, 231, **217**	12	24	9	**333**, 315, 297, 289, 257, 253, **217**
RvE3	13.5	**333**, 315, 297, 289, 271, 259, 253, **245**, 201	13.5	5	5	**333**, 315, 297, 289, 271, **245**, 201
AA bioative Metabolome						
LXA_4	11.2	**351**, 333, 307, 289, **279**, 271, **251, 235,** 233, 207, 189, **145**	11.2	7	5	**351**, 333, 307, 289, 271, **251, 235**, 189
LXB_4	10.5	**351**, 333, 315, 307, 289, 271, **251**, 243, **235, 221**, 207, 189, 177, 159, **115**	10.5	7	5	**351**, 333, 315, 307, 289, 271, **221**, 189, **115**
5S,15S-diHETE	13.2	**335**, 317, 299, 291, 273, **263**, 255, **235**, 201, 191, 173, **115**	13.2	7	4	**335**, 317, 299, 291, 273, **235**, 201, 191, 173, **115**
$15R\text{-}LXA_4$	11.4	**351**, 333, 307, 289, **279**, 271, **251, 235**, 217, 207, 199, **145**	11.4	8	8	**351**, 333, 307, 289, **279**, 271, **251, 235, 145**
$15R\text{-}LXB_4$	11.2	**351**, 333, 307, 297, 271, 261, **251, 235, 221,** 215, 203, 177, **115**	11.2	5	8	**351**, 333, 307, 261, **235, 221**, 203, 189, **115**
LTB_4	13.7	**335**, 317, 299, 291, 273, 255, **219**, 205, **195**, 179, 177, 161, 151, 133, **115**	13.7	69	9	**335**, 317, 299, 291, 273, 255, 205, **195**, 179, 161, **115**
5S,12S-diHETE	13.8	**335**, 317, 299, 291, 273, **219**, 205, **195**, 179, **115**	13.8	7	4	**335**, 317, 299, 291, 273, **219**, 205, **195**, 179
6-trans-LTB_4	13.1	**335**, 317, 291, 255, **219, 195**, 179	13.1	19	7	**335**, 317, 291, 273, 255, **219, 195**, 179
6-trans-12-epi-6-trans- LTB_4	13.3	**335**, 317, 299, 291, 273, 255, **219, 195**, 161, **115**	13.3	28	7	**335**, 317, 299, 291, 273, 255, **219, 195,** 161, **115**
20-OH-LTB_4	8.7	**351**, 333, 315, 289, 271, **195**, 179, 177, 151	8.7	88	8	**351**, 333, 315, 289, 271, **195**, 179, 177, 151
20-COOH-LTB_4	8.3	**365**, 347, 329, 321, 303, 285, **249**, 205, **195**, 177, 161, **115**	8.3	60	8	**365**, 347, 329, 321, 303, 285, **249**, 205, **195**, 177, 161, **115**
LTC_4	10	**626**, 608, 582, **497**, 479, **319, 308**, 301, 171	10	17	7	**626**, 608, 582, **319, 308**, 301, 171
LTD_4	9	**497**, 479, 301, **189, 179,** 135, **116**	9	30	8	**497**, 479, 301, **189, 179**, 135, **116**
LTE_4	10.4	**440**, **319**, 301, 283, **189, 131**	10.4	15	6	**440**, **319**, 301, 283, **189, 131**
PGD_2	10.6	**351**, 333, 307, 289, **279**, 271, **251**, 233, 217, 189	10.6	37	8	**351**, 333, 307, 289, **279**, 271, 233, 189
PGE_2	10.5	**351**, 333, 315, 289, **279**, 271, **251**, 235, 189	10.5	27	6	**351**, 333, 315, 289, **279**, 271, 235, 189
PGF_{2a}	10.8	**353**, 335, 317, **309**, 291, **281**, 273, 263, 235, 219, 193, 191, 173	10.8	8	8	**353**, 335, 317, **309**, 291, 273, 263, 193, 191
TxB_2	10	**369**, 325, 307, 289, **269**,195, 177, **169**, 125	10	8	8	**369**, 325, 307, 289, **269**, 195, 177, **169**, 125

RT = Retention time, S/N = Signal to Noise. Ions marked in **Bold** denote fragments

Appendix 6. Baseline peripheral blood lipid mediator profiles in RA patient.

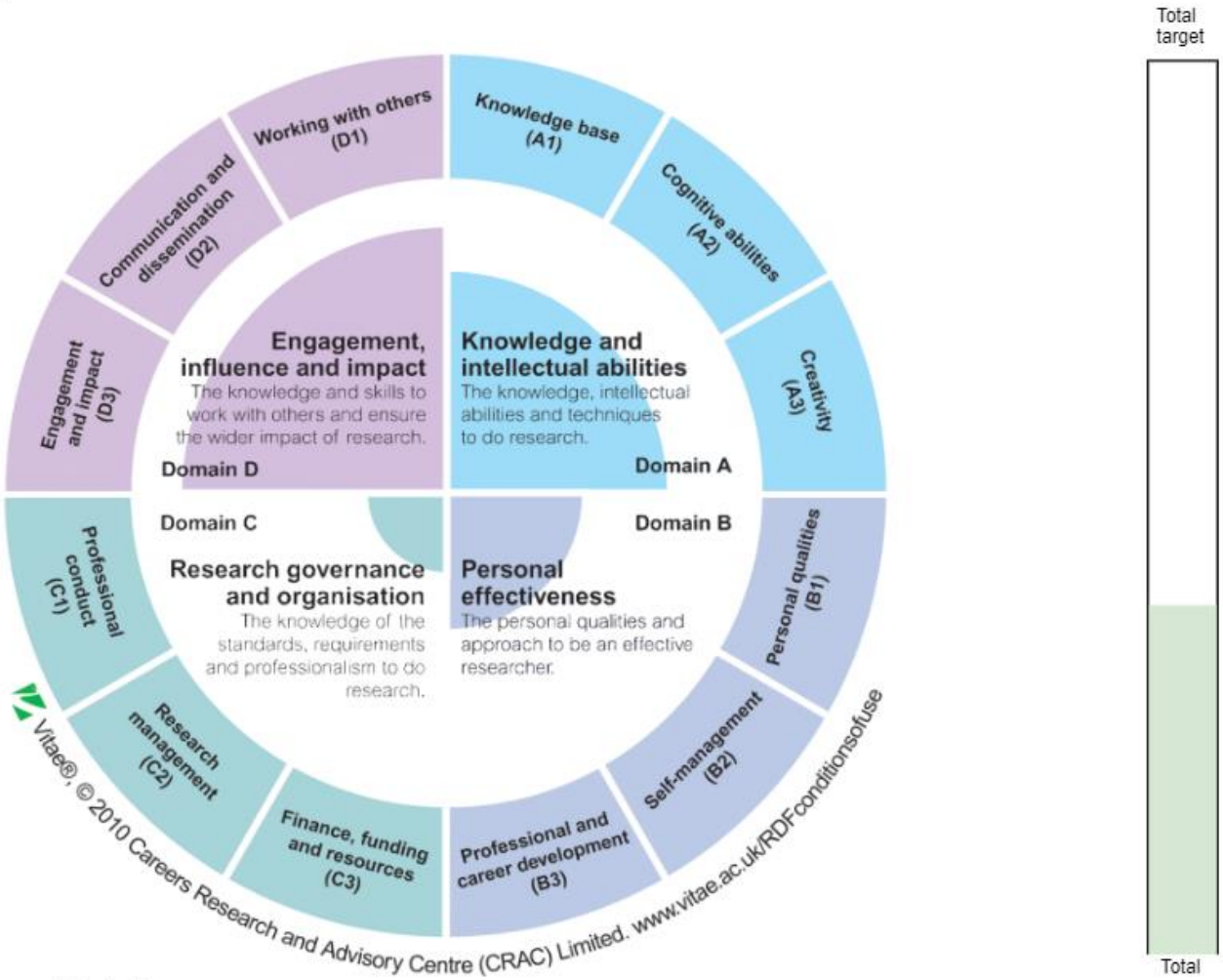
	DMARD Responders			DMARD Non-Responders		
DHA bioactive metabolome	**Mean**	**±**	**SEM**	**Mean**	**±**	**SEM**
RvD1	0.92	±	0.44	2.20	±	0.61
RvD2	1.75	±	0.73	1.52	±	1.52
RvD3	0.52	±	0.15	1.06	±	0.24
RvD4	1.40	±	0.53	5.70	±	1.19
RvD5	7.90	±	1.06	8.17	±	1.34
RvD6	1.82	±	0.66	1.30	±	0.40
17R -RvD1	2.43	±	1.62	3.69	±	1.92
17R -RvD3	0.13	±	0.05	0.60	±	0.20
PD1	0.78	±	0.27	0.99	±	0.33
PDX	0.63	±	0.21	1.39	±	0.46
17R-PD1	0.31	±	0.13	0.74	±	0.40
22-OH-PD1	2.36	±	2.36	1.91	±	1.46
PCTR1	0.13	±	0.13	0.00	±	0.00
PCTR2	1.48	±	0.65	2.10	±	1.80
PCTR3	16.87	±	6.74	4.75	±	4.63
MaR1	0.00	±	0.00	0.00	±	0.00
MaR2	0.47	±	0.20	0.56	±	0.31
7S,14S-diHDHA	3.21	±	1.57	3.65	±	1.73
4S,14S-diHDHA	0.98	±	0.48	4.29	±	1.13
22-OH-MaR1	4.64	±	3.77	5.31	±	3.98
14-oxo-MaR1	0.73	±	0.38	2.65	±	0.78
MCTR1	0.16	±	0.16	0.38	±	0.21
MCTR2	1.42	±	0.46	1.44	±	1.04
MCTR3	8.45	±	2.67	6.57	±	2.92
n-3 DPA bioactive metabolome						
RvT1	0.92	±	0.69	0.66	±	0.31
RvT2	2.97	±	1.31	2.26	±	0.71
RvT3	0.49	±	0.17	0.60	±	0.19
RvT4	0.48	±	0.13	0.78	±	0.18
$RvD1_{n\text{-}3\ DPA}$	4.73	±	0.78	4.50	±	0.58
$RvD2_{n\text{-}3\ DPA}$	3.64	±	1.22	2.32	±	0.96
$RvD5_{n\text{-}3DPA}$	3.39	±	1.32	3.06	±	1.26
$PD1_{n\text{-}3\ DPA}$	0.72	±	0.18	1.26	±	0.38
10S, 17S-diHDPA	0.69	±	0.33	2.11	±	0.49
$MaR1_{n\text{-}3\ DPA}$	0.07	±	0.07	1.25	±	0.49
EPA bioactive metabolome						
RvE1	1.98	±	0.56	1.94	±	0.83
RvE2	0.30	±	0.25	0.71	±	0.34
RvE3	2.89	±	0.78	3.96	±	0.95
AA bioactive metabolome						
LXA_4	0.72	±	0.28	0.96	±	0.60
LXB4	5.14	±	3.99	3.05	±	1.47
15R-LXA_4	0.97	±	0.26	5.71	±	2.27
15R-LXB_4	3.90	±	1.09	9.13	±	2.60
5S,15S-diHETE	36.90	±	7.28	46.68	±	7.59
LTB_4	65.59	±	13.57	115.09	±	37.71
5S,12S-diHETE	11.42	±	9.03	12.69	±	4.52
6-trans-LTB_4	2.93	±	0.57	5.37	±	1.82
12-epi-6-trans-LTB_4	3.56	±	0.73	7.89	±	2.10
20-OH-LTB_4	61.80	±	20.51	147.89	±	50.54
20-COOH-LTB_4	12.86	±	5.40	22.68	±	7.65
LTC_4	12.34	±	3.16	18.59	±	8.10
LTD_4	7.42	±	2.20	6.93	±	2.71
LTE_4	32.99	±	7.84	72.85	±	16.13
PGD_2	9.00	±	2.47	15.22	±	7.22
PGE_2	13.39	±	5.30	18.75	±	5.76
PGF_{2a}	29.26	±	15.96	27.65	±	7.49
TXB_2	33.46	±	13.46	83.09	±	25.63

Results are expressed pg/ml. n = 30 for DMARD Responders and n = 22 DMARD Non-Responders.

Appendix 7: Transferable skills (via Centre for Academic and Professional Development) Training Record.

Mr EA Gomez Cifuentes (170531937)

Progress



Personal Details

Full Name:	Esteban Alberto Gomez Cifuentes	**Gender:**	Male
Username:	hhy530	**Email:**	e.a.gomezcifuentes@qmul.ac.uk
Telephone:		**Mobile:**	+447741093970
Enrolment Status:	R-E-E	**Programme:**	RRPF-QMWHRN1 PhD FT WHRI (Non-Clinical)
Course Name:	PhD FT William Harvey Research Institute (Non-Clinical)	**Award Code:**	RP
Start Date:	01-Oct-2019	**Expected End Date:**	01-Oct-2023
Route:	RSWHN		
Faculty:	Medicine and Dentistry	**School:**	William Harvey Research Institute
Department:	William Harvey Research Institute		

Supervisors

Title	Given Names	Last Name	Telephone	Email	Active
Dr	Jesmond Paul	Dalli		j.dalli@qmul.ac.uk	true
Prof	Conrad	Bessant		c.bessant@qmul.ac.uk	true

Points Summary

Year	Type	Pts:	A	B	C	D	Total	Cap:	A	B	C	D	Total
1st	**Conference attendance sub-total**		**12.0**	**8.0**	**0.0**	**0.0**	**20.0**						
	Basic IT/Software course		3.0	0.0	0.0	0.0	3.0						
	Core research knowledge or methods course (e.g. LTCC, IALS courses, masters lectures)		5.0	0.0	0.0	0.0	5.0						
	Course/event attendance sub-total		**8.0**	**0.0**	**0.0**	**0.0**	**8.0**						
	External funding application <£2,000		0.0	0.0	4.0	2.0	6.0						
	Funding application sub-total		**0.0**	**0.0**	**4.0**	**2.0**	**6.0**						
	Journal Club/Reading Group/lab meeting Presentation		14.0	0.0	0.0	14.0	28.0						
	Giving presentations sub-total		**14.0**	**0.0**	**0.0**	**14.0**	**28.0**						
	Journal Club/Reading Group/lab meeting/mentoring group - attendance		9.5	0.0	0.0	0.0	9.5						
	Meeting/club/reading group attendance sub-total		**9.5**	**0.0**	**0.0**	**0.0**	**9.5**						
	Seminar attendance (series)		2.0	0.0	0.0	0.0	2.0						
	Seminar attendance sub-total		**2.0**	**0.0**	**0.0**	**0.0**	**2.0**						
	Mentoring/supervising of Project Student		2.0	1.0	0.0	2.0	5.0						
	Teaching sub-total		**2.0**	**1.0**	**0.0**	**2.0**	**5.0**						
	Refereed Publication (Journal Paper, Book chapter, not abstract) acceptance		2.0	0.0	0.0	8.0	10.0						
	Refereed Publication (Journal Paper, Book chapter, not abstract) submission		2.0	0.0	0.0	8.0	10.0						
	Written publications sub-total		**4.0**	**0.0**	**0.0**	**16.0**	**20.0**						
	Year 1 Total (with caps applied)		**51.5**	**9.0**	**4.0**	**34.0**	**98.5**						
Total	*Conference Attendance (Four days)*		12.0	8.0	0.0	0.0	20.0						
	Conference attendance sub-total		**12.0**	**8.0**	**0.0**	**0.0**	**20.0**		**18.0**	**12.0**			**30.0**
	Basic IT/Software course		3.0	0.0	0.0	0.0	3.0		5.0				5.0
	Core research knowledge or methods course (e.g. LTCC, IALS courses, masters lectures)		5.0	0.0	0.0	0.0	5.0		100.0				100.0
	Course/event attendance sub-total		**8.0**	**0.0**	**0.0**	**0.0**	**8.0**						
	External funding application <£2,000		0.0	0.0	4.0	2.0	6.0				8.0	4.0	12.0
	Funding application sub-total		**0.0**	**0.0**	**4.0**	**2.0**	**6.0**						
	Journal Club/Reading Group/lab meeting Presentation		6.0	0.0	0.0	6.0	12.0		6.0			6.0	12.0
	Giving presentations sub-total		**6.0**	**0.0**	**0.0**	**6.0**	**12.0**						
	Journal Club/Reading Group/lab meeting/mentoring group - attendance		9.5	0.0	0.0	0.0	9.5		30.0				30.0
	Meeting/club/reading group attendance sub-total		**9.5**	**0.0**	**0.0**	**0.0**	**9.5**						
	Seminar attendance (series)		2.0	0.0	0.0	0.0	2.0		30.0				30.0
	Seminar attendance sub-total		**2.0**	**0.0**	**0.0**	**0.0**	**2.0**		**30.0**				**30.0**
	Mentoring/supervising of Project Student		2.0	1.0	0.0	2.0	5.0		8.0	4.0		8.0	20.0
	Teaching sub-total		**2.0**	**1.0**	**0.0**	**2.0**	**5.0**						
	Refereed Publication (Journal Paper, Book chapter, not abstract) acceptance		2.0	0.0	0.0	8.0	10.0		4.0			16.0	20.0
	Refereed Publication (Journal Paper, Book chapter, not abstract) submission		2.0	0.0	0.0	8.0	10.0		4.0			16.0	20.0
	Written publications sub-total		**4.0**	**0.0**	**0.0**	**16.0**	**20.0**						
	Total (with caps applied)		**43.5**	**9.0**	**4.0**	**26.0**	**82.5**						
	Grand Total (with caps applied)		**43.5**	**9.0**	**4.0**	**26.0**	**82.5**						
	Target		60.0	20.0	15.0	30.0	210.0						