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

---

18 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 **3**
- Rheumatoid arthritis and treatments 3
- Specialised pro-resolving mediators in RA 5
- Genetic variants and SPM-related genes in RA and inflammatory disorders 7
- Bioinformatic strategies to identify biomarkers 8
- Objectives 10

Chapter 2: Methods **12**
- Materials 12
- Pathobiology of Early Arthritis Cohort 12
- Lipid mediator profiling 12
- Data 15
- Machine learning models 15
- Model evaluation 17
- Feature selection and model improvement 18
- Learning curves 18
- Pathotypes analyses 18
- Network analyses 19
- Enzyme transcript expression 19
- Candidate gene analysis 19
- Statistical analysis 21
- Data and software availability 21

Chapter 3: Results **22**
- Plasma SPM concentrations can predict the response to DMARD treatment 22
- Improved machine learning models identify specific SPM that are predictive of DMARD treatment response 27
- Increased eicosanoid and SPM concentrations in peripheral blood from DMARD non-responders 28
- Baseline plasma SPM concentrations are linked with distinct disease pathotypes in RA patients 28
- Decreased SPM concentrations in DMARD non-responder 6 months after treatment initiation 32
- Candidate gene analysis identifies genomic variants associated with RA in SPM-related genes 35

Chapter 4: Discussion **40**
- Future work 44

References **45**

Appendix **55**

Chapter 1: Introduction

Rheumatoid arthritis and treatments

Inflammation is a primordial protective process mounted by the body to ward off infections and repair damaged tissues. An uncontrolled inflammatory response can lead to the development and progression of inflammatory diseases, such as rheumatoid arthritis (RA), inflammatory bowel disease (BID), cardiovascular conditions (such as heart attack), etc [1,2]. These diseases represent a social and economic burden around the world; especially RA affects 17.6 million people worldwide, reducing their quality of life; and, only in the UK, generating costs around £560 million a year [3].

RA is a chronic autoimmune inflammatory disease characterized by 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 [4].

RA is known to affect primarily the synovium [5]. The progression of the disease can be separated into three stages: the non-specific inflammation, the amplification in the synovium and the chronic inflammation [6].

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 predisposition and environmental factors not yet fully understood [6,7]. 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. Cluster of differentiation (CD)4^+ and $CD8^+$ T cells, B cells, and, especially, macrophages and synovial fibroblasts will accumulate to promote synovial hyperplasia, known as pannus [8].

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 [8,9].

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 [10].

Although there is no 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 [11,12]. 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 [8,9,12]; however, synthetic DMARDs are still preferred, with low-dose of methotrexate (MTX) as the most common drug used [11].

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 for the purine/pyrimidine synthesis (and therefore the DNA and RNA synthesis), through the inhibition of the dihydrofolate reductase (DHFR) [13,14]. 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 [13,14]. 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 [13–15].

Currently, MTX in combination with other DMARDs, such as hydroxychloroquine and sulfasalazine, is the first regime used in the treatment of patients with RA. 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 [11]; 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 [16].

Similar concerns arise when patients are treated with biologic agents; particularly, anti-TNF therapies, which have shown to be successful in the treatment of RA in cases where synthetic DMARDs treatment fails. These therapies can control the signs and symptoms of the disease, lead to functional improvement and reduce the

time RA patients need to stay away from work due to the illness [17]. Moreover, anti-TNF in combination with MTX was found to have greater benefits than the treatment with only one of them [17,18]. However, and even though the mechanisms of action of the anti-TNF are more understood than the classical DMARDs treatments (neutralisation of TNF interactions, regulation of cytokine network, apoptosis, cell recruitment, reduction of proinflammatory molecules, such as metalloproteinases and IL-1 [18]), around one-third [19,20] of patients do not respond to this treatment, with side effects associated with an increased risk of infection and malignancy being reported [19].

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 [21]. It is therefore 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 inflammation and prevent disease chronicity [22]. 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 [23]. The first group of SPM discovered, 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 [24]. Since then, other families of SPM have been discovered including resolvins, protectins and maresins [23].

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). Each of the fatty acids gives rise to structurally distinct SPM families (Figure 1). The SPM pathways start with the insertion of oxygen into the backbone of the fatty acid, a reaction primarily carried out by lipoxygenases (ALOX), cyclooxygenase-2 (COX-2) and cytochrome P450 (CYP450) [23,25].

DHA is converted by ALOX15 and/or ALOX5 to produce D-series Resolvins (RvDs), protectins (PDs) and Protectin conjugates in tissue regeneration (PCTRs); while ALOX12 converts this fatty acid to Maresins (MaRs) and Maresin conjugates in

tissue regeneration (MCTRs). When COX-2 is acetylated in the presence of aspirin, Aspirin-triggered resolvins (17R-RvDs) and protectins (17R-PDs) are produced [26].

Resolvins counter regulate proinflammatory mediator actions and promote resolution via regulating the ability of macrophages to uptake debris and clear microbes. Protectins display anti-inflammatory, neuroprotective and antiapoptotic actions; while maresins promote macrophage phagocytosis and play a role in pain reduction and tissue regeneration [27].

n-3 DPA is the precursor of 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 [27,28]. $RvD_{n\text{-}3\ DPA}$ are produced through the enzymatic action of ALOX15 and ALOX5, whereas the biosynthesis of $PD_{n\text{-}3\ DPA}$ is initiated by ALOX15. 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 $MaR_{n\text{-}3\ DPA}$ reduce neutrophil infiltration in peritonitis models. These mediators also counter regulate the production of inflammatory cytokines, enhancing macrophage phagocytosis and also reducing neutrophil chemotaxis and adhesion to endothelial cells [27–29].

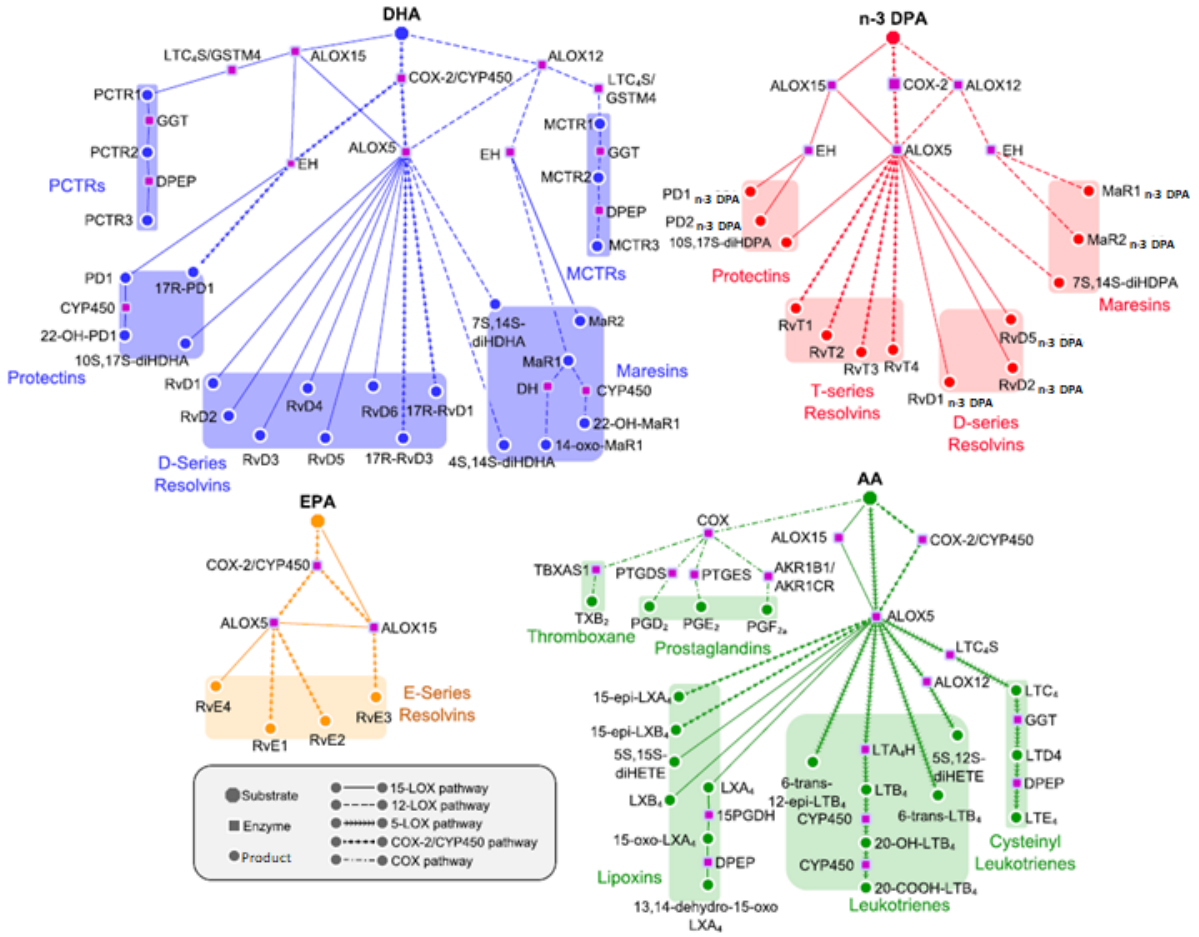


Figure 1. Biosynthesis of SPM. Biosynthetic pathways leading to the production of the different SPM families via the enzymatic conversion of the essential fatty acids DHA, n-3 DPA, EPA and AA.

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 [23,25].

AA, in addition of producing proinflammatory mediators such as prostaglandins (PGs), leukotrienes (LKs) and thromboxane (TX), is the substrate for Lipoxins (LXs), produced by either COX-2 or ALOX15 and ALOX5. LXs inhibit chemotaxis and the production of proinflammatory cytokines by neutrophils, monocytes and macrophages. LXs help in the recruitment of neutrophils in the site of inflammation when they are needed, through promoting the extrusion of neutrophil's pseudopods *via* the increase of cytosolic calcium. LXs also clear these cells during the resolution step [30].

In the inflammation process, as in the case of RA, pro-inflammatory lipid mediators such as PGE_2, LTB_4 and PGD_2 play a role in mediating pain, heat and swelling, while SPM help to control the ongoing inflammation [31,32]. It has been observed that an increase of SPM in plasma is linked with a reduction in joint inflammation and promotes joint protection in RA patients: an increase of RvE2 in the synovial fluid has been associated with decreased joint pain, while during ongoing joint inflammation there is a downregulation of several SPM including RvDs and aspirin-triggered RvDs [22,33]. An increase of RvTs production was associated with a decrease of disease activity in the joint and reduction of leukocyte migration and activation [21]. It has been demonstrated that the presence of the pro-inflammatory PGE_2 serves as a feedback signal to switch from inflammation to resolution, mediating the production of LXA_4 and 15R-LXB_4, essentials to restore the homeostasis after inflammation in RA [28,34]. The evidence of resolution activity of SPM in RA suggests that these mediators can be considered as possible new therapies, and, on the same note, as biomarkers of the disease.

Genetic variants and SPM-related genes in RA and inflammatory disorders

The fact that the biological actions of SPM are stereospecific indicates that these molecules interact with receptors activating the mechanisms of resolution of inflammation [35]. The G-protein-coupled receptors (GPCRs) are a class of receptors associated with SPM activity [36,37]. The activation of one or several GPCRs by SPM induces cellular functions that promote resolution of inflammation [38]: GPR37 mediates the protective actions of PD1 (increased of phagocytosis and alteration of cytokine release); while GPR32 and GPR18 regulate the biological actions of RvD1 and RvD3 (counter-regulation of proinflammatory mediator actions and enhance of resolution via macrophages activity) [39–41]; recently, GPR101 was found to mediate

the biological actions of $RvD5_{n-3\ DPA}$, enhancing the protective mechanisms of macrophages, neutrophils and monocytes [35].

Several studies in genetically modified mice, in which one of these receptors was knockdown, demonstrated a reduction or uncoupling of the protective actions of SPM. For example, *ALX/FPR2* null mice, the receptor for LXA_4 and RvD1, display an unrelenting inflammatory response and a defective resolution [42]; while *GPR101* knockdown in mice reversed the joint protection of $RvD5_{n-3\ DPA}$ during inflammatory arthritis and its protective mechanisms during bacterial infections [35].

Even though the use of knockdown mouse models to study the function of the enzymes and receptors associated with SPM has been helpful to understand the relationship between these genes and lipid mediators, the discoveries in animal models do not always translate to the human level [43]. Moreover, little is known about the impact of loss-of-function polymorphisms in these enzymes and receptors in SPM dysregulation, immune cell biology and the progression of RA and inflammatory disorders. Some studies have demonstrated the correlation of single nucleotides polymorphisms (SNPs) in SPM-related genes with different diseases such as colon and rectal cancer (SNPs in *COX-1* and low DHA intake were associated with a higher risk in colon cancer), obesity (a genetic variant in the *ERV1/ChemR23* gene was associated with enhancing leukocyte responsiveness to RvE1) and bleeding [17-19]. Therefore, it is important to understand the possible effects of these genetic variants in SPM production and activity in a human population. Moreover, these associations can be correlated with the progression of RA and other inflammatory conditions, especially knowing that there is a higher risk of having a cardiovascular disease if you suffer from an autoimmune inflammatory condition such as RA [17].

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 [44]. 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 [45], 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 [45–48].

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) [46]. Machine learning, different from univariate association analyses, can address some of the issues associated with biological data while also providing strategies to consider unknown associations between the different variables present in the samples [49].

Focusing on RA, 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 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.

For the analysis of genetic variants, genome-wide association analysis (GWAS) has been a strategy widely used for the analysis of millions of SNPs across the entire genome to identify associations with specific phenotypes [50]. Although around 50000 SNPs have been associated with different diseases and traits, GWAS presents several limitations, including a high number of misleading associations due to population stratification and the identification of variants and genes with no biological relevance leading to limited clinical predictive value [51]. Another approach, called candidate gene (CG) association studies, even though it has a lower success rate of identifying associations than classical GWAS [51,52], is a valid strategy for discovering SNPs with a direct biological effect on the disease of interest, and it can lead to the identification of potential biomarkers and to a better understanding of the condition. More than 1000 associations in RA have been identified using GWAS but none of those SNPs has been located in SPM-related genes. I hypothesized that a focused CG approach can highlight relevant genetic variants in RA and other inflammatory conditions that are lost in the GWAS noise.

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, I expect to establish novel tools to interrogate the lipid mediator biosynthetic pathways and explore the utility of combining lipid mediator profiles. I will also perform a CG association analysis using the UK Biobank dataset to provide new insights into molecular mechanisms that may be linked with differential SPM regulation in RA, the response to DMARD treatments and other related inflammatory conditions.

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*: I 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:* to gain insight into the regulation of specific SPM families by distinct therapeutics, I will use machine learning to interrogate lipid mediator profiles described before. Here I will build classification models with optimal predictive capability and then I 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 as to evaluate their prediction performance. This to build 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 taking 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:* I will use these biosynthetic networks to assess which mediator pathways become dysregulated in RA patients with distinct disease pathotypes.

- *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 RA and other inflammatory conditions*: I will identify SNPs in SPM-related genes associated with RA and other inflammatory conditions and run CG analysis between the genotypes and the different responses to treatment with the aim of trying to understand the role of the genetic variance in the progression of the disease. I plan to run quality control over the identified SNPs and create logistic regression models to find genetic associations. False discovery rate and other relevant artefacts when working in CG 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, I plan of requesting samples from those patients to evaluate how genetic variants can have an effect on 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) [10]. All patients 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 of MeOH+IS/sample was 4:1 (vol/vol); in this case, 0.5 mL of sample was added to 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 of the conditioning of the C18 columns with MeOH, followed 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 an 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 [28].

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 and a peak with a minimum of 5 data points. The identity of each mediator was further confirmed by matching of at least 6 diagnostic ions to that of the reference standard, with a minimum of one backbone fragment being identified in at least one of the samples from the sample set.

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, I 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 from 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 and 12 max US Synovial Thickness and US Power Doppler scores. Age, sex and clinical parameters not mentioned were not considered for this first approximation of creating a model able to classify the response of RA patients to DMARD treatment.

A second patient cohort of 58 patients (36 responders and 22 non-responders to DMARDs therapy) was obtained from the PEAC study and was used as the test dataset for the lipid mediator profiling and Clinical Score models, and also as the training dataset for the improved models based on specific biomarkers.

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 [53]. 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 pro-inflammatory lipid mediators.

Four supervised machine learning methodologies were used to create the classifier models: Bayesian generalized linear model, Elastic net regression [54], Support Vector Machine (SVM) [55] and random forest [56].

A Bayesian classifier is a statistical approach based on applying Bayes theorem: the probability of having a first condition due to the presence of a second condition can be used to predict the second condition based on having the first one [57]. The Bayesian methods take prior data (the classification in the training dataset) and combine it with the likelihood of new data being assigned to a specific class (features values associated with a class) to produce a posterior distribution (the probability of class given the provided data) and determine what is the most probable classification of new cases [58].

The Bayesian generalized linear models were created using the R Packages “caret” (https://cran.r-project.org/web/packages/caret/index.html) and “arm” (https://cran.r-project.org/web/packages/arm/index.html). I used bootstrapping (new datasets are created from the original data by randomly sampling with replacement) as the validation strategy. No tuning parameters were required for the creation of these models.

Elastic net is a regularized regression that combines the constraints of the LASSO (Least Absolute Shrinkage and Selection Operator) and Ridge methods [59]. While LASSO performs feature selection based on a penalization system to shrink the most number of coefficient values to zero, translating to a reduction of the number of predictors to improve the accuracy of the model; Ridge regression penalizes the features to be small without forcing them to be zero, this with the purpose of decrease model complexity while keeping all the variables [54]. Elastic net solves the LASSO’s disadvantage of eliminating important variables that are correlated with each other, and also addresses Ridge’s downside of keeping uninformative variables.

I implemented the Elastic net methodology using the R Packages “caret” and “glmnet” (https://cran.r-project.org/web/packages/glmnet/index.html). I used bootstrapping to identify the optimal penalty parameters lambda (λ), a numeric value

defining the amount of shrinkage, and alpha (α), the mixing parameter between ridge (α = 0) and LASSO (α = 1). Optimal penalty parameters are defined as those that generate the most stable model reducing the variance of the prediction error.

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 [60]. The algorithm systematically increases the dimensionality of the data using a kernel function until it finds the optimal hyperplane that serves as a threshold between the groups [54]. The kernel function used here was a nonlinear kernel 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 [54,61].

The SVM models were created using the R Package "classyfire" (https://cran.r-project.org/src/contrib/Archive/classyfire/), which uses bootstrapping as the validation strategy. To identify the best model, I 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 giving us stable models for all groups tested. Furthermore, I also used the inbuilt automatic optimization step that includes minimization of the bootstrapping error to improve and validate the models [61].

Random forest operates by getting the consensus of weak decision tree classifiers [60]. 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 [62]. It's useful that the algorithm does not use all the available features to create each tree because this allows the identification of the stronger predictors in the data set and permits further analysis to improve the prediction models [63].

In the present studies, I 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. I 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, I 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 I 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.

Model evaluation

Receiver operating characteristic curves (ROC curves) were built to evaluate the prediction accuracy of the models when predicting between DMARDs responder and

DMARDs non-responder in a test dataset. ROC curves are created by plotting the true positive rate against the false positive rate, showing the sensitivity and specificity of the model when the discrimination threshold changes. The area under the curve (AUC) is calculated as the prediction performance of the models. ROC curves were created using the R package "pROC" (https://cran.r-project.org/web/packages/pROC/index.html). AUC values close to 1 (AUC > 0.8) refer to good classifier models.

Alongside ROC curves, other statistics such as the percentage of correctly classified samples (% accuracy score), specificity and sensitivity were also calculated.

Feature selection and model improvement

As random forest showed the best scores during the evaluation step, the model improvement was based on the random forest methodology. Different models were created using only the most relevant lipid mediators from the "importance" function of the "randomForest" package. This function organizes the model's features by relevance based on the model's decreased mean accuracy when the specific feature is not present. The % accuracy score and AUC (ROC) were calculated for all the models and, according to the results, the best model and the most relevant biomarkers for the classification of the DMARD-responder and DMARD non-responder patients were selected.

Learning curves

In order to establish the ideal number of samples to create our random forest models; and thereby confirm that I used enough samples to have stables models, I produced a learning curve of the model using either all the lipid mediators or the model created using the most relevant lipid mediators based on the "importance" analysis from the "randomForest" package. Machine learning models were created using random forest (best *mtry* value and *ntree* of 10,000) and increasing the number of samples used starting with 10. I recorded the final accuracy score for each model and the standard deviation of the out of the bag error across the 10,000 trees. I added one sample for the responders and one for the non-responders each time. Samples from the responders group were added one by one after the available number of non-responder was reached.

Pathotypes analyses

All the data (training and test cohorts) was separated based on the specific pathotype shown for the patients: pauci-immune/fibroid (n = 28), lympho-myeloid (n = 27) and diffuse-myeloid (n = 31). This was made to seek better classification

models and see if specific lipid mediators were responsible for the different manifestations of the disease. The models were built using random forest and different statistical scores were calculated for the validation of each model.

Network analyses

Statistical differences between the normalised concentrations (expressed as the fold change) of the lipid mediators from the DMARD non-responder and DMARD responder groups were determined using a two-sided t-test followed by a multiple comparison correction using Benjamini–Hochberg procedure. Based on these differences, lipid mediator biosynthesis networks were constructed using Cytoscape v3.7.1 [64]. The different pathways were illustrated using different colours and line shapes, while up or downregulated mediators were denoted using upward and downward facing triangles, respectively, and on changes of the node's size. Bolded mediators represent statistical differences between the two groups. The comparison between DMARD non-responders and DMARD responders was made with pre and post-treatment data.

Enzyme transcript expression

RNA sequencing data from naive DMARD patients were obtained from Humby F, *et al* [10]. Normalization of the RNA-seq raw data and differential gene expression analysis between DMARD responder and DMARD non-responder were performed using the quasi-likelihood method of the Bioconductor R package "edgeR" (https://bioconductor.org/packages/release/bioc/html/edgeR.html). Results are expressed as the log counts per million of each gene.

Candidate gene analysis

For CG analysis, I used the phenotypic and genomic data of participants from the UK Biobank project (April 2020 release) [65]. The UK Biobank project is a cohort study compiling phenotypic (including health-related information) and genomic data of ~500000 participants from across the UK.

I identified as RA patients individuals from the UK biobank reporting the disease code (1464 from UKBID 20002) associated with RA. For the control group, I selected healthy volunteers who do not report inflammatory conditions, including RA, cardiovascular conditions, etc. For this first approximation, I focused on British and South Asian participants older than 60 after checking the demographic of RA and controls and finding similar distributions between the two groups.

I first identified 4606 and 110 participants with RA for the British and South Asian population, respectively. Next, I applied filters to remove low-quality variants and

samples: exclude SNPs with more than 20% of missing genotype, with substantial differences in missing genotype between groups, with a Hardy–Weinberg equilibrium *p*-value < $1x10^{-3}$, with minor allele frequency (MAF) < 1% for common variants and MAF between 0.5 and 1% for rare variants. To filter low-quality samples I remove individuals with more than 10% of missing genotypes and I exclude duplicate and related (up to third-degree) samples. For related samples, I used the information provided by the UK biobank. Quality controls were performed using PLINK1.9 toolset [66] (http://pngu.mgh.harvard.edu/purcell/plink/). After the quality control steps, the number of RA samples reduced to 4121 and 106 for the British and South Asian population, respectively.

CG analysis was focused on SPM-related genes, including enzymes associated with SPM biosynthetic pathways: *ALOX5, ALOX12, ALOX15, ALOX15B, GSTM4, PTGS2, EPHX2, LTC4S, DPEP1, DPEP2, DPEP3, mGST2, mGST3, GGT1 and GGT2*; and SPM receptors: *LGR6, GPR18, GPR32, GPR37, GPR99, GPR101, CMKLR1, CYSLTR1, CYSLTR2, ALX/FPR2 and LTB4R*. I used PLINK1.9 to create logistic regressions considering covariates and genetic models (additive, dominant, recessive, genotypic and hethom) to identify SNPs associated with RA in the previously mentioned genes. All models included sex, age, body mass index (BMI), genotyping array, smoking status, alcohol consumption, meat intake, UK biobank centre where the sample was collected and the first 10 principal components associated with population stratification as covariates. For each model, I calculated the genomic inflation factor (ƛ) and created Q–Q (quantile-quantile) plots to verify how well the models fit the dataset (ƛ close to 1 is considered a good fit). GC analyses were performed independently for the British and South Asian population.

As these analyses are candidate gene approach and not GWAS, instead of using the *p*-value threshold of $5x10^{-8}$, I considered statistically significant associations those SNPs with a consensus adjusted *p*-value < 0.05 using different multiple comparison corrections methodologies (Bonferroni, Holm, Sidak single-step, Sidak step-down and Benjamini & Hochberg) as suggested by Clarke GM, *et al* [67].

I used gnomAD (https://gnomad.broadinstitute.org/) to predict the outcome of the candidate SNPs from the CG analysis. I also used the GWAS catalogue (https://www.ebi.ac.uk/gwas/home) and PhenoScanner (http://www.phenoscanner.medschl.cam.ac.uk/) to verify if those SNPs were associated with other diseases. I used HaploReg v4.1 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) to investigate patterns of linkage disequilibrium (non-random associations between alleles from different loci). Finally, I explore the expression quantitative trait loci (eQTLs), alleles that explain differences in the mRNA expression of specific genes (cis-eQTLs when the loci is located near to the affected gene, and trans-eQTLs when the loci is distant to the gene) [68]. I used QTLbase (http://mulinlab.org/qtlbase/index.html) to discover

whether candidate SNPs were eQTLs and explain gene expression phenotype in SPM-related genes.

Statistical analysis

I performed all statistical analyses and data derivation using R (https://www.r-project.org/), SIMCA 14.1 (Umetrics, Umea, Sweden) and MetaboAnalyst 4.0 [63,69]. 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 analyses were performed using partial least squares-discrimination analysis (PLS-DA) or 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. PLS-DA is based on a linear multivariate model that identifies variables that contribute to the class separation of observations (e.g. treatment response) based on 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. Permutation tests (1000 permutations) were performed to indicate how significant the PLSA-DA results were. 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, I measured the plasma lipid mediator concentrations in a matched, deeply phenotyped early RA patient cohort before treatment initiation (Appendix 4 to see patients information). Plasma lipid mediators were identified using LC-MS/MS strategy and different criteria that include matching the retention time of lipid mediators in liquid chromatography when compared against standards (Appendix 5 to see the criteria in detail). In RA patient plasma, I identified lipid mediators from all four major essential fatty acid metabolomes: DHA, n-3 DPA, EPA and AA (Appendix 6). I next used OPLS-DA, a multivariate analysis that generates a regression model based on continuous variables to discriminate between two conditions [69], to define the difference between DMARD responder and DMARD non-responder patients based on the lipid mediator profile. OPLS-DA was able to identify two clusters representing each of the patient groups; the classification using the original data was better in most cases when a permutation test was performed (Figure 2A, B).

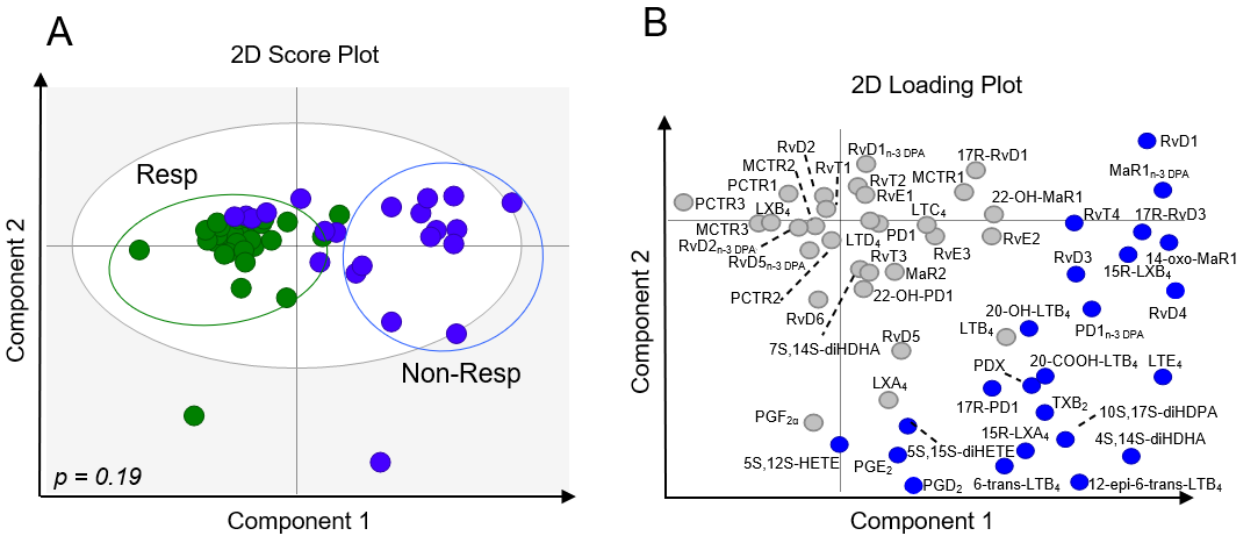


Figure 2. Plasma SPM can differentiate between DMARD-responders and DMARD-non-responders in RA patients. Plasma was collected from RA patients prior to the initiation of treatment with DMARDs and lipid mediator concentrations established using LC-MS/MS based lipid mediator profiling (see methods for 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 scores greater than 1 are highlighted in blue and upregulated in Non-Resp. *p*-value is calculated as the proportion of times that class separation based on permutation obtains similar results as the obtained on the original data. Results are representative of n = 30 Resp and n = 24 Non-Resp.

I then applied machine learning methodologies to create predictive models of the responsiveness to DMARD treatment in RA patients using plasma lipid mediator profiles. I applied four machine learning methodologies: Bayesian classifier, Elastic net regression, SVM and random forest; to build models based on either all the lipid mediators, SPM from each of the essential fatty acid metabolomes or the clinical score. First, I 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, ensemble iteration values in the SVM models and the penalty parameters, λ and α, in the Elastic net 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 was found to be a stable configuration for random forest and SVM, respectively, and gave us the most accurate models (Figure 3A, B). Since 70 resamplings were efficient for the SVM models, I used this number as the number of bootstrap interactions to identify the best combination of λ and α to get the most stable Elastic net model (Figure 3C).

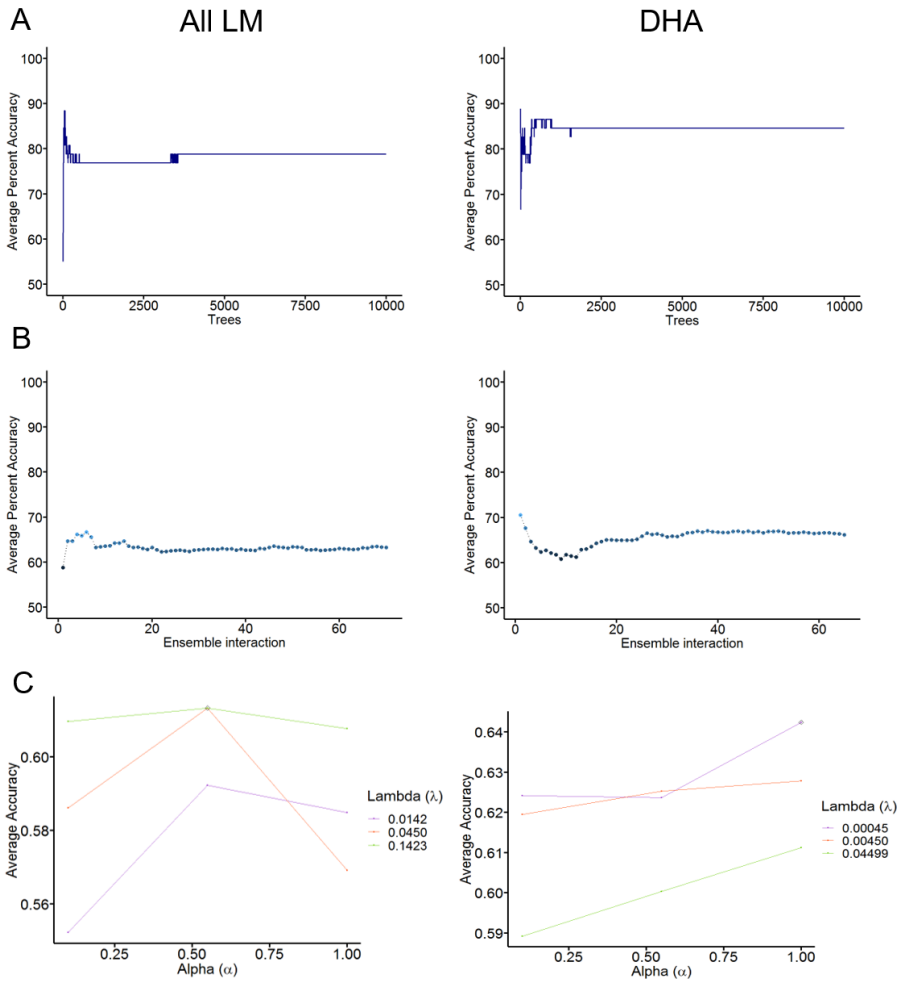


Figure 3. Establishing the optimal parameters for the random forest, SVM and Elastic net models. Representative results obtained for average accuracy for (**A**) random forest, (**B**) SVM and (**C**) Elastic net models by increasing the number of (A) decision trees (*ntree* value), (B) ensemble interactions or (C) selecting the best combination of penalty parameters (λ and α). 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 4 and Table 1). Specifically, I 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 ~81% (~64% for Elastic net, ~62% for SVM and ~59 for Bayesian) and that of the n-3 DPA metabolome was ~69% (~61% using SVM, Elastic net and Bayesian). These values were higher than those obtained using a combination of clinical parameters, including disease duration and RF titre. The accuracy score for the model with all the lipid mediators was ~73%, also higher when compared with the clinical score model. Of note, when creating a learning curve for all the lipid mediators model using the random forest methodology, I observed that the accuracy score of the model stabilized after using around 40 samples, with a trend around 73% and a reduction of the variation of the out of the bag error of all the created trees; this suggesting that the number of samples used is a suitable representation to create a stable model (Appendix 7).

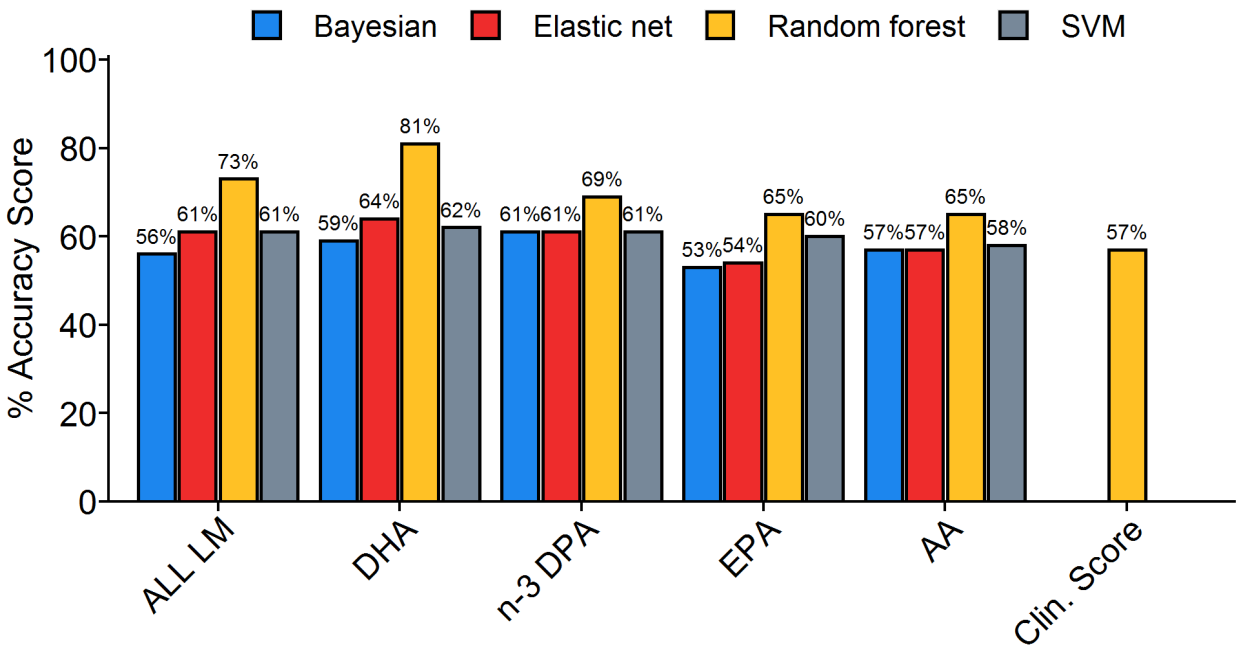


Figure 4. Baseline plasma lipid mediator profiles are predictive of treatment responsiveness in DMARD naive RA patients. Peripheral blood was collected in RA patients prior to DMARD treatment initiation and lipid mediator profiles were established using LC-MS/MS. % accuracy score of prediction models 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 a bayesian generalized linear ("caret" and "arm" packages from R), elastic net regression ("caret" and "glmnet" packages from R), random forest ("randomForest" package from R) or 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 different machine learning methodologies.

Machine Learning methodology	Samples	Model	Number of variables	% Accuracy Score	Model validation						Model evaluation						
					TPR	TNR	TP	FP	TN	FN	TPR	TNR	TP	FP	TN	FN	AUC
arm (Bayesian)	All samples (1st cohort)	Four metabolomes	54	56	0.6	0.5	60	50	50	40	0.33	0.5	33	50	50	67	0.56
arm (Bayesian)	All samples (1st cohort)	DHA metabolome	23	59	0.67	0.49	67	51	49	33	0.47	0.32	47	68	32	53	0.61
arm (Bayesian)	All samples (1st cohort)	n-3 DPA metabolome	10	61	0.73	0.45	73	55	45	27	0.89	0.41	89	59	41	11	0.72
arm (Bayesian)	All samples (1st cohort)	EPA metabolome	3	53	0.82	0.16	82	84	16	18	0.89	0.18	89	82	18	11	0.56
arm (Bayesian)	All samples (1st cohort)	AA metabolome	18	57	0.7	0.39	70	61	39	30	0.64	0.32	64	68	32	36	0.68
glmnet (Elastic net)	All samples (1st cohort)	Four metabolomes	54	61	0.77	0.39	77	61	39	23	0.94	0.5	94	50	50	6	0.72
glmnet (Elastic net)	All samples (1st cohort)	DHA metabolome	23	64	0.62	0.68	62	32	68	38	0.83	0.32	83	68	32	17	0.58
glmnet (Elastic net)	All samples (1st cohort)	n-3 DPA metabolome	10	61	0.74	0.43	74	57	43	26	0.89	0.32	89	68	32	11	0.65
glmnet (Elastic net)	All samples (1st cohort)	EPA metabolome	3	54	0.8	0.19	80	81	19	20	0.92	0.18	92	82	18	8	0.56
glmnet (Elastic net)	All samples (1st cohort)	AA metabolome	18	57	0.71	0.36	71	64	36	29	0.89	0.41	89	59	41	11	0.66
Classyfire (SVM)	All samples (1st cohort)	Four metabolomes	54	61	0.63	0.54	63	46	54	37	0.94	0.05	94	95	5	6	0.53
Classyfire (SVM)	All samples (1st cohort)	DHA metabolome	23	62	0.65	0.56	65	44	56	35	0.78	0.18	78	82	18	22	0.54
Classyfire (SVM)	All samples (1st cohort)	n-3 DPA metabolome	10	61	0.63	0.54	63	46	54	37	0.94	0.23	94	77	23	6	0.66
Classyfire (SVM)	All samples (1st cohort)	EPA metabolome	3	60	0.62	0.52	62	48	52	38	0.92	0.05	92	95	5	8	0.58
Classyfire (SVM)	All samples (1st cohort)	AA metabolome	18	58	0.6	0.48	60	52	48	40	0.92	0.09	92	91	9	8	0.66
randomForest (RF)	All samples (1st cohort)	Four metabolomes	54	73	0.87	0.55	87	45	55	13	0.58	0.64	58	36	64	42	0.72
randomForest (RF)	All samples (1st cohort)	DHA metabolome	23	81	0.9	0.68	90	32	68	10	0.39	0.55	39	45	55	61	0.44
randomForest (RF)	All samples (1st cohort)	n-3 DPA metabolome	10	69	0.83	0.5	83	50	50	17	0.78	0.32	78	68	32	22	0.58
randomForest (RF)	All samples (1st cohort)	EPA metabolome	3	65	0.8	0.45	80	55	45	20	0.53	0.32	53	68	32	47	0.6
randomForest (RF)	All samples (1st cohort)	AA metabolome	18	65	0.77	0.5	77	50	50	23	0.83	0.77	83	23	77	17	0.89
randomForest (RF)	All samples (1st cohort)	Clin. Score	10	57	0.7	0.41	70	59	41	30	0.39	0.38	39	62	38	61	0.56
randomForest (RF)	All samples (2nd cohort)	RvD4, 10S,17S-diHDPA, 15R-LXA$_4$, MaR1$_{n\text{-}3\ DPA}$	4	83	0.92	0.68	92	32	68	8	0.83	0.59	83	41	59	17	0.8
randomForest (RF)	All samples (2nd cohort)	RvD4, 10S,17S-diHDPA, 15R-LXA$_4$, 5S,12S-diHETE, 4,14-diHDHA, MaR1$_{n\text{-}3\ DPA}$	6	86	0.94	0.73	94	27	73	6	0.87	0.64	87	36	64	13	0.79

TPR sensitivity, *TNR* specificity, *TP* true positives, *FP* false positives, *TN* true negatives, *FN* false negatives, *AUC* area under the curve.

Since the best predictors are the ones with the fewest number of features, I further evaluated the ability of the DHA metabolome-based model (best score random forest, SVM and Elastic net) to accurately categorize patients using the resulting confusion matrix of the model. Here I found that the model based on concentrations of DHA derived mediators was able to correctly classify ~90% of responders in the appropriate category (Figure 5).

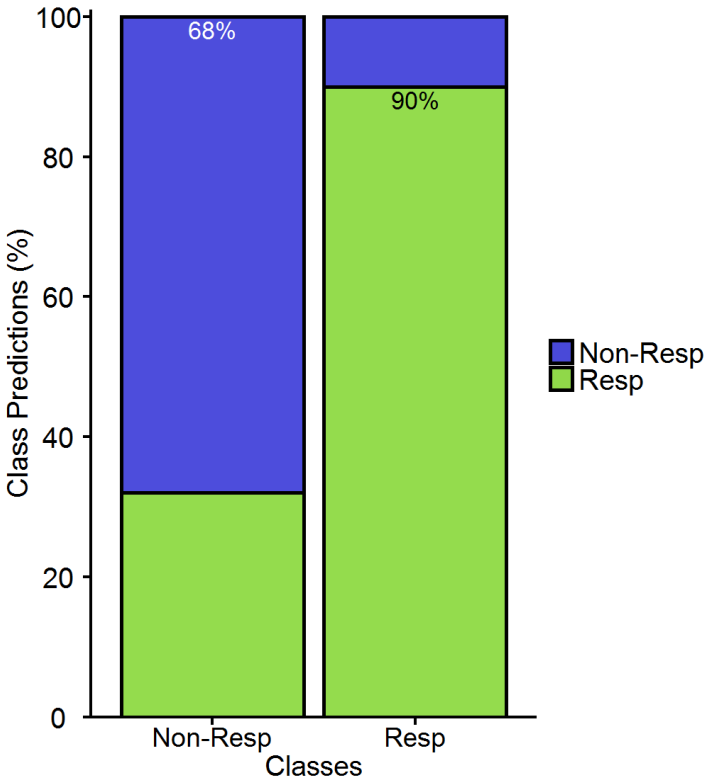


Figure 5. The DHA metabolome-based model can classify DMARD-responder patients. Peripheral blood was collected in RA patients prior to DMARD treatment initiation and lipid mediator profiles were established using LC-MS/MS. Classification predictions for each class (sensitivity and specificity) of the 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.

In order to validate the robustness of our models, I used peripheral blood lipid mediator profiles from a second cohort of DMARD naive patients composed of 36 responders and 22 non-responders and tested whether the models created using all the lipid mediators or mediators from each metabolome were able to correctly classify responders and non-responders (Appendix 8 and 9 for patient clinical information and lipid mediator concentrations). For this purpose, I assessed the ROC curve, which evaluates the diagnostic potential of a classifier by varying its discrimination threshold. Assessment of the area under the ROC curve demonstrated that the random forest model with all the lipid mediators gave an AUC of 0.72, while the DHA metabolome model gave an AUC of 0.44, whereas the n-3 DPA metabolome model gave an AUC of 0.58 (Figure 6 and Table 1). AUC values in the same range were found for the Bayesian, SVM and Elastic net models; with AUC between 0.56 and 0.72 for all the lipid mediator models, an AUC between 0.54 and 0.61 for the DHA model and an AUC between 0.64 and 0.72 for the n-3 DPA model (Table 1). These results indicate that baseline peripheral blood lipid mediator profiles are linked with DMARD treatment outcome and can predict the responsiveness of this therapy in RA patients with more accuracy than classical clinical scores.

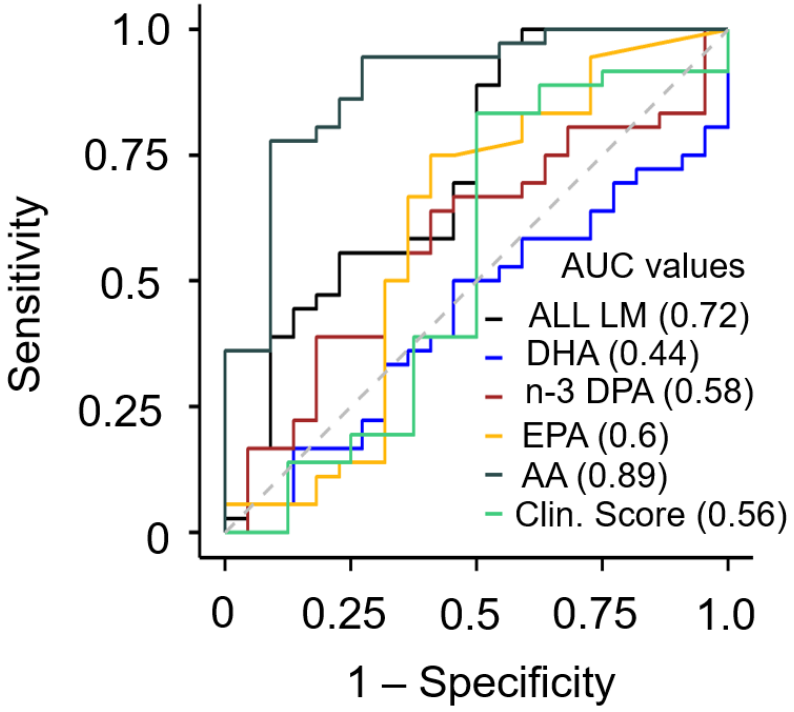


Figure 6. SPM-based models can predict with certain accuracy the responsiveness to DMARD treatment in a new and independent RA patient cohort. Peripheral blood was collected in RA patients prior to DMARD treatment initiation and lipid mediator profiles were established using LC-MS/MS. ROC curves and AUC values (provided in brackets) for predictive models based on all the lipid mediators (ALL LM), individual fatty acid metabolomes or clinical scores (Clin Score). All the models were created using the random forest methodology ("randomForest" package from R) since these models obtained the best accuracy scores.

Improved machine learning models identify specific SPM that are predictive of DMARD treatment response

Knowing that lipid mediator profiles are linked to the response to DMARD treatment in RA patients, I next raise the question of whether I could identify specific SPM that may be useful as biomarkers for treatment responsiveness and can improve the prediction accuracy of the machine learning models. For this purpose, I applied the random forest "importance" analysis from the R package "randomForest", which identifies the relevance of every mediator in the performance of the model based on the prediction accuracy. Here I found that the DHA-derived RvD4 and 10S,17S-diHDPA were the most important mediators in predicting treatment responsiveness, with 4S,14S-diHDHA (also DHA-derived), AA-derived 15R-LXA_4 and 5S,12S-diHETE, and n-3 DPA-derived Maresin 1 ($MaR1_{n\text{-}3DPA}$) also displaying a marked contribution, although to a lesser extent than RvD4 and 10S,17S-diHDPA (Figure 7A). Having identified potential candidate biomarkers, I next built machine learning models using the random forest methodology and concentrations of either RvD4, 10S, 17S-diHDPA, 15R-LXA_4 and $MaR1_{n\text{-}3\ DPA}$ or RvD4, 10S, 17S-diHDPA, 15R-LXA_4, 5S,12S-diHETE, 4S,14S-diHDHA and $MaR1_{n\text{-}3\ DPA}$ from the second cohort of DMARD naive patients. Here I found that the model with the top six lipid mediators was able to predict DMARD responsiveness with an accuracy score of ~86% while

the model with the top four lipid mediators had a prediction score of ~83% (Figure 7B and Table 1). I next evaluated the prediction ability of our models using a different group of DMARD naive patients. ROC curve analysis demonstrated that the model built using the four mediators gave an AUC of 0.80, whereas the top six derived model gave an AUC of 0.79 (Figure 7C and Table 1). The learning curve of the top four lipid mediators indicates a stable model (no overfitting or underfitting), with a constant accuracy score and a reduction of the standard deviation of the out of the bag error of the trees used to create the models (Appendix 7). These models were markedly better in predicting DMARD response than those obtained using all the lipid mediators, the DHA-derived mediators and clinical scores (Figure 4 and 7), indicating that I was able to identify specific SPM as biomarkers of DMARD responsiveness in RA patients.

Increased eicosanoid and SPM concentrations in peripheral blood from DMARD non-responders

To have a better understanding of the mechanism determining the responsiveness of patients to DMARD treatment, I conducted SPM biosynthetic pathways to identify which pathways were differentially regulated between the DMARD responder and non-responder groups. Pathway analyses were made for the four essential fatty acids metabolomes, demonstrating an upregulation of specific SPM, including the DHA-derived RvD4 and the n-3 DPA-derived $MaR1_{n\text{-}3DPA}$ in DMARD non-responders. SPM increases were coupled with an upregulation of proinflammatory eicosanoids, including PGD_2 and PGE_2, in DMARD non-responders when compared with DMARD responders (Figure 8). To determine whether the differences in SPM expression were associated with changes in a transcriptional level of enzymes involved in SPM production, I analysed RNA sequencing data and measured the gene expression of ALOX enzymes in peripheral blood from these two patient groups. ALOX5, ALOX12, ALOX15 and ALOX15B transcript levels were similar between the DMARD responders and DMARD non-responders (Figure 9). These results suggest that regulation of SPM biosynthetic pathways may be linked to the regulation of enzyme activity, either through protein translation or post-translational modifications.

Baseline plasma SPM concentrations are linked with distinct disease pathotypes in RA patients

Based on molecular and histological features in the synovium, RA can be classified into three distinctive pathotypes: lympho-myeloid, diffuse-myeloid and pauci-immune-fibroid. These pathotypes are associated with distinct responses to DMARD treatment.

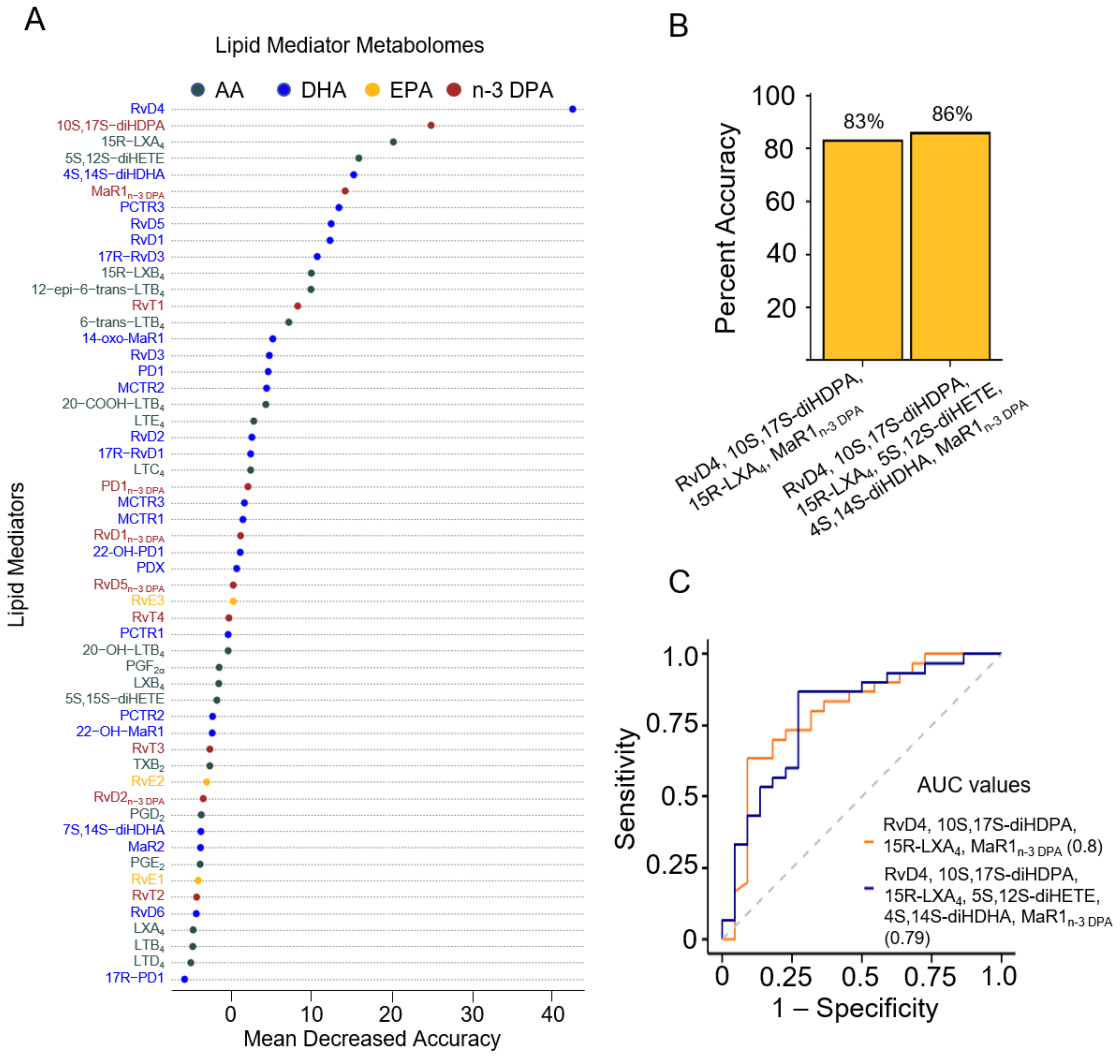


Figure 7. Identification of specific SPM that are predictive of treatment outcome. Plasma was collected from RA patients prior to the initiation of treatment with DMARDs and lipid mediator concentrations established using LC-MS/MS-based lipid mediator profiling. (**A**) Relevance of lipid mediators in the prediction performance of the "ALL LM" model based on decreasing accuracy. (**B**) Percent accuracy score of models using the top four or top six lipid mediators from the random forest "importante" analysis. (**C**) ROC curves and AUC values for predictive models based on the indicated SPM. All the models were created using the random forest methodology ("randomForest" package from R).

I explored then if the differences observed within the synovium between the pathotypes were also extended into the systemic circulation. For that purpose, lipid mediator profiling was conducted to assess whether peripheral blood lipid mediator concentrations in RA patients from each of these three groups were distinct prior to the initiation of DMARD therapy. PLS-DA analysis using lipid mediator profiles identified three distinct clusters for each of the pathotypes (Figure 10A). When I explored the VIP scores, which identify the contribution of each mediator in the observed separation between clusters, I identified upregulation of SPM, including 15R-LXA_4 and MCTR2 in the pauci-immune-fibroid pathotype, RvE2 and 4S,14S-diHDHA in the lympho-myeloid pathotype, and PCTR3 and RvD2 in the diffuse-myeloid pathotype. Proinflammatory mediators, such as PGD_2 and TXA_2, measured as its stable further metabolite TXB_2, were also found upregulated in plasma from the pauci-immune-fibroid pathotype patients (Figure 10B).

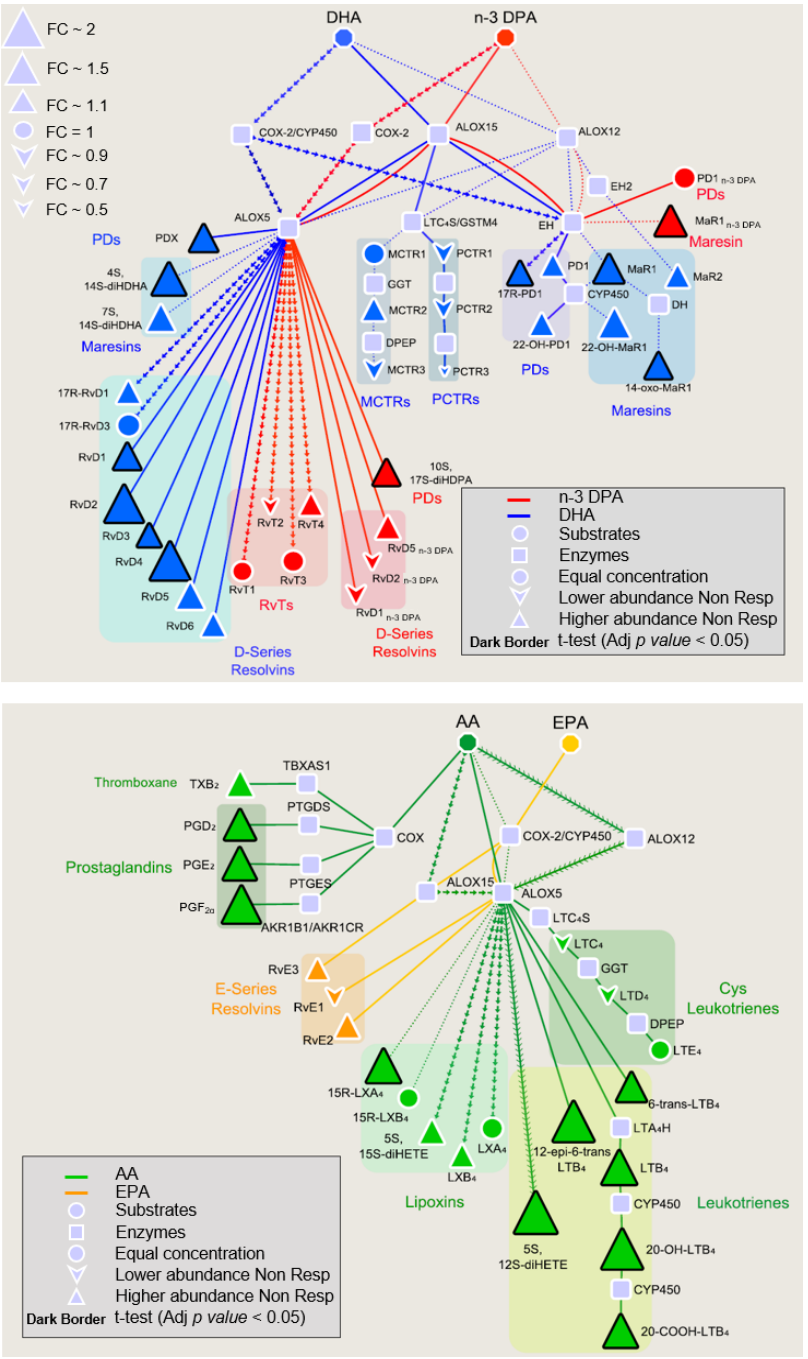


Figure 8. Upregulation of baseline peripheral blood lipid mediator concentrations in DMARD non-responders when compared with responders. Peripheral blood was collected in patients DMARD responders (Resp) and DMARD non-responders (Non-Resp) prior to DMARD treatment initiation. Peripheral blood lipid mediator concentrations were established using LC-MS/MS-based lipid mediator profiling. Pathway analysis for the differential expression of mediators from the (*top panel*) DHA and n-3 DPA, and (*bottom panel*) EPA and AA bioactive metabolomes in Non-Resp when compared to Resp. Statistical differences between the normalised concentrations (expressed as the fold change) of the lipid mediators from the Non-Resp and Resp groups were determined using a two-sided *t*-test followed by a multiple comparison correction using Benjamini–Hochberg procedure. Up- or downregulated mediators are denoted using upward and downward facing triangles, respectively, and on changes of the node's size. Bolded mediators represent statistical differences between the two groups when adjusted *p*-value <0.05. Results are representative of $n =$ 66 for Resp and $n =$ 44 Non-Resp.

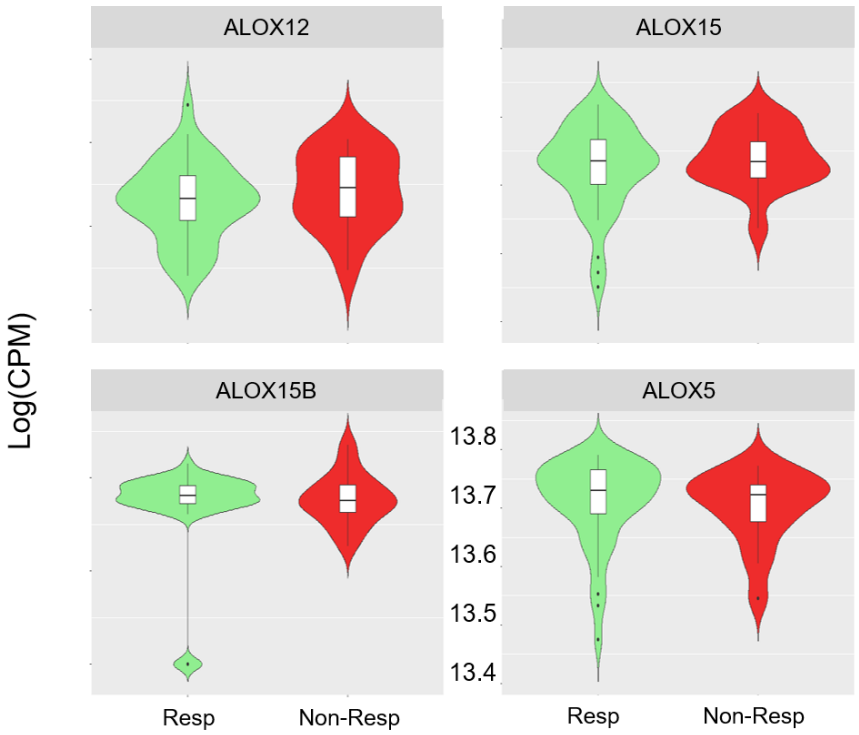


Figure 9. Baseline expression of ALOX enzymes is equal between DMARD responders and DMARD non-responders. RNA count normalization and differential gene expression analysis were performed using the quasi-likelihood method of the Bioconductor R package "edgeR". The data is expressed as the logarithm of counts per million (Log(CPM)), which indicates the number of reads mapping to a gene scaled by the number of reads you sequenced times one million. Results are represented as boxplots where the middle line is the median, the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends from the hinge to the largest value no further than 1.5 × IQR from the hinge (where IQR is the interquartile range) and the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge, while data beyond the end of the whiskers are outlying points that are plotted individually (as indicated by the geom_boxplot from the R package "ggplot2"). Results are representative of n = 43 for Resp and n = 15 for Non-Resp.

Having identified distinct clusters for each pathotype, I next explored the differences between lipid mediator profiles in DMARD responders and non-responders for each of these three categories. Pathway analysis was performed and I found an increase in ALOX5 products from both the n-3 DPA and DHA metabolomes in non-responders with lympho-myeloid or pauci-immune-fibroid pathotype when compared with the responders from the same group. These included significant increases in the DHA-derived RvD4 and PDX and the n-3 DPA-derived $MaR1_{n\text{-}3\ DPA}$ (Figure 11A). Pathway analysis for the AA and EPA-derived lipid mediators identified upregulation of 15R-LXA_4 in non-responders with lympho-myeloid and pauci-immune-fibroid pathotypes that were coupled with upregulation of prostaglandins, including PGD_2, PGE_2 and $PGF_{2\alpha}$, and leukotrienes, such as 5S,12S-diHETE and LTB_4 (Figure 11B).

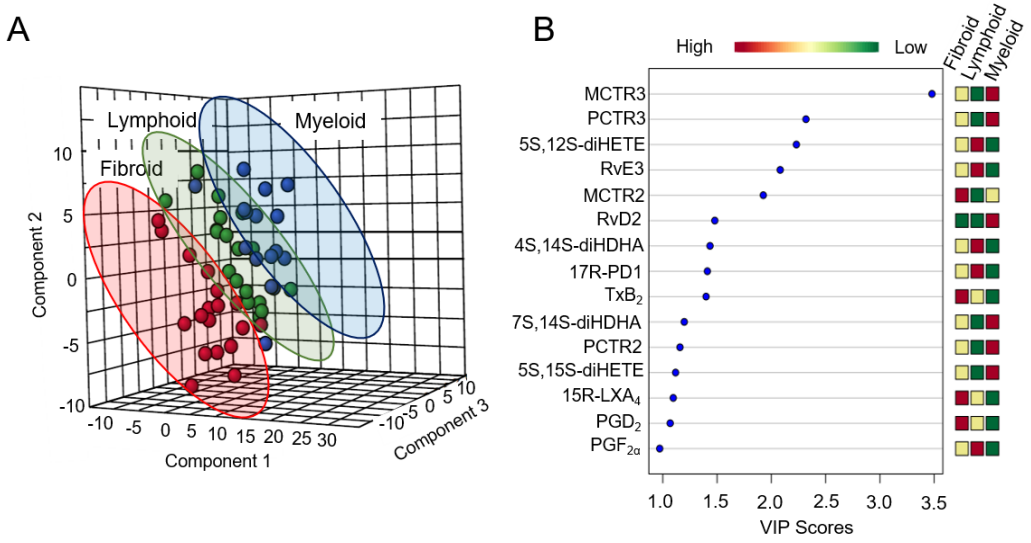


Figure 10. Plasma SPM can differentiate between disease pathotypes in RA patients. Plasma was collected from RA patients prior to the initiation of treatment with DMARD and lipid mediator concentrations were established using LC-MS/MS-based lipid mediator profiling. PLS-DA analysis was performed using peripheral blood lipid mediator concentrations from lympho-myeloid (lymphoid), diffuse-myeloid (myeloid) and pauci-immune-fibroid (fibroid) pathotypes in RA patients. (**A**) 3-dimensional score plot. (**B**) Variable importance in projection (VIP) scores of 15 lipid mediators with the greatest differences in concentrations between the three groups. Results are representative of n= 18 for Fibroid n = 17 for myeloid and n = 19 for lymphoid.

Since lipid mediator concentrations were different between these patient groups, I next assessed whether combining disease pathotypes with the biomarkers identified using the machine learning methodologies would further enhance the predictiveness of our models. The machine learning model using concentrations from RvD4, 10S,17S-diHDPA, 15R-LXA$_4$ and MaR1$_{n-3\ DPA}$ demonstrated an increase of its predictiveness when pathotypes were considered, with the ability of the models to correctly classify responders up to ~89% (Figure 12 and Table 2). Altogether, these results demonstrated a link between plasma lipid mediator concentrations and different disease pathotypes in RA. Moreover, the previously identified SPM biomarkers were able to correctly classify DMARD responders for each of the pathotypes.

Decreased SPM concentrations in DMARD non-responder 6 months after treatment initiation

Knowing that there were differences in baseline peripheral blood lipid mediator concentrations between DMARD responders and non-responder, I investigated whether these differences persist in patients 6 months after treatment initiation. OPLS-DA analysis using plasma lipid mediator profiles from DMARD responders and non-responders 6 months after initiation of treatment was able to identify a distinct cluster for each group (Figure 13A). Besides that, VIP scores identified 22 mediators that were relevant for the separation between DMARD responders and non-responders, such as DHA-derived PCTR2, RvD2, RvD3, EPA-derived RvE2 and AA-derived 15R-LXA$_4$ and 15R-LXB$_4$ (Figure 13B).

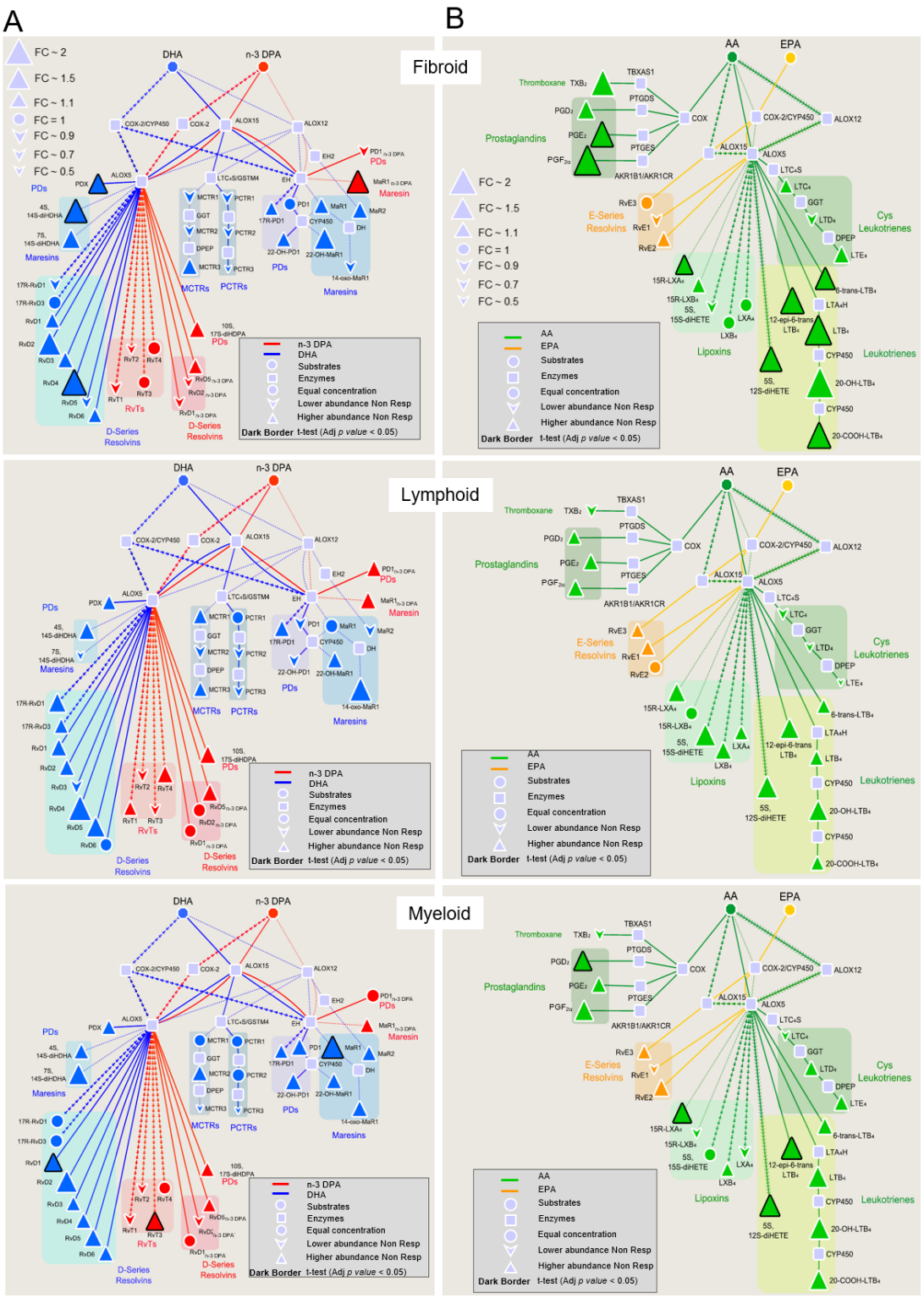


Figure 11. Differential regulation of essential fatty acids metabolomes in peripheral blood of DMARD responders and DMARD non-responders from RA patients with distinct pathotypes. Plasma was collected from RA patients prior to the initiation of treatment with DMARD. Lipid mediators were identified and quantified using LC-MS/MS-based lipid mediator profiling and the differential expression of mediators from DHA and n-3 DPA (**A**) and EPA and AA (**B**) metabolomes were analysed for each RA pathotype in DMARD non-responders (Non-Resp) when compared to DMARD responders (Resp). Statistical differences between the normalised concentrations of the lipid mediators from the Non-Resp and Resp groups were determined using *t*-test followed by a multiple comparison correction using Benjamini–Hochberg procedure. Up- or downregulated mediators are denoted using up and down facing triangles, respectively, and on changes of the node's size. Bolded mediators represent statistical differences between the two groups when adjusted *p*-value <0.05. Results are representative of *n* = 18 for fibroid Resp, *n* = 15 for fibroid Non-Resp, *n* =19 for lymphoid Resp, *n* = 10 for lymphoid Non-Resp, *n* = 22 for myeloid Resp, *n*= 15 for myeloid Non-Resp.

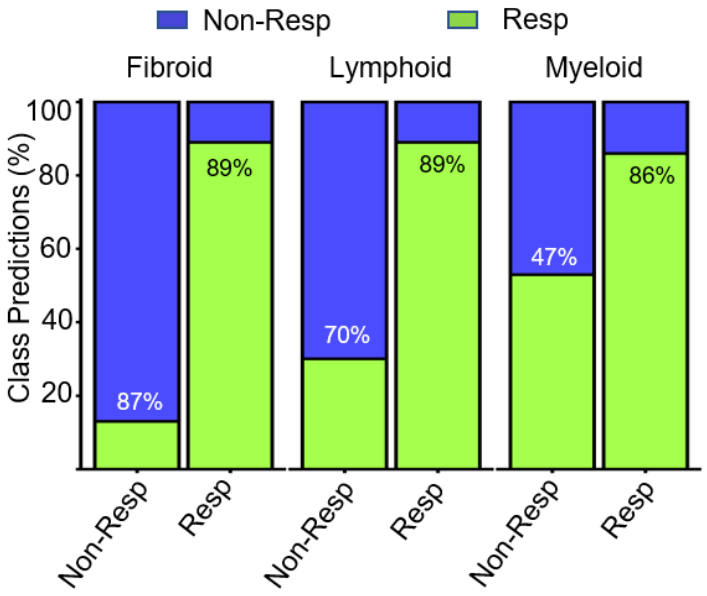


Figure 12. Combining disease pathotypes and select SPM concentrations enhances model predictiveness. Plasma was collected from RA patients prior to the initiation of treatment with DMARD and lipid mediator concentrations were established using LC-MS/MS-based lipid mediator profiling. Classification accuracies for each class (sensitivity and specificity) of the RvD4, 10S,17S-diHDPA, 15R-LXA$_4$ and MaR1$_{n\text{-}3\ DPA}$ model created using the specific dataset for the different pathotypes (fibroid, lymphoid and myeloid). Green indicates the samples that were predicted as Resp while blue indicates predicted Non-Resp. Percentages indicate true positives (Resp class) and true negatives (Non-Respon class). All the models were created using the random forest methodology ("randomForest" package from R).

Next, to have a better understanding of the mediator pathways that were differentially regulated between the two groups, I built biosynthetic pathways for each of the essential fatty acid metabolomes. This analysis identified the upregulation of ALOX5 and ALOX15-derived mediators from the DHA metabolome, such as RvD1, PDX, 17R-PD1 and 14-oxo-MaR1 in the DMARD non-responder group when compared with DMARD responders. Beyond that, I observed a generalized decrease of DHA, n-3 DPA and EPA-derived SPM levels in DMARD non-responders 6 months after treatment initiation. This was coupled with an observed increase of AA-derived ALOX5 products, which include the potent leukocyte chemoattractant LTB4 and the ionotropic cysteinyl leukotrienes [70] (Figure 14). These results indicate that 6 months after treatment initiation plasma SPM concentrations in DMARD responders were higher when compared with DMARD non-responder, suggesting a possible role of these lipid mediators in the efficacy of DMARDs therapy.

Table 2: Summary of RvD4, 10S,17S-diHDPA, 15R-LXA$_4$ and MaR1$_{n\text{-}3\ DPA}$ prediction model created using the specific dataset for the different disease pathotypes.

Machine Learning methodology	Samples	Model	Number of variables	% Accuracy Score	Model validation					
					TPR	TNR	TP	FP	TN	FN
randomForest (RF)	Fibroid samples (1st & 2nd cohort)	RvD4, 10S,17S-diHDPA, 15R-LXA$_4$, MaR1$_{n\text{-}3\ DPA}$	4	88	0.89	0.87	89	13	87	11
randomForest (RF)	Lymphoid samples (1st & 2nd cohort)	RvD4, 10S,17S-diHDPA, 15R-LXA$_4$, MaR1$_{n\text{-}3\ DPA}$	4	83	0.89	0.7	89	30	70	11
randomForest (RF)	Myeloid samples (1st & 2nd cohort)	RvD4, 10S,17S-diHDPA, 15R-LXA$_4$, MaR1$_{n\text{-}3\ DPA}$	4	70	0.86	0.47	86	53	47	14

TPR sensitivity, *TNR* specificity, *TP* true positives, *FP* false positives, *TN* true negatives, *FN* false negatives.

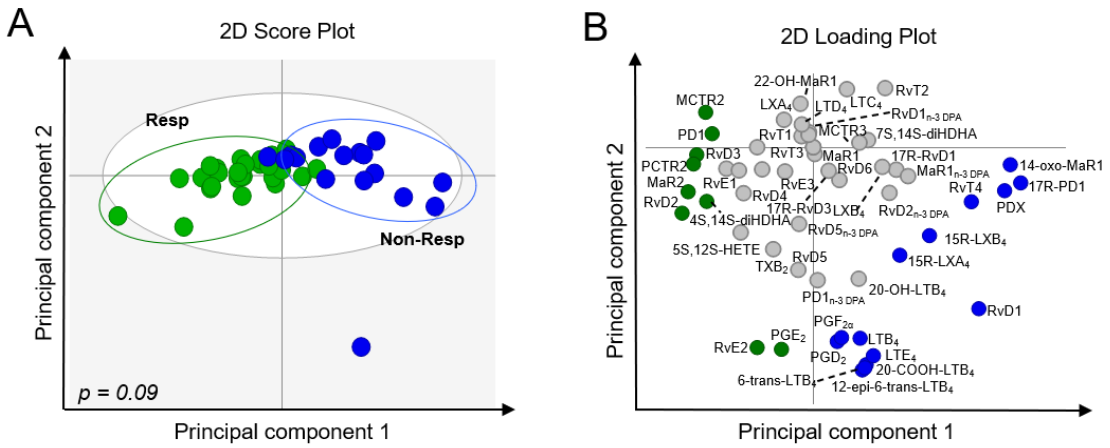


Figure 13. 6 months post-treatment plasma SPM can differentiate between DMARD-responders and DMARD-non-responders in RA patients. Peripheral blood was collected in patients that displayed reduced joint disease (DMARD responders, Resp) and those that did not (DMARD non-responders; Non-Resp) 6 months after treatment initiation. Peripheral blood lipid mediator profiles were established using LC-MS/MS-based lipid mediator profiling. OPLS-DA of lipid mediator profiles from Resp and 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 scores greater than 1 are highlighted in green and blue and correspond with either Resp or Non-Resp, respectively. *p*-value is calculated as the proportion of times that class separation based on permutation obtains similar results as the obtained on the original data. Results are representative of n = 27 Resp and n = 17 Non-Resp.

Candidate gene analysis identifies genomic variants associated with RA in SPM-related genes

After identifying SPM biomarkers associated with RA and the response to DMARD treatments, I decided to investigate if there was a connection between genetic variants located in SPM-related genes and the development of RA. This to explore the link between these mutations and dysregulated lipid mediator profiles, and have a better understanding of the molecular mechanisms associated with RA. I performed CG association analysis in two populations: British and South Asian (SA) from the UK Biobank dataset and focusing on enzymes and receptors involved in SPM production and their biological function (Appendix 10 for representative ƛ and Q-Q plot of the created models). In the British population, I identified the common SNP (MAF > 1%) rs2267707 (odds ratio 1.09, *p*-value $3.7\text{x}10^{-4}$) located in the intronic region of *GPR37* (PD1 receptor) and rs12984081 (odds ratio 0.87, *p*-value $1.4\text{x}10^{-4}$) located upstream from *ALX/FPR2* (RvD1 an RvD3 receptor). No information was found from those SNPs in GWAS catalog and PhenoScanner. Several other SNPs were found in linkage disequilibrium with rs2267707, all of them located in the intronic region of *GPR37*. From QTLbase, I discovered that rs12984081 has a statistically significant effect on the expression of *ALX/FPR2* in blood (Figure 15 and Table 3). Of note, other SNP with low *p*-value that didn't reach statistical significance after multiple comparison correction were identified in intronic regions from *FPR2, PTGS2* (PGs and RvEs pathway), *EPHX2* (MaR2 pathway) and *GPR32* (RvD1 and RvD3 receptor); among them, rs12609373 (odds ratio 0.91, *p*-value $5.4\text{x}10^{-4}$) in *ALX/FPR2* and rs72475893 (odds ratio 5.23, *p*-value $5.8\text{x}10^{-4}$) in *EPHX2* have been associated with an effect in the expression of their respective genes (Table 3).

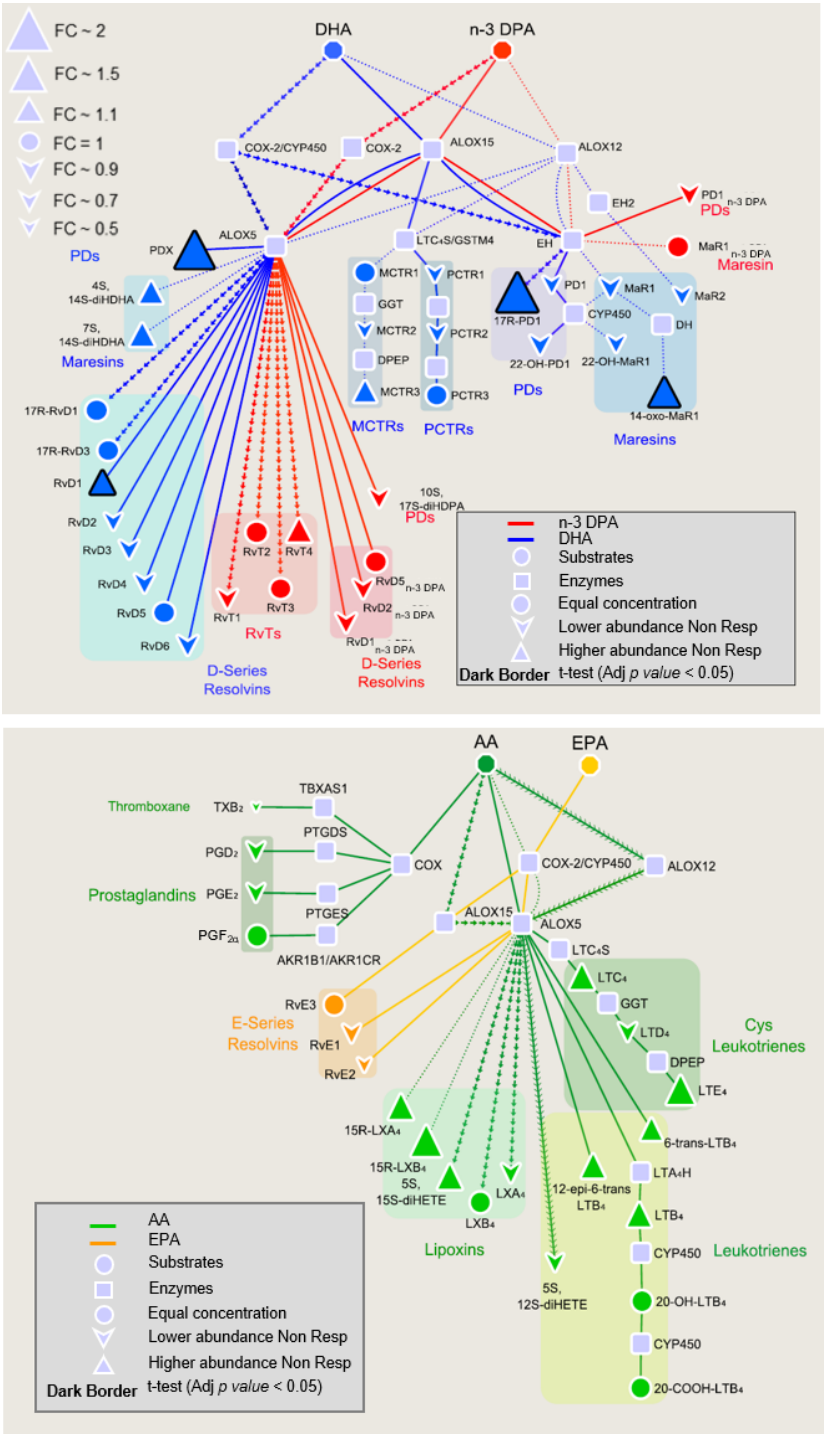


Figure 14. Decreased SPM levels in DMARD non-responders 6 months after treatment initiation. Peripheral blood was collected in patients that displayed reduced joint disease (DMARD responders, Resp) and those that did not (DMARD non-responders; Non-Resp) 6 months after treatment initiation. Peripheral blood lipid mediator profiles were established using LC-MS/MS-based lipid mediator profiling. Pathway analysis for the differential expression of mediators from the (*top panel*) DHA and n-3 DPA, and (*bottom panel*) EPA and AA bioactive metabolomes in Non-Resp when compared to Resp. Statistical differences between the normalised concentrations (expressed as the fold change) of the lipid mediators from the Non-Resp and Resp groups were determined using a two-sided *t*-test followed by a multiple comparison correction using Benjamini–Hochberg procedure. Up- or downregulated mediators are denoted using upward and downward facing triangles, respectively, and on changes of the node's size. Bolded mediators represent statistical differences between the two groups when adjusted *p*-value <0.05. Results are representative of $n = 27$ for Resp and $n = 17$ Non-Resp.

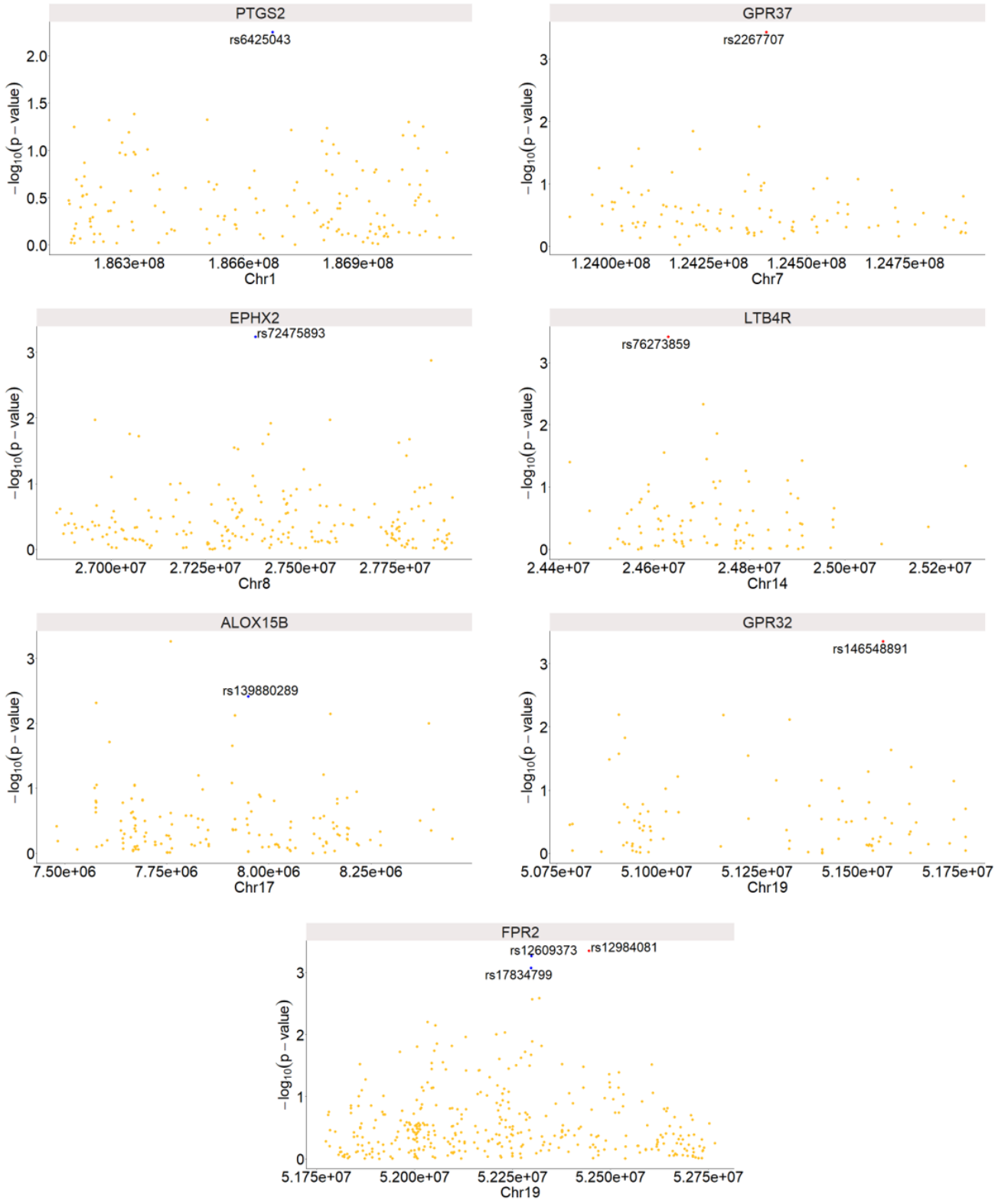


Figure 15. Genetic variants associated with RA identified in SPM-related genes in a British population. Candidate gene association analysis using RA patients and healthy controls from the UK Biobank were performed. Manhattan plot of the association statistics (*p* values) and the chromosome location for each SNP are displayed for the following genes as indicated in the figure: *PTGS2, GPR37, EPHX2, LTB4R, ALOX15B, GPR32 and ALX/FPR2*. The red dots indicate statistical significance SNP after consensus of different multiple comparison corrections methodologies (Bonferroni, Holm, Sidak single-step, Sidak step-down and Benjamini & Hochberg) (adjusted *p*-value < 0.05). Blue dots indicate candidate SNP (based on QTLbase and gnomAD) passing a suggested significance threshold of *p*-value < 0.005.

I then explored rare variants (MAF between 0.5 and 1%) and identified two SNPs: rs146548891 (odds ratio 2.21, *p*-value $4.5x10^{-4}$) nearby to *GPR32* and rs139880289 (odds ratio 2.36, *p*-value $3.8x10^{-4}$) in *ALOX15B*. rs146548891 is a missense variant predicted to be a damaging mutation (based on gnomAD database), although located in an adjacent gene to *GPR32* (*KLK13*) and associated with degenerative nervous diseases; while rs139880289 in *ALOX15B* is a stop-gain variation predicted to be a loss-of-function mutation (Figure 15 and Table 3).

Table 3: Summary of candidate SNP identified in SPM-related genes using a candidate gene association approach*.

Study	Chr	gene	PB	SNP	Variant type	Model	OR	p value	Adj p value	Effect	Predicted outcome	PhenoScanner	Linkage disequilibrium	cis-eQTL summary[&]
British	7	GPR37	1.24E+08	rs2267707	Common	HET	1.09	0.00037	0.0376	Intron Variant	NA	NA	rs2267707, rs2402749, rs12673697, rs147319418, rs33935352, rs2299903, rs35683122, rs4731208, rs34204767	Adrenal gland and lung GPR27 (p < 2.18e-7)
British	19	FPR2	5.24E+07	rs12984081	Common	DOM	0.87	0.00014	0.0501	Upstream Variant	NA	NA	rs12984081, rs12463058	Blood FPR2 (p < 4.54e-29)
British	19	FPR2	5.23E+07	rs12609373	Common	ADD	0.91	0.00054	0.1678	Intron Variant	NA	High density lipoprotein	rs200883779, rs12609429, rs59662074	Blood FPR2 (p < 4.15e-17)
British	19	FPR2	5.23E+07	rs17834799	Common	ADD	0.91	0.00085	0.2425	Intron Variant	NA	NA	rs12609373, rs200883779, rs12609429, rs59662074	NA
British	1	PTGS2	1.87E+08	rs6425043	Common	ADD	1.17	0.00564	0.6908	Intron Variant	NA	NA	NA	NA
British	8	EPHX2	2.74E+07	rs72475893	Common	REC	5.23	0.00058	0.1215	Intron Variant	NA	NA	NA	Blood EPHX2 (p < 0.0025)
British	19	GPR32	5.16E+07	rs146548891	Rare	ADD	2.21	0.00045	0.0453	Upstream Variant	NA	bronchiectasis and degenerative nervous disease	NA	NA
British	17	ALOX15B	7.95E+06	rs139880289	Rare	ADD	2.36	0.00386	0.4431	Stop Gained	Loss of function (pLoF)	NA	NA	NA
SA	17	ALOX15B	7.95E+06	rs9895916	Common	DOM	2.14	0.00016	0.0398	Missense Variant	Benign (Polyphen)	NA	rs9901281, rs9900937, rs61711499, rs8070967, rs147529751	Blood-Monocyte ALOX15B (p < 0.01)
SA	1	PTGS2	1.87E+08	rs4651331	Common	ADD	1.71	0.00021	0.0325	Intron Variant	NA	NA	Multiple SNP in intronic regions of the genes	Blood-Neutrophils PTGS2 (p < 0.03)
SA	12	CMKLR1	1.09E+08	rs73201532	Common	DOM	0.31	0.00022	0.0448	Upstream Variant	NA	Monocyte Count	rs4964679, rs7963734	Blood-Monocyte CMKLR1 (p < 1.43e-10)
SA	13	GPR99	9.79E+07	rs6491347	Common	DOM	0.50	0.00049	0.0593	Intron Variant	NA	NA	Multiple SNP in intronic regions of the genes	Thyroid Gland GPR99 (p < 0.0005)
SA	1	PTGS2	1.87E+08	rs5273	Common	DOM	3.01	0.00112	0.2750	Missense Variant	Benign (Polyphen) / Deleteirous (SIFT)	NA	NA	Blood PTGS2 (p < 7.06e-14)

Chr Chromosome, *SNP* single nucleotide polymorphism, *PB* pair bases, *OR* odd ratios, *HET* hethom model, *DOM* dominant model, *ADD* additive model, *REC* recessive model.

* CG analysis for the British population was performed using 4121 RA patients and 202906 healthy controls. For the South Asian (SA) population, I included 106 RA patients and 3771 healthy controls. Highlighted cells in red indicate adjusted *p* value < 0.05. Yellow cells indicate probably damaging genetic mutation, while green cells indicate benign mutation and purple cells indicate loss-of-function mutation.

[&] *p* values associated with cis-eQTL were obtained from QTLbase.

Finally, I explored genetic variants in SPM-related genes in the SA population from the UK Biobank. Here I identified five common variants; four of them reach statistical significance after multiple comparison correction: rs9895916 (odds ratio 2.14, *p*-value $1.6x10^{-4}$) in *ALOX15B*, rs4651331 (odds ratio 1.71, *p*-value $2.1x10^{-4}$) in *PTGS2*, rs73201532 (odds ratio 0.31, *p*-value $2.2x10^{-4}$) in *CMKLR1* (RvE1 and RvE2 receptor) and rs6491347 (odds ratio 0.50, *p*-value $4.9x10^{-4}$) in *GPR99* (LTE_4 receptor). From those, rs9895916 in *ALOX15B* is a missense variant predicted to be benign, while the other three are intron and upstream variants. rs5273 in *PTGS2* is a genetic variant that didn't reach statistical significance after adjusting the *p*-value for multiple comparisons (odds ratio 3.01, *p*-value $1.1x10^{-3}$), but is a missense variant predicted to be deleterious (Figure 16 and Table 3). Of note, all the five variants were cis-eQTL of their respective genes (Table 3).

Altogether, these results indicate a possible link between genetic variants in SPM-related genes and RA. Moreover, I was able to identify candidate SNPs that may play a role in the dysregulation of SPM during RA.

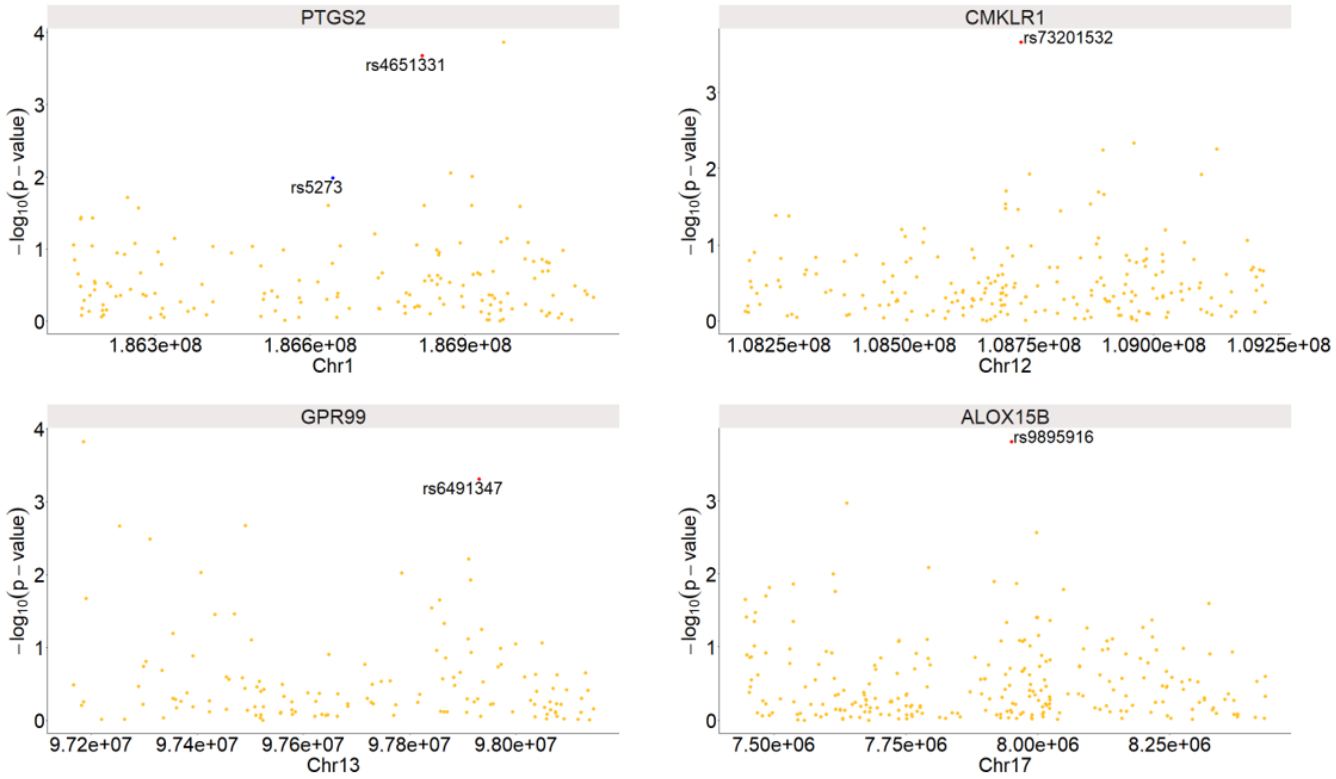


Figure 16. Genetic variants associated with RA identified in SPM-related genes in a South Asian population. Candidate gene association analysis using RA patients and healthy controls from UK Biobank were performed. Manhattan plot of the association statistics (*p* values) and the chromosome location for each SNP are displayed for the following genes as indicated in the figure: *PTGS2, CMKLR1, GPR99, and ALOX15B*. The red dots indicate statistical significance SNP after consensus of different multiple comparison corrections methodologies (Bonferroni, Holm, Sidak single-step, Sidak step-down and Benjamini & Hochberg) (adjusted *p*-value < 0.05).

Chapter 4: Discussion

Identifying in RA patients those who will respond to DMARD treatment or not represents a challenge. Results presented in this thesis demonstrate a role for SPM as predictive biomarkers to DMARD responsiveness in RA. Using lipid mediator profiling in combination with machine learning methodologies, I demonstrated that baseline plasma concentrations of specific SPM were predictive of response to DMARD treatment, identifying novel functional biomarkers that were tested in an independent RA patient cohort. Moreover, I also found persistent differences in plasma SPM concentrations between responders and non-responders 6 months after DMARD treatment initiation. Candidate gene association analyses were also assessed to identify genetic variants in SPM-related genes; here, I found candidate SNP in enzymes and receptors involved in lipid mediator production and activity, given a glimpse of the possible role of genetic mutations in the dysregulation of SPM during RA.

Since RA is characterized by an uncontrolled inflammation due to the inability of the host immune response to regulate it, it is more than natural to hypothesize the involvement of SPM in RA. Previous studies implicate a role for altered resolution mechanisms in the onset and propagation of RA. An increase of synovial RvE2 concentration was associated with a decrease of joint pain, while plasma SPM were negatively related to erythrocyte sedimentation rate in inflammatory arthritis [71]. Downregulation of RvD3 during ongoing joint inflammation was observed in arthritic mice model [33], while an increase of RvTs, through the administration of atorvastatin or pravastatin, was associated with a decrease of disease activity [28]. Several reports have demonstrated an upregulation of LXA_4 and 15R-LXB_4, important mediators in the restoration of homeostasis, in arthritis, supporting an anti-inflammatory role during chronic inflammation [34,72,73]. In this study, I found an upregulation of selected SPM concentrations, such as RvDs and MaRs, at baseline in DMARD non-responder that was coupled with an increase of the proinflammatory eicosanoids PGs and LKs. Moreover, 6 months after treatment initiation, the differences between responders and non-responders persist with higher concentrations of selected SPM in the non-responder group. Since one of the key biological roles of SPM is to counter-regulate eicosanoid production [22,27], which are promoters of the pathogenesis of RA [74–76], these results suggest a fault in SPM activity in DMARD non-responders, such as, for example, a compromised interaction between SPM and their receptors. Similar findings have been reported in progressive multiple sclerosis, with an increased concentration of RvD5 impaired with a low expression of its receptor GPR32 [77]. In diabetes, the signalling downstream of the RvE1 receptor, ERV1/ChemR23, was found to be altered, reducing the ability of RvE1 to regulate peripheral blood leukocyte responses from diabetic patients [78].

The identification of biomarkers of different diseases has become a priority in the last decade. The heterogeneity of complex diseases makes necessary the identification of biomarkers and the development of specific treatments designed to address specific manifestations of these conditions. Not going very far, RA is recognized as a highly heterogeneous disease presenting different pathotypes originated by genetic and environmental factors that still need to be studied in more detail [6,79,80]. Using machine learning methodologies, I demonstrated that selected baseline plasma SPM concentrations were able to predict the response of DMARD treatments in two RA patient cohorts. Machine learning models were created using Bayesian classifiers, Elastic net regression, SVM and random forest methodologies. Although Bayesian models are a simple approach with the advantage to be fast and easy to implement, they are very limited since they assume that all the features are independent variables, which rarely applies to biological data [81]. Elastic net models are well established for feature selection; however, their performances can be affected when multicollinearity is highly present, since the flexibility of the model at the moment of choosing the tuning parameters can lead to overfitting [82]. On the other hand, SVM allows misclassification (which is useful in biological datasets) but requires more computational power and is very sensible to parameters manipulation [54]. Since random forest is the methodology that works best with large datasets with a high number of features [56], allowing the identification of specific lipid mediators that improve prediction accuracy, it was not surprising to obtain the best predictive models using this strategy, with prediction accuracy of up to ~89%. Moreover, specific plasma SPM concentrations prior to DMARD treatment initiation were also diagnostic of distinct joint disease pathotypes. These newly identified biomarkers that were predictive of DMARD responsiveness are very likely to be specific for this group of treatments since most of the patients in both cohorts were on a wide range of other medications, although there were no significant differences in any of these parameters between responders and non-responders. In addition, since SPM production can be considered a signature of leukocyte activation status [27,83,84], these findings indicate that peripheral blood SPM concentrations are potential functional biomarkers for both patient stratification and predicting treatment responsiveness to DMARDs.

The identification of specific pathotypes has provided insights into RA and the potential responsiveness to different treatments. Cell lineage-specific gene expression showed low levels of all types of immune inflammatory cells in the fibroid pathotype, while the opposite was observed in the lymphoid pathotype with high levels of immune inflammatory cells including macrophages, T cells and B cells. The myeloid pathotype, similar to the lymphoid pathotype in a prevalent macrophage infiltration, lacks T cells and B cells [85,86]. In synovium blood, the lymphoid pathotype was linked with differentially expressed genes associated with type I interferon response. Pathway enrichment analysis in lymphoid samples associated this pathotype with B cell, plasma cell and macrophage infiltration. Myeloid samples were linked with proinflammatory pathways, such as phosphoinositide 3-kinase (PI3K) and

nuclear factor of activated T-cells (NFAT) signalling, known to be part of the infiltration of neutrophils and macrophages into the synovium. Fibroid samples were associated with pro-fibroblast pathways with low presence of immune inflammatory cells pathways [85]. In this study, in the DMARD non-responder group I found high concentrations of RvDs in the myeloid and fibroid group, which are known to inhibit cytokines production, such as IFN-α [87]. This, however, coupled with high concentrations of eicosanoids. These results were not observed in the lymphoid pathotype where there weren't statistical differences between responders and non-responders. These suggest a possible correlation between SPM production and molecular features in the different RA pathotypes, not only observed in the synovium but in the systemic circulation, that can be further studied in more detail. Moreover, SPM concentrations were able to identify distinct joint disease pathotypes and predict DMARD responsiveness to the same extent, or better, than predictive models created using gene expression data [85,88].

SPM are produced by the stereoselective conversion of essential fatty acids by distinct enzymes that include members of lipoxygenase and cyclooxygenase families. These enzymes are also involved in proinflammatory eicosanoids production. During tissue injury, prostaglandins and leukotriene play a role in the migration of leukocyte traffic to the site of injury, promoting inflammation. Here, the migration of neutrophils to the tissue stimulates the switching from the production of eicosanoids to lipoxins, marking the end of the acute inflammatory response [89,90]. This switching relies on the regulation of these biosynthetic enzymes at both transcriptional and translational level, as well as via post-translational modifications. Studies have found, for example, that nitric oxide promotes the *S*-nitrosylation of COX-2, increasing its activity and up-regulating RvTs production. The phosphorylation of ALOX5 *via* the calcium-calmodulin-dependent protein kinase II-p38 promotes the translocation of the enzyme from the cytosol (where, together with ALOX15, produce SPM) to the nuclear membrane where will produce LTB_4 with the help of phospholipase A_2 and LTA_4 hydrolase [91]. It has been demonstrated that Toll-like receptor 7, important in pathogen recognition during tissue injury, stimulates the initiation of DHA-derived biosynthetic pathways producing RvDs, specially PD1 and RvD1 in human peripheral blood monocytes [92]. In blood coagulation, an increase of prothrombotic and proinflammatory mediators is observed in early time intervals, while SPM were associated with later times of the process [89]. In atherosclerosis, an increase in the expression of phosphorylated ALOX5 and a decrease in the RvD1 to LTB_4 ratio has been observed during ongoing inflammation [93]. In the present study, I found that gene expression of ALOX5, ALOX12, ALOX15 and ALOX15B were similar between DMARD responders and non-responders. In addition, the switching of eicosanoids to SPM production was not observed in the non-responder group at baseline, and most SPM concentrations were either similar or reduced in non-responders after 6 months of treatment initiation. All these results suggest that the SPM biosynthetic pathways become uncoupled in the non-responders probably due to post-translational modifications of the enzymes.

The heritability of RA has been estimated to be about 60%. Major histocompatibility complex (MHC) class II genes have been associated with the pathogenesis of RA, although non-related MHC genes have also been linked to RA. Besides identifying genes associated with the development of the diseases, other genes have been linked to disease outcome, prognosis and response or non-response to therapies [20,94,95]. GWAS, together with animal studies, have helped to identify all classes of genes associated with susceptibility to RA. Apart from MHC genes, which have a great influence on immune diseases, strong associations have been found with *PTPN22*, *PADI4*, *CTLA4*, among other genes [94]. PTPN22, the protein tyrosine phosphatase N22, is now considered to play a role in several autoimmune diseases that include, besides RA, systemic lupus erythematosus, lupus and type I diabetes [96]. The underlying mechanism of the presence of genetic variants in *PTPN22* and the susceptibility to autoimmune diseases is still not very well understood; although the expression of this gene has exhibited regulatory activities for T cells, B cells and cytokines profiles [97]. Other genes, such as *IL23R*, the receptor of the proinflammatory IL23, and *TRAF1*, the receptor of TNF family members, with their influence in proliferation, apoptosis, cytokine activation and proinflammatory process have also been identified to have genetic variants associated with RA [97]. These findings indicate the likelihood of identifying genetic variants in genes associated with inflammation and disease progression. Since, in the present study, I observed a dysregulation of SPM most likely linked to the activity of their enzymes and receptors, I applied a CG analysis approach with the purpose to explore possible genetic mutations in SPM-related genes associated with RA. Here, using the UK Biobank dataset, I found candidate SNPs in *GPR37*, *FPR2*, *EPHX2* and *ALOX15B* in a British population. Furthermore, since South Asian individuals are more susceptible to suffering RA [98], I explored SNPs in this population identifying candidate variants in *ALOX15B*, *PTGS2*, *CMKLR1* and *GPR99*. Even though some of those SNPs didn't reach statistical significance, most of them were identified as cis-eQTL of the genes where they were found [99–101]. In addition, the SNPs identified in *ALOX15B* are a missense variant and stop-gain variation predicted to be a loss-of-function mutation. As mentioned previously, uncoupled SPM concentrations and the expression of their respective receptors can be linked with several diseases, as it has been seen in multiple sclerosis [77] and obesity [102]. Moreover, genetic variants in *PTGS2*, *ALOX5*, *ALOX12*, *ALOX15* have been associated with cancer and other diseases [103–105], with near-null variants identified in *ALOX15* correlated with loss of catalytic activity and risk of coronary artery disease [106,107]. Together, these analyses identified candidate SNPs associated with RA in SPM-related genes; although further analysis and experimental validation is required to corroborate these results.

In summation, the present study identifies novel functional biomarkers, including RvD4, 10S,17S-diHDPA, 15R-LXA_4 and $MaR1_{n\text{-}3\ DPA}$, that predict both treatment response to DMARDs as well as joint disease pathotype. In addition, I identified candidate SNPs, located in SPM-related genes such as *GPR37*, *FPR2* and

ALOX15B, to be associated with RA that are worth exploring in more detail as possible biomarkers of RA. Thus, these biomarkers may be clinically useful in identifying patients who are unlikely to respond to conventional DMARD therapy and would benefit from being fast-tracked to the next level of RA therapeutics. This would in turn help minimise or even prevent further structural damage to the joints together with disease progression and disability, thereby improving quality of life.

Future work

I am to further explore the utility of SPM as biomarkers that predict responsiveness to treatment. I will focus on Anti-TNF treatment to see if I can also create machine learning models using plasma SPM concentrations and explore common biomarkers in predicting biological and synthetic DMARD treatment. Our candidate gene approach has some limitations, starting with the lack of validation in vitro or in vivo of the effect of the genetic variants that I identified; and, in the case of the SA population, the small number of samples that may reduce the statistical power of my findings. I am in the process of getting access to the East London Gene and Health dataset and samples, which will allow me to explore the genetic mutations of the SA population further. I also want to improve the association models (including more possible covariates and exploring other genetic models) and extend the analysis to other inflammatory conditions similar to RA, such as cardiovascular conditions; and compare between DMARD responders and non-responders, since SNPs have been associated to treatment response in small datasets [20,95].

References

1. Ginwala, R., Bhavsar, R., Chigbu, D. I., Jain, P. & Khan, Z. K. Potential role of flavonoids in treating chronic inflammatory diseases with a special Focus on the anti-Inflammatory activity of apigenin. *Antioxidants (Basel)* **8**, (2019).
2. Hunter, P. The inflammation theory of disease. The growing realization that chronic inflammation is crucial in many diseases opens new avenues for treatment. *EMBO Rep.* **13**, 968–970 (2012).
3. Bowman, S. J. & Guest, L. The national clinical audit for rheumatoid and early inflammatory arthritis. *Clin. Med.* **16**, 500–501 (2016).
4. Sayah, A. & English, J. C. Rheumatoid arthritis: A review of the cutaneous manifestations. *Journal of the American Academy of Dermatology* vol. 53 191–209 (2005).
5. 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).
6. 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).
7. 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).
8. 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).
9. Firestein, G. S. Evolving concepts of rheumatoid arthritis. *Nature* **423**, 356–361 (2003).
10. 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' by Humby. *Ann. Rheum. Dis.* (2019)

doi:10.1136/annrheumdis-2019-215903.

11. Whittle, S. L. & Hughes, R. A. Folate supplementation and methotrexate treatment in rheumatoid arthritis: a review. *Rheumatology* **43**, 267–271 (2004).
12. 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).
13. Cronstein, B. N. & Aune, T. M. Methotrexate and its mechanisms of action in inflammatory arthritis. *Nat. Rev. Rheumatol.* (2020) doi:10.1038/s41584-020-0373-9.
14. Friedman, B. & Cronstein, B. Methotrexate mechanism in treatment of rheumatoid arthritis. *Joint Bone Spine* vol. 86 301–307 (2019).
15. 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* vol. 112 2509–2514 (2015).
16. Aletaha, D. & Smolen, J. S. Diagnosis and management of rheumatoid arthritis. *JAMA* vol. 320 1360 (2018).
17. Geiler, J., Buch, M. & McDermott, M. F. Anti-TNF treatment in rheumatoid arthritis. *Curr. Pharm. Des.* **17**, 3141–3154 (2011).
18. Nam, J. L. *et al.* Current evidence for the management of rheumatoid arthritis with biological disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of RA. *Ann. Rheum. Dis.* **69**, 976–986 (2010).
19. Bongartz, T. *et al.* Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: systematic review and meta-analysis of rare harmful effects in randomized controlled trials. *JAMA* **295**, 2275–2285 (2006).
20. Liu, C. *et al.* Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis. *Mol. Med.* **14**, 575–581 (2008).
21. Wasserman, A. Rheumatoid Arthritis: common questions about diagnosis and

management. *Am. Fam. Physician* **97**, 455–462 (2018).

22. Perretti, M., Cooper, D., Dalli, J. & Norling, L. V. Immune resolution mechanisms in inflammatory arthritis. *Nature Reviews Rheumatology* vol. 13 87–99 (2017).
23. 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).
24. 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).
25. Leuti, A., Maccarrone, M. & Chiurchiù, V. Proresolving lipid mediators: endogenous modulators of oxidative stress. *Oxid. Med. Cell. Longev.* **2019**, 1759464 (2019).
26. 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).
27. Serhan, C. N. & Levy, B. D. Resolvins in inflammation: emergence of the pro-resolving superfamily of mediators. *J. Clin. Invest.* **128**, 2657–2669 (2018).
28. 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).
29. Dalli, J., Colas, R. A. & Serhan, C. N. Novel n-3 immunoresolvents: structures and actions. *Sci. Rep.* **3**, 1940 (2013).
30. Sharma-walia, N. & Chandrasekharan, J. Lipoxins: nature's way to resolve inflammation. *Journal of Inflammation Research* 181 (2015).
31. Serhan, C. N., Krishnamoorthy, S., Recchiuti, A. & Chiang, N. Novel Anti-inflammatory-pro-resolving mediators and their receptors. *Current Topics in Medicinal Chemistry* vol. 11 629–647 (2011).
32. Emanuela Ricciotti, G. A. F. Prostaglandins and Inflammation. *Arterioscler. Thromb. Vasc. Biol.* **31**, 986 (2011).

33. Arnardottir, H. H. *et al.* Resolvin D3 Is Dysregulated in arthritis and reduces arthritic inflammation. *J. Immunol.* **197**, 2362–2368 (2016).
34. 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).
35. Flak, M. B. *et al.* GPR101 mediates the pro-resolving actions of $RvD5_{n\text{-}3\ DPA}$ in arthritis and infections. *J. Clin. Invest.* **130**, 359 (2019).
36. Dalli, J. & Serhan, C. N. Identification and structure elucidation of the pro-resolving mediators provides novel leads for resolution pharmacology. *Br. J. Pharmacol.* **176**, 1024–1037 (2019).
37. Sugimoto, Y. & Narumiya, S. Prostaglandin E Receptors. *Journal of Biological Chemistry* vol. 282 11613–11617 (2007).
38. Hansen, T. V., Vik, A. & Serhan, C. N. The protectin family of specialized pro-resolving mediators: potent immunoresolvents enabling innovative approaches to target obesity and diabetes. *Front. Pharmacol.* **9**, 1582 (2018).
39. Krishnamoorthy, S. *et al.* Resolvin D1 binds human phagocytes with evidence for proresolving receptors. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 1660–1665 (2010).
40. Dalli, J. *et al.* Resolvin D3 and aspirin-triggered resolvin D3 are potent immunoresolvents. *Chem. Biol.* **20**, 188–201 (2013).
41. Chiang, N., Dalli, J., Colas, R. A. & Serhan, C. N. Identification of resolvin D2 receptor mediating resolution of infections and organ protection. *Journal of Experimental Medicine* vol. 212 1203–1217 (2015).
42. Recchiuti, A., Mattoscio, D. & Isopi, E. Roles, actions, and therapeutic potential of specialized pro-resolving lipid mediators for the treatment of inflammation in cystic fibrosis. *Front. Pharmacol.* **10**, (2019).
43. Narasimhan, V. M. *et al.* Health and population effects of rare gene knockouts in adult humans with related parents. *Science* **352**, 474–477 (2016).
44. 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).

45. Bishop, C. M. Pattern recognition and machine learning. (Springer, 2016).
46. Zitnik, M. *et al.* Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. *Inf. Fusion* **50**, 71–91 (2019).
47. 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).
48. Williams, F. M. K. Biomarkers: in combination they may do better. *Arthritis Research & Therapy* vol. 11 130 (2009).
49. Valletta, J. J. & Recker, M. Identification of immune signatures predictive of clinical protection from malaria. *PLoS Comput. Biol.* **13**, e1005812 (2017).
50. Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.* **11**, 424 (2020).
51. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019).
52. Zondervan, K. T. & Cardon, L. R. Designing candidate gene and genome-wide case-control association studies. *Nat. Protoc.* **2**, 2492–2501 (2007).
53. Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Min.* **10**, (2017).
54. James, G., Witten, D., Hastie, T. & Tibshirani, R. An Introduction to Statistical Learning: with Applications in R. (Springer Science & Business Media, 2013).
55. Bennett, K. P. & Campbell, C. Support vector machines. *ACM SIGKDD Explorations Newsletter* vol. 2 1–13 (2000).
56. Breiman, L. Classification and Regression Trees. (Routledge, 2017).
57. Casella, G. & Berger, R. L. Statistical Inference. *Biometrics* vol. 49 320 (1993).
58. Bolstad, W. M. Introduction to Bayesian Statistics. (2004).
59. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of*

the Royal Statistical Society: Series B (Statistical Methodology) vol. 67 301–320 (2005).

60. 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* vol. 40 2681–2693 (2018).
61. 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).
62. Lötsch, J. *et al.* Machine-learning based lipid mediator serum concentration patterns allow identification of multiple sclerosis patients with high accuracy. *Scientific Reports* vol. 8 (2018).
63. 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* vol. 19 556–557 (2014).
64. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
65. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
66. Purcell, S. *et al.* PLINK: A tool set for Whole-Genome Association and population-based linkage analyses. *The American Journal of Human Genetics* vol. 81 559–575 (2007).
67. Clarke, G. M. *et al.* Basic statistical analysis in genetic case-control studies. *Nat. Protoc.* **6**, 121–133 (2011).
68. Alexandra C. Nica, E. T. D. Expression quantitative trait loci: present and future. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, (2013).
69. Chong, J., Yamamoto, M. & Xia, J. MetaboAnalystR 2.0: from raw spectra to biological insights. *Metabolites* vol. 9 57 (2019).
70. Rådmark, O., Werz, O., Steinhilber, D. & Samuelsson, B. 5-Lipoxygenase, a key enzyme for leukotriene biosynthesis in health and disease. *Biochim. Biophys. Acta*

1851, 331–339 (2015).

71. Barden, A. E. *et al.* Specialised pro-resolving mediators of inflammation in inflammatory arthritis. *Prostaglandins Leukot. Essent. Fatty Acids* **107**, 24–29 (2016).
72. Hashimoto, A. *et al.* Antiinflammatory mediator lipoxin A4 and its receptor in synovitis of patients with rheumatoid arthritis. *J. Rheumatol.* **34**, 2144–2153 (2007).
73. Krönke, G. *et al.* 12/15-Lipoxygenase counteracts inflammation and tissue damage in arthritis. *The Journal of Immunology* (2009).
74. McCoy, J. M., Wicks, J. R. & Audoly, L. P. The role of prostaglandin E2 receptors in the pathogenesis of rheumatoid arthritis. *J. Clin. Invest.* **110**, (2002).
75. Fattahi, M. J. & Mirshafiey, A. Prostaglandins and rheumatoid arthritis. *Arthritis* vol. 2012 1–7 (2012).
76. Ma, Y., Hong, F.-F. & Yang, S.-L. Role of prostaglandins in rheumatoid arthritis. *Clin. Exp. Rheumatol.* **39**, 162–172 (2021).
77. Kooij, G. *et al.* Specialized pro-resolving lipid mediators are differentially altered in peripheral blood of patients with multiple sclerosis and attenuate monocyte and blood-brain barrier dysfunction. *Haematologica* **105**, 2056–2070 (2020).
78. Freire, M. O., Dalli, J., Serhan, C. N. & Van Dyke, T. E. Neutrophil resolvin E1 receptor expression and function in Type 2 diabetes. *J. Immunol.* **198**, 718–728 (2017).
79. Nerviani, A. *et al.* A pauci-Immune synovial pathotype predicts inadequate response to TNFα-Blockade in rheumatoid arthritis patients. *Front. Immunol.* **11**, (2020).
80. Weyand, C. M., Klimiuk, P. A. & Goronzy, J. J. Heterogeneity of rheumatoid arthritis: from phenotypes to genotypes. *Springer Semin. Immunopathol.* **20**, (1998).
81. Cao, Y., Fang, X., Ottosson, J., Näslund, E. & Stenberg, E. A comparative study of machine learning algorithms in predicting severe complications after bariatric surgery. *J. Clin. Med. Res.* **8**, (2019).
82. Pavlou, M., Ambler, G., Seaman, S., De Iorio, M. & Omar, R. Z. Review and evaluation of penalised regression methods for risk prediction in low‐dimensional data with few events. *Stat. Med.* **35**, 1159 (2016).

83. Dalli, J. & Serhan, C. N. Specific lipid mediator signatures of human phagocytes: microparticles stimulate macrophage efferocytosis and pro-resolving mediators. *Blood* **120**, e60–72 (2012).
84. Gao, Y. *et al.* Female-specific downregulation of tissue polymorphonuclear neutrophils drives impaired regulatory T Cell and amplified effector T Cell responses in autoimmune dry eye disease. *J. Immunol.* **195**, 3086–3099 (2015).
85. Lewis, M. J. *et al.* Molecular portraits of early rheumatoid arthritis identify clinical and treatment response phenotypes. *Cell Rep.* **28**, 2455–2470.e5 (2019).
86. Ouboussad, L., Burska, A. N., Melville, A. & Buch, M. H. Synovial tissue heterogeneity in rheumatoid arthritis and changes with biologic and targeted synthetic therapies to inform stratified therapy. *Front. Med.* **6**, 45 (2019).
87. Li, C. *et al.* Role of Resolvins in the inflammatory resolution of neurological diseases. *Front. Pharmacol.* **11**, (2020).
88. Orange, D. E. *et al.* Identification of three rheumatoid arthritis disease subtypes by machine learning integration of synovial histologic features and RNA sequencing data. *Arthritis Rheumatol* **70**, 690–701 (2018).
89. Norris, P. C. & Serhan, C. N. Metabololipidomic profiling of functional immunoresolvent clusters and eicosanoids in mammalian tissues. *Biochem. Biophys. Res. Commun.* **504**, 553–561 (2018).
90. Serhan, C. N. Pro-resolving lipid mediators are leads for resolution physiology. *Nature* **510**, 92–101 (2014).
91. Colas, R. A. *et al.* Proresolving mediator profiles in cerebrospinal fluid are linked with disease severity and outcome in adults with tuberculous meningitis. *FASEB J.* **33**, 13028–13039 (2019).
92. Koltsida, O. *et al.* Toll-like receptor 7 stimulates production of specialized pro-resolving lipid mediators and promotes resolution of airway inflammation. *EMBO Mol. Med.* **5**, 762–775 (2013).
93. Fredman, G. *et al.* An imbalance between specialized pro-resolving lipid mediators and

pro-inflammatory leukotrienes promotes instability of atherosclerotic plaques. *Nat. Commun.* **7**, 12859 (2016).

94. Kurkó, J. *et al.* Genetics of rheumatoid arthritis - a comprehensive review. *Clin. Rev. Allergy Immunol.* **45**, (2013).
95. Senapati, S. *et al.* Genome-wide analysis of methotrexate pharmacogenomics in rheumatoid arthritis shows multiple novel risk variants and leads for TYMS regulation. *Pharmacogenet. Genomics* **24**, 211–219 (2014).
96. Hinks, A. *et al.* Association between the *PTPN22* gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population: Further support that *PTPN22* is an autoimmunity gene. *Arthritis & Rheumatism* vol. 52 1694–1699 (2005).
97. Schulz, S. *et al.* rs2476601 in *PTPN22* gene in rheumatoid arthritis and periodontitis—a possible interface? *Journal of Translational Medicine* vol. 18 (2020).
98. Siebert, S. *et al.* Characteristics of rheumatoid arthritis and its association with major comorbid conditions: cross-sectional study of 502649 UK Biobank participants. *RMD Open* **2**, e000267 (2016).
99. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
100. Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *Cold Spring Harbor Laboratory* 447367 (2018) doi:10.1101/447367.
101. Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, (2013).
102. López-Vicario, C. *et al.* Association of a variant in the gene encoding for ERV1/ChemR23 with reduced inflammation in visceral adipose tissue from morbidly obese individuals. *Sci. Rep.* **7**, 15724 (2017).
103. Habermann, N. *et al. PTGS1, PTGS2, ALOX5, ALOX12, ALOX15, and FLAP* SNPs: interaction with fatty acids in colon cancer and rectal cancer. *Genes Nutr.* **8**, 115–126 (2013).
104. Goodman, J. E., Bowman, E. D., Chanock, S. J., Alberg, A. J. & Harris, C. C.

Arachidonate lipoxygenase (ALOX) and cyclooxygenase (COX) polymorphisms and colon cancer risk. *Carcinogenesis* **25**, (2004).

105. Zheng, Z. *et al.* The biological role of arachidonic acid 12-lipoxygenase (ALOX12) in various human diseases. *Biomed. Pharmacother.* **129**, 110354 (2020).
106. Horn, T. *et al.* Molecular basis for the catalytic inactivity of a naturally occurring near-null variant of human ALOX15. *Biochim. Biophys. Acta* **1831**, 1702–1713 (2013).
107. Assimes, T. L. *et al.* A near null variant of 12/15-LOX encoded by a novel SNP in ALOX15 and the risk of coronary artery disease. *Atherosclerosis* **198**, 136–144 (2008).

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
$RvD1_{n\text{-}3\ DPA}$	*143*	-33	-10	-20.2	-10
	215	-33	-10	-26.8	-10

$RvD2_{n\text{-}3DPA}$	*261*	-33	-10	-24	-10
	233	-33	-10	-27	-10
$RvD5_{n\text{-}3DPA}$	*199*	-33	-10	-22.5	-10
	263	-33	-10	-19.3	-10
$PD1_{n\text{-}3\ DPA}$	*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
$MaR1_{n\text{-}3\ DPA}$	*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: Cohort 1 RA patient 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-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-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

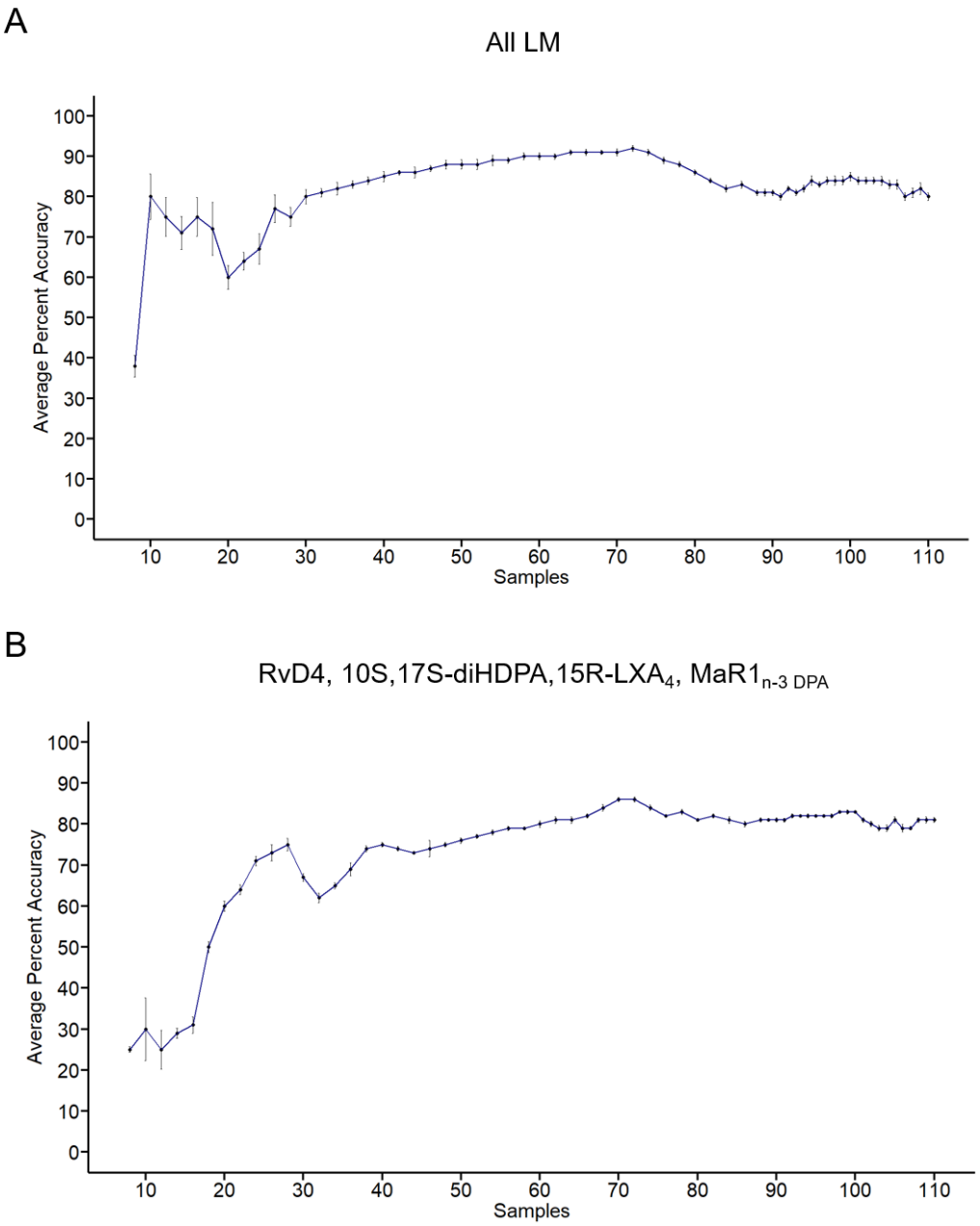
RT = Retention time, S/N = Signal to Noise. Ions marked in **Bold** denote fragments

Appendix 6. Baseline peripheral blood lipid mediator profiles in RA patients from cohort 1.

	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 = 24 DMARD Non-Responder.

Appendix 7. Learning curves for the models created using the random forest methodology.



Appendix 7. Establishing the minimum number of samples to create the all lipid mediator model using the random forest methodology. Machine learning models were created using the random forest methodology. Learning curves were made increasing the number of samples used to create the models. Results represent the average percentage of accuracy for the (**A**) all lipid mediator model and (**B**) the top four lipid mediator model. Samples from the first cohort and the second cohort were used to create the learning curves. Error bars represent the standard deviation of the out of the bag error of all trees used to create the model for each set of samples used. All LM = all lipid mediator; Top four lipid mediator = RvD4, 10S,17S-diHDPA,15R-LXA$_4$, MaR1$_{n\text{-}3\ DPA}$. All the models were created using the random forest methodology ("randomForest" package from R).

Appendix 8: Cohort 2 RA patient demographics and clinical information.

	DMARD Responders (n=36)	DMARD Non-Responders (n=22)
Pathotype (n)	Lymphoid (9), Fibroid (8), Myeloid (13), Ungraded (6)	Lymphoid (2), Fibroid (7), Myeloid (9), Ungraded (4)
Ethnicity	Caucasian (18), Asian (3), Indian (1), Chinese (1), Black Caribbean (1), Bengali (1), Somalian (1), Black African (2), Afro-Caribbean (1), Bangladeshi (2), Sudanese (1), Pakistan (1), Korean (1), Caribbean (1)	Caucasian (8), Asian (5), Black African (1), Bangladeshi (2), Mixed Greek (1), Black (1), Pakistani (1), British (1), African (1)
Gender (n)	Female (13), Male (22)	Female (19), Male (3)
Age at Recruitment – years	51 (±14)	50 (±10.4)
Onset	4.8 (±2.7)	4.8 (±3.1)
Currently smoking (%)	7 (20)	6 (22.3)
Co-Morbs Baseline (n)	Osteoarthritis (1), Acne (1), Anaemia (1), Asthma (5), Depression (3), Hypothyroidism (2), Gastro-oesophageal reflux disease (1), anal fistula (1), Hay fever (1), Varicose veins (1), Hypertension (10), Glaucoma (2), Psoriasis (1), Gastric ulcers (1), Diverticulitis (1), Allergy to penicillin (1), Hypercholesterolemia (5), Erectile dysfunction (1), Ischaemic heart disease (2), Diabetes (3), Coronary heart disease (1), Enlarged prostate (1), Carpal tunnel syndrome (1), Low vitamin D (1), GORD (1), Ca prostate (1)	Osteoarthritis (3), Asthma (4), Depression (2), Hypertension (6), Hypercholesterolemia (6), Diabetes (5), Osteopenia (1), Low back pain (1), Anaemia (1), Psoriasis (3), Gout (1), heart block (1), hearing loss (1), laiden deficiency (1), Polcystic ovaries (1), Chest infection (1), Gastritis (1), Fibromyalgia (1), Renal tubular acidosis (1),
Concomitant Med. Baseline (n)	NSAIDs (1), Tetracycline (1), Ferrous sulfate (2), Fluoxetine(1), Fluticasone (1), Salbutamol (1), Lansoprazole (2), Levothyroxine (2), Naproxen (5), Gaviscon (1), Movicol (1), Ibuprofen (5), Amlodipine (6), Timolol (1), Latanoprost (1), Omeprazole (2), Tramadol (1), Diclofenac (2), Perindopril (2), Etoricoxib (1), Indapamide (1), Calcium (2), Irbesartan (1), Vitamin D (2), Sibicol (1), Bendroflumethiazide (2), Simvastatin (6), Co-codamol (5), Metformin (3), Ramipril (2), Repaglinide (1), Insulin (2), Sildenafil (1), Aspirin (3), Dipyridamole (1), Isosorbide (1), Calci-chew (1), Doxazosin (1), Prednisolone (1), Lisinopril (2), Atenolol (2), Gabapentin (1), Symbicort (1), Paracetamol (1), Solgar (1), Glucosamine (1), Multivitamin (1), Senokot (1), Amitriptyline (1), Flixotide (1), Ventolin (1), Phyllocontin (1), Tamsulosin (1), Arcoxia (1)	Salbutamol (2), Lansoprazole (3), Levothyroxine (1), Naproxen (2), Ibuprofen (3), Amlodipine (2), Omeprazole (2), Tramadol (2), Diclofenac (3), Irbesartan (1), Vitamin D (1), Bendroflumethiazide (1), Simvastatin (2), Co-codamol (3), Metformin (7), Ramipril (2), Aspirin (1), Doxazosin (2), Lisinopril (1), Gabapentin (1), Paracetamol (4), Ventolin (1), Arcoxia (2), Co-dydramol (3), Gliclazide (3), Atorvastatin (4), Pioglitazone (1), Trimethoprim (1), Lipitor (1), Allopurinol (1), Voltarol (1), Buprenorphine (1), Dihydrocodeine (2), Glucophage (1), Lanus (1), Novorapid (1), Amoxycillin (1), Clarithromycin (1), Losartan (1), Frusemide (1), Vitamin B12 (1), Humalog (2), Exenatide (1), Glucaxide (1), Glargine (1), Citalopram (2), Colecalcif (1), Metatone (1), Paroxitene, (1), Fenobrate (1), Ferrous fu (3), Bisacodyl (1), Pregabalin (1),
Recent Steroid Therapy (n)	No (30), Yes (5)	No (20), Yes (2)
Steroid Treatment (n)	Depo-Medro (2), Pred (2), Fluticason (1)	Methylpred. (1), Depo-Medro (1)
DMARD Treatment (n)	MTX (1), MTX/ASA (10), MTX/ASA/HCQ (5), ASA (1), ASA/HCQ (2), HCQ (1), MTX/HCQ/LEF (1), MTX/HCQ (12)	MTX (4), MTX/ASA/HCQ (2), MTX/ASA (6), MTX/HCQ (2), ASA (2), ASA/HCQ (1), HCQ (1), ASA/LEF (1)
ESR (mm/hr)	42 (±34)	45 (±26)
CRP (mg/l)	22 (±34)	31 (±35)
CCP (UI/L)	197 (±218)	164 (±200)
RF (UI/L)	119 (±187)	272 (±227)
Tiredness VAS (1-100)	49 (±31)	51 (±32)
Pain VAS (1-100)	61 (±28)	55 (±26)
Pt. VAS Global Health (1-100)	62 (±29)	62 (±30)
Physician VAS Global Assess. (1-100)	60 (±24)	60 (±25)
Tender Joints Baseline (Number/28)	15 (±8)	13 (±8)
Swollen Joints Baseline (Number/28)	9 (±6)	6 (±5)
HAQ (Max 38)	1.46 (±0.75)	1.85 (±0.68)
DAS28 (0-10)	6.03 (±1.48)	5.67 (±1.20)
Delta DAS28 (difference between 2 time points	-2.59 (±1.18)	0.14 (±1.0)
Response (1)	Responder (36)	Non-responder (22)
Response (2)	Good (19), Moderate (17)	Non-responder (22)
US Synovial Thickness (12max) (Number/36)	2 (±0)	

US Power Doppler (12max) (Number/36)	1.5 (±0.5)	
US Synovial Thickness (Biopsied Joint) (Number/3)	13 (±1)	
US Power Doppler (Biopsied Joint) (Number/3)	4 (±1)	
Radiographic Erosion (n)	No (20), Yes (6)	No (11), Yes (4)

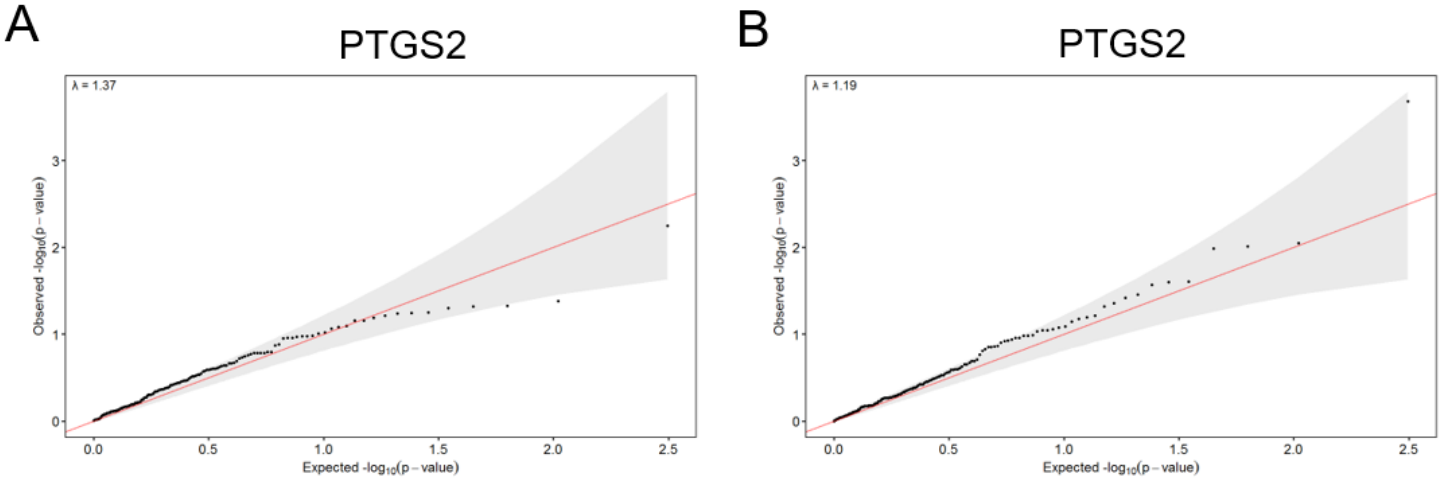
Erythrocyte sedimentation rate = ESR, C-reactive protein = CRP, Anti-cyclic citrullinated peptide = CCP, rheumatoid factor =RF, VAS = visual analogue scale, HAQ = Health Assessment Questionnaire, DAS28 = disease activity score-28, US = ultrasound, MTX = methotrexate, ASA = aspirin, HCQ = hydroxychloroquine,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

Appendix 9. Baseline peripheral blood lipid mediator profiles in RA patients from cohort 2.

	DMARD Responders			DMARD Non-Responders		
DHA bioactive metabolome	**Mean**	**±**	**SEM**	**Mean**	**±**	**SEM**
RvD1	0.31	±	0.19	0.50	±	0.17
RvD2	0.08	±	0.04	16.96	±	11.42
RvD3	0.01	±	0.01	0.10	±	0.05
RvD4	0.93	±	0.27	1.98	±	0.51
RvD5	0.15	±	0.04	2.39	±	1.41
RvD6	0.18	±	0.03	1.72	±	0.78
17R -RvD1	0.49	±	0.12	0.68	±	0.23
17R -RvD3	0.01	±	0.00	0.02	±	0.02
PD1	0.38	±	0.06	0.92	±	0.28
PDX	0.12	±	0.02	1.47	±	0.60
17R-PD1	0.04	±	0.01	0.65	±	0.34
22-OH-PD1	0.09	±	0.03	0.66	±	0.35
PCTR1	0.00	±	0.00	0.00	±	0.00
PCTR2	0.17	±	0.06	0.13	±	0.08
PCTR3	0.36	±	0.14	0.00	±	0.00
MaR1	0.44	±	0.14	3.27	±	0.81
MaR2	0.08	±	0.03	1.43	±	0.74
7S,14S-diHDHA	0.67	±	0.17	2.70	±	0.70
4S,14S-diHDHA	0.65	±	0.09	3.79	±	2.12
22-OH-MaR1	0.74	±	0.26	6.55	±	1.98
14-oxo-MaR1	0.04	±	0.02	0.02	±	0.01
MCTR1	0.00	±	0.00	0.00	±	0.00
MCTR2	0.44	±	0.12	2.15	±	0.81
MCTR3	0.29	±	0.12	1.83	±	1.14
n-3 DPA bioactive metabolome						
RvT1	0.48	±	0.45	0.00	±	0.00
RvT2	2.07	±	0.18	2.40	±	1.36
RvT3	0.32	±	0.11	0.43	±	0.13
RvT4	1.74	±	1.03	0.93	±	0.34
$RvD1_{n\text{-}3\ DPA}$	0.78	±	0.21	0.74	±	0.50
$RvD2_{n\text{-}3\ DPA}$	0.36	±	0.25	0.17	±	0.06
$RvD5_{n\text{-}3DPA}$	0.49	±	0.22	2.69	±	0.90
$PD1_{n\text{-}3\ DPA}$	0.06	±	0.03	0.10	±	0.06
10S, 17S-diHDPA	0.08	±	0.03	0.12	±	0.06
$MaR1_{n\text{-}3\ DPA}$	0.05	±	0.03	1.53	±	0.69
RvE1	0.24	±	0.08	0.01	±	0.01
RvE2	1.58	±	0.27	5.37	±	1.62
RvE3	0.21	±	0.06	0.18	±	0.10
AA bioactive metabolome						
LXA_4	0.02	±	0.01	0.41	±	0.16
LXB4	0.80	±	0.49	1.86	±	0.83
15R-LXA_4	0.65	±	0.31	3.82	±	0.82
15R-LXB_4	1.39	±	0.30	0.44	±	0.17
5S,15S-diHETE	2.05	±	0.44	5.87	±	3.78
LTB_4	18.04	±	4.20	66.04	±	14.51
5S,12S-diHETE	0.15	±	0.07	10.51	±	2.95
6-trans-LTB_4	0.72	±	0.14	4.79	±	0.98
12-epi-6-trans-LTB_4	0.63	±	0.17	9.97	±	2.10
20-OH-LTB_4	10.92	±	3.42	34.65	±	7.93
20-COOH-LTB_4	1.43	±	0.48	7.26	±	1.89
LTC_4	51.99	±	17.43	14.45	±	3.07
LTD_4	14.88	±	4.12	8.14	±	1.89
LTE_4	91.54	±	23.83	32.76	±	6.85
PGD_2	0.87	±	0.15	5.84	±	1.39
PGE_2	1.30	±	0.31	7.13	±	1.72
PGF_{2a}	2.04	±	0.62	23.01	±	4.64
TXB_2	53.28	±	13.57	45.37	±	16.21

Results are expressed pg/ml. n = 36 for DMARD Responders and n = 22 DMARD Non-Responders

Appendix 10. Genomic inflation factors (ƛ) and Q-Q plots for CG association models.

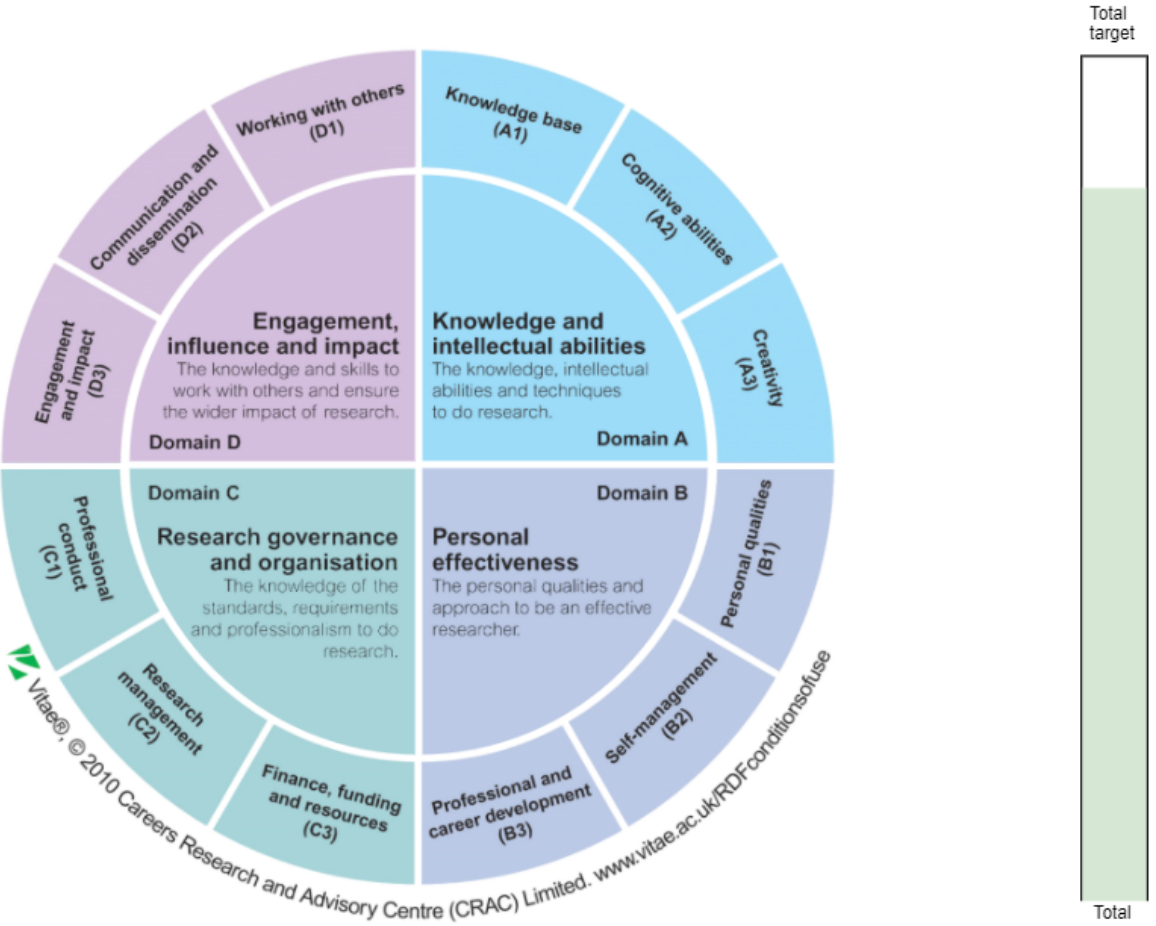


Appendix 10. Genomic inflation factors (ƛ) and Q-Q plots for CG association models of SNP in PTGS2. Representative ƛ and Q-Q plot obtained for CG association models of SNP in PTGS2 from the (**A**) British and (**B**) South Asian population to verify how fit the model is to each of the dataset used.

Appendix 11. 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		

Points Summary

Year	Type	Pts:	A	B	C	D	Total	Cap:	A	B	C	D	Total
0	**Course/event/seminar attendance sub-total**		**0.0**	**0.0**	**3.0**	**0.0**	**3.0**						
	Year 0 Total (with caps applied)		**0.0**	**0.0**	**3.0**	**0.0**	**3.0**						
1st	*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**						
	Basic IT/Software/Programming course		4.0	0.0	0.0	0.0	4.0						
	Core research knowledge or methods course (e.g. LTCC, IALS courses, masters lectures)		5.0	0.0	0.0	0.0	5.0						
	Doctoral College event/course		0.0	2.0	3.0	2.0	7.0						
	Health and Safety		0.0	0.0	1.0	0.0	1.0						
	School/Institute Induction < or = 3 hours		0.0	1.0	1.0	0.0	2.0						
	Course/event/seminar attendance sub-total		**9.0**	**3.0**	**5.0**	**2.0**	**19.0**						
	External funding application <£2,000		0.0	0.0	4.0	2.0	6.0						
	Funding/Intellectual Property 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)		**52.5**	**12.0**	**9.0**	**36.0**	**109.5**						
2nd	*Conference Attendance (One day)*		6.0	4.0	0.0	0.0	10.0						
	Conference attendance sub-total		**6.0**	**4.0**	**0.0**	**0.0**	**10.0**						
	Ethics/Research Integrity/Good Clinical Practice		0.0	0.0	8.0	0.0	8.0						
	Other CPD course		2.0	0.0	5.0	0.0	7.0						
	Other career training or development		0.0	2.0	0.0	0.0	2.0						
	Course/event/seminar attendance sub-total		**2.0**	**2.0**	**13.0**	**0.0**	**17.0**						
	Ethical Approval (NHS)		5.0	0.0	15.0	5.0	25.0						
	Ethical approval sub-total		**5.0**	**0.0**	**15.0**	**5.0**	**25.0**						
	Internal Presentation (>30 mins)		4.0	2.0	0.0	2.0	8.0						
	Journal Club/Reading Group/lab meeting Presentation		1.0	0.0	0.0	1.0	2.0						
	Giving presentations sub-total		**5.0**	**2.0**	**0.0**	**3.0**	**10.0**						
	Journal Club/Reading Group/lab meeting/mentoring group - attendance		1.0	0.0	0.0	0.0	1.0						
	Meeting/club/reading group attendance sub-total		**1.0**	**0.0**	**0.0**	**0.0**	**1.0**						
	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		4.0	0.0	0.0	16.0	20.0						
	Written publications sub-total		**6.0**	**0.0**	**0.0**	**24.0**	**30.0**						
	Year 2 Total (with caps applied)		**25.0**	**8.0**	**28.0**	**32.0**	**93.0**						
Total	*Conference Attendance (Four days)*		12.0	8.0	0.0	0.0	20.0						
	Conference Attendance (One day)		6.0	4.0	0.0	0.0	10.0						
	Conference attendance sub-total		**18.0**	**12.0**	**0.0**	**0.0**	**30.0**		**18.0**	**12.0**			**30.0**
	Basic IT/Software/Programming course		4.0	0.0	0.0	0.0	4.0		30.0				30.0
	Core research knowledge or methods course (e.g. LTCC, IALS courses, masters lectures)		5.0	0.0	0.0	0.0	5.0		120.0				120.0
	Doctoral College event/course		0.0	2.0	3.0	2.0	7.0						
	Ethics/Research Integrity/Good Clinical Practice		0.0	0.0	8.0	0.0	8.0				30.0		30.0
	Health and Safety		0.0	0.0	4.0	0.0	4.0				10.0		10.0
	Other CPD course		2.0	0.0	5.0	0.0	7.0						
	Other career training or development		0.0	2.0	0.0	0.0	2.0			60.0			60.0
	School/Institute Induction < or = 3 hours		0.0	1.0	1.0	0.0	2.0			3.0	3.0	3.0	9.0
	Course/event/seminar attendance sub-total		**11.0**	**5.0**	**21.0**	**2.0**	**39.0**						
	Ethical Approval (NHS)		5.0	0.0	15.0	5.0	25.0		5.0		15.0	5.0	25.0
	Ethical approval sub-total		**5.0**	**0.0**	**15.0**	**5.0**	**25.0**						
	External funding application <£2,000		0.0	0.0	4.0	2.0	6.0				8.0	4.0	12.0
	Funding/Intellectual Property application sub-total		**0.0**	**0.0**	**4.0**	**2.0**	**6.0**						
	Internal Presentation (>30 mins)		4.0	2.0	0.0	2.0	8.0		12.0	6.0		6.0	24.0
	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		**10.0**	**2.0**	**0.0**	**8.0**	**20.0**						
	Journal Club/Reading Group/lab meeting/mentoring group - attendance		10.5	0.0	0.0	0.0	10.5		30.0				30.0
	Meeting/club/reading group attendance sub-total		**10.5**	**0.0**	**0.0**	**0.0**	**10.5**						
	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		4.0	0.0	0.0	16.0	20.0		4.0			16.0	20.0
	Refereed Publication (Journal Paper, Book chapter, not abstract) submission		4.0	0.0	0.0	16.0	20.0		4.0			16.0	20.0
	Written publications sub-total		**8.0**	**0.0**	**0.0**	**32.0**	**40.0**						
	Total (with caps applied)		**66.5**	**20.0**	**40.0**	**51.0**	**177.5**						
	Grand Total (with caps applied)		**66.5**	**20.0**	**40.0**	**51.0**	**177.5**						