TNF gene promoter DNA methylation level in early Rheumatoid Arthritis

Thomas James Sargent

Submitted in accordance with the requirements for the degree of Master of Science by Research (MScRes)

The University of Leeds
School of Medicine and Health
March 2023

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

First, I give my gratitude and thanks to my project supervisors, Dr. Frederique Ponchel (FP|) and Dr. Kulveer Mankia (KM), for their guidance, support and advice throughout the PhD. Secondly, I thank Mrs Rekha Parmar (RPa) for training me in the laboratory techniques necessary to conduct this project, conducting DNA extraction, bisulfite conversion and qMSP for a proportion of the samples and being an additional source of advice and support with the project. I also send my thanks to Dr. Rujiraporn Pitaksalee (RPi), a PhD student who developed the qMSP assay I utilised in this project and conducted DNA extraction, bisulfite conversion and qMSP for a proportion of the samples. The processing of blood samples and separation and freezing of the PBMCs was performed by the tissue bank staff, allowing me to analyse TNF gene DNA-methylation results for 312 patients. Thank you also to Miss Helen Ng (HN) and Miss Chin Liu (CL), two medical students whom I worked alongside to retrieve and process the clinical data from patient records.

This research project has therefore been carried out as a team effort as part of a wider biomarker programme of work headed by Dr Ponchel. My own contributions, fully and explicitly indicated in the thesis, have been the undertaking of some of the DNA extraction, bisulfite conversion, qMSP, sorting and merging of clinical data, data curation (including DAS28 calculations), as well as the descriptive and modelling statistical analysis of the data.

Abstract

Background: To improve disease outcomes, Rheumatoid Arthritis (RA) should be treated early as possible. The role of biomarkers is essential to facilitate diagnosis and early access to treatment. Epigenetic modification is an important mechanism that could act as a biomarker in early RA. The value of differential methylation of the TNF gene was explored as a biomarker using the quantitative methylation specific qPCR (qMSP).

Hypothesis: There will be significant difference in the percentage of DNA methylation of the TNF gene in patients with RA.

Objective: To measure percentage DNA methylation of the TNF gene from PBMC samples of patients with early drug naïve inflammatory arthritis using a qMSP assay to measure difference in TNF gene methylation between RA and non-RA in patient from an early inflammatory arthritic clinic and see whether it can be utilised as a biomarker to predict the diagnosis of RA.

Methods: A qMSP assay was developed to measure DNA methylation from blood samples. PBMC samples from the IACON and RADAR cohorts (n=312). Logistic regression was used to establish the added value of the assay compared to clinical data only.

Results: Percentage DNA-methylation of the TNF gene was significantly lower in RA (median 3.13) compared with other forms of inflammatory arthritis (median 6.61), with a p-value $4.1x10^{-9}$ by MWU, distinguishing early RA from other forms of inflammatory arthritis. Using Forward logistic regression, the qMSP assay performed well, with an OR of 1.840 (95%CI: 1.567-2.162, p<0.0001) and an AUROC of 0.826 (0.771-0.881) for predicting RA. The reference model predicted 87.8% of cases with an AUC=0.950 *(P=≤0.0001*). Adding the TNF DNA methylation levels increased the prediction accuracy by +1.7% (89.4%) accuracy and an AUC=0.967 (*P=≤0.0001*).

Conclusion: The outcome is the characterisation of a novel biomarker of early RA which could be utilised to improve the management of RA.

Table of Contents

Chapter 4: RESULTS

Chapter 5: DISCUSSION

Figures and Tables

Abbreviations & Units

1. Abbreviations

Ankylosing Spondylitis (AS)
Anti-Citrullinated Protein Antibody (ACPA)
Anti-Inflammatory Macrophage (M2)
Anti-Nuclear Antibody (ANA)
Antibody (Ab)
Antigen (Ag)
Antigen Presenting Cells (APCs)
Area under the Receiver Operator Curve
Body Mass Index (BMI)
C Reactive Protein (CRP)
Crohn's Disease (CD)
Cytotoxic T cell (Tc)
Disease Activity Score (DAS)
Disease Modifying Anti-Rheumatic Drugs (DMARDS)
Early Arthritis Clinic (EAC)
Erythrocyte Sedimentation Rate (ESR)
Extracellular Matrix (ECM)
Fibroblast-like Synoviocytes (FLS)
Helper T Cell (Th)
Human Leukocyte Antigen (HLA)
Immune Mediated Inflammatory Disease (IMID)
Interleukin (IL)
Juvenile Idiopathic Arthritis (JIA)
Kappa B (RANK)
Inflammatory Arthritis (IA)
Immunoglobulin (Ig)
Inflammation Related Cell (IRC)
Inflammatory Arthritis Continuum (IAC)
Major Histocompatibility Complex (MHC)
Matrix Metalloproteinases (MMPs)
Methotrexate (MTX)
Monoclonal Antibody (MAb)
Neutrophil Extracellular Traps (NETs)
Non-steroidal Anti-inflammatory drugs (NSAIDS)
Peripheral Blood Mononuclear Cells (PBMCs)
Polymerase Chain Reaction (qPCR)
Pro-Inflammatory Macrophage (M1)
Psoriatic Arthritis (PsA)
Quantitative Methylation Specific Polymerase Chain Reaction (qMSP)
Reactive Arthritis (ReA)
Receiver Operator Curve (ROC)
Receptor Activator of Nuclear Factor

2. Units

Millilitre (mL)
Microlitre (uL)
Nanolitre (nL)
Milligram (mg)
Microgram (ug)
Nanogram (ng)

Chapter 1: Introduction

1 The Presentation and Management of Rheumatoid Arthritis

1.1 Epidemiology and Clinical Presentation

Rheumatoid Arthritis (RA) is an immune-mediated arthropathy where inflammation, swelling and pain in the joints are the predominant manifestation of the disease[1,2]. RA impacts approximately 1% of the population of most developed nations, making it a particularly prevalent autoimmune disease[3]. RA is more frequent in females, with a ratio approximately 3:1 females to males[4]. The disease is most prevalent in the age range of 40 to 60 years old and has both genetic and environmental contributions to disease manifestation and progression[5].

RA most commonly impacts the metacarpal phalangeal joints and proximal interphalangeal joints of the hands and metatarsal phalangeal joints of the feet, though inflammation and damage to larger joints, such as the shoulders, elbows, knees and ankles are a frequent clinical presentation[2,6]. The disease tends to impact the joints symmetrically and impacts more than 5 joints, described as polyarthropathy[2]. Morning stiffness is a key feature of the disease, and an improvement of symptoms tends to occur later in the day or through physical activity[6]. Progression of the disease in the joints can lead to deformity and disability. In the hands, manifestations of deformity include swan neck deformity, buttonhole deformity, ulnar deviation and Z-thumb deformity. In the feet, manifestations of deformity include hammer toe deformity, claw toe deformity and mallet toe deformity. These deformities result from the hyperextension, flexion and subluxation of the joints of the hands and feet[6]. Bone erosions, osteopenia and osteoporosis occur, resulting in loss of structural integrity of the bone and deformity[7]. Pathological presentations of RA can be clinically identified through imaging, usually by ultrasound but also MRI in clinical trials, and has served as a key tool in aiding diagnosis and monitoring disease progression[8,9].

RA can also present systemically, where multiple organs may be affected[6]. Rheumatoid Nodules can develop on the skin, a very common extra-articular presentation of RA[10]. Chronic inflammation in RA can also lead to interstitial lung disease, fibrosis of the lungs and pleural

effusion[11]. Inflammation and damage to the heart tissue or blood vessels, and the build-up of atheromatous plaques, increase the risk of cardiovascular disease, such as heart attack and stroke[12]. Patients with RA may also develop a triad of conditions, where low blood cell count, splenomegaly and the core joint manifestations of RA occur, described as Felty Syndrome[13].

1.2 Diagnosis

Diagnosis of RA is based on fulfilling classification criteria following the guidance of the the American College of Rheumatology (ACR) and European League Against Rheumatism (EULAR) criteria for RA, updated in 2010[2]. This scoring system is a guideline and not definitive, nor do patients need to meet all the criteria for diagnosis but serves to guide the probability of correct diagnosis based on progress made regarding biomarkers of disease and our understanding of disease progression. The guidelines score points based upon clinical presentation of joint involvement, serology and duration of symptoms (see Table 1). Counting tender joint count (TJC) and swollen joint count (SJC) at clinic and through imaging is one metric used and also serves to confirm symmetrical polyarthropathy. The presence of the autoantibodies in serology, such as rheumatoid factor (RF) and anti-citrullinated peptide antibodies (ACPA), are also scored. High levels of the inflammatory markers C-reactive protein (CRP) and erythrocyte sedimentation (ESR) in the blood are also contributors towards a diagnosis but are by no means indicative, occurring in a plethora of inflammatory diseases and infection as generalised markers of inflammation. A high score, in combination with the patient's family history, clinical history, symptom presentation and duration can collectively aid towards a diagnosis at the discretion of the clinician[2].

Joint Involvement	
1 large joint	0
2-10 large joints	1
1-3 small joints	2
4-10 small joints	3
>10 joints	5
Serology	
Negative RF and negative ACPA	0
Low-positive RF or low-positive ACPA	2
High-positive RF or high positive ACPA	3
Acute Phase reactants	
Normal CRP and normal ESR	0
Abnormal CRP or abnormal ESR	1
Duration of symptoms	
<6 weeks	0
≥6 weeks	1

Table 1 – EULAR 2010 Classification Criteria for Rheumatoid Arthritis[(2)]

The criteria of classification of Rheumatoid Arthritis, as outlined by the European League Against Rheumatism 2010. This involves a scoresheet based on the categories of joint involvement, presence of anti-citrullinated autoantibodies (ACPA) and Rheumatoid Factor (RF) by serology, presence of C-Reactive Protein (CRP) and Erythrocyte Sedimentation Rate (ESR) and duration of symptoms. The higher the score, the greater the likelihood of Rheumatoid Arthritis.

1.3 Management and Treatment

Monitoring of RA after diagnosis usually involves imaging by ultrasound to compare progress in tissue swelling, bone erosions, structural changes and deformity[8,9]. This is used in combination with the disease activity score (DAS28), a composite score for RA that is used to monitor changes in activity as outlined by the National Institute for Health and Care Excellence (NICE) guidance[14]. This composite score combines the total number of tender joints (TJC) and swollen joints (SJC) impacted, the concentration of CRP and ESR in the blood and a Visual Analogue Scale of General Health (VAS-GH, a scale rating from 0-100) by the patient, regarding the extent of symptoms such as pain and mobility over the last 7 days, which is used as a subjective measure of disease impact. Several cut-off points are used to define clinical stages of disease for remission (DAS28<2.6), low disease activity (DAS28<3.2), and high disease activity (DAS28>5.1)[14].

The standard treatment regime for RA usually starts with a conventional synthetic disease modifying anti-rheumatic drug (csDMARD) as first-line treatment, most commonly methotrexate. Leflunomide or sulfasalazine may be used if a contraindication for methotrexate is present or added upon methotrexate treatment when LDA is not achieved with MTX alone after 6 months. Systemic corticosteroids (usually glucocorticoids), and NSAIDs may also be used for symptomatic treatment during this time and for glucocorticoids as a bridging therapy until the more gradual effects of the csDMARD has been established during initial therapy[15,16]. This overall strategy is now called the Treat to Target (T2T) approach, the target being the achievement of clinical remission defined by DAS28<2.6[17].

If either LDA or remission is achieved, a gradual tapering of csDMARDs is implemented, either by dose reduction or by longer intervals between doses. However, if significant disease progression is still occurring, with concern for progression of cumulative disability, an addition of a targeted monoclonal antibody therapeutic, known as biologics or biological disease modifying anti-rheumatic drug (bDMARDs), can be combined with a csDMARD after unsuccessful csDMARD therapy alone (defined by a DAS28>5.1), known as combination therapy[15,16]. bDMARDs offer a targeted mechanism of action but are more expensive than csDMARDs[18]. Numerous bDMARDs target the excessive production of pro-inflammatory cytokines secreted by immune cells. TNFα inhibitors Adalimumab and Infliximab and IL6 inhibitor Tocilizumab have been highly successful therapeutics in the treatment of RA[16,19].

Other biologics used to treat RA include Anakinra, an IL1 inhibitor, and Abatacept and Rituximab, which supress T-cell and B-cell activity, respectively[19]. However, biologics are usually only prescribed upon unsuccessful response or contraindication from a csDMARD because bDMARDs are more likely to cause side effects, such as increased risk of opportunistic infection[16].

Janus kinase inhibitors (JAKis), such as Tofacitinib and Baricitinib, are inhibitor molecules of the antiviral intracellular JAK-STAT pathway recently brought to market[20]. Their mechanism of action is based upon inhibiting the excessive production of interferons. JAKis, considered tsDMARDS, may carry the same heightened risk of infection as bDMARDS and therefore they are also currently advised to be reserved until unsuccessful treatment of a csDMARD[16].

Advice is continually being updated about the most appropriate tapering of therapeutics, side effects and cost-benefit analysis of various treatment strategies. Guidelines based on these discussions are continually updated from sources such as EULAR and American College of Rheumatology (ACR) guidance[15,16]. Difficult-to-treat RA occurs in a proportion of patients with RA and poses challenges regarding our existing therapeutics[21,22]. Therefore, research is being conducted to develop biomarkers that predict therapeutic response to optimise treatment and prognosis for RA patients[23].

Unsuccessful therapeutic intervention or delayed diagnosis can result in substantial disease progression[24]. Fortunately, prognosis for patients with RA has improved drastically over the years due to the increased attainment of remission or LDA because of earlier diagnosis and intervention and the improvements in targeted therapeutics on the market[25]. Nonetheless, patients with long-lasting sequalae of RA have a significantly reduced quality of life and a population of patients still have little option in terms of therapeutics, having exhausted all available drugs[21,26,27]. Mortality also remains lower in patients with RA compared with the general population, which may be predominantly due to increased cardiovascular disease risk amongst patients with RA, which can lead to earlier risk of stroke and myocardial infarction[12]. However, co-morbidities including pulmonary disease and increased risk of malignancy has also been associated with RA and so may also cause early mortality[28,29]. Research therefore needs to be carried to further improve prognostic outcomes for patients with RA.

2 The Pathogenesis of Rheumatoid Arthritis

2.1 Disease Associated Risk Factors

As a polygenic disease, the cause of RA is complicated, but known to be immune mediated[30,31]. Individuals who develop RA are likely to have a family history of the disease, however the probability of inheriting RA in both monozygotic twins is about 15-30%, indicating factors beyond genetics[32,33]. Known genetic mutations associated with RA include alleles of human leukocyte antigen (HLA) genes, such as the HLA-DRB1 shared epitope (SE). These alleles code for protein sequences for the major histocompatibility complex (MHC) type 2, which have the function of antigen presentation to CD4+T-cells. This can stimulate T-cells to become effector cells capable of inflammatory immune responses or eliciting tolerance, depending upon the antigen and co-stimulatory signals involved. Variations to this allele cause alterations in the folding of the MHC protein that may impact the binding affinity to peptide presented to the T-cell and their response[34].

Comprehensive epigenome-wide association studies (EWAS) have revealed many non-HLA mutations of genes associated with RA, including many T-cell related genes that lead to alteration of co-stimulatory proteins (CD2, CD28, CD40, CD56 and CTLA-4) and T-cell receptor (TCR) genes[35-37]. PTPN22 is a strongly associated gene, which is involved in the activation threshold of both T-cells and B-cells[37,38]. Furthermore, alterations to genes associated with inflammatory responses, such as STAT4 of the JAK-STAT pathway, have also been implicated[39]. Genes associated with PAD enzymes, such as PADI4, have been associated with RA[37,40]. PAD enzymes are involved in the process of citrullination, a post-translational modification implicated in RA through associated autoimmune reaction by ACPA to citrullinated proteins[41]. Lastly, TRAF1-C5 has been associated with rheumatoid arthritis, a loci of chromosome 9 that encodes tumour necrosis factor receptor 1 and complement C5[42]. Interestingly, ACPA+ and ACPA– disease do not share the same genetic predisposition suggesting distinct mechanisms are at play in these two subgroups of RA patients[122].

2.2 Environmental Risk Factors

Environmental factors have also been linked to the development of RA, such as smoking, high BMI, dysbiosis of the gut microbiome and periodontal disease[43,44]. An environmental mechanism associated with RA is through the citrullination of proteins, a post-translational modification of arginine to citrulline. Increased citrullination is associated with the generation of anti-citrullinated protein antibodies (ACPA) that target the citrullinated proteins, particularly in the joint in RA, such as to fibrinogen, vimentin and α-enolase, which contributes to pathology[41]. Free citrullinated proteins from dysregulated apoptosis can be phagocytosed and presented to adaptive immune cells on MHC class 2 receptors to trigger immune responses and induce B-cell maturation that switch to an IgG-ACPA repertoire[45,46]. Many environmental factors implicate the citrullination of proteins in their mechanism of action. Smoking has been associated with the prevalence of citrullinated proteins, shared epitope and ACPA antibodies, through the mechanism of action is uncertain[46-48]. The bacterium which commonly causes periodontal disease, *Porphyromonas Gingivalis,* has been associated with RA by releasing arginine-gingipain enzymes, which is able to cleave proteins to reveal arginine residues that are the target of citrullination by PAD enzymes expressed by the pathogen and the host[49,50]. This may contribute to the increase of citrullinated proteins in the mouth during periodontal disease, a region of the body critical in generating tolerance to exogenous antigens and ultimately causes a build-up of ACPA. Further studies should investigate whether pathobionts in the gut and lung may contribute to a protein citrullination burden by similar mechanisms, which could explain the association between gut and lung microbiome dysbiosis and RA.

Further microbial associations could be explained due to dysbiosis of the gut and oral microbiome, where a reduction of species diversity and species richness in the gut that has been associated with inflammation and a variety of non-communicable diseases, including some associations with RA[44,51]. The role of the microbiome in regulating the mucosal immune system is a current topic of research. Dysbiosis may lead to immune dysregulation, aberrant signalling and transcriptomic changes and vice-versa, disease-related change in diet and lifestyle may cause the dysbiosis, so whether this is a true cause of the disease or a consequence of having RA remains to be established[52]. There has been associations between gut dysbiosis and RA with pathobionts such as *Prevotella Copri,* and potentially other pathobionts of the *prevotella* genus, though the mechanism behind this association is not clear[53]. *Porphyromonas Gingivalis* also produces the enzyme α-enolase, which displays 82% homology to host α-

enolase, raising the possibility of molecular mimicry, as autoantibodies to α-enolase have been observed in patients with RA[54].

Additional environmental associations include vitamin D deficiency, where patients with RA and numerous other autoimmune diseases have been found far more prevalent in populations in the northern hemisphere with deficient vitamin D. The mechanism behind this association is inconclusive and poorly understood[55]. Lastly, chronic exposure to environmental toxins, including organic particulates, silica and mineral dust has been linked with the development of RA, perhaps through the irritation and cell damage induced stress by these exogenous particles, which may lead to chronic inflammation, particularly in the lung[56,57].

3 Inflammatory Arthritis Continuum

3.1 The Concept of a Continuum

The presence of autoantibodies, particularly ACPA, can appear years before clinical symptoms and the diagnosis of RA[58,59]. This indicate a slow progression of disease, now fully acknowledged as the Inflammatory Arthritis Continuum (IAC)[60]. Here, distinct phases of disease before clinical RA have been characterised (Figure 3).

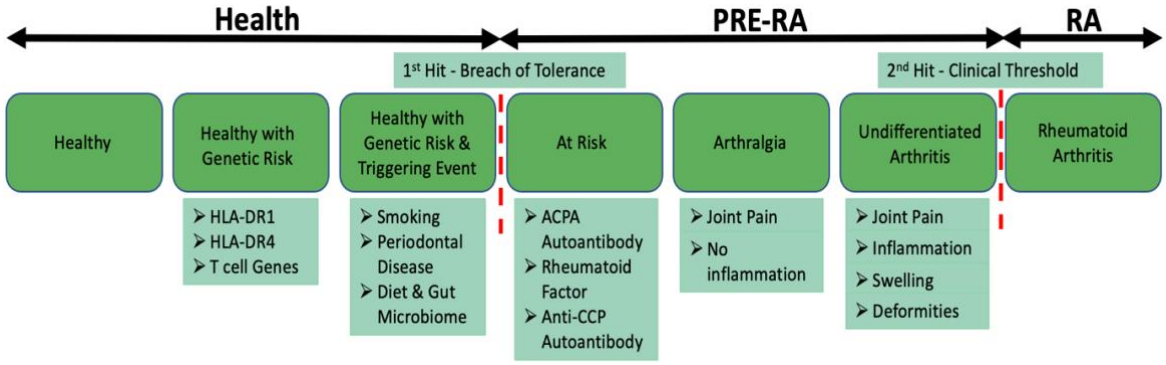


Figure 1 – The Inflammatory Arthritis Continuum

The IAC is a model of disease progression. Healthy individuals at risk with genetic susceptibility for RA (shared epitope, T-cell related genes) may also experience environmental triggers including smoking and periodontal disease, which increases the chance of the "first hit", an event describing progression of disease towards a breach of tolerance and subsequent autoimmunity, expressing autoantibodies such as ACPA and RF. Patients may then clinically progress to experience arthralgia. Subclinical synovitis may be present. Patients may further progress to show clinically observable inflammation and swelling and may show early signs of joint deformity. However, patients at this stage may not meet the diagnostic criteria for RA, being diagnosed as having undifferentiated arthritis. However, many will progress to meet the classification criteria for RA and therefore may be clinically diagnosed with RA.

The first phase involves healthy patients that have genetic and environmental factors that are associated with RA, described as at-risk individuals. This includes factors mentioned earlier that contribute to pathogenesis, such as the genetic mutations (HLA-DR4, HLA-DRB1, HLA-DR1,) and environmental factors such as smoking and periodontal disease[33,47,50,53,61]. These factors increase susceptibility to RA and may be associated with loss of tolerance and the production of autoantibodies in RA, such as ACPA or anti-CarP[58,62].

Once this triggering event has occurred, described as the first hit, the patient may gradually develop idiopathic pain without the presence of clinical signs of pathology. This phase is described as arthralgia, where symptoms are present but there is no SJC despite some TJC[63]. However, subclinical synovitis may be detectable with ultrasound or MRI imaging[64].
Another triggering event, the second hit, may occur that leads towards progression into undifferentiated arthritis (UA), usually with clinical synovitis (SJC) and other pathological involvement without yet meeting the EULAR classification criteria[65-67]. At this stage, the patient usually has clear signs of disease that may involve inflammation markers such as CRP, tender joints, swelling present and pathology in the joints particularly upon clinical assessment by ultrasound[65-67]. Some patients remain classified as undifferentiated arthritis due to the complex and atypical presentation of their disease and many will progress further to meeting EULAR criteria of RA classification[2,68]. This phase of disease is often described as the 'window of opportunity' because patients can meet diagnostic criteria at this stage whilst early intervention with therapeutics has proven critical for their long-term prognostic outcomes[69].

However, there is not a single path to RA, and many patients never present with autoantibodies while with clear IA symptoms[60,70,71]. This represents approximately a third of the patients developing RA seen in the Early Arthritis Clinic, experiencing delays in diagnosis (for up to 24 months sometimes) and therefore access to treatment.

3.2 The Application of the IAC

The establishment of the IAC and evidence of stages of progression to RA can be very useful for clinicians and researchers in the diagnosis and prognosis of patients with inflammatory arthritis. Firstly, the model proposes that mechanisms of autoimmunity in RA may not be solely responsible for the pathogenesis and progression of the disease due to the presence of

autoantibodies in patients 'at risk' who do not have clinically noticeable inflammation and those with clinical symptoms that progress along the IAC and match EULAR criteria who are seronegative for autoantibodies[70,72]. Furthermore, the model indicates a 'window of opportunity' that is critical for desirable prognostic outcomes that may frequently be missed in seronegative patients[70,71]. This is reflected by the findings that patients who are ACPA negative and develop RA by virtue of meeting the other EULAR criteria have different genetic risk, whilst development and progression to RA in these circumstances can still be predicted[73,74].

4 Immunological Mechanisms in Rheumatoid Arthritis

4.1 T-cells

T-cells originate in the thymus, where they undergo positive selection in the cortex and negative selection in medulla. This process is aided by antigen presenting cells (APCs) and structural cells of the thymus, thymocytes. Positive selection encourages adequate affinity to foreign antigen and negative selection depletes autoreactive T-cells[75]. Thymic development is therefore critical in governing the healthy immune responses of T-cells and have been a focus of research in autoimmune disease. T-cell receptor excision circles (TRECs) are small circles of DNA found in naïve T-cells that function in T-cell receptor (TCR) rearrangement to provide the diversity of TCR repertoire for antigen found in T-cell populations. TRECs are not inherited during cell division, so as the population of peripheral T-cells expand, the concentration of TRECs that can be detected dilute. TREC content can therefore indicate thymic activity and/or the extent of proliferation and differentiation of T-cells[76]. TREC content has been shown to be depleted in the T-cell populations of RA patients compared to healthy controls, particularly within the naïve helper T-cell subset, suggesting the abnormal thymic activity, in association with defective IL7 producton by the thymus epithelium cells[76,77].

Abnormal maturation of T-cells have been noted in patients with RA. This appears to result in the loss of CD62L homing receptor expression specifically on naïve CD4+ T-cells, defining a subpopulation of naïve T-cells termed inflammation related cells (IRCs) not seen in health[77]. These cells could evade trafficking to lymph nodes through their lack of CD62L expression,

accumulating in the periphery[77]. Furthermore, they can enter IL6-expressing tissue and hence home for inflamed sites and my group hypothesise that this is the results of chronic exposure of naïve cells to IL6[77].

Activation and polarisation of naïve CD4+T-cells is a core aspect of their function (Figure 1). Upon stimulation from antigen presenting cells (APCs), naive T-cells mature and polarise towards effector phenotypes. This is governed by transcription factors, where T-Bet polarise naïve T-cells into Th1 cells, GATA3 polarise naïve T-cells into Th2 cells and RORγT polarise naïve T-cells into Th17 cells[78]. In RA, Th1 cells are not polarising appropriately, with no full engagement of Tbet[79,80]. Research has also demonstrated an increased polarisation of CD4+T-cells into a mixed, pro-inflammatory Th1/Th17 cell phenotype[81].

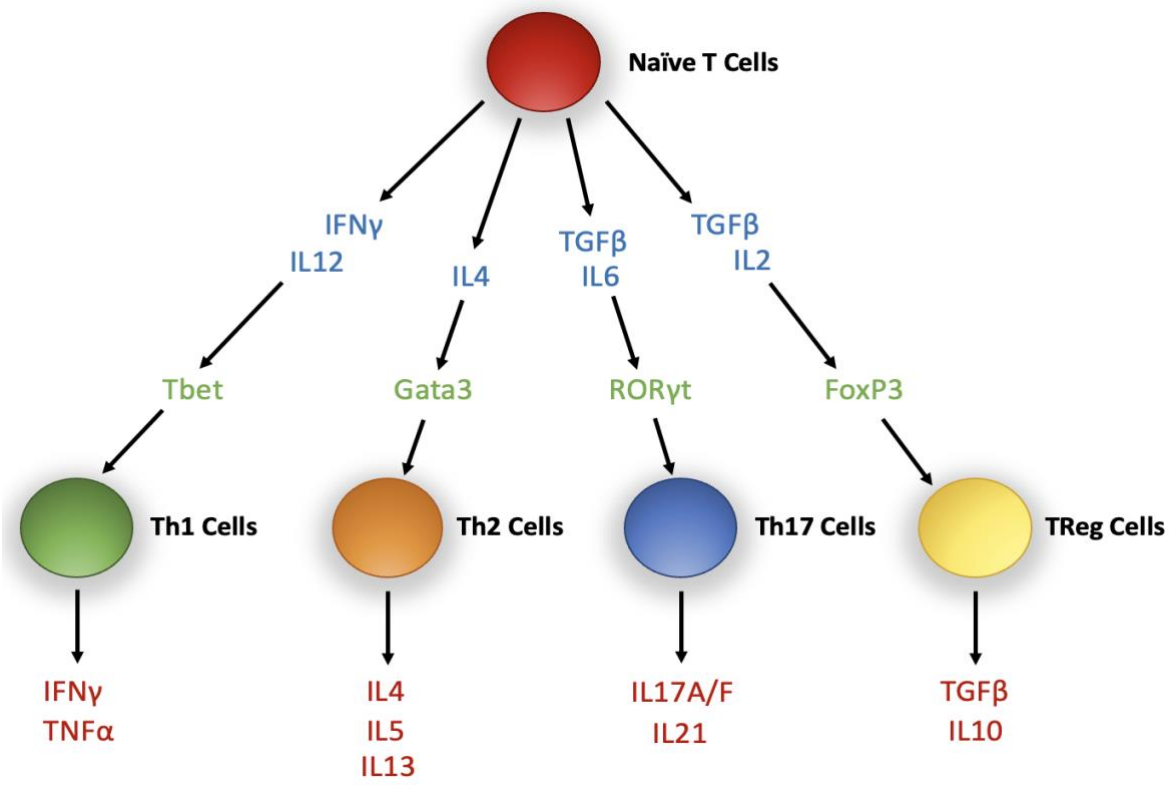


Figure 2 – The Polarisation of CD4+ Helper T-Cells

***In blue**: key cytokines involved in polarisation of naïve T-cells into specialised CD4+ helper T-cells. **In green**: the transcription factor implicated in governing epigenetic change of the naïve T-cell into the respective CD4+ helper T-cells. In red: key cytokines produced by polarised CD4+ helper T-cells.*

Regulatory T-cells (Tregs) are another important subset of CD4+T-cells that are generated in the thymus, characterised by the transcription factor FOXP3[62]. They can also be generated from normal effector T-cells as a product of polarisation[62] (inducible Treg, Figure 1). They function to regulate and resolve pro-inflammatory responses to avoid chronic inflammation[82]. Abnormalities in Tregs have been noted in RA, such as a deficiency in particular phenotypes of Tregs, including lower frequencies[83]. High concentrations of TNFα was shown to be partially responsible for diminishing the regulatory capacity of Tregs in RA. TNFα appears to alter the Treg functional capacity by reducing the activity of the transcription factor FOXP3, which is critical for the differentiation of naïve CD4+T-cells into Tregs in the periphery[84].

Other subsets of T-cells have been associated with chronic inflammation and extra-articular manifestations in RA, such as memory CD28- CD4+T-cells[85,86]. Th9 and Th22 subsets have also been identified and associated with autoimmune disease, secreting IL9 and IL22, respectively, though research is ongoing regarding their potential pro-inflammatory roles[87,88]. Th9 and Th22 cell polarisation is possibly driven by transcription factors PU.1 and AhR, respectively[87,89]. The flexibility between these phenotypes demonstrates that CD4+T-cells can be very adaptable in their functions depending on the environment and that the disease environment in RA may drastically alter T-cell activity to favour inflammatory roles that consequentially contribute to pathology.

Altogether, naïve CD4+T-cells are not maturing appropriately in RA, where a high number of T-cells express markers of both naïve (CD45RA) and mature (CD45RO) surface markers simultaneously and losing CD62L[77]. These cells (termed IRC) appear early in RA (notably in pre-clinical RA), predicting progression to disease[74] and seem to remain present even in remission[77,90]. These cells are also hyper-responsive to stimulation by antigen and mitogens and therefore may have a lower activation threshold towards inflammatory phenotype[77]. Furthermore, subsets of CD4+T-cells in RA also seem to upregulate chemokine surface markers inappropriately, such as CXCR4, CCR4 and CCR5[73]. This may contribute to increased homing of these cells to sites of inflammation, such as the joint[77].

The expansion of knowledge regarding T-cell subset dysregulation in RA has been exploited to develop biomarkers for RA progression based on cell subsets in the blood. Using flow cytometry, quantifying cell subset frequencies is possible and loss of naïve CD4+T-cells, abnormal differentiation into IRC and loss of Treg precede and predict the development of

symptoms in the at-risk stage of the inflammatory disease continuum, allowing for the classification of RA patients (notably ACPA-negative patients) early at the undifferentiated arthritis stage[91-94]. Biomarkers based on cell subsets have also had success in predicting response to methotrexate and flare of disease upon tapering of csDMARDs[91-94].

4.2 B-cells and Autoantibodies

B-cells are involved in activating T-cells by presenting antigens along with co-stimulatory signals[95]. This is particularly relevant considering the success of B-cell depleting therapy in RA (using anti-CD20 notably, such as Rituximab) that results in B-cell loss, but that do not affect plasma cells, hence the production of antibodies[96].

However, B-cells are also associated with pathogenesis in RA through other direct mechanisms also, such as the production of RANKL, inflammatory cytokines such as IL-6 and autoantibodies[97,98].

Rheumatoid Factor (RF) is an autoantibody found in approximately 70% of RA patients. RF is an IgM autoantibody that binds to the patients IgG to form immune complexes which can deposit in the joint[99]. Depositions of RF enhance the activity of macrophages, including increasing the production of TNFα[100,101]. Deposition of immune complexes can also trigger the complement cascade, contributing to further inflammation[99].

ACPA are autoantibodies that bind to proteins that have been citrullinated as part of physiological responses, such as fibrinogen, fibronectin, vimentin, α-enolase; among many other proteins in the joint[102]. This causes immune-mediated response and results in damage to connective tissue, likely also through activating innate mechanisms such as the complement cascade and activating innate immune cells through Fc receptor signalling[103]. ACPA has also been associated with the increased production of TNFα through engaging macrophages[103].

Anti-carbamoylated protein antibodies (anti-CarP) have also been observed in patients with RA, which are autoantibodies that bind to proteins that have been carbamoylated[62,104]. Carbamoylation has also been associated with increased bone destruction, though the precise mechanism of his association is not clear[62]. Additional post-transcriptional modifications

(acetylation, phosphorylation and more) also generate autoantibodies and have been reported in RA (albeit in lower frequency) and appear to link autoimmune responses to particular attacks on tissue or proteins[105].

4.3 Macrophages

Resident macrophages in RA polarise to express an inflammatory M1 phenotype, which can be initiated by cytokines and possibly ACPA[106,107]. Macrophages can promote inflammation in RA that contribute to osteoclastogenesis and bone erosion[108,109]. They secrete pro-inflammatory cytokines, such as IL6, IL17 and TNFα, which engage fibroblasts, fibroblast-like synoviocytes and osteoclasts to initiate cartilage degradation and bone erosion, respectively[110-113]. Altogether, they are heavily contributing to perpetuating inflammation through their highly active production of inflammatory mediators.

4.4 Resident Cells

Resident cells of the joint in RA, such as fibroblasts, fibroblast-like synoviocytes and chondrocytes, express a phenotype very different than in health, favouring high production of proteinases, such as MMPs, that contribute to extracellular matrix and cartilage tissue degradation[114-116]. The progressive destruction and deformity of joints by these cells is driven in part from the inflammatory signalling of macrophages and T-cells, such as IL6, IL17 and TNFα[110,117]. IL1, IL6 and TNFα has been shown to induce the expression of RANKL in fibroblast-like synoviocytes, promoting osteoclastsogenesis[118]. T-cells and B-cells also express RANKL, which communicate by binding with RANK on osteoclasts to promote local bone erosion[98,119].

4.5 Cytokine Dysregulation and Immune Mediated Inflammatory Disease

Pro-inflammatory cytokines are excessively secreted by various immune and resident stromal cells in the synovial joint of patients with RA, leading to chronic synovitis and joint destruction[120]. Cytokines frequently implicated include IL1, IL6, IL17, IL23, IFNγ and TNFα[121]. This inflammatory network maintains chronic inflammation, however, development of cytokine dysregulation may change depending on stage of disease (early, established, chronicity)[121]. The spatial and temporal dependent crosstalk between cell types demonstrates the complexity of disease pathophysiology in RA, involving adaptive immune cells (Th1 cells, Th17 cells, B-cells), innate cells (macrophages, neutrophils) and resident stromal cells (fibroblasts, fibroblast-like synoviocytes, chondrocytes, osteoclasts) in propagating the disease environment of RA[121]. The heterogeneity of disease, including the importance of innate and stomal cells in disease progression, indicate the disease is not simply propagated through classical mechanisms of autoimmunity, but also through mechanisms of dysregulated cytokine signalling and inflammatory cell phenotype that overlaps significantly with research exploring polygenic autoinflammatory diseases[122]. This observation of the overlap of polygenic autoimmune disease pathophysiology with that of polygenic autoinflammatory diseases suggests that they may be treatable with similar therapeutics[31]. This concept has allowed for increased collaboration of research, therapeutic insight and broadening of scope into pathogenesis of these conditions, using an umbrella term of immune mediated inflammatory diseases (IMIDs), which can be seen as a spectrum of non-communicable conditions that cause damage to self from those predominantly autoimmune in nature by adaptive immune mechanisms to those predominantly autoinflammatory by innate dysfunction[31,122]. Common culprits involved in cytokine dysregulation, such as IL1, IL6, IL17, IL23, TNFα, IFNγ as mentioned, are implicated across the spectrum of IMIDs and therefore these conditions may share common origins in early pathophysiology and progression[31,122]. There has even been documented cases of patients matching the criteria of RA who display seemingly predominant autoinflammatory mechanisms of disease, exemplifying the fluid nature of immune-mediated diseases[123]. Understanding this overlap and targeting dysregulated cytokines with novel biologics is therefore a critical step forward regarding the focus of research for RA and beyond.

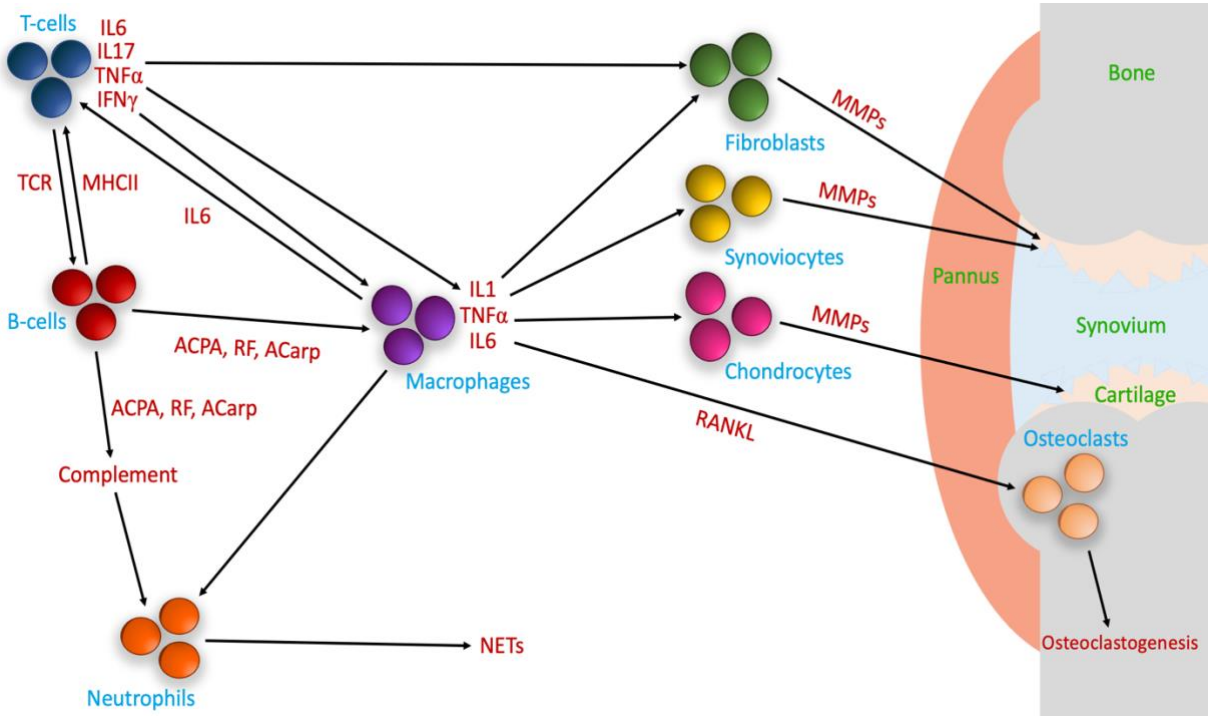


Figure 3 – Cytokine Dysregulation in Rheumatoid Arthritis

In the synovium of the joint of a patient with RA, there is dysregulation of cytokine signalling and cell phenotype. For example, the expression of high titres of autoantibodies such as ACPA and RF by B-cells, cytokines such as IL6 and TNFα *by T-cells and macrophages, MMPs by fibroblasts, fibroblast-like synoviocytes and chrondrocytes that lead to cartilage destruction and signalling of RANKL to osteroclasts that leads to bone absorption and erosions by osteoclastogenesis.*

Overall, the pathophysiology of RA involves both innate and adaptive immune responses[121,124]. Early RA may implicate T-cell driven events, whether antigen specific is not clear, leading to activation and polarisation of other cells in a self-maintaining network of pro-inflammatory signals, towards the development of a chronic stage that can no longer be returned to normal by regulatory mechanisms[90,121,125]. Hence the need for early diagnosis, to be able to treat before this chronic stage has been fully established.

5 Epigenetic Changes in Rheumatoid Arthritis:

5.1 An introduction to Epigenetics

Epigenetics refers to the ability of cells to regulate the expression of genes by mechanisms independent of transcription, but which modulate the accessibility of the DNA to the transcriptional machinery by regulating the structure of the chromatin. The phenotype of cells is therefore not solely based on the genetic material that codes for proteins but also by a program that allows or restricts access to these genes. However, this program is highly flexible and can be modulated based on cues from the environment. DNA is wrapped around histone proteins, which is required to keep a compact DNA structure inside the nucleus of cells, forming chromatin. However, the configuration of DNA wrapped around histones can be altered and can selectively allow genes to be expressed. Two core methods of epigenetic regulation include chromatin remodelling and DNA modification[126,127].

5.2 Chromatin Remodelling

Chromatin remodelling is one way in which epigenetic regulation occurs. Chromatin structural complexes are responsible for packing nucleosomes close together to form heterochromatin or unpacking nucleosomes and spreading them apart to unpack DNA into a euchromatin configuration (Figure 4). When the genetic material is in heterochromatin configuration, the strand of DNA is tightly wrapped around histone proteins, preventing the access of transcription factors to the promotors of entire sets of genes. When the DNA is in euchromatin configuration, where the strand of DNA is loosened away from the histone proteins, access of transcription factors to the gene promotors is possible. The histones can regulate this process based on harbouring post-translational modifications mediated by specific enzymes that modify their capacity to wrap the DNA around them. Histone acetylation occurs on lysine residues of histone tails. Lysine is positively charged and attracted to the DNA backbone and other nucleosomes and acetylation results in less attraction of the histones to the DNA, promoting chromatin unfolding. Histone methylation is another mechanism of chromatin remodelling that occurs on lysine or arginine residues of histone tails, mediated by histone methyl transferases (HMTs)[126,127].

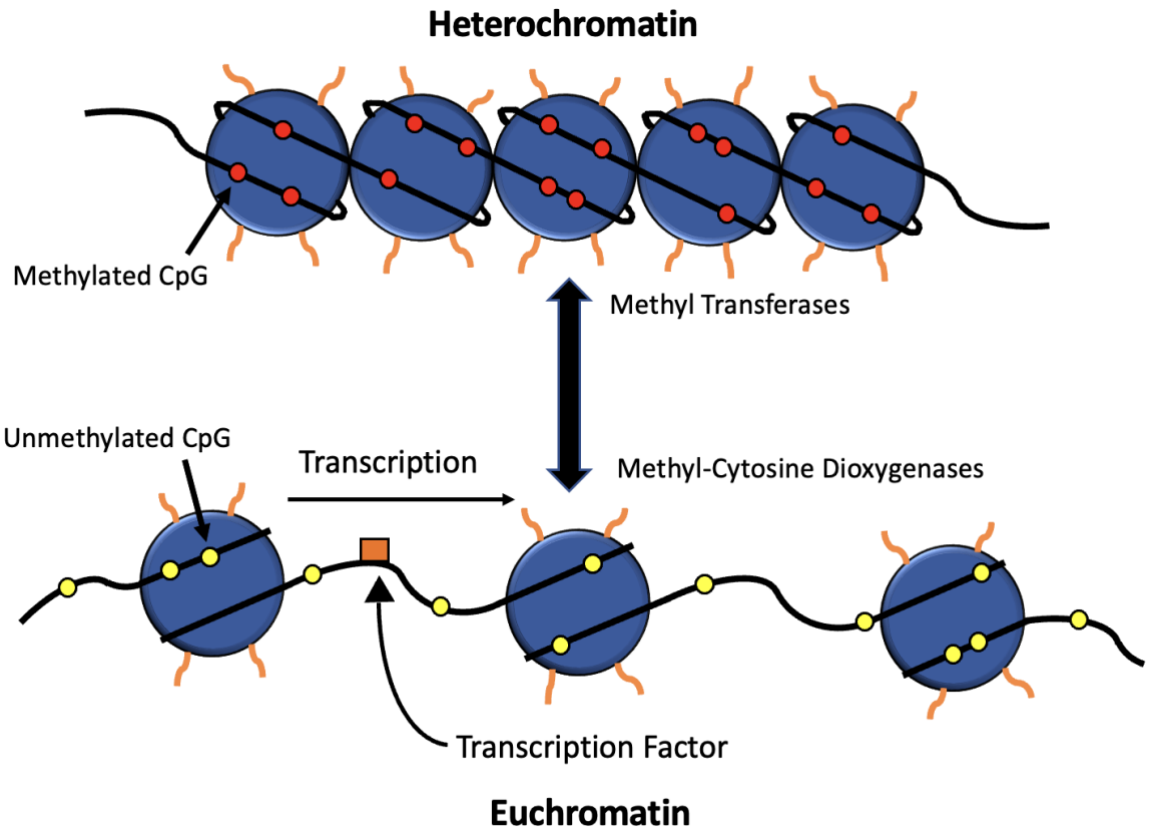


Figure 4 – Epigenetic Modification of the Chromosome and DNA

An example of epigenetics – heterochromatin confers a closed configuration, whilst euchromatin confers an open configuration, where transcription can then occur by transcription factors on revealed promotors. Post-translational modification by methylation is governed by methyl transferase enzymes and methyl-cytosine dioxygenase enzymes. An increase in DNA and chromosomal methylation leads to a closed configuration (heterochromatin) of genes and demethylation leads to an open configuration (euchromatin).

5.3 DNA Modification

DNA modification is another mechanism by which epigenetic regulation occurs. The promotor is the region of any gene that provides the binding site for transcription factors and these regions of the DNA also undergo inhibition or activation, altering the regulation of the corresponding genes. This occurs by methylation of cytosine in a dinucleotide sequence with guanine (Figure 4), which do not alter the genetic code but modulate the structure and accessibility of the chromatin. CpG are often grouped in islands, which are small regions of DNA particularly rich in CpG dinucleotide pairs. A CpG island is defined by having over 50% CpG pairs for longer than 200 base pairs. These islands occur in over 60% of gene promotors and are rarely found outside of promotors. DNA methylation occurs on the cytosine residues of CpGs and therefore although DNA methylation can occur all over the genome, it particularly occurs in the promotor region of genes. DNA methyl transferases (DNMTs) are enzymes which covalently attach a methyl group to cytosine residues on carbon 5, converting cytosine into 5-methyl-cytosine, which can occur at CpG islands of gene promoters. Demethylases are the enzymes which remove the methyl groups from cytosine residues, utilising the enzymatic proteins ten-eleven translocation methyl-cytosine dioxygenases (TET) and activation-induced cytidine deaminase (AID), which can reverse the effects of gene silencing at promotors. DNA methylation of the promotor region of a gene typically results in the repression of gene transcription. This is achieved through methylated DNA attracting methyl-CpG-binding domain proteins (MBDs), which bind the methylated-CpG and prevent access of the transcription factor. MBDs can also allow the binding of HDACs, which can then function to remove acetyl groups from histones and promote the condensing of DNA into heterochromatin in combination with the action of chromatin remodelling complexes, preventing transcription factors from expressing those genes. Both DNMTs and MBDs bound to the methylated-DNA can also directly methylate histones in conjunction with DNA-methylation[128].

5.4 Measuring DNA Methylation

An initial approach towards studying DNA methylation is through a process of discovery. This is achieved by indiscriminately exploring a wide area of the genome for potential differential methylation signatures of interest, known as Epigenome-Wide Association Studies (EWAS). There are a number of ways this can be achieved, such as through large microarrays (450k, EPIC), whole genome bisulphite sequencing (WGBS) or reduced representation bisulphite sequencing (RRBS) that provide high throughput of data[129]. Comprehensive bioinformatic pipelines can then be used to interpret large datasets[129]. Such studies can be used as a process of discovery for areas of intrigue in large cohorts of patients or patient data from in-silico datasets. This can often be a starting point for hypothesis generation or determining areas of intrigue for deeper analysis of a particular loci or CpG region of interest[129].

Pyrosequencing is a next-generation sequencing method which involves special dinucleotides (dNTPs) of each base containing pyrophosphate as part of their backbone, which emits light each time a new base is added to the sequence catalysed by DNA polymerase. Primers for target sequences can be designed using MethPrimer[130]. The technique is based around the release of pyrophosphate when a dinucleotide is bound to the sequence, where pyrophosphate is converted into ATP by the enzyme ATP sulfurylase. The ATP is then utilised by luciferase, which generates the photon of light when the enzyme luciferase converts luciferin with oxygen into oxyluciferin. The enzyme apyrase removes unincorporated dNTPs. dNTPs are added one at a time into a flow cell, which means we can determine the sequence of the DNA strand working through base pairs one at a time. Subsequent repeat dNTPs will multiply the amplification of the signal. By sequencing DNA that is both untreated and treated with bisulphite conversion into a purpose-made machine, you can compare both strands as a pyrograph with software to look for the proportion of cytosine dinucleotides in the methylated strand that has been converted into uracil in the unmethylated strand. By deducting the number of unmethylated cytosines from the total number of cytosines in the strand, you can calculate the proportion of methylated cytosines as a percentage methylation of that strand[131].

Quantitative Methylation Specific Polymerase Chain Reaction (qMSP) is a technique based upon qPCR that utilises primers for both methylated DNA and unmethylated DNA sequences. The primers are typically designed using MethPrimer for a particular CpG target sequence[130].

The methylated sequence will retain cytosine nucleotides at CpG regions, as the modification is not impacted by bisulphite conversion, whereas the unmethylated sequence will have any cytosine nucleotides at CpG sites converted into uracil. The mismatch between the two sequences will be noticeable as the unmethylated sequence will produce amplification during qPCR, whereas the methylated sequence will not. Samples can be run and compared with 100% methylated and 100% unmethylated controls to determine percentage methylation of the total cytosines at the target CpG region of your sample.[132].

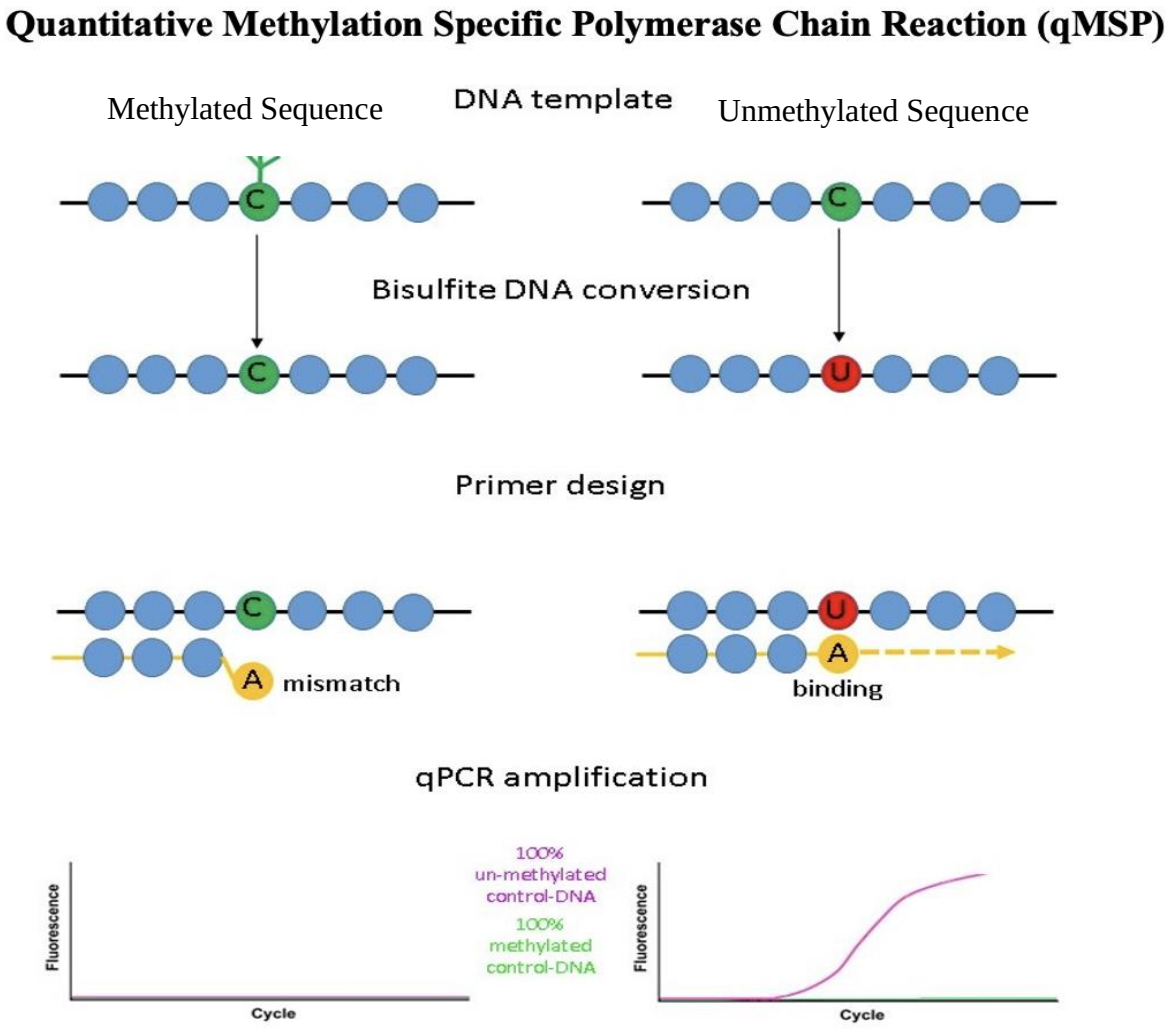


Figure 5 – The Process Behind qMSP

On the left: *A methylated cytosine, unaffected by bisulphite conversion of DNA.*

On the right: *An unmethylated cytosine, converted from cytosine to uracil and causing a mismatch compared with the methylated DNA strand. The unmethylated strand will then be amplified upon qPCR, whereas the methylated strand will not.*

5.5 Hypomethylation in Rheumatoid Arthritis

Epigenome-wide association studies (EWAS) are a promising avenue of research for RA. Previously, epigenome-wide association studies (EWAS) have been used to identify numerous loci of various non-HLA genes, such as T-cell related genes, associated with RA[35,36,42]. More recently, epigenome-wide association studies looking at DNA methylation is becoming invaluable regarding the study of epigenetic changes in many diseases[133].

Patients with RA have shown a tendency of genome-wide hypomethylation in various cell types (T-cells, fibroblasts, fibroblast-like synoviocytes)[134-136]. Our group explored DNA methylation in 3 cells type in early, drug naïve patients (CD4+T-cell naïve and memory as well as monocytes), where we have shown the most significant difference in naïve T-cells[137], with over 600 gene being differentially methylated (with 55% hypomethylation) while only 400 genes were identified in memory cell (90% hyper-methylated) and less than 50 genes in monocytes. Many differentially methylated genes in naïve T-cells were cytokine/chemokine and their receptors (including IL6/IL6R and TNF/TNF-Rs), while an overall network analysis of the 600 genes pointed to a central role for STAT3/JAK/IL6 signalling with links to TNF/related genes and Th17 cells[137].

Others have shown particular effects on many genes individually, especially within T-cell associated genes. For example, hypomethylation of the CTLA-4 gene has been observed in patients with RA compared with healthy controls, associated with Treg dysfunction[138]. Hypomethylation of the IL6 gene was also confirmed in patients with RA compared with healthy controls, which also led to over-expression of the IL6 gene and was associated with excessive inflammation and disease progression through IL6 signalling[139].

Epigenetic regulation of genes early in RA could have a critical role in RA pathogenesis and may explain the non-genetic element of disease shown in monozygotic twin studies[32,140].

6 Diagnostic Biomarkers of Rheumatoid Arthritis

6.1 Heterogeneity in Rheumatoid Arthritis

RA is a heterogeneous disease, where different pathophysiology can be seen between patients due to the complex and variable mechanisms involved in the disease course over many years[60,141]. This includes the different presentation of cell subsets involved in disease, serology, joint pathology, systemic complications, relapse/remission and resistance to current therapeutics[21,73,90,92,124,141].

Establishing endotypes of disease, where subgroups of RA patients can be clustered based on the characteristics measured by biomarkers present in disease, is the goal of precision medicine approaches[142]. This allows for clinical and therapeutic interventions to be tailored for the patient based upon their endotype to optimise long-term favourable prognosis, including therapeutic response[70]. However, in RA, evidence suggests that earlier diagnosis and intervention are crucial for improving the prognosis of patients with RA before the extensive heterogeneity of established disease has manifested, adding an important temporal element[23]. It is critical to avoid worsening of pathology and cumulative disability by predicting and treating RA early within the 'window of opportunity'[69]. With the support of the IAC model, numerous biomarkers of disease endotype and progression have been developed, allowing for the ability to give confident diagnosis, prognosis and monitoring of disease early in the development of RA[69]. There is also an overlap between pre-RA and the clinical presentation of other forms of inflammatory arthritis, so establishing biomarkers specific to RA also requires the ability to distinguish RA from other forms of early inflammatory arthritis[143]. Many biomarkers have been proposed and some have made it to clinical practice, whilst others have failed to demonstrate significant clinical utility but provided information about the pathophysiology of disease progression.

6.2 Serological Biomarkers in Rheumatoid Arthritis

ACPA is present in approximately 50-60% of early IA patients developing RA and shows more specificity (~97-99%) than RF and many other serological biomarkers[58,144,145]. The level of ACPA can indicate severity of disease, including increased likelihood of active disease, pathological progression and bone erosions[146,147]. However, with ~40-50% of RA patients ACPA negative, this poses issues regarding early diagnosis, where prognostic outcomes for ACPA negative individuals can be worsened substantially due to delayed diagnosis[71]. Further biomarkers need to be established in conjunction with ACPA to further improve classification of RA, with focus centred towards establishing biomarkers for effective diagnosis of RA in those with ACPA negative disease.

RF also appears frequently in patients with RA (60-70%); however, it has worse predictive power for diagnosis of RA than ACPA[148]. It has a low specificity (~30%)[148], being associated with many inflammatory and non-inflammatory conditions, infections and also with ageing[149].

Anti-Carp autoantibodies are an alternative serological biomarker that has high specificity for RA but even lower sensitivity than ACPA and RF, only being present in ~10% of early RA patients[104]. Our group showed that it has nonetheless additional value in early IA for RA classification[150]. In addition, anti-Carp autoantibodies have been shown to be a useful biomarker in predicting worse prognosis regarding joint damage[62].

6.3 Other Clinical Biomarkers in Rheumatoid Arthritis

Markers of systemic inflammation, such as CRP and ESR, appear in a myriad of inflammatory diseases, including a plethora of communicable and non-communicable conditions, making them poor diagnostic biomarkers due to low specificity for RA[151,152]. Normal ranges of CRP and ESR are also frequently observed in patients with early disease, meaning lower than ideal sensitivity, leading to missed diagnosis[152].

Clinical assessment can be standardised and utilised as biomarkers, such as with individual metrics of disease (TJC, SJC) or the DAS28 compound score[153]. These metrics are important and widely utilised regarding the monitoring of disease severity and remission in patients[14,153].

Imaging has proven vital in detecting damage and disease progression in the joint and associated tissues[9]. Subclinical synovitis, such as grey scale synovitis, are detectable by ultrasound early in disease and being investigated as potential biomarkers of diagnosis and disease progression[64,154]. The predictive power of ultrasound imaging early in disease is potentially advanced further using power doppler scale ultrasound (PDUS)[66]. Though ultrasound is cheaper and more accessible as a technology, alternative imaging techniques, such as MRI and PET, have also been investigated with promising results, but remain impractical due to limitations in cost, accessibility and tolerability of radiation for some patients[9].

6.4 T-cell Subsets as Biomarkers

In our group, the presence of T-cell subsets have previously been explored before my project as biomarkers for predicting diagnosis in individuals pre-RA and for prognosis across the IAC. Populations of naïve T-cells, Th17 cells, Tregs and inflammation related cells (IRCs) have been found to be significantly different between those with pre-RA and healthy controls[83,94,155]. In a comprehensive study of more than 700 participants, Treg cells were diminished early in the IAC, having a more important role in the at-risk stage of disease progression. Naïve T-cells had the strongest prediction value of prognosis towards progression to RA from at-risk stage, prediction of disease flare and successful MTX induced remission[155]. This was sufficiently strong to inform clinicians towards rapid therapeutic treatment for those progressing swiftly across the IAC into clinical RA and a clinical trial is currently stratifying early RA for treatment choice, which may lead towards more informed decisions regarding the likely success of drug tapering. On the other hand, both Tregs and naïve T-cells were found to be diminished in pre-RA, but IRCs were found to be significantly raised early in disease, providing insight into the second hit event being able to trigger naïve CD4+T-cells to differentiate into IRCs[77,83]. In another study, IRCs were also found during remission, but lacking chemokine receptors and their inflammatory potential diminished, no longer hyper-responsive. A higher number of these IRCs during remission predicted the likelihood of unsuccessful tapering of medication and inability to sustain remission, meaning this T-cell subset has potential clinical utility as a biomarker for relapse risk[90].

7 Preliminary Findings

7.1 Changes in DNA Methylation

In a previous paper by our group before my project, involving the work of PhD graduate Dr. Rujiraporn Pitaksalee, it was shown that multiple CpG regions are differentially methylated (DM) in early drug naïve RA patients[137]. The DNA methylation of CpG regions was compared in monocytes, naïve CD4+ T-cells and CD4+ memory T-cells in 10 patients with RA and 6 healthy controls. Multidimensional scaling (MDS) and in-house code was utilised to analyse samples, including +/-1500 base pairs from CpGs. The naïve T-cells in RA had 18,020 DM CpGs ($p \leq 0.01$), more than 14197 for memory T-cells and 6490 for monocytes. When corresponding to genes, T-cells displayed 648 DM genes, memory T-cells displayed 605 DM genes and monocytes only displayed 58 DM genes. The monocytes displayed isolated DM-CpGs, whereas memory CD4+ T-cells displayed larger groups of DM CpGs but spread over greater regions that weren't numerous. CD4+ naive T-cells, however, displayed numerous DM-CpGs clustered in CpG islands[137].

Many DM genes in naïve CD4+ T-cells were associated with cytokines, such as IL6 and TNFα and their receptors. The TNF gene displayed hypomethylation, with significant clustering of DM CpGs observed in the CD4+ naïve T-cells, whilst no DM of TNF was observed in the monocytes or CD4+ memory T-cells.

Using DMRcate, we investigated DM regions (DMR) for the most highly DM genes[156]. The dataset was filtered to 355 candidates, where 262 were hypomethylated and 93 were hypermethylated in CD4+ naïve T-cells. This included TNF, STAT5, IFN signalling genes and HLA-related genes.

Utilising two gene expression datasets for CD4+ T-cells, a list of differentially expressed (DE) genes was aggregated between early drug naïve RA patients and healthy controls. The most DE genes between RA and healthy controls were JAK1, STATs, TNF family genes and IFN signalling genes. DM and DE genes showed close associations[137].

We then investigated whether these DM genes in CD4+ naïve T-cells could indicate specific pathways associated with pathogenesis using the STRING database network analysis. We built an interaction model that displayed three JAK1/STAT nodes, which were IL6-IL6R/JAK1/STAT3, JAK1/STAT2, and IL2R/IL15R/JAK1/STAT5. Further inflammatory pathways deriving from IL6 involved the TNF/TNFR family. An additional inflammatory node for IL18/JAK1/STAT4 also indicated additional members of the TNF family, linking IL4R, IL13 and STAT5[137].

7.2 DNA Methylation for the TNF Gene Candidate

The TNF gene was selected for further validation by our group before my project, involving the work of Dr. Rujiraporn Pitaksalee, based on its known relationship in RA pathogenesis. A region -850/+2000 base pairs from the transcriptional start of the TNF gene was analysed and showed partial demethylation, with a β-value of 50% methylation in CD4+ naïve cells (Figure 5). This region was almost fully demethylated (average 8%) in CD4+ memory T-cells and almost fully methylated (average 88%) in monocytes, but on a shortened scale of -175/+343 base pairs.

A sequencing assay was then developed to be utilised on DNA from early RA patients and healthy controls. A 273 base pair region that contained 8 CpGs of the region of interest (Figure 6, highlighted in the pink box) was sequenced from CD4+ T-cell DNA. In healthy controls, 50%/50% methylated and demethylated DNA was observed, suggesting two subpopulations of CD4+ T-cells, one with methylated DNA and the other with unmethylated DNA. In RA (n=9), 90% of the DNA was unmethylated for all CpGs.

We decided to develop a quantitative methylation specific polymerase chain reaction (qMSP) assay that could be utilised to measure percentage difference of DNA methylation of the TNF gene between groups. This assay was developed by PhD graduate Dr. Rujiraporn Pitaksalee, and my project consisted of further contributing to the generation and analysis of data generated with the assay to validate the assay as a clinical biomarker, across the IAC, between those eventually diagnosed with RA and other forms of inflammatory arthritis (non-RA)[137,157].

Chapter 2: Hypothesis & Objectives

1. Hypothesis

My hypothesis was that there will be a significant difference in percentage of DNA-methylation of the TNF gene in patients with RA compared with patients with an alternative diagnosis (the non-RA group).

2. Objectives

The first objective of my project was:

- To measure the percentage DNA methylation of the TNF gene from PBMC samples of patients with early drug naïve inflammatory Arthritis using the qMSP assay.

- To establish if there is a difference in TNF gene methylation between those classified as Rheumatoid Arthritis and other forms of inflammatory arthritis (non-RA group).

- To evaluate the utility of TNF gene methylation as a biomarker for disease outcome (i.e., diagnosis) above and beyond the currently used criteria for classification

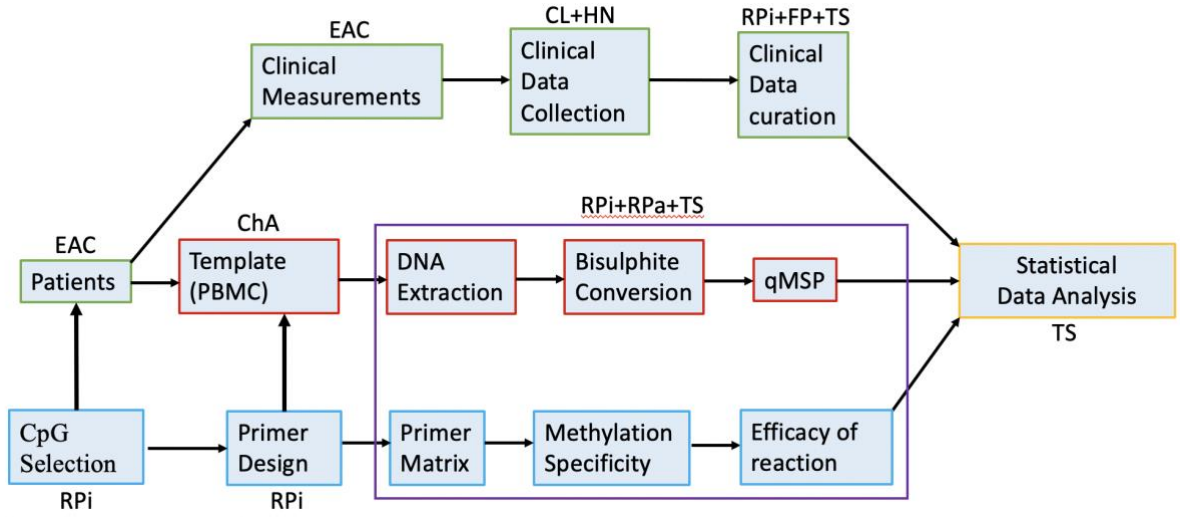


Figure 6 – The Workflow of the Project

In green: *The clinical aspects of the project, involving patient recruitment, clinical measurements, clinical data collection and formatting****. In red:*** *the core development of the TNF assay from the utilisation of patient bloods, involving PBMC isolation, DNA extraction, bisulphite conversion and the qMSP assay measurements.* ***In blue:*** *the preparations required to validate the qMSP assay, including selection of the target CpG region of the TNF gene, primer design, quantification of the assay through primer matrix, methylation specificity and dilution series****, In yellow:*** *the statistical analysis of data, including descriptive statistics of the cohort, DAS28 calculations, TNF methylation comparisons and modelling statistics, including logistic regression and AUROC.*

EAC = Early Arthritis Clinic, ChA = Chapel Allerton Hospital; All other components of the project occurred at the Brenner Building of St James University Teaching Hospital.

TS = Mr. Thomas Sargent, RPi = Dr. Rujiraporn Pitaksalee, RPa = Mrs. Rekha Parmar, CL = Miss Chin Liu, HN = Miss Helen Ng, FP = Dr.. Frederique Ponchel.

Chapter 3: Methods

1 Ethics

This project has been approved by the Leeds (West) Research Ethics Committee (Ethically approved REC: 09/H1307/98). Blood samples were obtained from patients recruited from the Early Arthritis Clinic at Chapel Allerton Hospital in Leeds, forming the cohorts IACON and RADAR. All patients involved in this project gave informed consent through a form regarding how their data and samples would be utilised for the study and gave the option of the right to withdraw. All clinical data necessary was extracted from research databases and NHS servers by researchers with permission to obtain and utilise such data. The clinical data was anonymised on these records by allocating each patient a study-ID number and identifiable data was removed, (such as date of birth), notably as it was not required for the multivariate and statistical analysis required for this project. Data was only shared through university systems onto an encrypted, shared drive.

2 Participant Recruitment and Data Collection

Participants were recruited into the IACON and RADAR cohorts from the Early Arthritis Clinic. Participants were undiagnosed patients suspected to have inflammatory arthritis by a general practitioner (GP), with symptoms such as arthralgia, inflammation and swelling of the joints, who were referred to the EAC. Participants underwent clinical assessment, which involved the recording of swollen and tender joint counts (SJC, TJC), serology for inflammatory markers (CRP, ESR), serology for autoantibodies (ACPA, RF), a visual analogue scale for patient wellbeing (VAS-GH) and a question regarding smoking status (had they ever previously smoked?). The date since they experienced joint pain, inflammation and swelling was also recorded and the duration of symptoms until first visit was calculated. These participants were then assessed at 0 months (baseline), 6 months and 12 months. A diagnosis of RA, if fulfilling EULAR classification criteria, another form of inflammatory arthritis (e.g., PsA, Reactive Arthritis) or undifferentiated arthritis (UA) was established at baseline for suitability for the project and then re-evaluated during follow-up visits at 6 months and 12 months. All participants were drug-naïve upon baseline visit (0 month) when bloods were taken for the project, and data used for a prediction model of final diagnosis given at 12 months.

3 Assay Development and Optimisation

3.1 qMSP Primer Design

Differential methylation of the TNF gene was selected as a candidate from an illumina genome-wide discovery as an ideal CpG target by a bioinformatic analysis in R Studio/Bioconductor and a search of genomic databases and also based on the bisulphite sequencing data. This work was performed by PhD graduate Dr. Rujiraporn Pitaksalee.

Based on the results of the assay developed for validation, utilising bisulphite sequencing of DNA, the overall region around the candidate CpG was demethylated in RA and the qMSP assay design could assume that the whole region was demethylated. Design of a novel primer/probe for the qMSP occurred from a sequence which has high demethylation. This resulted in lower complexity once converted with bisulphite required, so that the assay was designed toward measuring methylated DNA, as needed to obtain a suitable TM for primers. This was also performed by Dr. Rujiraporn Pitaksalee before I joined the group. Her design is as follows:

Gene	Position	F/R/Probe	Sequence	Product size
TNF	Chr 6: 31,543,091-31,543,211	F 5' to 3'	TTT**CG**GAAT**CG**GAGTAGGGAG	121
		R 5' to 3'	ACCCTACACCTTCTATCT**CG**ATTTCTT	
		Probe	T**CG**TTTT**CGCG**ATGGAG	
GAPDH	Chr 12: 6,645,449-6,645,570	F 5' to 3'	TTGGGTAGTTTTGGAGTTTTTAGTTG	122
		R 5' to 3'	AATACAACATCTCCTTACCCCCAA	
		Probe	AGTTAGGTTAGTTTGGTAGGGAA	

Table 2 - qMSP Assay Primer & Probe

A table demonstrating the position on the chromosome, forward and reverse primer and probe and product size of the primers and probes for the CpG target region of the TNF gene and GAPDH housekeeping control gene.

3.2 qMSP Primer Matrix

Surfaces & equipment (pipettes, racks, filter tip boxes) were cleaned thoroughly before use with ethanol & DNase.

Reverse primers, forward primers and probes for GAPDH and TNF were kept on ice with a lid to avoid light. Commercially available 100% methylated and 100% un-methylated control-DNA were obtained. The appropriate volume of below reagents for Master Mix and primers and probe was calculated as displayed in Table 1 and mixed and loaded in a 96 well plate of PCR-optical grade plastic. Methylated and unmethylated DNA template or nuclease free water (as negative control) was then pipetted into each wells using the format shown in table 2. The PCR plate was then centrifuged at 700g for 10 minutes, room temperature. The same procedure was then carried out for the forward and reverse primer for GAPDH. then the PCR reactions were carried out on a QS5 qPCR Machine and analysed using QuantStudio.

3.3 qMSP Standard Curve

Surfaces & equipment (pipettes, racks, filter tip boxes) were cleaned thoroughly before use with ethanol & DNase.

Reverse primers, forward primers and probes for GAPDH and TNF were kept on ice with a lid to avoid light. The appropriate volume of reagents for primer mix, master mix and plating were calculated for the TNF and GAPDH forward reverse primer and pipetted into Eppendorf tubes respectively. The plates were then centrifuged at 700g for 10 minutes, room temperature, and then run on a QS5 qPCR Machine and analysed using QuantStudio.

4 Preparing Samples

4.1 Isolation of PBMCs

The samples were analysed from IACON and RADAR came from the department tissue bank and were processed and stored between 2010-2019 by the facility staff.

Phosphate buffer saline (PBS) was made from commercial tablets in 1 litre of water. 30ml of blood was collected from participants into 9ml green heparin tubes. The blood was transferred into a 50ml falcon tube and diluted with 20ml PBS. 15ml Ficoll-Paque solution (Axis Shield) was added to two separate 50ml falcon tubes. The blood/PBS was carefully layered onto 15ml Ficoll-Paque. The falcon tubes were centrifuged at 2400rpm, 20 minutes, no accelerator and no brake, room temperature. The layer of PBMCs was carefully extracted from both 50ml falcon tubes and transferred to a separate 50ml falcon tube. The falcon tube was then topped up to 50ml with PBS for wash and centrifuged at 1800rpm, 10 minutes, brake on, room temperature. The supernatant was discarded, and the pellet was resuspended in 2ml PBS. The solution was then transferred to a 15ml falcon tube and topped up with PBS to 15ml. The 15ml falcon tube was then centrifuged at 1500rpm, 10 minutes, brake on, room temperature. The supernatant was discarded, and the pellet was resuspended in 2ml freezing medium (pure foetal calf serum (FCS), 10% dimethyl sulfoxide (DMSO)). The PBMCs were then transferred over to -150 ^{o}C freezers and stored untouched until used.

4.2 Unfreezing PBMCs

Frozen vials containing PBMCs were suspended in a water bath at 36^{o}C and removed when they almost fully defrosted. 10ml PBS was added to 15ml falcon tubes and defrosted cells were transferred to the falcon tube. Each tube was centrifuged at 1600rpm for 10 minutes, room temperature, brake and acceleration on. The pellet was kept on ice for DNA extraction.

5 qMSP Assay

5.1 DNA Extraction

DNA extraction was performed using the DNA Extraction Kit by Kyogen and the protocol included for the kit was followed and optimised for this process. Surfaces & equipment (pipettes, racks, filter tip boxes) were cleaned thoroughly before use with ethanol & DNase. A heat block was set to 56^{o}C.

200uL PBS was pipetted up and down to resuspend the cell pellets from the above steps. 200uL of each sample was then transferred to a sterile Eppendorf tubes. 20uL of Kyogen Protease was added to each Eppendorf tube. 200ul Lysis Buffer (AL) was also added. Each Eppendorf tube was then mixed by vortex for 15 seconds. The tubes were then incubated in the heat block at 56^{o}C for 10 minutes. The tubes were then centrifuged at full speed 13,000 rpm (machine name and company), 10 seconds, room temperature to remove condensation. 200uL ethanol was pipetted to each tube and this solution was mixed by vortex for 15 seconds. The tubes were then centrifuged at full speed, 10 seconds, room temperature. The samples were then transferred into a column sitting in collection tubes. The collection tubes were then centrifuged at 10,000RPM for 1 minute, room temperature. The columns were transferred to fresh collection tubes. 500ml of Wash Buffer 1 (AW1) was pipetted into the columns. The collection tubes were centrifuged at 10,000RPM for 1 minute, room temperature. The columns were switched to fresh collection tubes. 500ml of Wash Buffer 2 (AW2) was pipetted into the columns and were then centrifuged at 12,000RPM for 3 minutes, room temperature. The columns were then transferred into new sterile Eppendorf tubes. The Eppendorf tubes were then centrifuged at 10,000RPM for 1 minute, room temperature. 100uL Elution Buffer (AE) was added to each of the columns in Eppendorf tubes. The columns were then allowed to soak at room temperature for 5 minutes. The Eppendorf tubes were then centrifuged at 10,000RPM for 1 minute, room temperature. The elute from the Eppendorf were collected (the column discarded) and the DNA concentration in the solution was measured by the Nanodrop1000 and analysed with the software ND1000 Nucleic Acid mode. Nuclease Free Water was used to calibrate the Nanodrop, and Elution Buffer (AE) was used as blank. The ng/uL and 260/280 values for each sample was recorded. DNA was stored at -20oC until used.

5.2 Bisulphite Conversion

Bisulphite conversion was performed using the Bisulphite Conversion Kit by Kyogen and the kit protocol was followed and optimised for this process. Surfaces & equipment (pipettes, racks, filter tip boxes) were cleaned thoroughly before use with ethanol & DNase. DNA samples were retrieved from freezers and left to thaw in the heat block. The volume needed for each sample was calculated from the DNA concentration obtained in the previous step. Complementary volume of nuclease free water was added to Axygen PCR Strips and the correct volume of DNA was transferred.

CT Reagent was made fresh each time, by using CT Reagent powder, 900ul of Nuclease Free Water, 300uL of M-Dilution Buffer and 50uL of M-Dissolving Buffer. The CT Reagent was then added to a shaker to mix for 10 minutes. 130uL CT Reagent was added to each PCR Strip. The PCR strips were then centrifuged at 700g for 5 minutes, room temperature. The PCR Strips were run on the PCR Machine under the setting recommended by Zymo Biconversion. The PCR strips were then centrifuged at 700g for 5 minutes, room temperature. 600uL of M-Binding Buffer was pipetted into IC columns sitting in collection tubes. Each sample from the PCR Strip was added into a column in the collection tube. The IC columns were centrifuged at full speed for 1 minute, room temperature. The columns were then placed into fresh collection tubes. 24ml 100% ethanol was added with 6ml M-Wash Buffer (4:1 ratio) into a Falcon tube. 100uL of the ethanol/M-Wash Buffer was added to each column in the collection tube. The collection tube was then centrifuged at full speed for 1 minute, room temperature. 200uLl of M-Desulphonation Buffer was then added to each column and the solution was incubated at room temperature for 20 minutes. The collection tubes were then centrifuged at full speed for 1 minute, room temperature. The columns were transferred into fresh collection tubes. 200uL of M-Wash Buffer was added to the column. The columns were centrifuged at full speed for 1 minute, room temperature. The columns were placed into sterile Eppendorf tubes and 20uL of M-Elution Buffer was added to the columns. The solution was then incubated at room temperature for 5 minutes. The columns in the Eppendorf tubes were then centrifuged at full speed for 2 minutes, room temperature. The converted DNA solution was measured by the Nanodrop1000 and analysed with the software ND1000 Nucleic Acid mode. Nuclease Free Water was used to calibrate the Nanodrop, and Elution Buffer (AE) was used as blank. The ng/uL and 260/280 values for each sample was recorded. DNA was stored at -20^{o}C until used.

5.3 qMSP

Surfaces & equipment (pipettes, racks, filter tip boxes) were cleaned thoroughly before use with ethanol & DNase. Bisulphite-converted DNA samples were retrieved and left to thaw at room temperature.

0.5uL forward primer and reverse primer for TNF was added to Eppendorf tubes and made up to 100uL with nuclease free water. 3.6uL forward primer and 10.8 reverse primer for GAPDH was added to Eppendorf tubes and both made up to 120uL. A master mix was prepared for both primers, consisting of 104uL diluted forward and reverse primer, 104uL TaqMan probe, 520uL universal master mix and 104uL nuclease free water into an Eppendorf tube. 18uL master mix for TNF and GAPDH was then pipetted into a qPCR plate. 10ng/uL of bisulphite converted DNA was calculated and made up in nuclease free water. 2uL of each sample was then pipetted into a qPCR plate, allowing for triplicates. Finally, methylated and unmethylated DNA, as well as nuclease free water as a control, was pipetted into the qPCR plate. The plates were then centrifuged at 700g for 10 minutes, room temperature, and then run on a QS5 qPCR Machine and analysed using QuantStudio.

Appropriate volume of reagents for primers/probe, master mix and water was calculated according to the final optimised qMSP conditions, as described below.

6 Data Processing

6.1 TNF Gene Methylation Data Processing

qMSP results were exported from the qPCR machine as Ct scores. These scores were then transferred into excel, organised by plate, and the Ct values for each duplicate repeat was averaged and converted into ΔCt scores.

Percentage methylation of the CpG region of the target gene, TNF, and housekeeping control gene, GAPDH, was determined by Ct values measured by the qPCR machine. Each PCR plate was processed through a formula that acted as a calibrator in excel, which was necessary for the results to be reproducible between each experiment.

The formula of the calibrator was as follows:

Percentage of methylation (%)	=	Relative level of methylation x 100
	=	$2^{-\Delta\Delta Ct}$ x 100
Where:		
ΔCt sample	=	Ct_{sample} target gene - Ct_{sample} internal control
ΔCt calibrator	=	$Ct_{calibrator}$ target gene - $Ct_{calibrator}$ internal control
ΔΔCt	=	ΔCt sample - ΔCt calibrator

Sample	=	DNA template from an individual patient after bisulfite conversion.
Calibrator	=	Control DNA template 100% methylated and bisulfite converted.
Target gene	=	Is the gene of interest for the assay (e.g., TNF gene) at the CpG chosen for the assay development with expected change in methylation status.
Internal control	=	Is the GAPDH gene used for normalization from a region independent of any methylation change.

Based upon the primer matrix and dilution series data to analyse the performance of each batch of primers and probes and consistency between the TNF and GAPDH genes, any plate variation of TNF methylation data due to small deviation in pipetting was appropriately adjusted between plates using the calibration corrections.

6.2 Clinical Data Retrieval and Processing

Clinical data was downloaded from the IACON and RADAR cohorts study database, which consisted of participants who had opted to take part in the biomarker sub-study and donated blood for research. The initial extraction of data from these study databases was conducted by Miss Helen Ng (HN) and Miss Chin Liu (CL), two medical students whom I worked alongside to retrieve and process the clinical data from patient records into standardised variables for data analysis. Clinical data, consisting of patient demographics, clinical presentation and serology, was gathered from the database, including hospital visits at baseline 6 and 12 months. The data was anonymised and given a consistent formatting and converted into excel tables.

Demographic variables were patient age, sex (male / female), smoking status (ever / never) and symptom duration. Clinical presentations measured were tender joint count (TJC), swollen joint count (SJC) and a Visual Analogue Scale of General Health (VAS-GH) score. Serological data included C-Reactive protein (concentration), Rheumatoid Factor (RF) (positive or negative) and Anti-Citrullinated Antibody (ACPA) (positive or negative). Time to diagnosis was calculated as difference between 1st visit and the visit at which a diagnosis was recorded. Categorical data (gender, smoking, autoantibodies) were coded into binary options to be utilised by the software IBM SPSS Statistics 27. Continuous data was rounded to 2 decimal places for consistency where relevant. Missing data was marked as -99 and that figure was set for IBM SPSS Statistics 27 to recognise as missing values. Variables with missing data was disregarded in the statistical analysis. If any one component of the DAS28 score, including TJC, SJC, CRP or VAS-GH was missing for a participant, the entire DAS28(4)CRP score was not calculated and also included as missing data.

DAS28(4)CRP was calculated for each patient from the individual components of the compound score, including TJC, SJC, CRP and VAS-GH using the following formula:

$$\text{DAS28(4)CRP} = 0.56 \times + 0.28 \times + 0.36 \times \ln(\text{CRP}+1) + 0.014 \times \text{VAS-GH} + 0.96$$

7. Statistics

7.1 Descriptive Statistics

A test of normality was performed on variables with continuous data and were all found to display a skewed distribution. Therefore, non-parametric tests were used throughout, including the Mann-Whitney U test. The range (including minimum and maximum values) and 25%/75% inter-quartile range was calculated and used in tables alongside the number of participants (n number) and number of missing data points. Categorical data was analysed for significance using the chi squared test.

7.2 Modelling Statistics

Multivariate analysis, combining clinical data (variables for demographics, clinical presentation and serology as outlined earlier) and TNF qMSP data, was conducted. Binary Logistical Regression was carried out to calculate the Odds Ratio (OR) of predicting diagnosis of RA for clinical variables without and then combined with the TNF qMSP data. An area under the receiver operator curve (AUROC) was plotted to visually represent the prediction of each model (univariate and multivariate) and to calculate their area under the curve (AUC). Percentage accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV) and negative predictive value (NPV) for predicting RA diagnosis was calculated.

Chapter 4: Results

1. Candidate CpG Selection

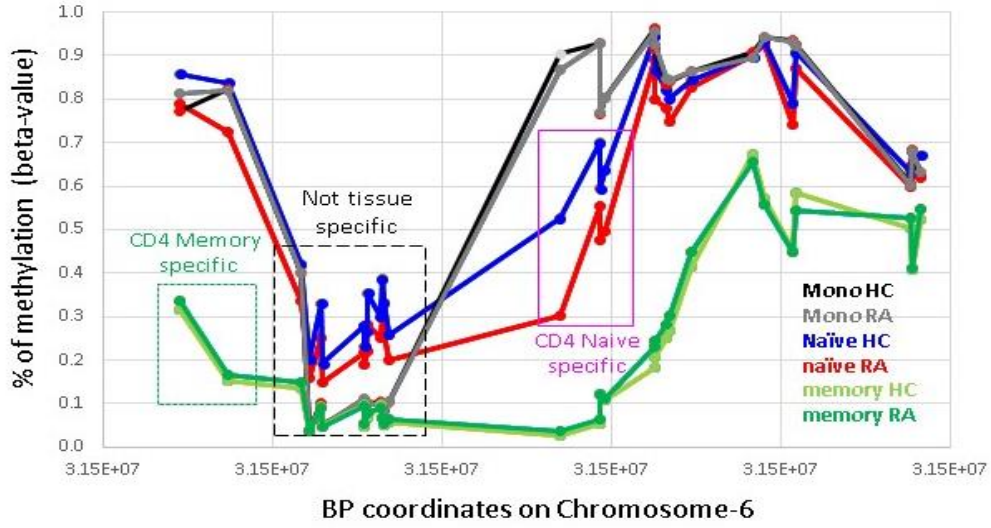
Data was analysed from a number of public datasets utilised by Dr. Rujiraporn Pitaksalee to complement our own dataset[137]. The data involved epigenome-wide methylation of RA patients and healthy controls from a variety of immune cell types, including naïve CD4+T-cells, memory CD4+T-cells and monocytes.

We began to filter the results for potential candidates based upon specific criteria. First, CpG regions were selected that were demethylated in both naïve CD4+T-cells and memory CD4+T-cells, with between a 0% and 50% a methylation β-value, while highly methylated in monocyte, with a β-value that was above 80%. We then filtered the results for CpG regions that displayed significant difference in differential methylation between RA patients and healthy controls, with a significance of $p=\leq 0.0001$. 26 candidates were chosen from this initial selection criteria. The candidates were then further filtered based upon the region surrounding the candidates, based upon methylation activity or demethylation 300-500bp surrounding the region of the candidate. After this filtering process had occurred, the genes that showed most promise were TNF and IFITM1.

Candidates were then selected from our dataset based on extent of the observed differential methylation between RA patients and healthy controls in various sources of sample (PBMCs, whole blood, isolated cells) and across a variety of immune cell types (CD4+T cell populations, B cells, monocytes, NK cells, granulocytes) because we wanted our assay to be adaptable across a number of sample sources and work with a mix of cell populations. The datasets of various sample sources and populations of immune cells were filtered for an $\Delta\beta$-value at least $\geq 10\%$ and a statistical significance of at least $p=\leq 0.01$. After this filtering process, 22 potential CpG candidates were selected. The genes that showed the most promise was HDAC4, IRF8 and MIR21.

With all the preliminary results considered from filtering our dataset and public online datasets for candidates, the gene candidates with the most potential that were selected to trial for the assay were TNF, HDAC, IRF4, IRF8, IFITM1 and MIR21[157].

a) Methylation levels from our dataset



b) Average methylation levels in blood cell types from several datasets (the dotted line represents naïve CD4+T-cells, plain line is memory cells)

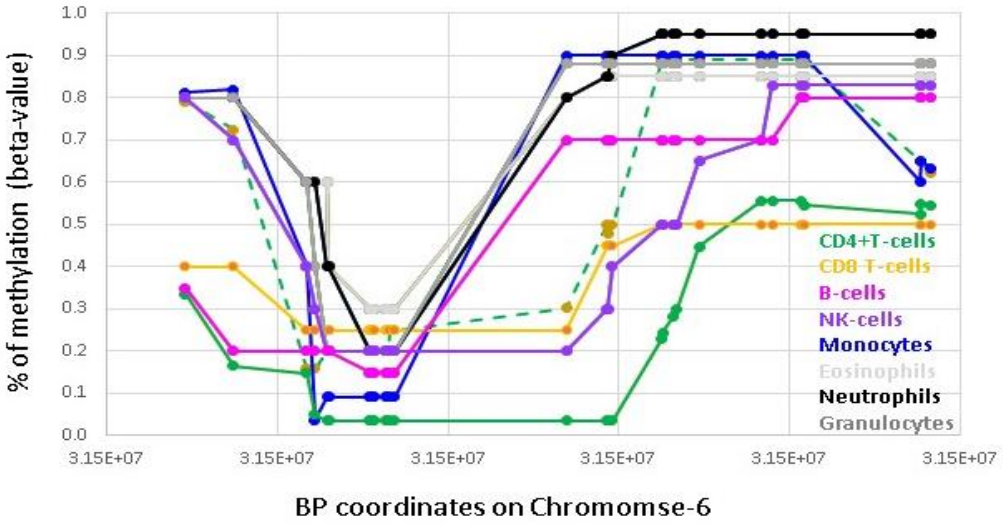


Figure 7 – Identification of the CpG Target

Percentage methylation of the region of DNA involving and surrounding the candidate CpG region of the TNF gene, A) Comparing naïve CD4+T-cells, memory CD4+T-cells and monocytes between RA patients and healthy controls, and B) In other immune cells that form PBMCs and whole blood, such as B cells, NK cells and granulocytes . The target sequence selected displayed differential methylation in naïve T cells between RA patients and healthy controls, with no demethylation in monocytes and full demethylation with no significant difference in methylation levels between RA patients and healthy controls in memory T cells.

2. Primer Design

Primer design was performed by Dr. Rujiraporn Pitaksalee for the candidate gene, TNF, and the housekeeping control gene, GAPDH. Primers were designed to run optimally at 59-60ºC, which is the optimal temperature for conventional qPCR.

Table 2, displayed earlier within the methods section, demonstrates the nucleotide sequence of primer pairs designed for the assay.

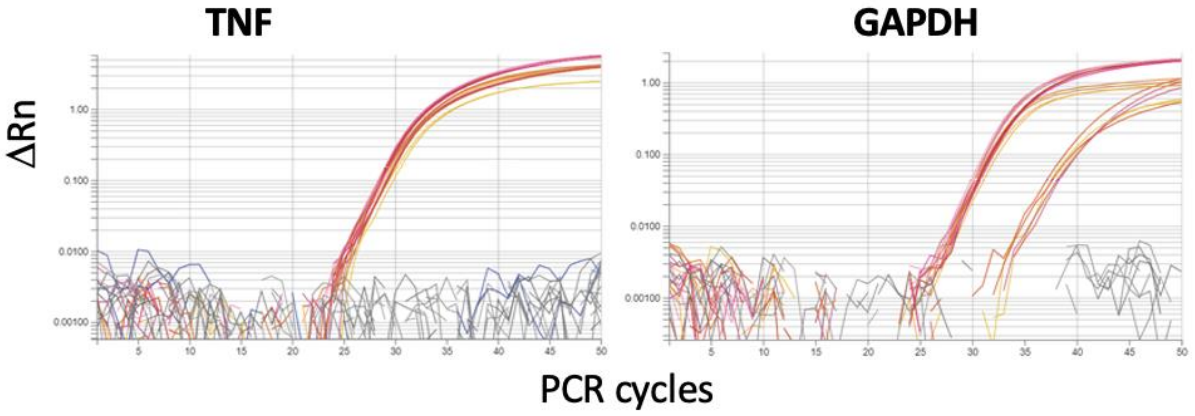


Figure 8** – **Amplification Plot of the TNF and GAPDH Genes

A qMSP assay was designed, using forward and reverse primers at different concentrations (50-900nM) to determine optimal amplification difference of the target genes Ct value compared with the GAPDH housekeeping gene.

There were no primers available for the sequences of interest for MIR21, eliminating this candidate gene of interest.

The designed primers were further optimised to find the best conditions for amplification using 10ng of 100% methylated control DNA. The aim was to find the lowest Ct value for the TNF gene primer that was consistent with the amplification for the housekeeping control gene, GAPDH. The TNF gene candidate showed consistency across all primer concentrations that were used. However, with the housekeeping gene, GAPDH, the reverse primer posed a bottleneck on amplification as the limiting reagent, regardless of the concentration of forward primer used[157].

3. Assay Optimisation

3.1. Primer Matrix

A primer matrix was conducted by Dr. Rujiraporn Pitaksalee to test for the activity of the forward and reverse primers for the gene of interest, TNF, and the housekeeping control gene, GAPDH. The optimum concentration of forward and reverse primer for both genes could then be determined based upon the closest consistency of Ct value between our candidate gene of interest, TNF, and the housekeeping control, GAPDH[157].

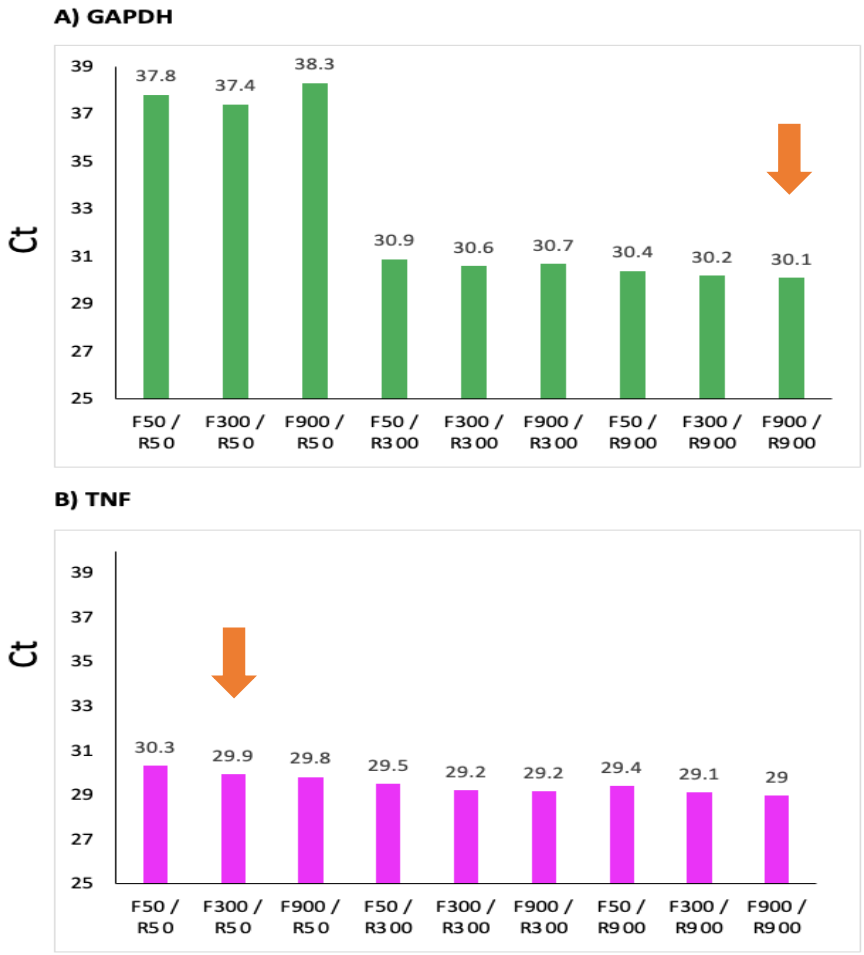


Figure 9 – Primer matrix for GAPDH and TNF Genes

Different combinations of primer concentrations between 50-900nM were tested for ***A)*** *The housekeeping control gene, GAPDH, and* ***B)*** *The candidate target gene, TNF. Primer concentrations of F300/R50 and F900/R900nM were chosen for TNF and GAPDH, as these concentrations displayed the closest consistency of Ct value between both genes.*

3.2. Specificity for Methylation

The candidate gene, TNF, was then tested for specificity to methylated DNA by Dr. Rujiraporn Pitaksalee for 100% unmethylated DNA. In turn, there should be no amplification for unmethylated DNA. The housekeeping gene, GAPDH, is required to be methylation independent to act as a housekeeping control. Specificity for methylation was observed for the candidate genes TNF, HDAC4 and IFITM1, but not for the other candidate genes, IRF4 and IRF8, where some amplification was observed for unmethylated DNA. This eliminated these genes as candidates for the qMSP assay and further validated the selection of the TNF gene candidate for the qMSP assay[157].

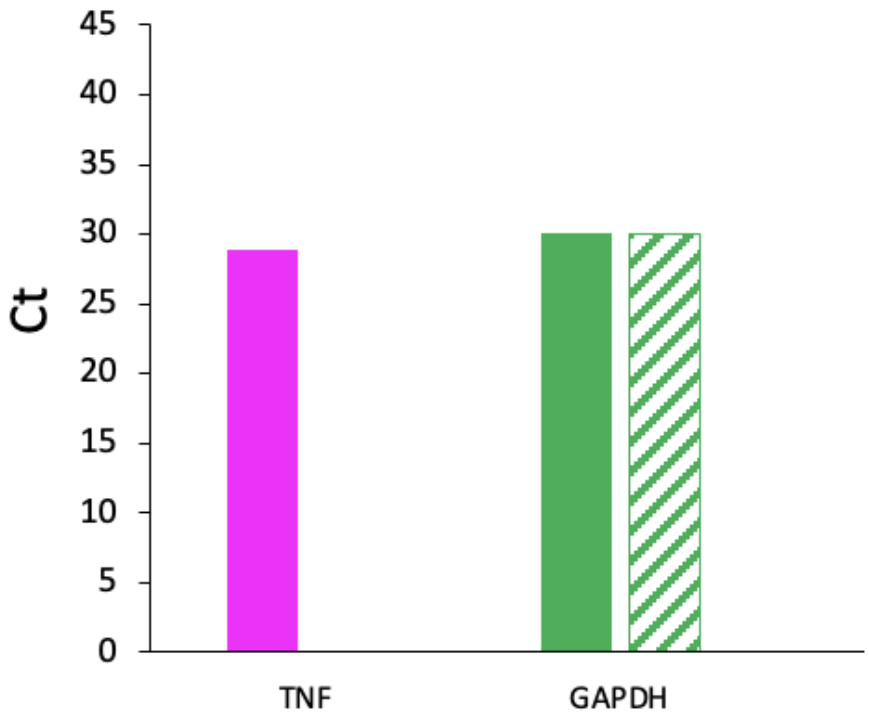


Figure 10 - Assay Specificity for Methylation

Solid bars: *Methylated DNA amplification.* ***Striped bars:*** *Unmethylated DNA amplification. The candidate target gene, TNF, displayed specificity for methylation, where no Ct value was observed using unmethylated DNA. The housekeeping control, GAPDH, displayed activity consistent and independent from methylation, making it an ideal control. The target gene, TNF, and the housekeeping gene, GAPDH, should also be consistent in observed amplification (Ct value). Both genes were investigated using a range of concentrations to establish the most efficient results for the qMSP assay.*

3.3. Dilution Series

A dilution series was conducted by Dr. Rujiraporn Pitaksalee, which tested for the consistency of the forward and reverse primers for the gene of interest, TNF, and housekeeping control, GAPDH. Through the testing of various concentrations of 100% methylated DNA and 100% unmethylated DNA by incremental gradients of concentrations ranging from 0.2ng to 50ng, consistency of the primers could be tested across the various dilutions by plotting a line of best fit. A consistent gradient, with closely matching Ct values between the gene of interest, TNF, and housekeeping gene, GAPDH, ensured consistency in function of the primers regardless of concentration of methylated DNA present and also helped to show that there was lack of impurities in the samples[157].

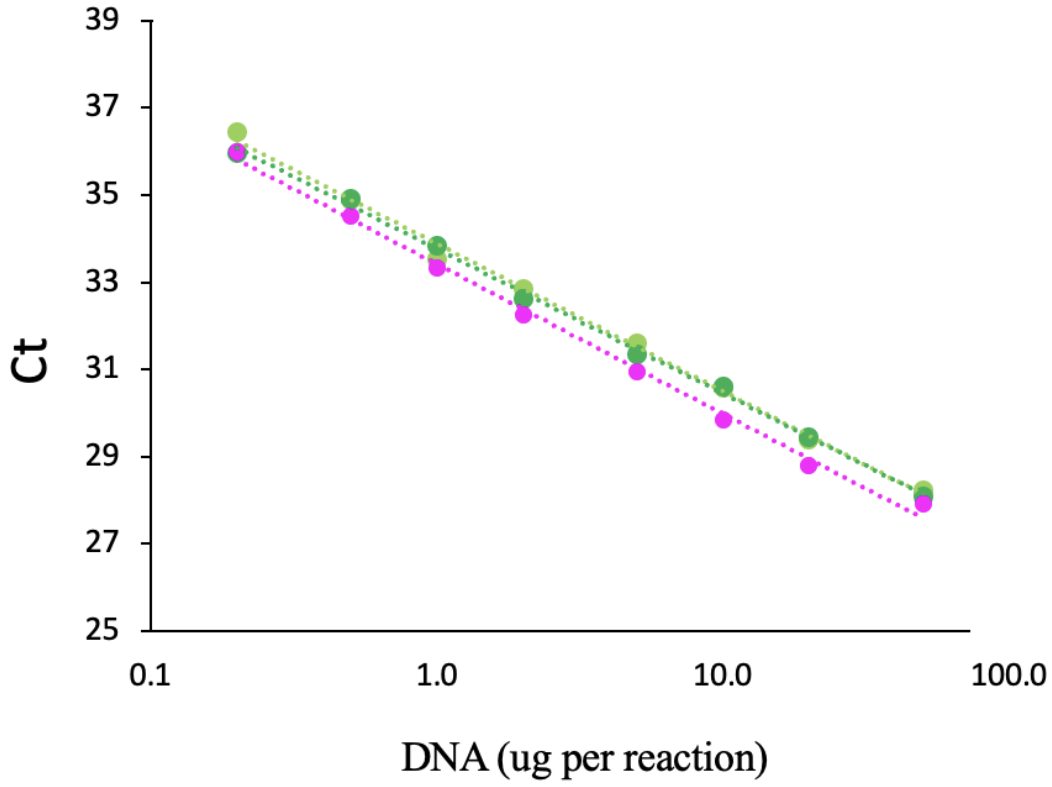


Figure 11 – Dilution Curve for TNF and GAPDH Genes

In green: *The line of best fit for the housekeeping gene, GAPDH, for methylated and unmethylated control DNA.* ***In pink:*** *The line of best fit for the candidate target gene, TNF, for methylated and unmethylated control DNA, where only the 100% methylated control DNA showing amplification. Both genes displayed consistency in the efficiency of amplification across the concentrations of methylated and unmethylated DNA.*

A linear regression model was used for the standard curve to plot and determine qPCR efficiency. The regression was linear for both the candidate gene of interest, TNF, and the housekeeping control gene, GAPDH, over the varying concentrations used, which ensured that the concentration of DNA used in the assay would not be a confounding factor. The assay was between 95.7% to 99.9% efficient, which indicated that there was a genuine 2-fold amplification for both the target gene, TNF, and the housekeeping control gene, GAPDH, for each PCR cycle.

The relative level of methylation for the target gene was calculated by comparing the Ct value of the candidate target gene, TNF, in proportion to the Ct value of the housekeeping control gene, GAPDH, and displayed as a percentage methylation (%). For HDAC4 and IFITM1, these candidate target genes ultimately didn't show meaningful amplitude differences between the RA and non-RA groups, meaning they failed at this stage as candidate genes for the assay.

Ultimately, the TNF gene candidate showed promise along each stage of validation, including a decent display of potential in the preliminary genome-wide association study data, having primers available for design with sequences of interest, performing well in the primer matrix without being a significant limiting factor, having methylation specificity and performing consistently in the dilution series alongside the housekeeping control, GAPDH. The gene also showed promise regarding differential methylation between the RA and healthy control group in the preliminary epigenome-wide associated study (EWAS) and bioinformatic analysis. This data meant that the TNF gene was selected for the target gene for the qMSP assay.

Variation in primer, probe, reagents and pipetting between plates could be a confounding factor that may impact the amplification of the qMSP assay. We therefore added a calibrator control for each plate and also added a reaction mix for 100% methylation DNA and 100% unmethylated DNA to act as positive and negative controls, alongside nuclease free water as a negative control, to ensure both the reproducibility and validity of the qMSP reactions.

With the different aspects of the qMSP assay design now validated for the candidate gene of interest, TNF, in relation with the housekeeping control gene, GAPDH, this qMSP assay could then be adopted on patient samples as a method to compare methylation of the candidate gene of interest TNF, to that of the GAPDH housekeeping control gene[157].

3.4 Source of Samples

As different sources of sample for DNA extraction matter, isolated CD4+T-cells, PBMCs and whole blood were all compared in both RA patients and healthy controls, to determine the optimal source of extracted DNA. This was performed by Dr. Rujiraporn Pitaksalee.

Significant differential methylation was observed between healthy controls and RA patients for isolated CD4+T-cells. A reduction in methylation of 2.6-fold was observed for the RA patients compared with healthy controls. For PBMCs, a reduction in methylation of 1.7-fold was observed for the RA patients compared with healthy controls. For whole blood, there was not a significant difference in differential methylation between patients with RA and healthy controls. The reason for this observed lack of significant difference in differential methylation between the RA patients and the healthy controls in whole blood may be explained by the fact whole blood incorporates a much wider repertoire of cells. CD4+ T-cells only consist of a fraction of total whole blood cells, meanwhile can consist of up to half of the cells that make up PBMCs. This means that the DNA for CD4+T-cells is diluted by the larger populations of cells with different DNA.

Isolated CD4+T-cells and PBMCs were sources of sample that displayed statistically significant differential methylation between patients with RA and healthy control. However, PBMCs were ultimately chosen over isolated CD4+T-cells because they are the least processed source of sample of the three which showed promising statistical significance in differential methylation between patients with RA and healthy controls. Therefore, PBMCs from early, drug-naïve inflammatory arthritis patients were chosen as the source of sample to use for our TNF gene qMSP assay[157].

4. Clinical Data Associations

A total of n=312 PBMC samples from patients registered into the IACON/RADAR study were obtained from our tissue bank. After 2-year follow-up, n=218 patients were classified as RA and n=94 were classified as non-RA

Table 3 describes the demographics and clinical data in both groups.

A)

Variable	Non-RA (n=94)				RA (n=218)				
	Median	Range	25% 75% IQR	Missing	Median	Range	25% 75% IQR	Missing	Mann Whitney U
Age	45	58(20-78)	19	0	58	68(19-87)	21	0	$7.85x10^{-12}$
Symptom Duration	6	146(1-146)	9	1	6	67(1-68)	8	6	0.256
Time to Diagnosis	6	24(0-24)	10	0	0	24(0-24)	0	4	0.000
TJC	3	28(0-28)	6	1	9	27(0-27)	12	0	$5.58x10^{-7}$
SJC	1	20(0-20)	4	1	4.5	50(0-50)	7	0	$5.80x10^{-9}$
CRP	2.5	78(0-78)	12	4	11.7	241(0-241)	26	9	$1.19x10^{-4}$
VAS-GH1	30	100(0-100)	45	5	50	100(0-100)	40	8	0.002117
DAS28(4)CRP	3.36	5.5(1.2-6.7)	1.8	10	4.92	6.9(1.0-7.8)	2.1	13	$1.65x10^{-9}$
% Methylation	6.16	19.78(0.88-20.66)	3.26	0	3.31	11.29(0.27-11.56)	2.02	0	0.000

B)

Variable	Non-RA (n=94)			RA (n=218)			
	Values	%	Missing	Values	%	Missing	X2
Sex (F/M)	65/29	69.15%	0	163/55	74.77%	0	0.304
Smoker (E/N)	42/46	47.73%	6	131/81	61.79%	6	0.025
RF (+/-)	10/82	10.87%	2	110/91	54.73%	17	$1.39x10^{-12}$
ACPA (+/-)	8/85	8.6%	1	123/78	61.20%	17	$3.25x10^{-17}$

Table 3 – Demographics of the Cohort for RA and Non-RA Patients

A) *Variables that displayed continuous data. Number of patients**,** median, range, interquartile range, number of missing data and P value by Mann Whitney U test is displayed.*

B) *Variables that displayed categorical data. Number of patients**,** value of each binary variable, percentage of first variable, number of missing data and P value by Chi Square (χ2) test is displayed.*

Binary variables: Sex (female, male); Smoker (ever smoked, never smoked), RF (positive, negative), ACPA (positive, negative)

Females were found to be more likely than males to have RA in both RA and the non-RA groups with no significant difference in gender skew (RA 74.77%, non-RA 69.15%, X^2 p=0.304). The median age of diagnosis was higher in RA compared with those with other forms of inflammatory arthritis (RA 58 (20-78), non-RA 45 (19-87), MWU $p=7.85x10^{-12}$). Those who had ever smoked were found to be more likely than those never smoked to have RA and this skew was much more pronounced than in the non-RA group with mild statistical significance (RA 61.79%, non-RA 47.73%, X^2 p=0.025). Symptom duration until assessment was roughly comparable between the two groups with no significant difference (RA 6 (1-146), non-RA 6 (1-68), MWU 0.256) but time to diagnosis was much later in the non-RA group (RA 0 months (baseline diagnosis), non-RA 6 months, MWU $p=\leq 0.0001$). Patients with RA had significantly higher ACPA titres in their serum compared with the non-RA group (RA 61.20%, non-RA 8.6%, MWU $p=3.25x10^{-17}$). Patients with RA had significantly higher titres of RF in their serum compared with the non-RA group (RA 54.73%, non-RA 10.87%, MWU $p=1.39x10^{-12}$). Patients with RA scored significantly higher in TJC score compared with the non-RA group (RA 9 (0-27), non-RA 3 (0-28), MWU $p=5.58x10^{-7}$). Patients with RA also scored significantly higher in the SJC score compared with the non-RA group (RA 4.5 (0-50), non-RA 1 (0-20), MWU $p=5.80x10^{-9}$). Patients with RA scored significantly higher in CRP concentrations in the blood compared with the non-RA group (RA 11.7 (0-241), non-RA 2.5 (0-78), MWU $p=1.19x10^{-4}$). Patients with RA scored significantly higher in the VAS-GH score compared with the non-RA group (RA 50 (0-100), non-RA 30 (0-100), MWU p=0.002). Patients with RA scored significantly higher in the overall DAS28(4)CRP compound score when compared with the non-RA group (RA 4.92 (1-7.8), non-RA 3.36 (1.2-6.7), MWU $p=1.65x10^{-9}$).

5. qMSP Assay

5.1 TNF Gene Methylation in Inflammatory Arthritis

Samples were collected and qMSP data acquired over more than 2 years due to the pandemic. Dr. Rujiraporn Pitaksalee performed over a half of the data acquisition and I then took over with Mrs. Rekha Parmer to continue data acquisition. DNA was extracted from the patient samples, bisulfite converted and utilised in the qMSP assay for the target gene, TNF, and housekeeping control gene, GAPDH. Results were plotted as percentage methylation and displayed as box plots and histograms.

Patients with RA had significantly lower percentage methylation (hypomethylation) of the TNF gene compared with the non-RA group (RA 3.31%, 95%CI 3.34- 3.78, non-RA 6.16%, 95%CI 5.95 - 7.33. MWU $p=\leq 4.1x10^{-9}$), confirming that the TNF-qMSP assay has the ability to discriminate patients progressing to RA from other forms of inflammatory arthritis.

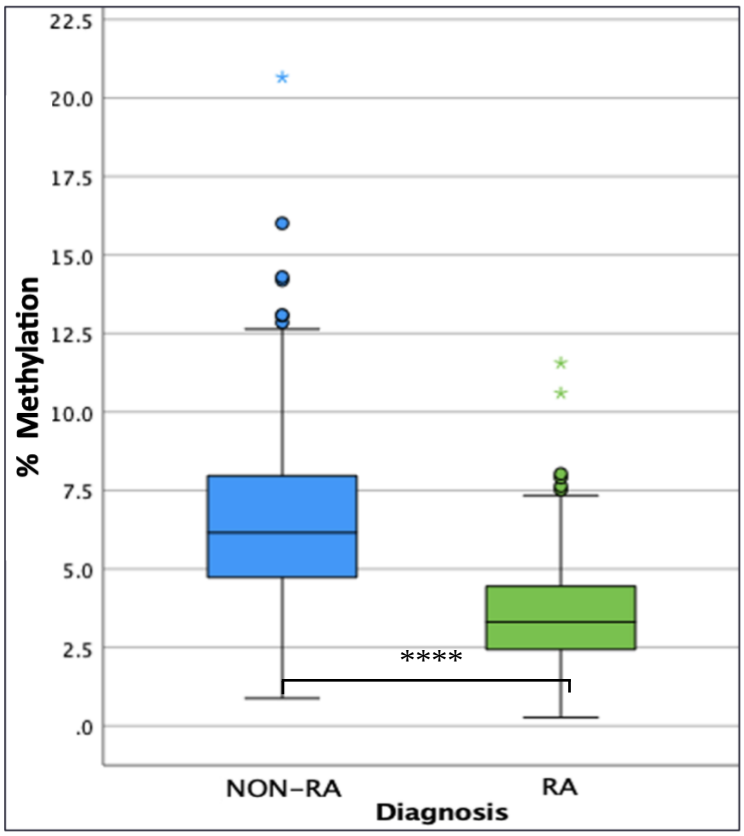
	RA	Non-RA	p-value
DNA methylation of TNF gene (%)	3.13	6.61	$P=4.1x10^{-9}$

Table 4 – Percentage DNA Methylation of the TNF Gene between RA and non-RA Cohort

The mean percentage of TNF gene methylation at the CpG target for the RA and non-RA groups, with difference in percentage TNF methylation reaching statistical significance at the genome-wide level, tested by Mann-Whitney ($P= 4.1x10^{-9}$).

The statistical significance of the site-specific differential percentage methylation of the CpG target at the TNF gene between the RA and non-RA group was found to surpass the threshold of genome-wide significance for epigenome wide association studies (EWAS) and extrapolations to other epigenome-based assay techniques ($P=3.6\times10-8$), as estimated by a recent paper by Saffari et al.[158]. This highlights the ability of this developed qMSP assay in distinguishing site-specific CpG differential methylation in RA towards the exploration of the application of this assay as a clinical biomarker for predicting RA diagnosis and towards potential uses beyond for other modelling parameters and clinical applications.

A)



B)

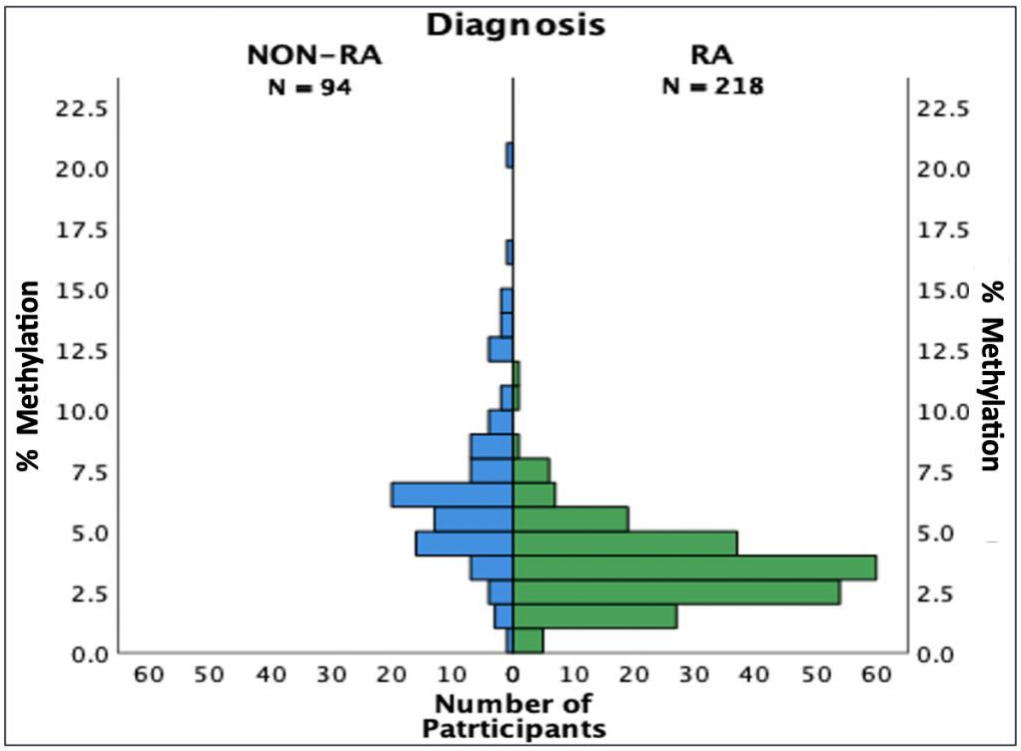


Figure 12 – TNF Gene Methylation between RA and Non-RA Cohort

A) *as a boxplot* ***B)*** *as a histogram. Median percentage TNF gene methylation between RA and non-RA groups were statistically significant by Mann Whitney U ($P=4.1x10^{-9}$).*

5.2. Modelling Statistics: Logistic Regression and AUROC

Once a statistically significant difference by Mann-Whitney U for percentage TNF methylation was observed between RA and the non-RA group ($P=4.1x10^{-9}$), the predictive potential for the TNF gene methylation qMSP assay was explored.

Logistic Regression was first performed and individually, percentage TNF methylation had an odds ratio (OR) of 1.840 (95%CI: 1.567-2.162, p<0.0001) and an AUROC of 0.826 (0.771-0.881) for predicting RA.

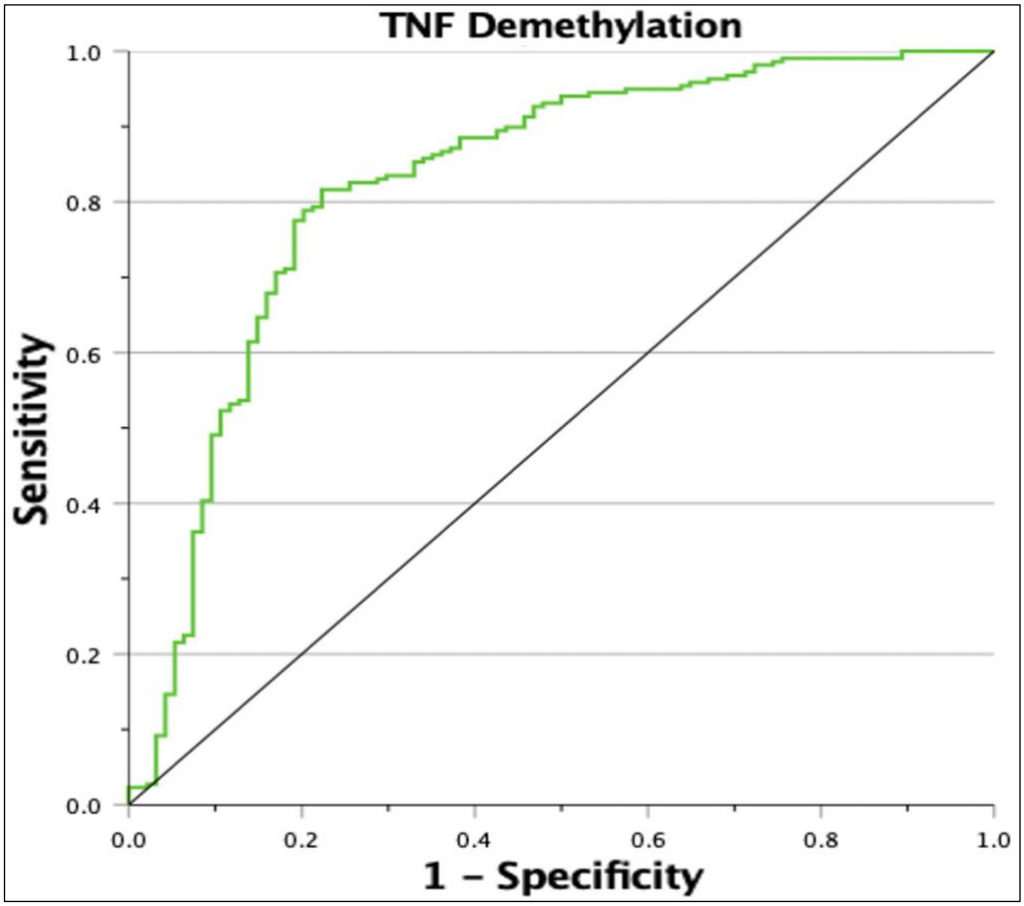


Figure 13 – Receiver Operator Curve of TNF Gene Methylation

Logistic regression was performed on TNF gene demethylation for the RA and non-RA groups towards predicting diagnosis of RA. TNF gene demethylation (extent of hypomethylation) showed high sensitivity and specificity for predicting diagnosis of RA, with an OR of 1.840 (95%CI: 1.567-2.162, p<0.0001) and an AUROC of 0.826 (0.771-0.881) for predicting RA diagnosis.

5.3. TNF Gene Methylation as a Biomarker of Diagnosis

After successfully showing a high sensitivity and specificity for diagnosis of RA with the TNF methylation qMSP assay, the utility of the assay was extended to compare the predictive power of the TNF methylation qMSP assay with the biomarkers that form the current standard of classification of RA according to the EULAR guidance 2010 for classification of RA[2].

Logistic regression was selected using a stepwise forward approach. Initially, a reference model, including all demographic and clinical parameters that form the EULAR classification criteria, was constructed. This notably included first ACPA, Age, SJC, and RF. Diagnosis of RA was predicted with this reference model with an accuracy of 87.8% and AUC=0.950.

Adding the percentage TNF methylation assay results as an additional variable to the reference model changed the priority of most significant variables of prediction of RA diagnosis. The model selected ACPA as the first step and the most associated biomarker with classification. Age was the second step. TNF methylation came third. RF was the fourth step. Swollen joint count the fifth step and lastly, gender was the sixth step.

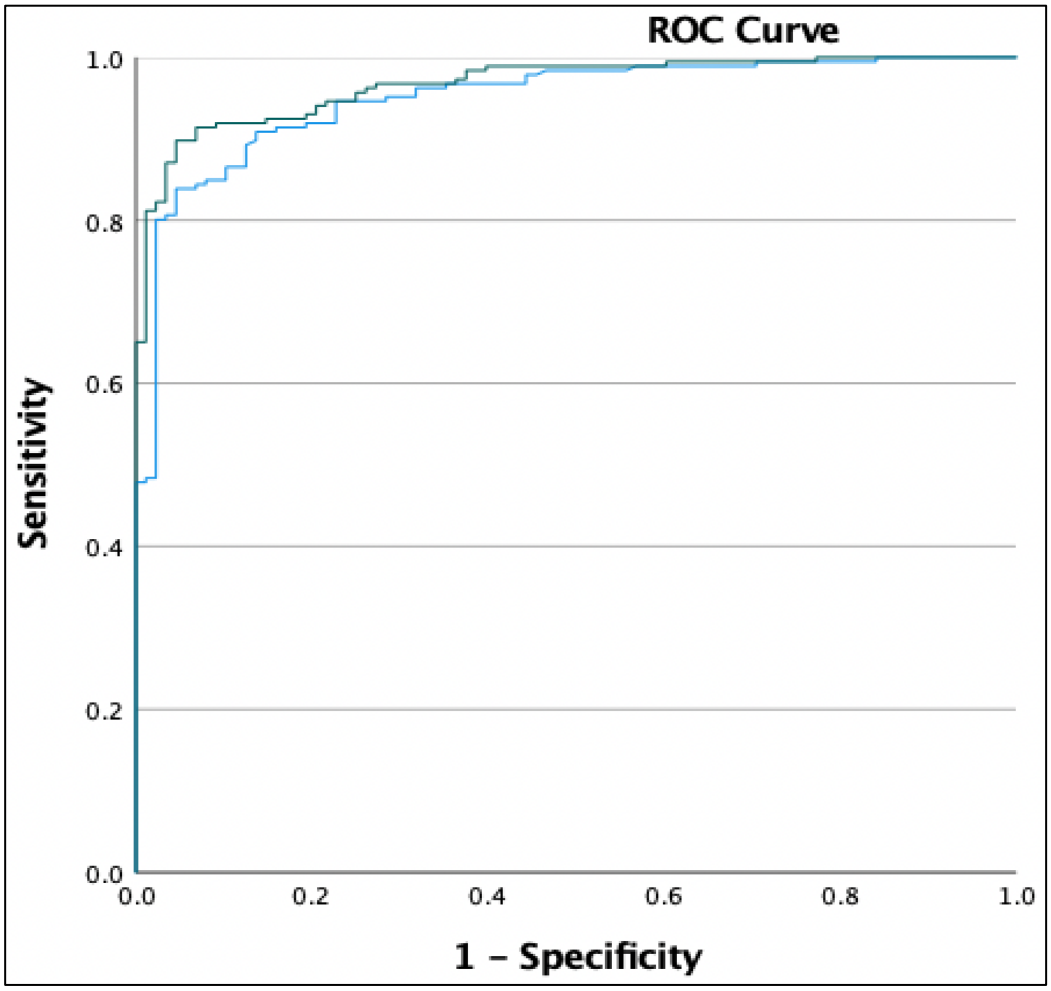


***Figure 14* – Receiver Operator Curve for the Clinical Model**

In blue: The clinical model alone, involving EULAR classification criteria biomarkers.

In green: The clinical model with the addition of the TNF methylation qMSP assay biomarker. Including the TNF methylation assay as a biomarker has an added value of specificity and sensitivity in addition to the established clinical biomarkers.

As mentioned previously, the reference model was able to predict the diagnosis of RA with 87.8% accuracy and an AUC=0.950 *(P=≤0.0001*). When the assay results of percentage TNF methylation was added to the reference model, this increased the accuracy of the model by +1.7% (89.4% total), with an AUC=0.967 (*P=≤0.0001*).

Model	Area under curve	Prediction	Gain in AUC
Clinical	0.950	87.8%	
Clinical + TNF	0.967	89.4%	+1.7%

Table 5 – AUROC and the Accuracy of the Models

AUROC and percentage accuracy towards predicting the diagnosis of RA from patients other non-RA, with the clinical model alone and with the addition of the TNF gene methylation assay. Lastly, the percentage gain in accuracy for prediction of RA from non-RA patents with the addition of the TNF gene methylation variable as an additional biomarker to the reference model is shown on the bottom right.

Chapter 5: Discussion

1. Summary

During my master project, I worked within the team to demonstrate the value of the TNF gene qMSP assay as a biomarker for RA classification. Individually, the TNF gene qMSP assay performed very well with high performance indexes. Combined with clinical data, it provided additional value for the classification of RA and importantly, this was reproduced in ACPA-patients, which are those more in need of a diagnostic biomarker.

2. Optimisation of the Assay

qMSP is a technique based upon polymerase chain reaction (PCR), which can amplify DNA that has been bisulphite converted[132]. Bisulphite sequencing uses sodium bisulphite to convert unmethylated cytosine into uracil. Unmethylated regions can then be distinguished from methylated regions, as methylated cytosines remain unaffected as cytosine[159]. qMSP is cost efficient and easy to perform, with ready-made kits that have established protocols available for purchase for DNA extraction and bisulphite sequencing. As such, it is a technology that could be used in routine hospital laboratories. qMSP is developing rapidly as a technology and may serve as a convenient, cost-efficient assay for testing the differential methylation of gene targets in clinic[132]. This technology could be utilised for clinical testing during assessment (for diagnosis, stratification for therapeutics) and monitoring of patients (progression, remission/flare). Differential methylation of the TNF gene may be a useful application of this technology that could be piloted in routine hospital service, with the potential to distinguish in clinic between pre-RA and other forms of early inflammatory arthritis from clinical samples.

Compared with pyrosequencing, which needs to be run on a specialist machine such as the PyroMark MD pyrosequencer to automate this niche sequencing technique, qMSP can be run on a typical qPCR machine, making the use-case of the qMSP assay far more adaptable in terms of existing infrastructure and training in the clinical setting[131]. Furthermore, the speed and ease of use of qMSP, with an already familiar process, will be most suitable to the clinical environment for rapid acquisition of results for tests[132]. qMSP also has a high sensitivity for results, avoiding false negatives[132].

The design of primers for TNF gene target and GAPDH control was established from data collected by illumina genome-wide methylation sequencing, bioinformatic analysis in R Studio with Bioconductor packages and a search of omics databases. The sequence selected needed to be specific to methylation and demethylation, not other factors that may generate a signal and confound the results, as this would cause false positive results. This was achieved for the TNF gene successfully, although other candidates also developed in the group failed mainly due to the sequence not allowing the design of primers at 60ºC, poor specificity for the methylated and unmethylated DNA target and a major difference in qPCR efficacy between the gene and the GAPDH control, therefore not allowing for a satisfactory design. This highlights the high dependency on local sequence for this type of assay.

3. The Cohorts

Collating a large amount of clinical data from patient records from an early arthritis cohort offered the opportunity to compare demographics and clinical metrics among patients early in disease, eventually diagnosed with RA or other forms of inflammatory arthritis. Patients who were diagnosed with RA were 3x more likely to be female, accounting for 74.77% of my RA cohort, which is comparable to past findings in the literature[160]. The reason for this female bias across IMIDs is uncertain, but hypotheses include hormonal factors, such as oestrogen, having an impacting in disease or X-linked chromosomal factors[4]. The median age of patients was older in those diagnosed with RA compared with other forms of inflammatory arthritis. Biomarkers of RA, such as RF, are associated with ageing, even within the healthy population, so mechanisms of inflammageing may contribute to the increased susceptibility to RA with age[161]. A history of ever smoking had a much greater risk of developing RA than other forms of inflammatory arthritis. This is possibly due to the closely associated role of citrullination in RA and the response of ACPA in disease, which has a high specificity for RA not commonly found in other forms of inflammatory arthritis. Other variables were less different due to the confounding symptoms between RA and other forms of inflammatory arthritis, including CRP TJC and SJC. Despite CRP being utilised as a non-specific biomarker for inflammation, there was clearly a higher titre of CRP associated with RA. Similarly, swollen joint counts were seen with more frequency in RA, whilst tender joint counts were less discriminative between groups. The DAS28 score is a composite score used routinely to monitor inflammatory arthritis, though variation does occur due to differing presentation between these conditions. Here, RA patients

did show a significantly higher DAS28 score than the non-RA group. The DAS28 score was therefore more specific for RA, which is not surprising, as it was developed for RA. All these parameters therefore align this cohort of patients with other similar groups of patients recruited in Leeds and the UK in early arthritis clinics, as well as in other countries with similar clinical set-ups (Netherlands and Sweden notably).

4. TNF Gene Methylation

4.1 TNF Gene Methylation in RA

Patients with RA had hypomethylation of the TNF gene when compared to other forms of inflammatory arthritis. The literature, including a previous study by our group, has shown genome-wide hypomethylation in RA patients compared to healthy controls[137]. The effect of methylation on gene expression varies depending on the region of the gene where the CpGs are (promoter, enhancer or regions with no regulatory role) or on other epigenetic marks, such as histone tails undergoing post-translational modification[126,127]. In the context of hypomethylation of CpG regions associated with the TNF gene, this was shown to affect gene transcription activity. Hypomethylation of the TNF gene suggests wider opening of the chromatin locally, enabling TNF gene expression, therefore resulting in over-expression of the TNF gene and a higher concentration of the TNFα cytokine. TNFα is well-recognised as a prominent inflammatory cytokine in RA pathophysiology, with CD4+ helper T-cells and macrophages being prominent producers of this cytokine[121]. We have shown that early epigenetic changes in RA alter the phenotype of CD4+ naïve T-cells in previous findings[137]. Here, we not only validate findings regarding the role of TNFα early in disease but also a role of epigenetic changes occurring at the TNF gene site. Because the CpG of choice was located in the Naive CD4+T-cell specific regulatory region, it may be hypothesised that the differences in methylation levels observed are reflecting signal coming mainly from naive and memory CD4+T-cells, whilst not fully excluding contribution from CD8+ T-cells and possibly NK cells, although those cells did not show differential methylation in publicly available data comparing HC and RA. Immune dysregulation, involving TNFα, has been known in patients with clinically established RA, whilst within very early RA, increase in circulating levels of protein in serum were observed but not predictive of diagnosis, maybe as coming from multiple

sources. This confirmed the specificity of the DNA-methylation changes in CD4+T-cells compared to other cell subsets.

TNFα is known to be both secreted by and interact with T cells and macrophages, which then upregulates genes that produce further inflammatory cytokines, such as IL1, IL6, IL17, IL23 and IFNγ, promoting further inflammation and cytokine dysregulation that contributes to synovitis and joint destruction in the joints of RA patients[120,121]. In the preliminary results of our lab regarding the epigenome-wide association study (EWAS) and bioinformatic analysis of early, drug naïve RA patients, the genes for IL6 and various IFN signalling genes showed promising differential methylation between RA and healthy controls[137]. These findings are important considering the temporal changes that may occur to cytokine dysregulation and immune pathophysiology across the IAC, depending on stage of disease (early, established, chronicity[121].

Due to the importance of IL6 and IFN known in disease pathophysiology of RA as mentioned, and the promising results found regarding the genes related to activity of these cytokines in our preliminary study, potential future direction for the project would be to explore the differential methylation of CpG sites of genes associated with IL6 and IFN signalling in a similar cohort of early, drug naïve RA patients to further test our qMSP assay and see whether these differential methylation of CpG targets of these genes could also be promising as clinical biomarkers of prediction for diagnosis, prognosis (predicting flare or remission) and treatment response (e.g., Methotrexate).

4.2 TNF Gene Methylation as a Biomarker

TNF gene methylation state in early RA was explored in this project to see whether it could be utilised as a clinical biomarker that can predict diagnosis of RA compared with other forms of inflammatory arthritis and observing whether it can compare clinically with the current EULAR 2010 criteria of classification, offering translational findings from bench to bedside[2]. We compared the ability to predict diagnosis of RA or non-RA against the current characteristics used in clinic, which involves autoantibody profile (ACPA, RF) inflammatory markers (CRP), duration of symptoms, clinical assessment (joint counts) and demographics such as age, gender and smoking status. Using binary logistic regression, I found that an

improvement of the performance index of the predictive model was seen for both accuracy and AUC. This demonstrates a potential clinical utility for TNF gene methylation as a biomarker for diagnosis in addition to current practices, using clinical characteristics. However, these results should be repeated and validated in additional cohorts. Epigenetic changes to the TNF gene by CpG methylation is one of the changes which occur in T-cells in RA. Elucidating how this occurs and at which particular time point across the IAC needs to be established in further studies to better understand the disease.

This also demonstrates the importance of epigenetic change early in disease. IL6 and IFN signalling may also play a predominant role in pathophysiology early in disease and both showed differential DNA-methylation in our original work in naïve CD4+T-cells. The efficacy and reliance upon anti-TNFα biologics in particular, but also IL6 inhibitor Tocilizumab and increasingly Jakis, show a vital role of these cytokine in established clinical RA, as outlined in the EULAR guidance to therapeutic treatment 2022[16]. The particular efficacy of anti-IL6 in early RA is still under investigation but would fit with our previous data putting this cytokine as central and possibly upstream of the change in TNF biology. It is possible that these events occur before clinical diagnosis and that with the genetic prevalence of T-cell related genes associated with disease, T-cells may be playing a critical role genetically and epigenetically, establishing cytokine dysregulation very early in disease. This potential divergence in the prominence of cytokine dysregulation early in disease between RA and other forms of inflammatory arthritis may serve to distinguish disease at the most pertinent stage to treat patients, before prominent pathology. Biomarkers that can be utilised in clinic for assessments and monitoring could be utilised to give accurate diagnosis and tailored intervention strategies.

Here, we have shown that the TNF gene is differentially methylated by statistically significant reduction of methylation levels (hypomethylation) compared to other forms of inflammatory arthritis, and particularly in ACPA- patients in need of novel biomarker. This difference can distinguish patients developing RA from other examples of overlapping presentations and therefore could be utilised as a tool for clinicians to support in separating the overlap between these pathologically related conditions. However, further studies will need to validate these findings, before this can be adopted widely.

ACPA is currently the most relied upon biomarker utilised in the diagnosis of RA, as it displays higher specificity for RA compared to RF and other biomarkers currently used in clinic[162]. As

the importance of mechanisms of citrullination in RA and consequential humoral autoimmune reaction to citrullinated proteins is unique to RA when compared with other forms of inflammatory arthritis, this affords a degree of specificity for ACPA. However, the sensitivity of ACPA for RA (~40-50%), whilst better performing than most markers and other pathological features, is still far from desirable[162]. This is evidenced by the presence of ACPA- (while possibly RF+ or anti-Carp+) individuals who never develop RA[145]. Furthermore, a large proportion of ACPA+ pre-RA individuals with arthralgia never develop RA and my group also showed recently that many (up to 10%) of OA patients can be ACPA+[91,93]. Furthermore, approximately one third of all patients eventually diagnosed with RA (by scoring highly on other metrics of the EULAR guidelines for diagnostic criteria) remain ACPA- (and again, often RF-)[163]. These patients desperately need a biomarker that can indicate disease early in the progression of RA because the over-reliance of ACPA for the diagnosis of RA and the lack of confidence in diagnosing ACPA- patients with RA leads to significantly delayed diagnosis. This delays therapeutic intervention and time to diagnosis has shown to be associated with worse disease outcomes. Here, we have shown the potential utilisation of a biomarker early in disease that can be utilised for RA patients, including those ACPA-. TNF gene hypomethylation was shown to occur in both ACPA+ and ACPA- patients, offering a potential biomarker which could be useful across these two endotypes.

5. Limitations

5.1 Limitations of Clinical Data

When the clinical data was compiled, some patient records did not contain all the required information for all variables collected, such as certain tests not conducted at patient visits. This included some patients not being asked their smoking status (if ever smoked) and did not have certain tests performed, such as for CRP concentrations, TJC, SJC or VAS-GH. This meant that missing data was apparent for a number of variables which may confound the output of the overall results or underestimate the significance. Though only a small number of variables contained any missing data in this study, this occurred most frequently with DAS28, due to the reliance of this composite score on all four variables of TJC, SJC, CRP and VAS-GH in the scores calculated. In these instances, when even one metric was missing, the DAS28 calculation had to be disregarded for that patient. In the future, we would hope there is more consistency in the data collected during patient triage and that the attainment of all data necessary for the project could be acquirable from patient records.

5.2 Technical Limitations of the Assay

The project contained a large cohort of participants, where bloods were collected, DNA extracted and qMSP performed over the course of a few years. Primer reagent that was made for each batch had slight alterations in performance, which risked confounding the results between batches of the project. This was mitigated by running a primer matrix and dilution series to analyse the performance of each batch of primers/probes and test for consistency between TNF and GAPDH. Any plate variation due to small deviation in pipetting was appropriately adjusted between plates using the calibration corrections. Nevertheless, this was an expected technical limitation of the technique when it was developed that could be circumvented entirely in subsequent iterations of a qMSP assay design.

6. Future Directions

6.1 Expanding Cohorts

TNF gene methylation could have a role clinically as a biomarker to distinguish patients with RA from other forms of inflammatory arthritis, but the results need to be repeated in a larger cohort to increase validity, especially from other countries for international validation to factor in different practices and demographics. This is important to establish that the assay can be clinically applicable worldwide. Validation of this data is also a necessary step before any work can be adopted worldwide. As such, early arthritis clinics similar to the one set up in Leeds exist in the UK and in other EU countries, notably Sweden and Netherlands, while in others, the organisation of care is different and prevent such work, for example in France, where patients are only seen at hospital very late in their disease course, when needing biologics. Any biomarker research programme is highly dependent on the selection criteria for the population studies and as such, it is very important to do this in early IA patients developing RA or other form of inflammatory arthritis and not comparing RA to health, as is done in many biomarker studies. The characteristics of the patients are also important and again, early IA needs to be defined as patients having IA, with a symptoms of less than 24 months of duration, drug naïve notably with respect to DMARDs. Oral Steroids may modify the biology and while routinely used in some countries (Netherlands notably), in others steroid are used locally (intramuscular injection, UK), this may also be an important confounding factor to consider.

At other points across the IAC, it would be important to clarify with greater precision the point in pre-RA disease progression this epigenetic modelling seems to occur. A cohort is being developed from those attending the CCP clinic, where participants are being gathered and bloods taken to be utilised to further expand on the results developed so far by our research team. By understanding the chronological events of immune dysregulation that develops in early RA, this may help us understand the pathogenesis of the disease and how we could utilise this in clinical practice as biomarkers and tackle the disease early with therapeutics and intervention. Going forward from here, it would be useful to establish which underlying mechanisms encourage such epimutations to occur and at which specific stage of the IAC this change begins in naïve CD4+ T cells. Therefore, carrying out similar studies from cohorts of patients identified earlier in disease progression will be an important direction for this work.

Those who attended the early arthritis clinic (EAC) had treatment was initiated and it would be important to establish whether this assay could provide a stratification biomarker for RA patients able to achieve remission with the 1st line recommended drug, methotrexate, whilst others should benefit more from biologics. The potential of TNF gene methylation by qMSP as a biomarker could be explored for the prediction of other outcomes. For example, likelihood of response to methotrexate at 6 months and 12 months of prescription would be the 1st outcome to investigate, where response to the therapeutic could be categorised as remission (if successful remission had been achieved as defined by EULAR guidelines) or non-remission (even if low disease activity had been achieved). The TNF gene methylation biomarker could therefore be tested utilising the qMSP assay to see if there is significant difference regarding TNF gene methylation and methotrexate response.

6.2 Future Utilisation of the Assay

The methylation of CpGs at other genes of interest identified from the genome-wide methylation sequencing, bioinformatic analysis and searching of omics databases, outlined in the pilot data, need to be further investigated. For example, differential methylation of the IL17A gene was observed and my group has established that a qMSP assay for this gene also has diagnostic value in EAC patients and for MTX response[94]. in contrast, the differential methylation of IFN related genes, which was associated with RA compared to healthy controls in the functional analysis of inflammatory nodes from a STING network, also suggested a prominent role of IFN-signalling. The methylation status of genes that code for STAT proteins, IFN proteins and related factors could therefore be an area of investigation. However, IFN-gene expression signature showed limited value in early RA (and none in established RA) as biomarker for any outcome, while a study suggested a possible role of IFN gene expression in arthralgia or pre-RA[164]. This suggests a precise timing of multiple events that needs further elucidation in pre-RA, early RA and for personalised medicine.

7. Conclusion

In conclusion, rheumatoid arthritis is a heterogeneous disease with complex, aetiology. The IAC model has been established to highlight distinct phases of disease progression, from at-risk individuals to patients with clinically established disease[60]. Current research is focussed on developing an understanding of RA pathogenesis and pathophysiology earlier along the IAC. This is particularly useful for developing biomarkers that have translational benefit from bench to bedside, such as predicting diagnosis sooner in the pre-RA disease course, predicting severity and pathological presentation of disease, predicting sustained remission and flare and also predicting successful therapeutic treatments. Pilot data utilising in-silico analysis and genome-wide methylation has revealed differential methylation of CpGs in early RA, which could also be apparent earlier in the IAC and needs further validation. This pilot data highlighted suspect genes, both in terms of differential methylation and also functional protein analysis using the STING database. Key genes implicated were those relating to IL6, TNFα, IL17 and IFN signalling. The TNF gene was selected for further analysis. In this thesis, a qMSP assay was developed to detect the methylation of the TNF gene in early RA patients obtained from the Early Arthritis Clinic, including the IACON and RADAR cohorts. We have shown a greater extent of hypomethylation of the TNF gene in patients with RA compared with other forms of inflammatory arthritis , especially early in disease. The predictive power of TNF gene methylation was then analysed as a biomarker to predict RA diagnosis and compared with the current standard of classification used in clinic, defined by the EULAR 2010 criteria for classification[2]. I have shown an even greater predictive value within this cohort when all metrics were combined with the TNF-qMSP data than with the current EULAR 2010 classification criteria alone.

Going forward from here, we need to establish where along the IAC TNF gene hypomethylation occurs and by which mechanisms this change occurs. In doing so, we could understand the early pathophysiology of RA and further establish the predictive potential of TNF gene methylation as a biomarker, by identifying if these epimutations occur earlier in the disease course for naïve T cells. This assay and the use of TNF gene methylation as a biomarker could therefore have the potential to diagnose patients earlier in the disease course and allow for earlier clinical intervention. In doing so, we may improve the prognosis of patients with RA and limit the extent of cumulative disability that develop due to the sustained progression of disease.

Reference List

1 Giannini D, Antonucci M, Petrelli F, Bilia S, Alunno A, Puxeddu I. One year in review 2020: pathogenesis of rheumatoid arthritis. Clin Exp Rheumatol. 2020;38:387-397.

2 Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham Iii CO, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD, *et al.* 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis & Rheumatism. 2010;62:2569-2581.

3 Alamanos Y, Voulgari PV, Drosos AA. Incidence and Prevalence of Rheumatoid Arthritis, Based on the 1987 American College of Rheumatology Criteria: A Systematic Review. Seminars in Arthritis and Rheumatism. 2006;36:182-188.

4 Gerosa M, De Angelis V, Riboldi P, Meroni PL. Rheumatoid Arthritis: A Female Challenge. Women's Health. 2008;4:195-201.

5 Myasoedova E, Davis J, Matteson EL, Crowson CS. Is the epidemiology of rheumatoid arthritis changing? Results from a population-based incidence study, 1985-2014. Ann Rheum Dis. 2020;79:440-444.

6 Grassi W, De Angelis R, Lamanna G, Cervini C. The clinical features of rheumatoid arthritis. European Journal of Radiology. 1998;27:S18-S24.

7 Schett G, Gravallese E. Bone erosion in rheumatoid arthritis: mechanisms, diagnosis and treatment. Nature Reviews Rheumatology. 2012;8:656-664.

8 Colebatch AN, Edwards CJ, Østergaard M, van der Heijde D, Balint PV, Agostino M-A, Forslind K, Grassi W, Haavardsholm EA, Haugeberg G, *et al.* EULAR recommendations for the use of imaging of the joints in the clinical management of rheumatoid arthritis. Annals of the Rheumatic Diseases. 2013;72:804.

9 Harnden K, Di Matteo A, Mankia K. When and how should we use imaging in individuals at risk of rheumatoid arthritis? Frontiers in Medicine. 2022;9.

10 Ziff M. The rheumatoid nodule. Arthritis & Rheumatism. 1990;33:761-767.

11 Kadura S, Raghu G. Rheumatoid arthritis-interstitial lung disease: manifestations and current concepts in pathogenesis and management. Eur Respir Rev. 2021;30:210011.

12 England BR, Thiele GM, Anderson DR, Mikuls TR. Increased cardiovascular risk in rheumatoid arthritis: mechanisms and implications. BMJ. 2018;361:k1036-k1036.

13 Owlia MB, Newman K, Akhtari M. Felty's Syndrome, Insights and Updates. Open Rheumatol J. 2014;8:129-136.

14 van Riel PL, Renskers L. The Disease Activity Score (DAS) and the Disease Activity Score using 28 joint counts (DAS28) in the management of rheumatoid arthritis. Clin Exp Rheumatol. 2016;34:S40-s44.

15 Garaffoni C, Adinolfi A, Bortoluzzi A, Filippou G, Giollo A, Sakellariou G, Sirotti S, Ughi N, Scirè CA, Silvagni E. Novel insights into the management of rheumatoid arthritis: one year in review 2022. Clin Exp Rheumatol. 2022;40:1247-1257.

16 Smolen JS, Landewé RBM, Bergstra SA, Kerschbaumer A, Sepriano A, Aletaha D, Caporali R, Edwards CJ, Hyrich KL, Pope JE, *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Ann Rheum Dis. 2023;82:3-18.

17 Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, Kvien TK, Navarro-Compán MV, Oliver S, Schoels M, *et al.* Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. Annals of the rheumatic diseases. 2016;75:3-15.

18 Joensuu JT, Huoponen S, Aaltonen KJ, Konttinen YT, Nordström D, Blom M. The cost-effectiveness of biologics for the treatment of rheumatoid arthritis: a systematic review. PLoS One. 2015;10:e0119683-e0119683.

19 Huang J, Fu X, Chen X, Li Z, Huang Y, Liang C. Promising Therapeutic Targets for Treatment of Rheumatoid Arthritis. Front Immunol. 2021;12:686155.

20 Morinobu A. JAK inhibitors for the treatment of rheumatoid arthritis. Immunol Med. 2020;43:148-155.

21 Nagy G, Roodenrijs NMT, Welsing PMJ, Kedves M, Hamar A, van der Goes MC, Kent A, Bakkers M, Pchelnikova P, Blaas E, *et al.* EULAR points to consider for the management of difficult-to-treat rheumatoid arthritis. Ann Rheum Dis. 2022;81:20-33.

22 Buch MH, Eyre S, McGonagle D. Persistent inflammatory and non-inflammatory mechanisms in refractory rheumatoid arthritis. Nature Reviews Rheumatology. 2021;17:17-33.

23 van der Helm-vanMil AHM, le Cessie S, van Dongen H, Breedveld FC, Toes REM, Huizinga TWJ. A prediction rule for disease outcome in patients with Recent-onset undifferentiated arthritis: How to guide individual treatment decisions. Arthritis & Rheumatism. 2007;56:433-440.

24 Jain A, Joseph S, James J, James TS, Kumar K, Raza K, Greenfield S, Shenoy P. Delay in diagnosis of rheumatoid arthritis: reasons and trends over a decade. Rheumatology International. 2023;43:503-508.

25 Aletaha D, Smolen JS. Diagnosis and Management of Rheumatoid Arthritis: A Review. JAMA. 2018;320:1360-1372.

26 Matcham F, Scott IC, Rayner L, Hotopf M, Kingsley GH, Norton S, Scott DL, Steer S. The impact of rheumatoid arthritis on quality-of-life assessed using the SF-36: A systematic review and meta-analysis. Seminars in Arthritis and Rheumatism. 2014;44:123-130.

27 Scott DL, Smith C, Kingsley G. What are the consequences of early rheumatoid arthritis for the individual? Best Practice & Research Clinical Rheumatology. 2005;19:117-136.

28 Simon TA, Thompson A, Gandhi KK, Hochberg MC, Suissa S. Incidence of malignancy in adult patients with rheumatoid arthritis: a meta-analysis. Arthritis Research & Therapy. 2015;17:212.

29 Antin-Ozerkis D, Evans J, Rubinowitz A, Homer RJ, Matthay RA. Pulmonary Manifestations of Rheumatoid Arthritis. Clinics in Chest Medicine. 2010;31:451-478.

30 Karami J, Aslani S, Jamshidi A, Garshasbi M, Mahmoudi M. Genetic implications in the pathogenesis of rheumatoid arthritis; an updated review. Gene. 2019;702:8-16.

31 Kuek A, Hazleman BL, Östör AJK. Immune-mediated inflammatory diseases (IMIDs) and biologic therapy: a medical revolution. Postgraduate Medical Journal. 2007;83:251.

32 MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, Silman AJ. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis & Rheumatism. 2000;43:30-37.

33 Viatte S, Barton A. Genetics of rheumatoid arthritis susceptibility, severity, and treatment response. Semin Immunopathol. 2017;39:395-408.

34 de Almeida DE, Ling S, Holoshitz J. New insights into the functional role of the rheumatoid arthritis shared epitope. FEBS Lett. 2011;585:3619-3626.

35 Raychaudhuri S, Thomson BP, Remmers EF, Eyre S, Hinks A, Guiducci C, Catanese JJ, Xie G, Stahl EA, Chen R, *et al.* Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Nature Genetics. 2009;41:1313-1318.

36 van der Linden MPM, Feitsma AL, le Cessie S, Kern M, Olsson LM, Raychaudhuri S, Begovich AB, Chang M, Catanese JJ, Kurreeman FAS, *et al.* Association of a single-nucleotide polymorphism in CD40 with the rate of joint destruction in rheumatoid arthritis. Arthritis & Rheumatism. 2009;60:2242-2247.

37 Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, Wolfe F, Kastner DL, Alfredsson L, Altshuler D, *et al.* Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. Am J Hum Genet. 2005;77:1044-1060.

38 Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM, Spoerke JM, *et al.* A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet. 2004;75:330-337.

39 Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, de Bakker PI, Le JM, Lee HS, Batliwalla F, *et al.* STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med. 2007;357:977-986.

40 Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, Suzuki M, Nagasaki M, Nakayama-Hamada M, Kawaida R, Ono M, *et al.* Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nature Genetics. 2003;34:395-402.

41 Curran AM, Naik P, Giles JT, Darrah E. PAD enzymes in rheumatoid arthritis: pathogenic effectors and autoimmune targets. Nature Reviews Rheumatology. 2020;16:301-315.

42 Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, Liew A, Khalili H, Chandrasekaran A, Davies LRL, *et al.* TRAF1–C5 as a Risk Locus for Rheumatoid Arthritis — A Genomewide Study. New England Journal of Medicine. 2007;357:1199-1209.

43 de Hair MJ, Landewé RB, van de Sande MG, van Schaardenburg D, van Baarsen LG, Gerlag DM, Tak PP. Smoking and overweight determine the likelihood of developing rheumatoid arthritis. Ann Rheum Dis. 2013;72:1654-1658.

44 Rooney CM, Mankia K, Emery P. The Role of the Microbiome in Driving RA-Related Autoimmunity. Frontiers in Cell and Developmental Biology. 2020;8.

45 Rombouts Y, Willemze A, van Beers JJBC, Shi J, Kerkman PF, van Toorn L, Janssen GMC, Zaldumbide A, Hoeben RC, Pruijn GJM, *et al.* Extensive glycosylation of ACPA-IgG variable domains modulates binding to citrullinated antigens in rheumatoid arthritis. Annals of the Rheumatic Diseases. 2016;75:578.

46 Alghamdi M, Alasmari D, Assiri A, Mattar E, Aljaddawi AA, Alattas SG, Redwan EM. An Overview of the Intrinsic Role of Citrullination in Autoimmune Disorders. J Immunol Res. 2019;2019:7592851.

47 Klareskog L, Stolt P, Lundberg K, Källberg H, Bengtsson C, Grunewald J, Rönnelid J, Erlandsson Harris H, Ulfgren A-K, Rantapää-Dahlqvist S, *et al.* A new model for an etiology of rheumatoid arthritis: Smoking may trigger HLA–DR (shared epitope)–restricted immune reactions to autoantigens modified by citrullination. Arthritis & Rheumatism. 2006;54:38-46.

48 Mahdi H, Fisher BA, Källberg H, Plant D, Malmström V, Rönnelid J, Charles P, Ding B, Alfredsson L, Padyukov L, *et al.* Specific interaction between genotype, smoking and autoimmunity to citrullinated α-enolase in the etiology of rheumatoid arthritis. Nature Genetics. 2009;41:1319-1324.

49 Wegner N, Lundberg K, Kinloch A, Fisher B, Malmström V, Feldmann M, Venables PJ. Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. Immunological Reviews. 2010;233:34-54.

50 Cheng Z, Meade J, Mankia K, Emery P, Devine DA. Periodontal disease and periodontal bacteria as triggers for rheumatoid arthritis. Best Practice & Research Clinical Rheumatology. 2017;31:19-30.

51 Belvoncikova P, Maronek M, Gardlik R. Gut Dysbiosis and Fecal Microbiota Transplantation in Autoimmune Diseases. Int J Mol Sci. 2022;23.

52 Wang L, Zhu L, Qin S. Gut Microbiota Modulation on Intestinal Mucosal Adaptive Immunity. J Immunol Res. 2019;2019:4735040.

53 Alpizar-Rodriguez D, Lesker TR, Gronow A, Gilbert B, Raemy E, Lamacchia C, Gabay C, Finckh A, Strowig T. Prevotella copri in individuals at risk for rheumatoid arthritis. Annals of the rheumatic diseases. 2019;78:590-593.

54 Lundberg K, Kinloch A, Fisher BA, Wegner N, Wait R, Charles P, Mikuls TR, Venables PJ. Antibodies to citrullinated α-enolase peptide 1 are specific for rheumatoid arthritis and cross-react with bacterial enolase. Arthritis & Rheumatism. 2008;58:3009-3019.

55 Harrison SR, Li D, Jeffery LE, Raza K, Hewison M. Vitamin D, Autoimmune Disease and Rheumatoid Arthritis. Calcif Tissue Int. 2020;106:58-75.

56 Wrangel O, Graff P, Bryngelsson IL, Fornander L, Wiebert P, Vihlborg P. Silica Dust Exposure Increases Risk for Rheumatoid Arthritis: A Swedish National Registry Case-Control Study. J Occup Environ Med. 2021;63:951-955.

57 Zhao N, Al-Aly Z, Zheng B, van Donkelaar A, Martin RV, Pineau CA, Bernatsky S. Fine particulate matter components and interstitial lung disease in rheumatoid arthritis. European Respiratory Journal. 2022;60:2102149.

58 Rantapää-Dahlqvist S, de Jong BAW, Berglin E, Hallmans G, Wadell G, Stenlund H, Sundin U, van Venrooij WJ. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis & Rheumatism. 2003;48:2741-2749.

59 van de Stadt LA, de Koning MHMT, van de Stadt RJ, Wolbink G, Dijkmans BAC, Hamann D, van Schaardenburg D. Development of the anti–citrullinated protein antibody repertoire prior to the onset of rheumatoid arthritis. Arthritis & Rheumatism. 2011;63:3226-3233.

60 Mankia K, Emery P. Review: Preclinical Rheumatoid Arthritis: Progress Toward Prevention. Arthritis & Rheumatology. 2016;68:779-788.

61 Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, Li Y, Kurreeman FAS, Zhernakova A, Hinks A, *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nature Genetics. 2010;42:508-514.

62 Shi J, Knevel R, Suwannalai P, van der Linden MP, Janssen GM, van Veelen PA, Levarht NE, van der Helm-van Mil AH, Cerami A, Huizinga TW, *et al.* Autoantibodies recognizing carbamylated proteins are present in sera of patients with rheumatoid arthritis and predict joint damage. Proc Natl Acad Sci U S A. 2011;108:17372-17377.

63 van Steenbergen HW, da Silva JAP, Huizinga TWJ, van der Helm-van Mil AHM. Preventing progression from arthralgia to arthritis: targeting the right patients. Nature Reviews Rheumatology. 2018;14:32-41.

64 Di Matteo A, Duquenne L, Cipolletta E, Nam JL, Garcia-Montoya L, Wakefield RJ, Mahler M, Mankia K, Emery P. Ultrasound subclinical synovitis in anti-CCP-positive at-risk individuals with musculoskeletal symptoms: an important and predictable stage in the rheumatoid arthritis continuum. Rheumatology (Oxford). 2022;61:3192-3200.

65 Quinn MA, Green MJ, Marzo-Ortega H, Proudman S, Karim Z, Wakefield RJ, Conaghan PG, Emery P. Prognostic factors in a large cohort of patients with early undifferentiated inflammatory arthritis after application of a structured management protocol. Arthritis & Rheumatism. 2003;48:3039-3045.

66 Freeston JE, Wakefield RJ, Conaghan PG, Hensor EMA, Stewart SP, Emery P. A diagnostic algorithm for persistence of very early inflammatory arthritis: the utility of power Doppler ultrasound when added to conventional assessment tools. Annals of the Rheumatic Diseases. 2010;69:417.

67 Novella-Navarro M, Plasencia-Rodríguez C, Nuño L, Balsa A. Risk Factors for Developing Rheumatoid Arthritis in Patients With Undifferentiated Arthritis and Inflammatory Arthralgia. Frontiers in Medicine. 2021;8.

68 Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M, Emery P, Ferraccioli G, Hazes JM, Klareskog L, *et al.* EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis. 2007;66:34-45.

69 Quinn MA, Emery P. Window of opportunity in early rheumatoid arthritis: possibility of altering the disease process with early intervention. Clin Exp Rheumatol. 2003;21:S154-157.

70 Pratt AG, Isaacs JD. Seronegative rheumatoid arthritis: Pathogenetic and therapeutic aspects. Best Practice & Research Clinical Rheumatology. 2014;28:651-659.

71 Coffey CM, Crowson CS, Myasoedova E, Matteson EL, Davis JM. Evidence of Diagnostic and Treatment Delay in Seronegative Rheumatoid Arthritis: Missing the Window of Opportunity. Mayo Clinic Proceedings. 2019;94:2241-2248.

72 Alpizar-Rodriguez D, Brulhart L, Mueller RB, Möller B, Dudler J, Ciurea A, Walker UA, Von Mühlenen I, Kyburz D, Zufferey P, *et al.* The prevalence of anticitrullinated protein antibodies increases with age in healthy individuals at risk for rheumatoid arthritis. Clinical Rheumatology. 2017;36:677-682.

73 Daha NA, Toes REM. Are ACPA-positive and ACPA-negative RA the same disease? Nature Reviews Rheumatology. 2011;7:202-203.

74 Burgers LE, van Steenbergen HW, ten Brinck RM, Huizinga TWJ, van der Helm-van Mil AHM. Differences in the symptomatic phase preceding ACPA-positive and ACPA-negative RA: a longitudinal study in arthralgia during progression to clinical arthritis. Annals of the Rheumatic Diseases. 2017;76:1751.

75 Kondo K, Ohigashi I, Takahama Y. Thymus machinery for T-cell selection. Int Immunol. 2019;31:119-125.

76 Koetz K, Bryl E, Spickschen K, O'Fallon WM, Goronzy JJ, Weyand CM. T cell homeostasis in patients with rheumatoid arthritis. Proc Natl Acad Sci U S A. 2000;97:9203-9208.

77 Ponchel F, Morgan AW, Bingham SJ, Quinn M, Buch M, Verburg RJ, Henwood J, Douglas SH, Masurel A, Conaghan P, *et al.* Dysregulated lymphocyte proliferation and differentiation in patients with rheumatoid arthritis. Blood. 2002;100:4550-4556.

78 Zhu J, Paul WE. Peripheral CD4+ T-cell differentiation regulated by networks of cytokines and transcription factors. Immunol Rev. 2010;238:247-262.

79 Ponchel F, Brown AK, Field SL, Quinn M, Conaghan P, Emery P, Isaacs JD. T-bet expression in rheumatoid arthritis patients with early, disease-modifying anti-rheumatic drug naïve disease is low and correlates with low levels of IL-7 and T-cell dysfunctions. Arthritis Research & Therapy. 2005;7:P18.

80 Kawashima M, Miossec P. mRNA Quantification of T-bet, GATA-3, IFN-γ, and IL-4 Shows a Defective Th1 Immune Response in the Peripheral Blood from Rheumatoid Arthritis Patients: Link with Disease Activity. Journal of Clinical Immunology. 2005;25:209-214.

81 Miao J, Zhang K, Lv M, Li Q, Zheng Z, Han Q, Guo N, Fan C, Zhu P. Circulating Th17 and Th1 cells expressing CD161 are associated with disease activity in rheumatoid arthritis. Scandinavian Journal of Rheumatology. 2014;43:194-201.

82 Lee GR. The Balance of Th17 versus Treg Cells in Autoimmunity. Int J Mol Sci. 2018;19.

83 Lawson CA, Brown AK, Bejarano V, Douglas SH, Burgoyne CH, Greenstein AS, Boylston AW, Emery P, Ponchel F, Isaacs JD. Early rheumatoid arthritis is associated with a deficit in the CD4 + CD25 high regulatory T cell population in peripheral blood. Rheumatology. 2006;45:1210-1217.

84 Nie H, Zheng Y, Li R, Guo TB, He D, Fang L, Liu X, Xiao L, Chen X, Wan B, *et al.* Phosphorylation of FOXP3 controls regulatory T cell function and is inhibited by TNF-α in rheumatoid arthritis. Nature Medicine. 2013;19:322-328.

85 Fasth AER, Cao D, Van Vollenhoven R, Trollmo C, Malmström V. CD28nullCD4+ T Cells – Characterization of an Effector Memory T-Cell Population in Patients with Rheumatoid Arthritis. Scandinavian Journal of Immunology. 2004;60:199-208.

86 Broadley I, Pera A, Morrow G, Davies KA, Kern F. Expansions of Cytotoxic CD4+CD28− T Cells Drive Excess Cardiovascular Mortality in Rheumatoid Arthritis and Other Chronic Inflammatory Conditions and Are Triggered by CMV Infection. Frontiers in Immunology. 2017;8.

87 Azizi G, Yazdani R, Mirshafiey A. Th22 cells in autoimmunity: a review of current knowledge. Eur Ann Allergy Clin Immunol. 2015;47:108-117.

88 Li Y, Yu Q, Zhang Z, Wang J, Li S, Zhang J, Liu G. TH9 cell differentiation, transcriptional control and function in inflammation, autoimmune diseases and cancer. Oncotarget. 2016;7:71001-71012.

89 Chang HC, Sehra S, Goswami R, Yao W, Yu Q, Stritesky GL, Jabeen R, McKinley C, Ahyi AN, Han L, *et al.* The transcription factor PU.1 is required for the development of IL-9-producing T cells and allergic inflammation. Nat Immunol. 2010;11:527-534.

90 Burgoyne CH, Field SL, Brown AK, Hensor EM, English A, Bingham SL, Verburg R, Fearon U, Lawson CA, Hamlin PJ, *et al.* Abnormal T cell differentiation persists in patients with rheumatoid arthritis in clinical remission and predicts relapse. Annals of the Rheumatic Diseases. 2008;67:750.

91 Ponchel F, Burska AN, Hunt L, Gul H, Rabin T, Parmar R, Buch MH, Conaghan PG, Emery P. T-cell subset abnormalities predict progression along the Inflammatory Arthritis disease continuum: implications for management. Scientific Reports. 2020;10:3669.

92 Ponchel F, Goëb V, Parmar R, El-Sherbiny Y, Boissinot M, El Jawhari J, Burska A, Vital EM, Harrison S, Conaghan PG, *et al.* An immunological biomarker to predict MTX response in early RA. Annals of the Rheumatic Diseases. 2014;73:2047.

93 Hunt L, Hensor EM, Nam J, Burska AN, Parmar R, Emery P, Ponchel F. T cell subsets: an immunological biomarker to predict progression to clinical arthritis in ACPA-positive individuals. Ann Rheum Dis. 2016;75:1884-1889.

94 Burska AN, Thu A, Parmar R, Bzoma I, Samans B, Raschke E, Olek S, Conaghan PG, Emery P, Ponchel F. Quantifying circulating Th17 cells by qPCR: potential as diagnostic biomarker for rheumatoid arthritis. Rheumatology. 2019;58:2015-2024.

95 Neefjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nature Reviews Immunology. 2011;11:823-836.

96 Edwards JC, Szczepanski L, Szechinski J, Filipowicz-Sosnowska A, Emery P, Close DR, Stevens RM, Shaw T. Efficacy of B-cell-targeted therapy with rituximab in patients with rheumatoid arthritis. N Engl J Med. 2004;350:2572-2581.

97 Wu F, Gao J, Kang J, Wang X, Niu Q, Liu J, Zhang L. B Cells in Rheumatoid Arthritis : Pathogenic Mechanisms and Treatment Prospects. Front Immunol. 2021;12:750753.

98 Yeo L, Toellner KM, Salmon M, Filer A, Buckley CD, Raza K, Scheel-Toellner D. Cytokine mRNA profiling identifies B cells as a major source of RANKL in rheumatoid arthritis. Ann Rheum Dis. 2011;70:2022-2028.

99 Brown PB, Nardella FA, Mannik M. HUMAN COMPLEMENT ACTIVATION BY SELF-ASSOCIATED IgG RHEUMATOID FACTORS. Arthritis & Rheumatism. 1982;25:1101-1107.

100 Laurent L, Anquetil F, Clavel C, Ndongo-Thiam N, Offer G, Miossec P, Pasquali J-L, Sebbag M, Serre G. IgM rheumatoid factor amplifies the inflammatory response of macrophages induced by the rheumatoid arthritis-specific immune complexes containing anticitrullinated protein antibodies. Annals of the Rheumatic Diseases. 2015;74:1425.

101 Mathsson L, Lampa J, Mullazehi M, Rönnelid J. Immune complexes from rheumatoid arthritis synovial fluid induce FcgammaRIIa dependent and rheumatoid factor correlated production of tumour necrosis factor-alpha by peripheral blood mononuclear cells. Arthritis Res Ther. 2006;8:R64.

102 van Beers JJBC, Schwarte CM, Stammen-Vogelzangs J, Oosterink E, Božič B, Pruijn GJM. The rheumatoid arthritis synovial fluid citrullinome reveals novel citrullinated epitopes in apolipoprotein E, myeloid nuclear differentiation antigen, and β-actin. Arthritis & Rheumatism. 2013;65:69-80.

103 Clavel C, Nogueira L, Laurent L, Iobagiu C, Vincent C, Sebbag M, Serre G. Induction of macrophage secretion of tumor necrosis factor α through Fcγ receptor IIa engagement by rheumatoid arthritis–specific autoantibodies to citrullinated proteins complexed with fibrinogen. Arthritis & Rheumatism. 2008;58:678-688.

104 Gan RW, Trouw LA, Shi J, Toes REM, Huizinga TWJ, Demoruelle MK, Kolfenbach JR, Zerbe GO, Deane KD, Edison JD, *et al.* Anti-carbamylated Protein Antibodies Are Present Prior to Rheumatoid Arthritis and Are Associated with Its Future Diagnosis. The Journal of Rheumatology. 2015;42:572.

105 Burska AN, Hunt L, Boissinot M, Strollo R, Ryan BJ, Vital E, Nissim A, Winyard PG, Emery P, Ponchel F. Autoantibodies to posttranslational modifications in rheumatoid arthritis. Mediators Inflamm. 2014;2014:492873.

106 Zhu W, Li X, Fang S, Zhang X, Wang Y, Zhang T, Li Z, Xu Y, Qu S, Liu C, *et al.* Anti-Citrullinated Protein Antibodies Induce Macrophage Subset Disequilibrium in RA Patients. Inflammation. 2015;38:2067-2075.

107 Tardito S, Martinelli G, Soldano S, Paolino S, Pacini G, Patane M, Alessandri E, Smith V, Cutolo M. Macrophage M1/M2 polarization and rheumatoid arthritis: A systematic review. Autoimmunity Reviews. 2019;18:102397.

108 Fukui S, Iwamoto N, Takatani A, Igawa T, Shimizu T, Umeda M, Nishino A, Horai Y, Hirai Y, Koga T, *et al.* M1 and M2 Monocytes in Rheumatoid Arthritis: A Contribution of Imbalance of M1/M2 Monocytes to Osteoclastogenesis. Front Immunol. 2017;8:1958.

109 Mulherin D, Fitzgerald O, Bresnihan B. Synovial tissue macrophage populations and articular damage in rheumatoid arthritis. Arthritis & Rheumatism. 1996;39:115-124.

110 Kung C-C, Dai S-P, Chiang H, Huang H-S, Sun W-H. Temporal expression patterns of distinct cytokines and M1/M2 macrophage polarization regulate rheumatoid arthritis progression. Molecular Biology Reports. 2020;47:3423-3437.

111 Hasegawa T, Ishii M. Pathological Osteoclasts and Precursor Macrophages in Inflammatory Arthritis. Front Immunol. 2022;13:867368.

112 Onodera S, Kaneda K, Mizue Y, Koyama Y, Fujinaga M, Nishihira J. Macrophage Migration Inhibitory Factor Up-regulates Expression of Matrix Metalloproteinases in Synovial Fibroblasts of Rheumatoid Arthritis *. Journal of Biological Chemistry. 2000;275:444-450.

113 Hashizume M, Hayakawa N, Mihara M. IL-6 trans-signalling directly induces RANKL on fibroblast-like synovial cells and is involved in RANKL induction by TNF-alpha and IL-17. Rheumatology (Oxford). 2008;47:1635-1640.

114 Okada Y, Nagase H, Harris ED, Jr. A metalloproteinase from human rheumatoid synovial fibroblasts that digests connective tissue matrix components. Purification and characterization. J Biol Chem. 1986;261:14245-14255.

115 Pap T, Van Der Laan WH, Aupperle KR, Gay RE, Verheijen JH, Firestein GS, Gay S, Neidhart M. Modulation of fibroblast-mediated cartilage degradation by articular chondrocytes in rheumatoid arthritis. Arthritis & Rheumatism. 2000;43:2531-2536.

116 Tolboom TCA, Pieterman E, van der Laan WH, Toes REM, Huidekoper AL, Nelissen RGHH, Breedveld FC, Huizinga TWJ. Invasive properties of fibroblast-like synoviocytes: correlation with growth characteristics and expression of MMP-1, MMP-3, and MMP-10. Annals of the Rheumatic Diseases. 2002;61:975.

117 Sato K, Suematsu A, Okamoto K, Yamaguchi A, Morishita Y, Kadono Y, Tanaka S, Kodama T, Akira S, Iwakura Y, *et al.* Th17 functions as an osteoclastogenic helper T cell subset that links T cell activation and bone destruction. J Exp Med. 2006;203:2673-2682.

118 Hashizume M, Hayakawa N, Mihara M. IL-6 trans-signalling directly induces RANKL on fibroblast-like synovial cells and is involved in RANKL induction by TNF-α and IL-17. Rheumatology. 2008;47:1635-1640.

119 Takayanagi H, Ogasawara K, Hida S, Chiba T, Murata S, Sato K, Takaoka A, Yokochi T, Oda H, Tanaka K, *et al.* T-cell-mediated regulation of osteoclastogenesis by signalling cross-talk between RANKL and IFN-γ. Nature. 2000;408:600-605.

120 Boissier M-C, Semerano L, Challal S, Saidenberg-Kermanac'h N, Falgarone G. Rheumatoid arthritis: From autoimmunity to synovitis and joint destruction. Journal of Autoimmunity. 2012;39:222-228.

121 McInnes IB, Buckley CD, Isaacs JD. Cytokines in rheumatoid arthritis — shaping the immunological landscape. Nature Reviews Rheumatology. 2016;12:63-68.

122 McGonagle D, McDermott MF. A proposed classification of the immunological diseases. PLoS Med. 2006;3:e297.

123 Savic S, Mistry A, Wilson AG, Barcenas-Morales G, Doffinger R, Emery P, McGonagle D. Autoimmune-autoinflammatory rheumatoid arthritis overlaps: a rare but potentially important subgroup of diseases. RMD Open. 2017;3:e000550.

124 Fang Q, Zhou C, Nandakumar KS. Molecular and Cellular Pathways Contributing to Joint Damage in Rheumatoid Arthritis. Mediators Inflamm. 2020;2020:3830212.

125 Ponchel F, Vital EM, Kingsbury SR, El-Sherbiny YM. CD4+ T-cell subsets in rheumatoid arthritis. International Journal of Clinical Rheumatology. 2012;7:37-53.

126 Kouzarides T. Histone methylation in transcriptional control. Current Opinion in Genetics & Development. 2002;12:198-209.

127 Dor Y, Cedar H. Principles of DNA methylation and their implications for biology and medicine. The Lancet. 2018;392:777-786.

128 Bender J. DNA METHYLATION AND EPIGENETICS. Annual Review of Plant Biology. 2004;55:41-68.

129 Campagna MP, Xavier A, Lechner-Scott J, Maltby V, Scott RJ, Butzkueven H, Jokubaitis VG, Lea RA. Epigenome-wide association studies: current knowledge, strategies and recommendations. Clinical Epigenetics. 2021;13:214.

130 Li L-C, Dahiya R. MethPrimer: designing primers for methylation PCRs. Bioinformatics. 2002;18:1427-1431.

131 Delaney C, Garg SK, Yung R. Analysis of DNA Methylation by Pyrosequencing. Methods Mol Biol. 2015;1343:249-264.

132 Ramalho-Carvalho J, Henrique R, Jerónimo C. Methylation-Specific PCR. In: Tost J, ed. DNA Methylation Protocols. New York, NY: Springer New York; 2018:447-472.

133 Villicaña S, Bell JT. Genetic impacts on DNA methylation: research findings and future perspectives. Genome Biology. 2021;22:127.

134 Richardson B, Scheinbart L, Strahler J, Gross L, Hanash S, Johnson M. Evidence for impaired t cell dna methylation in systemic lupus erythematosus and rheumatoid arthritis. Arthritis & Rheumatism. 1990;33:1665-1673.

135 Karouzakis E, Gay RE, Gay S, Neidhart M. Epigenetic control in rheumatoid arthritis synovial fibroblasts. Nat Rev Rheumatol. 2009;5:266-272.

136 Nakano K, Whitaker JW, Boyle DL, Wang W, Firestein GS. DNA methylome signature in rheumatoid arthritis. Annals of the Rheumatic Diseases. 2013;72:110.

137 Pitaksalee R, Burska AN, Ajaib S, Rogers J, Parmar R, Mydlova K, Xie X, Droop A, Nijjar JS, Chambers P, *et al.* Differential CpG DNA methylation in peripheral naïve CD4+ T-cells in early rheumatoid arthritis patients. Clinical Epigenetics. 2020;12:54.

138 Cribbs AP, Kennedy A, Penn H, Read JE, Amjadi P, Green P, Syed K, Manka SW, Brennan FM, Gregory B, *et al.* Treg Cell Function in Rheumatoid Arthritis Is Compromised by CTLA-4 Promoter Methylation Resulting in a Failure to Activate the Indoleamine 2,3-Dioxygenase Pathway. Arthritis & Rheumatology. 2014;66:2344-2354.

139 Nile CJ, Read RC, Akil M, Duff GW, Wilson AG. Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. Arthritis & Rheumatism. 2008;58:2686-2693.

140 Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. Nature Reviews Rheumatology. 2013;9:141-153.

141 Zhao J, Guo S, Schrodi SJ, He D. Molecular and Cellular Heterogeneity in Rheumatoid Arthritis: Mechanisms and Clinical Implications. Front Immunol. 2021;12:790122.

142 Aletaha D. Precision medicine and management of rheumatoid arthritis. Journal of Autoimmunity. 2020;110:102405.

143 Finckh A. Early inflammatory arthritis versus rheumatoid arthritis. Current Opinion in Rheumatology. 2009;21.

144 Nishimura K, Sugiyama D, Kogata Y, Tsuji G, Nakazawa T, Kawano S, Saigo K, Morinobu A, Koshiba M, Kuntz KM, *et al.* Meta-analysis: Diagnostic Accuracy of Anti–Cyclic Citrullinated Peptide Antibody and Rheumatoid Factor for Rheumatoid Arthritis. Annals of Internal Medicine. 2007;146:797-808.

145 Ponchel F, Duquenne L, Xie X, Corscadden D, Shuweihdi F, Mankia K, Trouw LA, Emery P. Added value of multiple autoantibody testing for predicting progression to inflammatory arthritis in at-risk individuals. RMD Open. 2022;8.

146 Grosse J, Allado E, Roux C, Pierreisnard A, Couderc M, Clerc-Urmes I, Remen T, Albuisson É, De Carvalho-Bittencourt M, Chary-Valckenaere I, *et al.* ACPA-positive versus ACPA-negative rheumatoid arthritis: two distinct erosive disease entities on radiography and ultrasonography. Rheumatol Int. 2020;40:615-624.

147 Mouterde G, Rincheval N, Lukas C, Daien C, Saraux A, Dieudé P, Morel J, Combe B. Outcome of patients with early arthritis without rheumatoid factor and ACPA and predictors of rheumatoid arthritis in the ESPOIR cohort. Arthritis Res Ther. 2019;21:140.

148 Shmerling RH, Delbanco TL. How useful is the rheumatoid factor? An analysis of sensitivity, specificity, and predictive value. Arch Intern Med. 1992;152:2417-2420.

149 Besada E, Nikolaissen C, Nossent H. Should rheumatoid factor in rheumatoid arthritis be sent to Davy Jones's Locker? Scandinavian Journal of Rheumatology. 2012;41:85-88.

150 Ponchel F, van Delft MAM, Xie X, Burska AN, Duquenne L, Trouw LA, Emery P. Anti-carbamylated protein antibodies: are they useful for the diagnosis of rheumatoid arthritis? Clin Exp Rheumatol. 2021;39:146-150.

151 Sproston NR, Ashworth JJ. Role of C-Reactive Protein at Sites of Inflammation and Infection. Frontiers in Immunology. 2018;9.

152 Orr CK, Najm A, Young F, McGarry T, Biniecka M, Fearon U, Veale DJ. The Utility and Limitations of CRP, ESR and DAS28-CRP in Appraising Disease Activity in Rheumatoid Arthritis. Frontiers in Medicine. 2018;5.

153 Wells G, Becker JC, Teng J, Dougados M, Schiff M, Smolen J, Aletaha D, van Riel PL. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. Ann Rheum Dis. 2009;68:954-960.

154 Horton SC, Tan AL, Wakefield RJ, Freeston JE, Buch MH, Emery P. Ultrasound-detectable grey scale synovitis predicts future fulfilment of the 2010 ACR/EULAR RA classification criteria in patients with new-onset undifferentiated arthritis. RMD Open. 2017;3:e000394.

155 Ponchel F, Burska AN, Hunt L, Gul H, Rabin T, Parmar R, Buch MH, Conaghan PG, Emery P. T-cell subset abnormalities predict progression along the Inflammatory Arthritis disease continuum: implications for management. Sci Rep. 2020;10:3669.

156 Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, Clark SJ, Molloy PL. De novo identification of differentially methylated regions in the human genome. Epigenetics & Chromatin. 2015;8:6.

157 Pitaksalee R, Parmar R, Hodgett R, Emery P, Ponchel F. DNA Hypomethylation in the TNF-Alpha Gene Predicts Rheumatoid Arthritis Classification in Patients with Early Inflammatory Symptoms. Cells. 2023;12.

158 Saffari A, Silver MJ, Zavattari P, Moi L, Columbano A, Meaburn EL, Dudbridge F. Estimation of a significance threshold for epigenome-wide association studies. Genetic Epidemiology. 2018;42:20-33.

159 Darst RP, Pardo CE, Ai L, Brown KD, Kladde MP. Bisulfite Sequencing of DNA. Current Protocols in Molecular Biology. 2010;91:7.9.1-7.9.17.

160 Da Silva JAP, Hall GM. The effects of gender and sex hormones on outcome in rheumatoid arthritis. Baillière's Clinical Rheumatology. 1992;6:193-219.

161 van Schaardenburg D, Lagaay AM, Otten HG, Breedveld FC. THE RELATION BETWEEN CLASS-SPECIFIC SERUM RHEUMATOID FACTORS AND AGE IN THE GENERAL POPULATION. Rheumatology. 1993;32:546-549.

162 Vander Cruyssen B, Peene I, Cantaert T, Hoffman IE, De Rycke L, Veys EM, De Keyser F. Anti-citrullinated protein/peptide antibodies (ACPA) in rheumatoid arthritis: specificity and relation with rheumatoid factor. Autoimmun Rev. 2005;4:468-474.

163 Li K, Wang M, Zhao L, Liu Y, Zhang X. ACPA-negative rheumatoid arthritis: From immune mechanisms to clinical translation. EBioMedicine. 2022;83:104233.

164 Lubbers J, Stadt LAvd, Vosslamber S, Wesseling JG, Oztürk S, Schaardenburg Dv, Verweij CL. A1.7 Interferon and B-Cell Gene Signatures Contribute to Diagnosis of Pre-Clinical Rheumatoid Arthritis. Annals of the Rheumatic Diseases. 2013;72:A3.