The Stratification Potential of a Novel Epigenetic Biomarker in Rheumatoid Arthritis

Caitlin Duncan
MSc, BSc (Hons)

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

College of Medical, Veterinary and Life Sciences
Institute of Infection, Immunity and Inflammation
University of Glasgow

February 2021

Abstract

Rheumatoid arthritis (RA) is a chronic autoimmune disease of the joints, that affects 0.5-1% of the population globally. While primarily affecting the joints, systemic inflammation impacts other organs and the disease has a significant socioeconomic burden. While there are a wide range of medications to pharmacologically manage RA, it is a largely heterogeneous disease and the current treatment strategy does not consider the heterogeneity between patients. As such, precision medicine approaches to treatment are desired. A 5-loop chromosome conformation signature (CCS) was identified that had 90% specificity at predicting non-response to methotrexate (MTX) in early RA. These epigenetic biomarkers offer a novel strategy for improving patient care, and provide insight into disease pathogenesis.

The aim of the work presented in this thesis was to further characterise this novel epigenetic biomarker. Investigation of this biomarker also offered the opportunity to hypothesise about underlying pathogenesis. A combination of molecular analysis of patient samples, and *in-silico* methodologies were applied to investigate these aims.

In the first instance, the CCS was validated as a biomarker for identifying MTX responders using bioinformatic tools. Preliminary work was also carried out to identify the optimal method for detecting chromosome loops from the signature in the lab. Quantitative PCR was thoroughly explored, but excluded as a reliable and robust method of loop detection for our signature of interest. It was also found that the CCS was MTX specific, and alternative signatures would be required for prediction of response to other csDMARDs. Further validation of the signature, using an independent clinical cohort, revealed that specific loops from the CCS held stratification potential while others did not. *In-silico* investigations revealed different epigenetic landscapes exist between loops associated with responders and non-responders to MTX. Specifically, data suggests loops associated with responders exist in an environment which enhances gene transcription, while loops associated with non-responders have an environment indicating potential for gene repression. Differences in chromatin architecture, revealed through a discovery

microarray, have indicated that 3D epigenetic endotypes exist within the early RA population. Further investigations suggested each endotype have different, unique pathways that are highly regulated. Furthermore, results revealed that there is a stable RA chromatin signature that exists, which highlights the importance of the 3D epigenome underpinning disease.

In summation, this body of work has shown CCS to be promising biomarker for the stratification of the early RA population. Furthermore, thorough investigation of this signature highlighted novel pathways that may be involved in disease pathogenesis. This work has exciting potential to contribute to improved RA treatment in the future.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to sincerely thank my supervisors, Professor Carl Goodyear, Professor Iain McInnes and Dr Ewan Hunter. Thank you, Carl for your scientific insight and support throughout my project. I would also like to thank you for the freedom and encouragement to attend engagement events throughout my PhD. It was a privilege to be able to talk about this work with the public, healthcare professionals, and importantly, patients with rheumatoid arthritis. Thank you, Ewan for all your support and insight with analysis and generously giving your time to this project. And thank you Iain for your infectious enthusiasm and encouragement throughout.

I must also thank Matt, Christina, Francis and Jane. Thank you, Matt for all your support in setting up protocols at the beginning of the project and for your helpful answers to my questions. Thank you, Christina for being so patient and helpful with my many analysis questions. Thank you to Francis for all the help and training on qPCR protocols and to Jane for helping with the transfer of samples on many an occasion. Thank you to everyone else at OBD who has helped with data processing or analysis and for making me feel so welcome when I visited.

I have to thank all of the Goodyear group past and present, I feel extremely lucky to have gone through the PhD journey with you all. Thank you to Heather and Louise for training in the early days and throughout. And thank you to Sabari for all help and conversation in the lab. Thank you to Sarah, Heather, Louise, Lewis, Katy, Lauren, Flavia, Aysin, Cecilia, Keiran, Andy, Maria and Patricia for the great office chat, coffee and cake breaks and of course, Friday pub!

Thank you to John for your all your help and enthusiasm with analysis. And thank you to Caron and Ashley for your help dealing with the collection or transport of clinical samples. Thank you also to Stefan, Georgia and Jill for helpful and encouraging conversations at each annual review.

To everyone on the precision medicine DTP, thank you for interesting scientific meetings and for the wonderful trip to Firbush. It was great to have you all as a support network throughout the programme.

To the Science Fam; Gemma, Gabby, Kirstin, Katy, Lauren, Kym and Heather, thank you for the daily chats and zoom writing sessions throughout lockdown, for Bake Off nights, weekend walks and trips away. To Marian, Hannah, Caitlin, Lindsey, Emma and Debbie, thank you for all of the fun times and support throughout the last 4 years.

To Kym, from the first day of undergrad together, to our PhDs two doors down from each other, you've been an amazing support and I couldn't have done it without you!

A BIG thank you to Thomas. A special recognition to you for supporting me through a thesis write-up during lockdown! Thank you for being my cheerleader, and making me laugh daily, no mean feat during the last year.

Thank you, Gregor for always being a phone call away and for supplying me with many boxes of rhubarb and custards! And Mum and Dad, I would simply not be where I am today without you both. Thank you for always believing in me.

I also truly thank all patients who generously donated samples, this work would be impossible without you.

Author's Declaration

I declare that this thesis is the result of my own work. No part of this thesis has been submitted for any other degree at The University of Glasgow, or any other institution.

Caitlin Duncan

Abbreviations

A	ACPA	Anti Citrullinated Protein/Peptide Antibodies
	ACR	American College of Rheumatology
	APC	Antigen presenting cell
B	BC	Buffy coat
	bDMARD	biological disease modifying anti-rheumatic drug
	BL	baseline
	bp	Base pair
C	CCS	Chromosome conformation signature
	CD	Cluster of differentiation
	CDAI	Clinical disease activity index
	CIA	Collagen induced arthritis
	ChIA-PET	Chromatin interaction analysis by paired end sequencing
	ChIP-Seq	Chromatin immunoprecipitation sequencing
	CRP	C-reactive protein
	csDMARD	conventional synthetic disease modifying anti-rheumatic drug
	Ct	Cycle Threshold
	CTCF	CCCTC-Binding Factor
	CXCL13	Chemokine (CXC-motif) ligand 13
D	DAS	Disease Activity Score
	DAS28	Disease activity score 28
	DC	Dendritic cell
	DHSs	Dnase Hypersensitive Sites
	DMARD	Disease modifying anti-rheumatic drug
	DMSO	Dimethyl sulphoxide
	DNA	Deoxyribose nucleic acid
	DNA-FISH	DNA- florescent in-situ hybridisation
E	EDTA	Ethylenediaminetetraacetic Acid

	eQTL	Expression quantitative trait loci
	ESR	Erythrocyte sedimentation rate
	EULAR	European League Against Rheumatism
	EWAS	Epigenome Wide association study
F	FACS	Fluorescence-Activated Cell Sorting
	FCS	Fetal calf serum
	FLS	Fibroblast-like synoviocyte
G	GWAS	Genome-Wide Association Study
	GI	Gastrointestinal
	GO	Gene ontology
H	H	Histone
	HCQ	Hydroxychloroquine
	HDA	High disease activity
I	IDT	Integrated DNA Technologies
	IFN	Interferon
	IFNAR1	Type 1 interferon receptor 1
	IGV	Integrated Genome Browser
	IL	Interleukin
	IL17A	Interleukin 17A
	IL21R	Interleukin 21 receptor
	IL23	Interleukin 23
	IR	Initial responder
J	JAK STAT	Janus kinase signal transducer and activator of transcription
	JIA	Juvenile Idiopathic Arthritis
K	K	Lysine
	kb	Kilobase

L	LCR	locus control region
	LDA	Low disease activity
M	mb	Megabase
	me	Methylation
	MHC	Major Histocompatibility locus
	MIQE	Minimum information for publication of quantitative real-time PCR experiments
	MMP	Matrix Metalloproteinases
	MTX	Methotrexate
N	NK Cell	Natural killer cell
	NR	Non-responder
	NSAID	Non-steroidal anti-inflammatory drug
	NTC	Non-template control
O	OBD	Oxford BioDynamics Plc
P	PBMC	Peripheral blood mononuclear cells
	PCR	Polymerase chain reaction
Q	QC	Quality control
	qPCR	quantitative Polymerase chain reaction
R	R	Responder
	RA	Rheumatoid arthritis
	RF	Rheumatoid Factor
	RT	Room temperature
	RTX	Rituximab
S	SERA	Scottish early rheumatoid arthritis
	SJC 28	Swollen joint count 28
	SF	Synovial Fluid

	SLE	Systemic lupus erythematosus
	SM	Synovial membrane
	SNBTS	Scottish national blood transfusion service
	SNP	Single nucleotide polymorphism
	SSZ	Sulphasalazine
T	TACERA	Towards a Cure for Early Rheumatoid Arthritis
	TAD	Topologically associated domain
	TAE	Tris-base Acetic Acid
	TF	Transcription factor
	TJC 28	Tender joint count 28
	TNF	Tumor necrosis factor
	T2T	Treat to target
V	VAS	Visual analogue scale
	3C	Chromosome conformation capture
	6m	6 months
	12m	12 months

Chapter 1 Introduction

1.1 Rheumatoid Arthritis

1.1.1 Prevalence

Rheumatic diseases are characterised by damage to joints and connective tissues, which is often accompanied by complications for multiple other vital organs[1]. Rheumatoid arthritis (RA) is a disease of this type, and is the most common inflammatory arthritis[2]. RA is prevalent in approximately 0.5-1%[3] of the population, and females are affected more than males with a ratio of 6.45:1[4]. The most common age of onset is between 45 and 65[5], however it can occur at any age, with under 16s receiving a diagnosis of juvenile idiopathic arthritis (JIA)[6]. RA is an autoimmune and inflammatory condition, with chronic inflammation originating from the synovium leading to eventual joint destruction[7].

RA presents with symptoms of pain, fatigue and symmetrical swollen joints, with the smaller peripheral joints, such as the hands and feet, affected first[8]. Systemic inflammation can impact the eyes, lungs and blood vessels[9] . If left uncontrolled, RA is a progressively disabling disease, which can impact on all parts of everyday life. Comorbidities such as heart disease and infection are often experienced, and can result in early mortality[10]. Cases of anxiety and depression are also known to be higher in those with RA, compared to healthy age and sex matched individuals[11]. In working-age adults with RA, the prevalence of depression has been found to be around 25%[12]. Interestingly, these comorbidities have not decreased with the better pharmacological management of RA[11]. It has also been found that those who experience psychiatric comorbidities are more likely to be impacted by other comorbidities that have been discussed[13]. Psychological symptoms have also been shown to negatively impact the likelihood of a person reaching remission after 12 months of treatment[14].

1.1.2 Development

The development of RA is caused by a loss of immune tolerance to the joints, the exact trigger of which is unknown. However, there are various well-known genetic and environmental factors that increase susceptibility to the development and severity of disease[15]. These contributions result in a phenotypically heterogeneous RA population. RA heritability has been demonstrated through twin studies; monozygotic twins have shown to both have an RA diagnosis in 12-15% of cases, 10% more than non-zygotic twins[2]. There are currently over 100 genetic loci known to be associated with RA[16]. The allele associated with the highest risk is *HLA-DR*, which is an allele part of the major histocompatibility (MHC) locus. Modifications in DR4 and DR14 are most common, with changes in these alleles present in approximately 90% of people with RA[2]. Non-MHC genes with well documented risk for RA susceptibly are protein tyrosine phosphatase non-receptor 22 gene (*PTPN22)* and peptidyl arginine deiminase 4 (*PADI4)*[17]. A suggested mechanism by which these genes are a risk for RA, is the increased citrullination[18].

The relatively low genetic susceptibility, illustrated by twin studies, indicates the considerable environmental component involved in RA development. It has been proposed that the environment can impact disease susceptibly via the microbiome, primarily the lungs, oral mucosa and gut[19]. Numerous studies have also proposed that mucosal sites are where RA begins. Cigarette smoking is one of the biggest contributing environmental risk factors for RA susceptibility and severity, and is known to impact these mucosal sites[20]. The risk of RA development in smokers is twice that of non-smokers, with a slightly higher risk in women than men. It is largely believed that the more somebody smokes, the higher the risk, yet even light smokers are known to have an increased chance of development. Some of the risks associated with smoking may be due to its effects on the immune system which include oxidative stress, inflammation and epigenetic changes[20]. The oxidative stress can be attributed to the free radicals contained in smoke, in addition to nicotine effects on mitochondria[21]. The effects of smoking extend to the ability to respond well to RA treatment, which has been largely documented in biologic therapy[22]. Periodontal disease is another well-known environmental risk factor for the development of RA[23]. This

also has the ability to cause epigenetic changes to the genome which results in the damage observed in the disease[24]. Aspects of the diet have also been shown to be an environmental risk factor for RA. These include low vitamin D, and high sugar and sodium intake[25-27]. The gap in the knowledge of heritability and environmental influence has meant the diagnosis, prognosis and pharmacological management of the disease is not always correct for the patient[28]. These potential changes to the genome caused by environmental influence remain to be thoroughly explored.

1.1.3 Socio-economic burden

Approximately 80% of working adults with RA experience pain, joint stiffness and reduced functionality, ultimately limiting the ability to work as normal, and carry out everyday tasks[29]. As such, RA is a disease with a considerable socio-economic burden. Work disability is a risk, even in early disease[30]. There are several risk factors for the development of workplace disability, which include the nature of the job, disease activity and age of onset[31]. When evaluating the economic burden of RA, three main cost categories are explored: direct, indirect and non-monetary[29]. Direct costs refer to costs to the healthcare system such as medication, and other costs which are incurred by the affected person and their families. This can include adapted transportation or living arrangements. Indirect costs are those which are incurred by employers for example. These include the cost of sick-leave or reduced productivity. Additionally, the large proportion of people who remain unresponsive to RA treatment add to this burden[32]. Non-responders (NR) to treatment often lose more days at work compared to responders (R), and are likely to have higher healthcare costs, due to side-effects or from continued inflammatory symptoms[32]. The cases of depression in working adults with RA is also a considerable contributor to the economic burden. One study from the US revealed that those with RA and depression were 20% more likely to be unemployed than those without depression, and had approximately £4000 more in healthcare costs[12]. It was also documented that those with depression were more likely to experience pain, which may have an impact on ability to attend work[12]. There are also the non-monetary costs, which are arguably the most important. Those are costs that impact quality of life for those affected by RA. Often these non-monetary costs

are left unmeasured in studies calculating the 'cost' of RA due to the difficulty in measuring this reliably[29].

1.1.4 Immunopathogenesis

The immunopathogenesis of RA is a complex, stepwise process ranging from pre-diagnosis to chronic inflammation[33]. This is driven by a host of innate and adaptive immune mechanisms, in addition to the stromal compartment[2]. The synovial membrane (SM) lines the joint capsules and secretes synovial fluid (SF) which is essential for maintaining joint function. In RA, the SM is one of the most affected tissues and undergoes significant structural changes. In healthy individuals, the synovial membrane is 1-2 cells thick and is comprised of blood vessels and adipocytes, as well as low levels of macrophages and fibroblasts[34]. The synovial fluid is comprised of nutrients and cytokines. In RA, the SM increases in thickness to form a pannus, around 10-12 cells thick with infiltrates from both innate and adaptive immune systems including macrophages, natural killer (NK) cells, B cells and T cells[2]. Fibroblast-like synoviocytes (FLS) are central players in synovial inflammation and mediate many of the processes in the RA synovium[35].

When FLS are activated they produce IL-6, which contributes to damage[36]. IL-6 is one of the major drivers of disease and exerts its effects on multiple other cells and molecules. IL-6 signals via an IL-6 specific receptor and surface glycoprotein named gp130. There are two types of IL-6 signalling, named classical and trans-signalling, which correspond to IL-6 binding to a surface or soluble receptor, respectively[36]. TNFα is another pro-inflammatory cytokine that drives disease. Macrophages in the joint are the main producers of this, but is it also released by activated T lymphocytes[37]. Two receptors exist for TNFα to bind to, these are named CD120a and CD120b[38]. TNFα is found in higher concentrations in the RA synovium compared to healthy, and studies have shown it to be associated with increased bone erosion[39].

The dysregulated synovium reduces the oxygen supply, resulting in a hypoxic environment, changing from around 8% oxygen to 1%[40]. This environment induces altered cellular metabolism, which in turn increases reactive oxygen species

which perpetuates the inflammation. In response to a hypoxic environment, immune cells activate proinflammatory signalling pathways such as NFκB[41] and hypoxia-inducible factors (HIF). HIF are transcriptional factors that are highly expressed in the RA synovium, and accumulating evidence suggests that HIFs helps to maintain the inflammatory environment in RA[42]. For example, HIF1α can upregulate TLR-4 on macrophages. Moreover, HIF1α has been shown to regulate mediators of angiogenesis, another known characteristic of RA pathogenesis[43]. This aggressive, inflamed synovial environment invades the proximal cartilage and bone, leading to destruction[40]. Synoviocytes become chronically activated and epigenetically changed, driving this inflammation further. These changes also occur in the infiltrating immune cells[40].

In RA specifically, the adaptive immune response is contributed to by autoantibodies[44]. These include rheumatoid factor (RF), an antibody against IgG, and anti-citrullinated protein antibodies (ACPA) that are mounted against post-translationally modified proteins[44]. Interestingly, these antibodies can exist before disease presentation. People with RA are classed as sero-positive or sero-negative, based on the presence or absence of these antibodies, respectively. Approximately 50-80% of RA patients are sero-positive to some extent[45]. The binding of these autoantibodies to the Fc receptor within synovium may trigger innate immune mechanisms such as the complement pathway.

The erosion of bone is another process which occurs during inflammation, and begins early, triggered by the inflamed synovial environment and proinflammatory cytokines. This results in the activation of the receptor activator NFκB ligand (RANKL). Fibroblasts expressing RANKL are stimulated by macrophage-colony stimulating factor (M-CSF) resulting in the differentiation of pre-osteoclasts into osteoclasts, which break down bone[46]. Monocytes in the synovial compartment can also differentiate into osteoclasts with stimulation from IL-6, TNFα and IL-17. This inflammatory environment can suppress the differentiation of osteoblasts, preventing the capacity to form new bone to counteract increased osteoclast activity[47]. Bone erosion in RA is irreversible, thus placing high importance on early intervention and appropriate treatment.

1.1.4.1 Peripheral Blood Mononuclear Cells

Peripheral Blood Mononuclear Cells (PBMCs) are circulating immune cells within blood, with a round nucleus[48]. PBMCs are comprised of members of both the innate and adaptive immune system[49]. In general, the majority of cells are T cells at around 70%, with B cells, monocytes and NK cells making up the other 15%, 10% and 5%, respectively[50]. The composition of cell subtypes within PBMCs can be influenced by several factors including hormone levels and state of inflammation[51]. The peripheral blood in those with RA has been shown to differ in numerous ways from those without, in factors such as cytokine production, cell phenotype and methylation status[52]. In the absence of an antigen, most PBMCs circulate as naïve cells without a specific function. The activation state and composition of PBMCs can reflect the wider immune condition of someone with RA, and thus present an appropriate population of cells to interrogate for insights into immunopathogenesis[53]. Moreover, this circulating population of cells are highly important from a research perspective, particularly when investigating the influence of pharmacological intervention[49].

1.1.4.2 Monocytes

Monocytes make up 10% of circulating peripheral blood cells[54] and they originate from the bone marrow[55]. Both in states of homeostasis and inflammation, monocytes migrate from the bloodstream into tissues[56]. In RA, monocytes can migrate into the synovium through interaction between CCR2 and CX3CR1 receptors and CCL2 and CX3CL1 ligands[55]. Three main populations of monocytes exist based on their surface markers. 90% of monocytes, otherwise known as the 'classical' type, belong to one population and are $CD14^{+}CD16^{-}$. The other populations of monocytes are $CD14^{+}CD16^{+}$ and $CD14^{-}CD16^{+}$, otherwise known as 'intermediate' and 'non-classical' subsets respectively[57]. Each subpopulation can mediate inflammatory processes in a different way. Monocytes play an important part in the innate immune system, with various roles including phagocytosis and antigen presentation[54]. While monocytes are great scavenger cells and protective in that way, they also have pathogenic mechanisms in inflammatory diseases such as atherosclerosis and RA[54]. Monocytes have the capacity to differentiate into macrophages and dendritic cells (DCs)[56]. In the RA joint, monocytes produce a variety of pro-inflammatory cytokines which result in

processes such as polarisation of $CD4^+$ T cells[56]. In addition, monocytes can also differentiate to osteoclasts, which contribute to the destruction of bone via dysregulated bone remodelling[58].

1.1.4.3 T cells

T cells make up approximately 70-90% of the PBMC population, most of which are $CD3^+$. Within the $CD3^+$ population, there are two main types of T cell: $CD4^+$ and $CD8^+$ T cells, the ratio of which is around 2:1[51]. In the normal state, T cells will circulate within the PBMC population as naïve or memory cells, without an effector function[51]. If a T cell encounters an antigen, it can differentiate and produce effector functions. T cells have been shown to play a significant role in RA pathogenesis, what type, and at what stage remains unclear as evidenced through risk genes which are involved in T cell activation[3]. Success of Abatacept, which targets T-cell co-stimulatory molecule CTLA4 in the treatment of RA, also demonstrates their pathogenic role[59]. $CD4^+$ T cells have a wide range of differentiated phenotypes including, T helper 1 (Th1), Th2, Th17 and Th22[51]. Th2 cells are known to induct antibody responses in B cells, influencing the generation of RF and ACPA[60]. Th17 cells also play a large role in RA pathogenesis, after being activated by antigen presenting cells (APCs) and cytokines including IL-6 and IL-21[61]. Th17 discovery shed new light onto RA pathogenesis, which was originally believed to be a Th1 driven disease. Th17 cells produce the cytokines *IL17A*, IL-17F, IL-21, IL-22 and IL-26[62]. Th17 cells can reside in the gut without eliciting pathogenic effects. However, if self-reactive Th cells are primed to become Th17 cells, they can become pathogenic and induce a pro-inflammatory response towards a specific organ, such as the joint[61]. Other pro-inflammatory cytokines IL-1 and *IL23* have been shown to regulate and enhance the Th17 response in animal models of autoimmune diseases[63].

1.1.4.4 B cells

B cells play several important immunological roles as part of the adaptive immune system; acting as antigen presenting cells (APCs), and producing antibodies are just two of these vital roles[64]. As with other immune cells, with the loss of self-tolerance, B cells can ultimately cause harm to the body. For example, B cells are known to have a pathogenic role in several autoimmune diseases, including systemic lupus erythematosus (SLE) and RA. Specifically, experiments with murine models, and the success of rituximab (causes B cell depletion) in the treatment of RA, highlight the crucial role of B cells in the pathogenesis [65,66]. Moreover, the production of cytokine and chemokines, such as

CXCL13 and *IL4* and *IL6* by B cells facilitates the infiltration of other immune cells into the joint[67]. Furthermore, B cells work closely with T cells to exert their pathogenic effects, such as the activation of autoreactive T cells[64].

1.1.5 Diagnosis

Over the last couple of decades, extensive research has demonstrated that early diagnosis of RA, and subsequent pharmacological intervention provides the optimal window for successful treatment and offers the best chance of achieving a good long-term outcome[68,69]. There is no diagnostic test which can diagnose RA, instead diagnosis is made by a clinician, based on a set of criteria initially set out in 1987. With an emphasis on the importance of early diagnosis, the 1987 RA classification criteria required updating. This was due to the fact the criteria were developed in those with established RA, and thus had poor sensitivity for those with early synovitis[70]. As such, the main RA clinical consortiums, The European League Against Rheumatism (EULAR) and the American College of Rheumatology (ACR) updated their diagnosis guidelines in 2010[71]. Through many validation studies, these new criteria have shown to have increased sensitivity from previous guidelines, from 31.9% to 72.3%. However, there was a reduction in specificity from 92.4% to 83.2%. The overall diagnostic ability is considered moderate[72]. The criteria for diagnosis considers 4 categories: 1) joint involvement, 2) the presence of serological markers, RF and ACPA, 3) measurements of the acute phase reactants, c-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) and 4) duration of symptoms[73] (Table 1.1). Each category has points assigned for severity. A total score of 6 or more results in a diagnosis of RA[74]. Capturing those who may have had symptoms for less than 6 weeks, but have other symptoms present, aims to capture people early and fulfil the aim to get treatment initiated as soon as possible. Of note, anybody presenting with bone erosions does not need the classification criteria to obtain an RA diagnosis, as any existing damage indicates presence of disease[74].

Table 1.1 2010 ACR/EULAR classification criteria
These criteria consider symptom information from 4 domains: A) joint involvement, B) the presence of serological markers, rheumatoid factor (RF) and anti-citrullinated protein antibodies (ACPA), C) measurements of the acute phase reactants, c-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) and D) duration of symptoms. A total score ≥6 results in an RA diagnosis. Adapted from Kay and Upchurch, 2012[74]
ACPA, anti-citrullinated protein antibodies; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; RF, rheumatoid factor

Domain	Category	Point Score
A	**Joint involvement (0-5 points)**	
	1 large joint	0
	2-10 large joints	1
	1-3 small joints	2
	4-10 small joints	3
	> 10 joints including at least 1 small joint	5
B	**Serology (at least one test needed classification (0-3 points)**	
	-RF and -ACPA	0
	Low +RF or low +ACPA	2
	High +RF or high +ACPA	3
C	**Acute phase reactants (at least one test needed for classification (0-1 point)**	
	Normal CRP and normal ESR	0
	Abnormal CRP or abnormal ESR	1
D	**Duration of symptoms**	
	< 6 weeks	0
	≥ 6 weeks	1

1.2 Treatment

1.2.1 Treatment Overview

RA treatment aims to control pain and inflammation, reduce joint destruction and achieve remission[75]. While there is currently no cure for RA, there are a wide variety of pharmacological therapies available to try and manage the disease. These include non-steroidal anti-inflammatory drugs (NSAIDs) and disease modifying anti-rheumatic drugs (DMARDs), both non-biologic and biologic. Non-biologic DMARDs offer broad immunosuppression functions, while biologic DMARDs (bDMARDs) target immune cells and mediators, such as B cells (Rituximab), IL-6 (Tocilizumab) and TNFα (Etanercept) involved in RA

pathology[76,77]. For a drug to be classed as a DMARD in the treatment of RA, it has to have demonstrated a reduction in the radiographic progression of disease[78]. Historically, NSAIDs were the only treatment option, and while these eased the pain of symptoms, they did not slow progression of disease[79]. The development of biologics over the past 2 decades has significantly improved disease outcome for many patients[80]. It should be noted that until the emergence of the biologics, many of the drugs used in RA were not created specifically for the disease and unsurprisingly this has contributed to the variation in efficacy of treatment in patients (Figure 1.1). Moreover, the heterogeneity between patients further influences the disparity in drug response.

The drugs themselves have not only developed over the last couple of decades, but the treatment strategy has also changed. RA clinics now adopt a treat to target strategy (T2T), developed in 2010[81] which aims to get patients into clinical remission, or at the very least, low disease activity (LDA). Disease activity is quantified by various clinical scores. The Clinical Disease Activity Index (CDAI) is widely used and considers the number of tender and swollen joints, as well the patient and clinician assessment of disease. Another widely use measure of disease activity is the Disease Activity Score (DAS) 28 of which there are several variations. The number of tender and swollen joints is always considered, and other inflammatory markers such as ESR and CRP can be interchanged. The T2T strategy involves changing treatment if disease activity is not improving within a designated duration[82]. The T2T approach can be adapted to any medication, which may vary from country to country, and it encourages an accelerated approach at treatment initiation, which has shown to be optimal in RA treatment for the long term[83] (Figure 1.2). The RA treatment regimen in newly diagnosed patients is particularly structured and during the periods of drug assessment, irreversible joint destruction often occurs in the individuals that are not responding to therapy. Currently, determining 1st line therapy does not consider underlying molecular mechanisms of disease but is based on clinical symptoms in addition to economics. bDMARDs are considerably more expensive than csDMARDs and are therefore only given in cases of poor efficacy or intolerability[76].

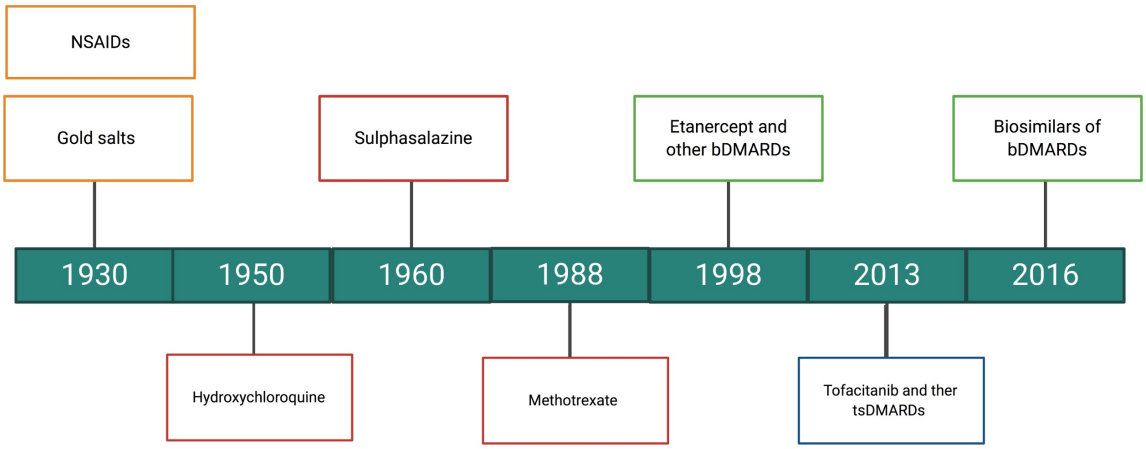


Figure 1.1 RA treatment history
Schematic illustrating the timeline over which drugs were implemented in the treatment of RA. Adapted from reflections on 'older' drugs: learning new lessons in rheumatology.[84]
bDMARDs, biological disease modifying anti-rheumatic drugs; NSAIDs, non-steroidal anti-inflammatory drugs; tsDMARDs, targeted synthetic disease modifying anti-rheumatic drugs

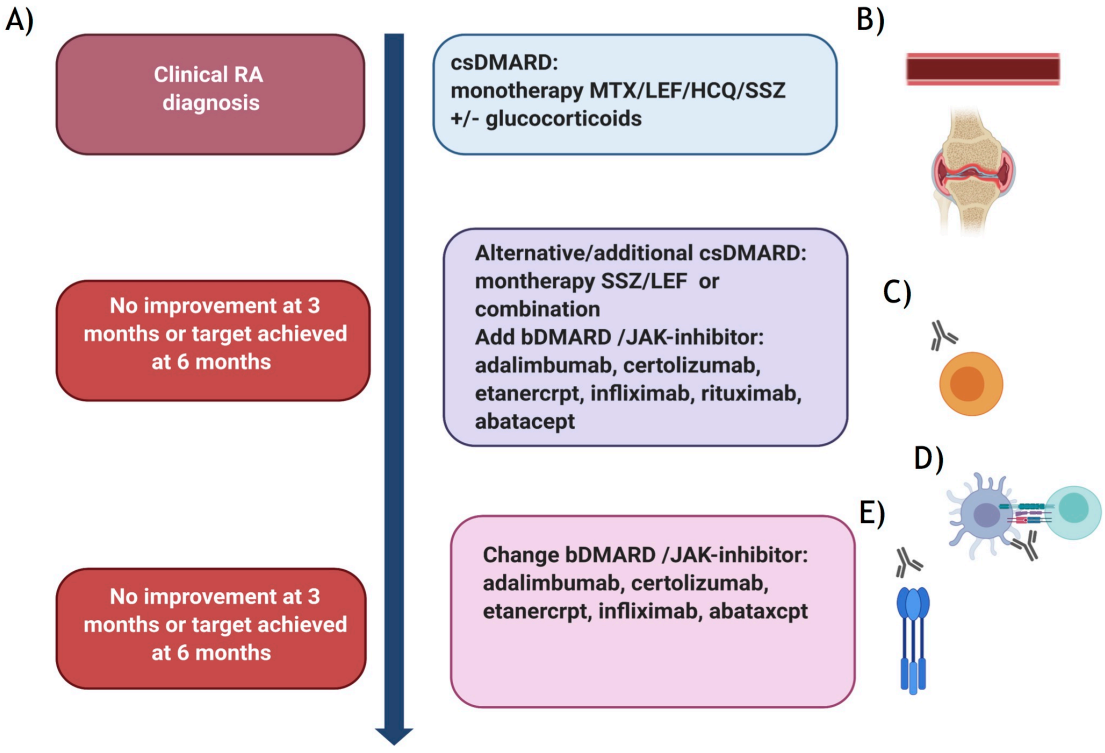


Figure 1.2 T2T Strategy
Schematic demonstrating current T2T strategy in the treatment of RA, A) T2T strategy, clinical diagnosis as defined by ACR/EULAR 2010 criteria, desired target is remission, or if remission not possible, low disease activity. B) Representation of targeting overall inflammation and lack of specific target. C) representation of rituximab targeting CD20 on B cells. D) Representation of etanercept targeting CTLA4 molecule. E) representation of anti-TNF targeting TNF molecule. Adapted from EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying anti-rheumatic drugs: 2019 update.[82]

bDMARD disease modifying anti-rheumatic drugs; csDMARDs disease modifying anti-rheumatic drugs; HCQ, hydroxychloroquine; JAK, janus kinase, LEF, leflunomide; MTX, methotrexate; RA, rheumatoid arthritis, SSZ, sulphasalazine

1.2.2 Current Treatment

1.2.2.1 csDMARDs

The nomenclature for RA drug treatment was updated in 2014. Conventional Synthetic DMARDs (csDMARDs) are the most common first-line therapy in the treatment of RA. This group of drugs comprises of methotrexate (MTX), hydroxychloroquine (HCQ), sulphasalazine (SSZ) and leflunomide (LEF). These can be given with or without glucocorticoids[85]. This group of drugs does not have a specific therapeutic target but aims to reduce overall inflammation.

1.2.2.2 Methotrexate

MTX, once known as amethopterin, has been used in the treatment of RA since FDA approval in 1986 and is now considered the 'anchor drug' in RA[86,87]. MTX can be found on the list of the WHO essential medications due to the number of uses for which it was not the original indication[87]. This drug was originally used as an anti-cancer agent and the exact mechanism in the treatment RA remains unclear, despite 40 years of use in this indication[88]. The mechanism of action in the treatment of Leukaemia, and the most documented mechanism of action in RA is folate antagonism, by inhibiting the enzyme dihydrofolate reductase (DFTR), essential for production of folate. For leukaemia treatment, MTX has to be given in high doses, however at lower doses it was found to be effective in RA[89]. This mechanism reduces purine synthesis and thus causes cell death. This has been shown to have effect on T cells[90]. Folic acid supplementation in RA is crucial, to compensate for the reduction in folate[91]. Many studies have explored the addition of folic acid on efficacy of MTX and it was found to have little effect, suggesting that folate inhibition is not the only mechanism by which inflammation is reduced[87]. Adenosine signalling and reduction of reactive oxygen species are other suggested mechanisms of MTX in the treatment of RA[88].

MTX has become an anchor drug due to its superior efficacy and tolerability compared to other csDMARDs[92] and is now the most popular 1st line treatment in the world[93]. It is often given as a monotherapy, but in some cases with the addition of other csDMARDs, HCQ or SSZ[94,95] and can be given either orally, or subcutaneously. While MTX is considered an effective therapy in RA, it is a drug with many side effects and some patients need to change drug due to intolerability. One of the most commonly documented side effect is problems with the gastrointestinal (GI) tract, with approximately 20-70% of people on MTX experiencing them[96]. Nausea is the most common GI tract side effect, with vomiting and abdominal pain also widely documented. Interestingly, Calasan and others have shown that MTX GI side effects can be associative or anticipatory[96]. Other known side effects include infections, as well as toxicity in the pulmonary, haematological and hepatic systems[97]. Side effects may contribute to non-compliance with medication, and thus contribute to the known efficacy of the drug. A recent study explored MTX adherence and found that often it is not optimal and various demographic factors lead to non-adherence[98].

While effective in a large proportion of recipients, MTX is not effective in every individual, with around 30% of patients switching drug within 12 months, some due to inefficacy, and some due to side effects[99]. A benefit of using MTX as first line, and an anchor drug with other medications, is the cost-effectiveness[100]. With MTX being unsuitable in approximately 30%, a biomarker for MTX response would be extremely valuable. Furthermore, as discussed previously, it is well established that the sooner effective therapy is initiated, the better the long-term outcome for the patient[77]. Currently, there is an assay that exists to try and determine MTX response in patients. This assay is used to genotype thiopurine S-methyltransferase (TPMT) which is involved in the MTX metabolism pathway. However, this assay has yet to achieve widespread clinical implementation[101]. This is due to evidence of interaction with other medications and varying results between laboratories[102].

1.2.2.3 Hydroxychloroquine

HCQ is another of the csDMARDs used in the treatment of RA and is another drug with a previous original indication. HCQ was first used in the treatment of

malaria. However, after it was found to have immunomodulatory properties, it was used in the treatment of autoimmune conditions such as systemic lupus erythematosus (SLE) and RA[103]. HCQ was developed though chemical adjustments to Atabrine, one of the first antimalarial drugs, after Atabrine caused severe side effects. HCQ has been found to be less efficacious than MTX or SSZ, and thus is rarely given as monotherapy, unless it's in cases of more mild disease or in combination with the other csDMARDs[103]. Similar to MTX, the mechanism by which HCQ reduces inflammation in RA is not fully understood[104]. HCQ is known to inhibit ion channels and Schroeder *et al* have shown that HCQ can inhibit calcium dependent potassium channels, which may lead to impaired inflammasome activity[104]. Another study has demonstrated the ability of HCQ to supress the inflammatory responses of class-switched B cells[105]. HCQ achieved this effect via TLR9, providing evidence of HCQ on toll-like receptors. T follicular helper cells have also been shown to be influenced by HCQ treatment[106]. Many studies have found benefits of HCQ on comorbidities of RA associated with the cardiovascular and metabolic systems[107]. In a mouse model of arthritis, HCQ had protective effects against atherosclerosis and a human population-based study demonstrated HCQ use was associated with a decreased risk of coronary artery diseases compared to non-users[108,109]. HCQ also has associated side effects, one of the main effects being retinopathy. Studies remain ongoing to determine an effective dose of HCQ that will not increase risk of retinopathy[110]. Like MTX, as HCQ is used in the early stages of disease, ability to predict drug response before use would be beneficial.

1.2.2.4 Sulphasalazine

SSZ is a csDMARD made in 1930 for use in the treatment of ulcerative colitis (UC) and RA. SSZ is a drug which combines the antibiotic sulphapyridine and the NSAID 5-aminosalicyliac acid[111]. It has been found to have similar efficacy to MTX, however is normally given in triple combination with MTX and HCQ[112]. One of the mechanisms of action of SSZ is inhibition of platelet thromboxane[111]. SSZ has been known to induce sulphonamide hypersensitivity reactions in people with RA, higher than that in the normal population[113]. Like the other csDMARDs in RA, the mechanism of action is not fully comprehended. One study has illustrated the effect of SSZ on endothelial cell chemotaxis[114]. This study also

demonstrated expression of *ICAM-1*, *IL-8* and *MCP-1*, known genes involved in angiogenesis[114]. While SSZ is efficacious in UC, the drug effects on the humoral immune system in RA is not clear. One study has suggested that in RA, SSZ exerts its effects systemically[115]. This study illustrated reduction of IL-6, in serum of patients treated with SSZ. It is thought this contributed to the serum reduction on IgA and IgM[115]. These results came after a previous study showed similar results which revealed a reduction in IL-6 levels, 4 months post-SSZ treatment, which correlated with a reduction in disease activity[116].

1.2.2.5 Combination therapy

The csDMARDs, MTX, HCQ and SSZ are sometimes given in conjunction to treat RA, referred to as triple therapy. If 1st line monotherapy MTX is unsuccessful, as an alternative to bDMARDs, a combination of csDMARDs can be used. This approach has been shown to be more efficacious that MTX alone or combination of HCQ and SSZ[117]. Several clinical trials have tested whether triple csDMARD therapy is better or worse than MTX with the addition of the bDMARD, etanercept and have found them to be comparable in treating RA[118]. In terms of drug cost, triple csDMARD therapy is superior. However, when assessing Quality Adjusted Life Years (QALYs), csDMARD triple therapy is inferior to the MTX-Etanercept combination[119]. These two treatment strategies have also been studied to understand whether other combinations are less likely to cause side effects such as infections and GI disturbances. A study by Quanch *et al* revealed that those treated with triple therapy were more likely to suffer GI side effects than those with MTX-etanercept, however they were less likely to suffer from infections[120]. Overall, the literature demonstrates that both these approaches work well for the treatment of RA, and in such a heterogeneous population, both approaches are beneficial. Yet there is still space for improvement as not all will respond to triple therapy.

1.2.2.6 bDMARDs

bDMARDs have been designed to target specific molecular structures involved in RA pathology[121]. The first bDMARD was developed against TNFα. A neutralising, chimeric monoclonal antibody was tested in a clinical trial and patients responded well to the treatment with considerable reductions in disease activity

measurements. The discovery of the efficacy of this type of treatment in RA paved the way for other bDMARDs targeting other molecules involved in inflammation[122]. Since then, 5 anti-TNFα medications have been licenced for use in the treatment of RA. Infliximab was the first of these to be used in patients[123]. This a monoclonal antibody that works by inhibiting TNFα from triggering the TNF receptor complex. Adalimumab is a monoclonal antibody that works in the same way. Both have been shown to have increased efficacy when given with MTX. Rituximab (RTX) is another widely used bDMARD used to treat RA. RTX is a monoclonal antibody against the CD-20 molecule, which is found on B cells. RTX uses a variety of mechanisms to deplete B cells, including apoptosis and complement-dependent cytotoxicity[124]. Complete B cell depletion occurs in the blood, but B cells in synovial tissue and bone marrow are only depleted in part. As such, response to RTX has been found to correlate with B cell levels in synovial tissue[125]. There is also a bDMARD that targets IL-6. Tocilizumab (TCZ) is a monoclonal antibody which targets soluble and membrane bound IL-6 receptor[126]. This stops IL-6 binding to the receptor and the signal transducer glycoprotein 130 complex, stopping downstream activation of the Janus Kinase signal transducer and activator of transcription (JAK-STAT) pathway. While bDMARDs have dramatically changed the treatment of RA, they can cost up to 10x more than csDMARDs[127]. As such, it is not cost-effective to initiate bDMARD therapy for everyone in early disease since csDMARDs are efficacious in many. This emphasises the need to find a way to establish who would benefit from csDMARD therapy at the beginning of treatment.

1.2.2.7 tsDMARDs

There is now a new drug class called the targeted synthetic DMARDs (tsDMARD), which target other small molecules involved in pathogenic pathways in RA. One such pathway is the JAK-STAT signalling pathway. This pathway is essential for many cytokines[128]. Upon cytokine stimulation, JAKs are activated and phosphorylate STATS. There are 2 drugs that target this pathway currently in clinical use. Tofacitinib primarily targets JAK1 and JAK3 family members *in vivo* [129] while baricitinib provides reversible inhibition of JAK1 and JAK2 family members[130]. Blocking of these kinases effects the downstream cytokines IL-2, IL-4, IL-9, IL-15 and IL-21. Tofacitinib has shown efficacy and safety and been

effective as monotherapy or in combination with MTX. The RA-BEAM clinical trial investigated the safety and efficacy of baricitinib compared to placebo and adalimumab[131]. Baricitinib achieved a reduction in disease activity, with more patients having an ACR20 response over placebo at 12 weeks. Baricitinib also inhibited radiographic progression of joint damage. The emergence of the tsDMARDs has provided an alternative to the conventional DMARDs and may be useful for treatment in those who are unresponsive to the csDMARDs.

1.3 The Human Genome

1.3.1 Genome Organisation

The size of the human genome is approximately 3100 million base pairs (bp)[132]. Packaging this DNA into 6μm of a cell nucleus presents a significant structural challenge[133]. As such, the DNA is packaged into a highly-organised structure (Figure 1.3)[134]. This intricate organisation not only serves a structural purpose, but a functional one. DNA has three main layers of organisation within the cell[135]. The baseline structure of DNA is the well-known double-helix, consisting of 4 bases, joined by a hydrogen bond with a sugar-phosphate backbone (Figure 1.3A)[136]. This double-helix is then wrapped around proteins known as histones. Multiple histones create nucleosomes, first described in 1974, which can be considered the core building block of the genome (Figure 1.3B). A nucleosome consists of 147 bp of DNA wound around an octamer of histone proteins[137]. This octamer is made from 2 of each type of histone protein: H2A, H2B, H3 and H4[138]. The DNA can then be wound into a higher order structure called chromatin (Figure 1.3C). This level is often referred to as 'beads-on-a-string', with the beads representing nucleosomes. The chromatin itself is then looped in a functional manner (Figure 1.3D). Chromosomes are then arranged into topologically associated domains (TADs), which facilitate increased DNA interaction between genes within a TAD by regulating enhancer-promotor contacts (Figure 1.3E)[139]. The boundaries of TADs are generally made up of highly-expressed genes. While TADs promote gene expression with genes in close-proximity, chromatin is also able to be regulated by features from a far topological distance. The dynamic chromatin structure allows regulatory factors to access the chromatin only when required and ensures there is no unrestrained

gene expression[137]. Finally, the DNA forms chromosome territories. The location of the chromosomes in these territories may resemble well-recognised positions[132]. A technique named chromosomal painting helped visualise these chromosome territories[140]. These studies demonstrated that genes on one chromosome interacted with genes on the same chromosome, more than they would interact with genes on another chromosome. In more recent years, these findings have been replicated with higher resolution technologies[141]. Euchromatin refers to chromatin in an open conformation, thus facilitating gene transcription, while heterochromatin is condensed often leading to gene repression[142].

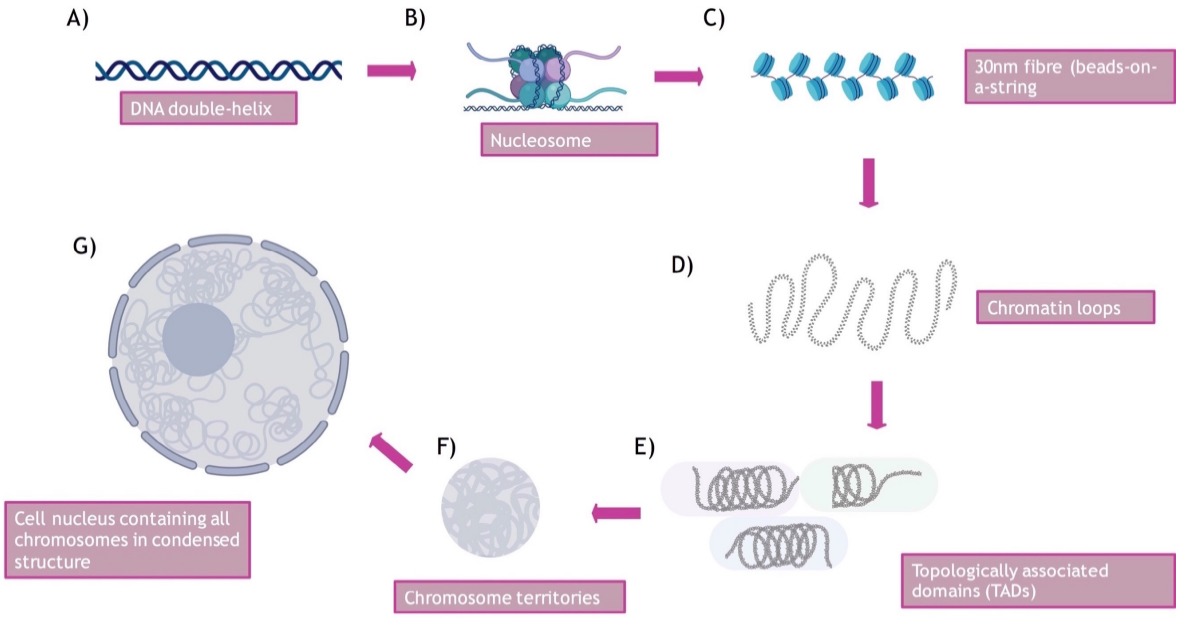


Figure 1.3 Schematic of Genome Organisation
Schematic illustrating the multiple layers of genome organisation within a cell. A) DNA double-helix. B) Nucleosomes composed of 147bp of DNA wrapped around 8 histone proteins (2x H2A, H2B, H3 &H4). C) 30nm fibre chromatin (beads-on-a-string). D) Chromatin loops. E) Topologically associated domains (TADs) consisting of chromosomes. F) Chromosome territories which form the 23 chromosomes in the human genome. G) DNA organised within the cell nucleus

1.3.2 Epigenome

The genome is packaged into every cell in the human body, yet cells can be phenotypically different. This can be explained in part by epigenetic mechanisms. Historically, epigenetics was defined as changes to the DNA that do not alter the DNA sequence itself, and that result in a stable, heritable

phenotype[143]. The definition of epigenetics has become more diverse and it is now considered that epigenetics encompasses changes to the chromatin, that may involve addition or removal of proteins, or changes to the chromatin structure itself[144]. Others have described epigenetics as 'the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states'[145]. It appears that the evolution of the definition focuses on the influence of structure, and consequential influence on gene function. In the 1990s, research on imprinting genes introduced DNA methylation into the 'epigenetics' definition, followed later by histone modifications[146]. Epigenetic changes include DNA methylation, histone modification and changes to the chromosome architecture[147].The tails of histone proteins are often susceptible to these epigenetic changes in the form of post-translational modifications such as methylation and phosphorylation. These modifications can have a direct effect on the structure of the chromatin which then results in the overall gene expression and thus phenotype, and studies have been able to illustrate this[148].

The epigenetic process is highly complex and is reflective of the environment's interaction with the genome [149]. Epigenetics have significant influence on cellular processes and often vary between different cell types[150]. With many unanswered questions left in genome research, many propose these questions could be answered from findings in the growing epigenomic research domain. Recently, epigenome wide association studies (EWAS), which are similar to genome wide association studies (GWAS) have been used to explore the genome for epigenetic impact in disease[151]. EWAS has an additional level of complexity over GWAS due to the dynamic, reversible nature of the epigenome. It is also believed that studies of the genome which have taken place already could be enhanced with additional epigenetic data[149]. Some epigenetic changes are associated with increased gene expression but in contrast some are associated with repression of gene expression. In general, both epigenetic 'writer' and 'eraser' proteins have been described that control these changes in gene expression. 'Writer' proteins cause changes to gene expression and transcription and 'eraser' proteins remove these epigenetic modifications[24].

Research has shown that epigenetic changes can be a contributory factor in the development of many autoimmune diseases[147]. In this case, epigenetic

modifications can influence processes such as immune cell function. Epigenetics could reveal a link between the known combined influence of genes and environment in RA[24]. The rheumatology field has seen considerable advances in epigenetics in recent years due to the development of many technologies allowing high-throughput analysis of data[152]. The epigenome is susceptible to changing characteristics of disease, as well as different therapies used to treat conditions such as RA. This presents a wide spectrum where epigenetics could be consulted for precision medicine application[153]. It is widely believed that understanding these mechanisms will contribute to the better management of RA in the future[154].

1.3.2.1 Methylation

DNA methylation, with the most technologies available to explore, is the best understood epigenetic modification[155]. Methylation is generally considered to be a stable epigenetic modification which is also heritable and can be a mechanism of regulation in cellular processes such as differentiation[153]. This epigenetic change has been well documented and has been observed in autoimmune diseases such as RA. DNA methylation is caused by the addition of a methyl group to the cytosine or adenine at position 5. Methylation of DNA is prevalent in several cell types involved in RA pathogenesis with synovial fibroblasts being one type affected[156]. The consequence of DNA methylation varies and can affect repression of transcription which can result in disease pathology. Glant *et al* performed one of the first studies into epigenetic modifiers in RA. This was a genome wide methylation profiling study which took place in PBMCs. The study indicated that methylation changes at the MHC locus increased the risk of developing RA[154]. It was shown that enzymes that can modify chromatin were found in genes that are known to be expressed in RA. These include acetyltransferases, methyltransfersases and histone kinases[154]. Another study has illustrated the major impact that one minor epigenetic change can have in RA. They showed that methylation at a single site in the promoter region for CTLA-4 in regulatory T cells (Tregs) could ultimately result in the failed activation of the immune modulatory kynurenine pathway[157]. It has also been recognised that methylation has an influence in RA by developing apoptosis resistant FLS[80]. Many of the methylation studies carried out to this day have lacked substantial numbers and have therefore been considered preliminary.

1.3.2.2 Histone Modification

Histone modifications refer to the post-translational addition or removal of proteins on the histone NH-2 terminus, or histone ‘tail’. A histone code hypothesis was proposed in 2000, which suggested patterns of these modifications could influence downstream biological processes in different ways[158]. Histone modifications include methylation, acetylation, phosphorylation, ubiquitination and sumoylation. Depending on the combination and number of these modifications, genes can become 1 of 4 states that are termed active, poised, bivalent or repressed. These states have been shown in

studies in $CD8^+$ T cells[159].

The functional implications of methylation on histones has been widely investigated, and has shown methylation can result in repression or activation of genes. Methylation occurs on lysine or arginine residues on the histone tail. Mono-, di or tri-methylation has been shown to indicate enhancers; in contrast trimethylation of H3K27 (H3K27me3) which is a known repressor mark. Histone methlytransferases are the enzymes responsible for facilitating the transfer of methyl groups. Acetylation of histone tails is widely considered to be a mark of gene activation. H3K27ac is a well characterised histone mark found enriched at active enhancer sites[160]. Histone modifications of the genome are regulated by 2 enzymes, histone acetyltransferase (HAT) and histone deacetylase (HDAC)[161]. Phosphorylation occurs to threonine or serine residues and is another modification known to be associated with activation[162]. Ubiquitination occurs to lysine residues and can be associated with either transcriptional repression or activation.

Many studies of histone modifications in RA have taken place in the synovial compartment. Studies in synovial fluid have illustrated reduced HDAC activity in RA compared to healthy controls[163]. Research in PBMCs has shown there is an alternative equilibrium of these enzymes in PBMCs of RA patients compared to the healthy population[161]. A study by Gillespie *et al* illustrated increased levels of HDAC in PBMCs[164]. Levels of these enzymes can give an indication of the levels of transcription of cytokines responsible for inflammatory responses in RA. Research by Toussirot *et al* has suggested that different RA therapies can exert varying epigenetic modifications in the form of histone acetylation and deacetylation[24]. In this study they showed that TNFα inhibitors such as Infliximab increased histone acetylation in the nucleus but alternatively, RTX, increased both acetylation and deacetylation enzymes[161]. Despite this research, the consequences of the changing levels of these enzymes with RA treatment remains to be fully investigated.

1.3.3.3 Chromatin loops

Chromatin architecture is the overarching epigenetic feature to the marks already described. As discussed, chromatin loops can be considered the third

level of organisation within the nucleus (Figure 1.3D). Chromatin loops offer an interesting mechanism to investigate the epigenome, as they can encompass methylation sites, histone modifications and miRNAs and can regulate how these features interact with each other. Study of chromatin loops clearly demonstrates the intricate relationship between genome structure and function. Simply, chromatin loops are formed when 2 parts of the genome, separated by an intermediate DNA sequence, are brought into close physical proximity[165]. Gene regulatory sequences are often not located beside the gene they control. When necessary, loops allow promotors and enhancers to be brought into proximity to a specific gene to permit activation and transcription[165]. Studies have demonstrated this, showing DNA is enriched with chromatin loops at active enhancer and promotor sites and are less are likely to be found at inactive sites or sites with histone modifications that cause repression[166]. Research into the drosophila genome indicated that loops were approximately 80kb in size, and comprised of 400 nucleosomes on average[167]. It was once considered that chromosome looping could only occur in *cis*, within a chromosome. However, studies have demonstrated regulation of a gene could occur from regulatory elements located several megabases (mb) away[168]. These larger distance, interchromosomal interactions are known as *trans*.

Investigations into the ß-globin cluster were the first to provide insight into distal regulation of the genome. The interest in the ß-globin loci, mainly due to its involvement in the blood disorder Thalassemia, provided the opportunity to discover the role of chromatin looping in human gene regulation[169]. Deletions in the DNA far away from the location of the ß-globin gene still resulted in the development of Thalasemmia alluding to the role of distal regulation[170]. It was through this finding that the locus control region (LCR) was found. This is a group on ß -globin gene arranged in a way on the chromosome that facilities development in a timely manner. A study by Cater *et al* was the first to show evidence of chromatin looping in LCR-ß-globin gene contact[171].

One of the first demonstrations of interchromosomal interactions was in the alternative expression of cytokine genes. The study by Charalampos *et al* revealed that dynamic chromatin organisation allowed the promotor of the *IFNy*, located on chromosome 10 to interact with regulatory elements of *IL-4*, located

on chromosome 11. This interaction has implications for the polarisation of $CD4^+$ T cells to Th1 or Th2[172,173].

Following on from these studies, proteins which are involved in the formation of chromatin loops were discovered. These are called boundary elements, or insulator proteins[174]. CCCTC-binding factor (CTCF) is one of the best documented insulator proteins[175]. It has been found to be a highly conserved nucleic acid binding protein, with approximately 40,000 binding sites throughout the human genome[176]. CTCF has been found to separate TAD boundaries[177]. The 11 zinc fingers of CTCF means it can interact with DNA in various ways. Studies have explored this binding and identified two motifs, M1 and M2. The M1 motif engages 4-7 zinc fingers, and M2 is found upstream with 9-11 zinc fingers. Where both these motifs can be found, it has been shown that CTCF is bound to DNA with very high affinity[178]. Along with insulator function, CTCF can facilitate chromatin looping. Tens of thousands of the CTCF sites throughout the genome have been found to be co-occupied by a protein complex called cohesin[179]. Cohesin is a ring-shaped complex made from multiple proteins that plays a role in DNA replication. The cohesin ring ensures chromosome segregation during mitosis and meiosis, protecting the genetic information that gets passed on[180]. It was also discovered that cohesin can bind to CTCF and facilitate chromatin looping[181].

In most cases, chromatin loops are believed to support gene transcription by priming genes to contact their promotors, however loops have also been known to play an inhibitory role[165]. Loops can also serve a purpose of bringing a promotor into proximity with its terminator. This has been demonstrated in a study with the breast cancer associated, *BRCA1* gene and the maternal *Igf2* gene. This should a chromatin loop can ensure the gene promotor is kept separate from its enhancer[182,183]. Regulated DNA architecture has also been shown to have a role in DNA repair[184]. Research has suggested that these chromatin loops, or when taken together known as chromatin conformation signatures (CCS), are more informative and stable epigenetic marks than other alterations to the genome[148].

There is evidence in the literature that demonstrates chromosome loops can both be stable and dynamic structures. Challenges in understanding these two positions

lies with the technologies currently available. While high-resolution 'C' technologies have their advantages, they can only provide a snapshot of the epigenome[132]. Studies with CTCF demonstrated that while CTCF and cohesin facilitate stable chromatin loop structures, they are dynamic and the conformation can be lost when CTCF and cohesin disassociate[185]. Studies of the human pituitary growth hormone have demonstrated patterns of stable of human gene expression, as facilitated by DNA architecture[186]. The reproducibility of TAD maps in studies has also strengthened the hypothesis that chromatin loops are stable. However, live cell imaging has illustrated the dynamic nature of loops. Overall, evidence suggests most chromatin is stable for a short period of time, before transforming. Many questions remain to be answered, particularly chromatin dynamics in the context of disease.

1.4 Precision Medicine

Precision medicine is a concept of basing clinical decision on measurable molecular biomarkers. It could be argued that medicine has always aimed to be 'precise', and this has been successfully demonstrated in blood transfusion and organ transplantation for decades. However, it has only been in recent years that the genome has been investigated for clues to prognosis of disease or treatment response. Treatment of the individual is the ultimate aim for clinicians, yet due to the nature of current clinical trials that are catered to a population, this can prove difficult[187]. The completion of The Human Genome project in 2003 significantly contributed to the explosion of genomic exploration [188]. The percentage of the genome that contributes to drug response is thought to range between 20% and 95%[150]. Studies involving this type of genetic exploration combine three important areas: the right population, the suitable technology, and finally, the collection of data[150]. Biomarkers that are identified can be incorporated into algorithms to predict prognosis or response to treatment for patients. Studies have shown that it is becoming increasingly straightforward to interrogate the genome but the translation of important findings into the clinic has proven challenging. This type of research has led to the current era of 'big data' with large datasets, which incorporate genomic information and patient characteristics. Some have described this time as a "biomarker revolution" and this has resulted in approved biomarkers in some

cancers. Despite this, it is thought that much of the germ line genetic variation available currently is not suitable for the implementation of precision medicine clinically. GWAS and identification of single nucleotide polymorphisms (SNPs) were one of the first types of study in precision medicine. These look to identify variant alleles which are associated with disease[189].

With the advancement of many genomic technologies, determining biomarkers is becoming easier than before. For a biomarker to be successful there is a certain number of criteria that must be met and studies must successfully illustrate this. For a biomarker to be clinically useful it must be consistently accurate, easily quantifiable, easily replicated and economically viable. Importantly, the biomarker to diagnose disease or predict treatment response must be superior to any existing methods[190]. Biomarker kinetics is an area of the precision medicine field that has not been investigated by many up until now. It is important to understand that biomarkers are dynamic and can change over time. Furthermore, enough statistical power is fundamental in biomarker studies. This means that specificity and sensitivity must be high enough i.e. as little as possible false negatives or false positives results, respectively[190].

1.4.1 Precision Medicine Technologies

1.4.1.1 Microarray

Microarrays have existed since 1995 when they were first documented by Schena *et al*[191]. Many technologies have been developed since, yet microarrays are still used today, and some consider them to be one of the fastest growing genomic technology[192].The DNA microarray provided a more straight-forward and high-throughput way to investigate the genome than normal sequence analysis, and could be termed sequencing by hybridisation[193]. Microarrays can be used to explore differences in gene expression, aiding biomarker research. The three main types of array are DNA, RNA and protein. The basis of the array is complementary hybridisation of DNA from a sample, to short complimentary probes printed in large numbers on a chip. The first microarrays conducted by Schena and others used complimentary DNA on a glass slide, however, now shorter oligonucleotides can be used which have a higher specificity[194]. This reaction creates images which can be analysed. There is both an *in-silico*, and

'wet lab' approach involved in microarray experiments. The *in-silico* part is involved in the array design, with the 'wet lab' part applying the desired samples to the microarray. One of the two main approaches of making DNA arrays are light-directed chemical synthesis and microarray spotting[195]. The light-directed chemical synthesis was first documented by the founder of Affymetrix in 1991 for a peptide array. This led to the generation of the first oligonucleotide array prepared in this way[193]. The first array was 1.25cm^2 in size and was printed with 256 oligonucleotides. Microarray technology is still being developed and it expected that the platforms for microarrays will be reduced in size, creating 'nanoarrays'[196].

1.4.1.2 3C technologies

In the last decade, analysis of the complex chromosome architecture and the influence it has on gene expression has increased our understanding of the epigenetic influence in drug response. Epigenetic research has benefited from the progress of genomic technologies, and genomic architecture can now be visualised in enhanced detail. Originally, loops had to be visualised through laborious, lower–throughput, methods such as electron or light microscopy. Fluorescence *in situ* hybridisation (FISH) offers an opportunity to view multiple loops at one time, however, the protocol for staining may impact the chromatin conformation[197]. One method of this enhanced visualisation, named chromosome conformation capture (3C), first described in 2002, has allowed loops in DNA to be investigated[197] (Figure 1.4). 3C is based on formaldehyde cross-linking of proteins and DNA. Cross-linking will be achieved for areas of the genome that are physically touching. 3C measures the frequency with which areas of the genome are cross-linked. A restriction enzyme is then applied to the cross-linked DNA, followed by ligation. The cross-linked DNA will be more likely to ligate over non-cross-linked, i.e. physically touching DNA will be ligated. These ligated fragments are subjected to a PCR reaction and gel electrophoresis allows visualisation of ligated fragments, which were once loops in the genome. 3C can be used to understand spatial organisation within the genome, as well as interactions between regulatory elements[198]. Since the first 'C' technology was published in 2002, there has been a rapid expansion of 'C' technologies. Often 3C is referred to as a one-to-one technology, and other C technologies have

allowed one-to-all (4C) and all-to-all (Hi-C)[199] visualisation. 3C was firstly carried out in yeast, followed by the mouse human ß-globin loci[200]. RA 3C has also been used to demonstrate how chromatin looping regulates expression of Th2 cytokines IL-4 and IL-13 in T cells[201]. While many other 'C' technologies have overlapped the original 3C, it still has place in biomarker discovery and precision medicine implementation. The protocol for 3C is less complex and laborious, and therefore economical.

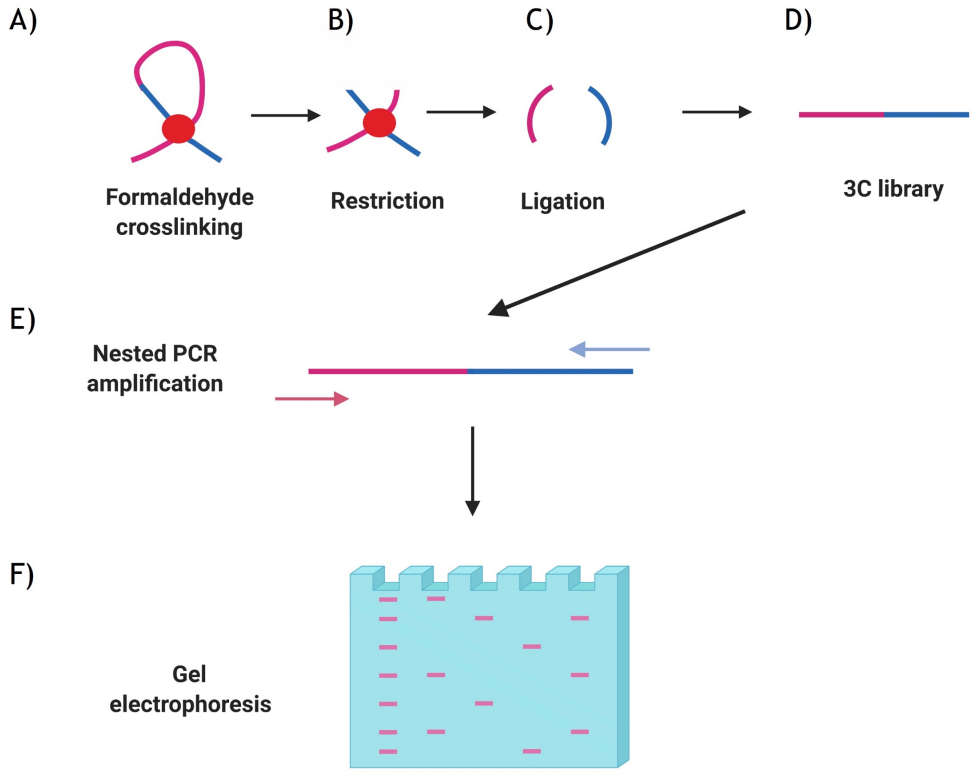


Figure 1.4 Schematic of Chromosome Conformation Capture (3C) protocol
Schematic representing the stages involved in the generation and visualisation of 3C DNA libraries. A) Formaldehyde is used to cross-link physically touching DNA. B) A restriction enzyme is used. C) a ligation enzyme is used to ligate the 2 DNA pieces together. D) a non-genomic, 3C template, representing a chromatin loop, is generated from the 2 physically touching DNA regions. E) Primers designed for both parts of the loop are used to amplify the template. F) PCR products are visualised on a gel, a band at the expected size represents the existence of the loop
3C; chromosome conformation capture; PCR, polymerase chain reaction

1.4.2 Liquid Biopsy

The 'liquid biopsy' is a concept most commonly known in the detection of cancer biomarkers, however is becoming more popular in other diseases such as Alzheimer's and autoimmune conditions[202,203]. Liquid biopsies use the blood as the source of genetic information. The blood holds a vast range of biomarker candidates including DNA, RNA, miRNA and circulating tumour DNA[204]. Blood samples as biomarkers provide advantages over tissue biopsies. These include a normally less invasive procedure and therefore more comfortable experience for the patient, and often less reliable on complex tissue imaging equipment[203]. Moreover, a blood sample offers the opportunity for additional testing which may lead to a more confident molecular analysis, and blood processing protocols are widely used across the globe in the clinical and industrial setting[202]. It has been suggested that liquid biopsies will provide a health economic benefit, for the ability to better provide earlier diagnoses and detect poor response to treatment[205]. In RA, it could be considered that the equivalent to a tumour biopsy is the synovial biopsy. This involves removal of a small part of the synovial lining in the joint. As with tumour biopsies, this is an invasive procedure and can cause discomfort, in an already inflamed joint. However, synovial biopsies have been used in precision medicine studies in RA. These have profiled lymphocytes, macrophages, FLS and cytokines from the synovium[206]. While these have yielded results, some researchers have transitioned to looking at liquid biopsy from the blood in the hope of identifying better predictive biomarkers. Circulating immune cells in RA can be considered as a liquid biopsy, and they have the potential to reveal much about the disease state. From an RA liquid biopsy, various methods have been applied such as gene expression profiling and immunophenotyping. Some promising results have been achieved from this work, such as the discovery that a group of interferon response genes could predict non-response to RTX[207] and the correlation of decreased circulating $CD28^+$ T cells with abatacept response[208]. Nevertheless, there is still no clinically validated biomarker being used to aid the treatment regimen in RA today.

1.4.3 Precision medicine in Rheumatoid Arthritis

It has been recognised that there are a significant lack of biological markers for RA[75]. Physicians are becoming increasingly aware that the genome or alternatively the epigenome could be consulted to improve the treatment regimen and long-term outcome for patients[80]. It is hoped that by increasing the understanding of the underlying molecular mechanisms of disease that biomarkers could be identified with the potential of predicting prognosis of disease or more excitingly the response to therapy. Ultimately, epigenetics alone is not responsible for the development of RA, it is a multifactorial disease influenced by the environment, risk genes and aging. RA is an extremely complex autoimmune disease and patients are subject to a unique combination of contributory factors which can alter their response to treatment, thus exemplifying the need for precision medicine. Despite the barriers and reservations to this, the transformational impact of precision medicine in oncology practices should hopefully pave the way for other disease areas which have the potential to see the benefit in the future. GWAS studies have successfully identified over 100 genetic loci that can be associated with RA. However, these loci do not always help to gain a better understanding of underlying disease mechanisms and therefore novel therapies are rarely produced. Moreover, cell types where changes in genetic loci exist cannot be identified through GWAS[209]. Although considerable GWAS have taken place in RA, there remains a large proportion of the heritable component of RA to be explained. The biggest challenge in genomic research in RA is linking different components together, *i.e.* matching the genomic data together in addition to potential proteomic and metabolomics data. Unique methods will need to be found to address this challenge. For precision medicine to be a success, researchers, healthcare professionals and industry representatives will have to collaborate successfully. The most plausible predictor of precision medicine clinically will not only include genetic information but will include other clinical markers and take into consideration epidemiological data.

Precision medicine approaches have been explored to predict prognosis, disease severity and treatment response. A recent study used '-omic' approaches to investigate a potential biomarker for pannus formation. They revealed

epigenetic alternations correlate with the aggressive RA phenotype, however exact mechanisms need to be clarified[210]. Predictive models using clinical and demographic characteristics have been investigated. Hyrich *et al* found males on concomitant MTX responded better to anti-TNFα therapy, etanercept and infliximab. Current smokers with high Health Assessment Questionnaire (HAQ) scores were predicted to be non-responders to anti-TNF therapies[22]. However, these characteristics alone would not be able to predict response with enough certainty. One study investigated the RA synovium for levels of TNFα transcripts. They found that high levels of TNFα transcripts correlated with high disease activity, and a worse response to first-line therapy[211]. This study alludes to the benefit of including synovial markers in the stratification of treatment in RA, however is dependent on standardisation of synovial biopsies. A more recent study by Humby *et al*. investigated cellular and molecular biomarkers from the synovium. They showed that in treatment of naïve patients, 3 synovial signatures existed in RA patients. These three subtypes were classed as lympho-myeloid, diffuse-myeloid and pauci-immune (few immune cells with dominant stromal cells)[212]. The discovery of these biomarkers was aided by immunohistochemistry methods, which is not the most high-throughput precision medicine tool.

Several studies have been conducted to find a biomarker for csDMARDs. One study analysed naïve T cell subsets in PBMCs from people with early RA. They found patients with a higher naive T cell frequency responded better to MTX than those with lower T cell frequencies. However, this study was limited by patient numbers[213]. A recent study combined demographic, clinical and psychological variables in an attempt to predict MTX non-response in the Rheumatoid Arthritis Medication Study (RAMS)[99]. This study aimed to capture the 'real-world' RA population. Limitations of this study include the high-level of non-response, which may be due to deviations from the normal RA treatment regimen of MTX escalation. The classification models did not achieve suitable sensitivity and specificity values[99]. Overall, the epigenetic research landscape in RA looks very promising. Costs of this research are decreasing dramatically, alongside increased throughput and resolution of genomic technologies[214]. It is quite likely the RA treatment regime will include DNA analysis in the future. This addition should ensure a much more positive outlook for RA patients worldwide.

1.4.4 EpiSwitch™

EpiSwitch™ is proprietary technology, developed by Oxford BioDynamics Plc (OBD) to facilitate the discovery of the blood based biomarker, specifically chromosome conformation. This platform has been used successfully to identify biomarkers in several cancers including thyroid cancer and melanoma, as well as neurodegenerative disorders such as Amyotrophic lateral Sclerosis (ALS) and Huntingtons disease[215 216-218]. This proprietary technology uses algorithms to predict sites in the genome where chromosomal loops are likely to occur. This differs to other CCS discovery, and by eliminating the need for a genome wide screen, allows more specific biological questions to be asked. The optimised discovery pipeline begins with the algorithmic approach for EpiSwitch™ sites, then identification of areas of the genome possibly implicated in the disease in question. A microarray platform is then utilised to observe chromosome conformation in samples of interest. This is followed by statistical analysis to inform about the most appropriate biomarker to take forward. Once these candidates are chosen, PCR primers are designed and PCR performed. Next, extensive statistical analysis is undertaken to find a CCS with the best potential for clinically relevant stratification. Finally, this can then be validated in an independent cohort. Importantly, this pipeline was successfully used to identify a 5-loop CCS with the ability to predict response or non-response to MTX in a treatment naïve RA cohort[219] (Figure 1.5). This technology aims to generate informative CCS biomarkers from discovery to clinical validation and ultimately clinical implementation. These biomarkers can be prognostic, diagnostic or predictive. If biomarkers can be used to stratify patients before entering into clinical trials, it is hoped that the success rate will substantially increase.

1.4.5 EpiSwitch™ in Rheumatoid Arthritis

The MTX response CCS is made up of 5 loci (*IL17A*, *CXCL13*, *IL21R*, *IL23* and *IFNAR1*)(Figure 1.5), known to be involved in RA pathogenesis[219]. These signature loci are primarily involved in cytokine and chemokine pathways. This biomarker could predict MTX non-response with 90% sensitivity. The discovery cohort was made up of 59 patients (30 responders and 29 non-responders), and the blinded-validation cohort, 19 patients. All patients were from the Scottish Early Rheumatoid Arthritis (SERA) cohort. This is a pan-Scotland, inception cohort of over 1,200 patients. Various clinical samples and information were taken and recorded from baseline, every 6 months. This biomarker was refined from a list of over 13,000 loop anchor sites across 309 genetic loci, many of which are known to be associated with RA. Statistical refinement reduced 100 to 30 loci. These were then reduced to the final 5. This study was the proof-of-principle that the structural epigenome could be used to predict MTX response in treatment naïve patients. This study opens the opportunity for investigating the relationship between chromatin conformation structure and function in RA, and a basis for validating this biomarker in other cohorts. While the biomarker discovery approach is considered robust, the sample number used in discovery and validation cohort could be considered small. Therefore, there is merit for exploring the signature in a higher number of patient samples. Moreover, the consequence or cause of these loops in RA patients is not known. Various methods could be used to shed light on this, which could reveal more about molecular biology underpinning MTX response and RA pathogenesis.

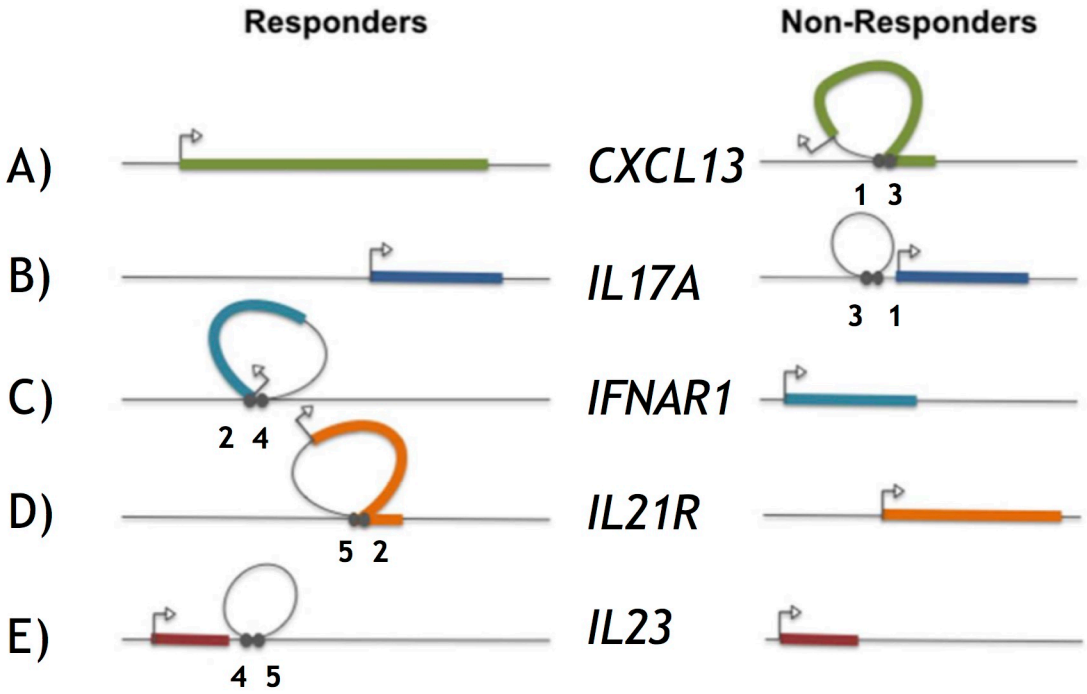


Figure 1.5 CCS for MTX response prediction
This 5-loci CCS has the capacity to differentiate responders and non-responders to MTX in treatment-naïve patients with RA with 90% sensitivity. Schematic of CCS is representing the signature in both response groups. Numbers represent EpiSwitch™ sites on the genome. A) *CXCL13* loop is present in non-responders, Chr 4. B) *IL17A* loop is present in non-responders, Chr 6.C) *IFNAR1* loop is present in responders, Chr 21. D) *IL21R* loop is present in responders, Chr 16. E) *IL23* loop is present in responders, Chr 12. For coordinates, see Appendix.

1.4.5.1 *IL17A*

IL17A is part of the *IL17* family, that has 5 other members (IL17B-F). *IL17A* signals through the *IL17* receptor on Th17 cells. The receptor exists as a heterodimer with *IL17C*[220].This heterodimer has been found on fibroblasts, endothelial and epithelial cells[221]. A number of cells from the adaptive immune compartment can produce *IL17A*, namely $CD3^+$, $CD4^+$ and $CD8^+$ T cells, NK cells, and Th17 cells[222]. *IL17A* production has several pathogenic implications in RA, including maturation of osteoclasts and fibroblast-like synoviocytes, as well as activation of macrophages, neutrophils and B cells[223]. Studies have suggested that presence of *IL17A* in RA synovium is a predictor of disease progression[224]. While *IL17A* blockade has been shown to be very effective in the treatment of autoimmune conditions such as psoriatic arthritis and psoriasis, blockade in RA has been less successful. Trials of secukinumab, an *IL17A* monoclonal antibody, did not have as profound therapeutic effects that other cytokine blockade

therapies have had, such as IL-6 and TNFα[223]. One study demonstrated that secukinumab was better than placebo in RA, but not superior to anti-TNFα therapy[225].

1.4.5.2 *CXCL13*

CXCL13 is a chemokine that belongs to the CXC family. This chemokine is chemotactic for B cells and interacts with the receptor CXCR5 on B cells to have its effect[226]. The role of this chemokine in B cell chemotaxis places it in position of RA pathogenesis. The levels of *CXCL13* in RA serum has been shown to high in both early and established RA[227]. Jones *et al* believe *CXCL13* to be correlated with RF in RA patients but show no correlation to other demographic or serological markers such as ACPA. One study has demonstrated that *CXCL13* works synergistically with CCL20 to recruit B cells to the synovium[228]. Lymphoid neogenesis is another process which implicates *CXCL13* in RA pathogenesis[229]. Several studies have been conducted which demonstrate *CXCL13* role in this process. *CXCL13* can also be produced by $CD4^+$ T cells, another cell type important in RA pathogenesis[230]. This study aimed to understand the mechanism by which *CXCL13* is produced by $CD4^+$ T cells. They found SOX4 was a fundamental transcription factor for this process and has associated this with the formation of FLS at inflammatory sites in human, such as synovium in people with RA. Similarly, Kobayashi *et al* demonstrated that $CD4^+$ T cells can produce *CXCL13* and are involved in ectopic lymphoid neogenesis at inflammatory sites[231]. The *CXCL13* receptor CXCR5, is also expressed on Tfh cells and it has been suggested that this essential for the development of RA. Interestingly, CXCR5 deficient mice are unable to develop Collagen induced arthritis (CIA)[232]. This study has shown the potential for targeting of the CXCR5 receptor for treatment in RA. As the only known ligand for the receptor is *CXCL13*, an antagonist for the receptor would have little pharmacological competition[232,233].

CXCL13 has shown promise of its predictive potential in several studies to date. Mainly, it has been identified as marker for predicting disease activity or potential outcome[234,28]. Largely, this may be due to the high levels of *CXCL13* in synovial tissue and fluid in individuals with RA[232]. Additionally, several studies

have attributed *CXCL13* to predictive capacity to TNFα inhibitors in the treatment of RA[235].

1.4.5.3 *IL21R*

IL21R belongs to the IL-2 cytokine family. The receptor is a heterodimer, consisting of an alpha and gamma chain which is shared by other cytokines including IL-2, IL-4 and IL-9[236]. The *IL21R* is structurally similar to IL-2R and IL-15R (Li et al., 2006). The IL-21 receptor can be found on multiple immune cells including DCs, NK cells, T cells and B cells[237]. IL-21 is mainly produced by $CD4^+$ T cells and NK cells and is proinflammatory in nature[238]. IL-21 signals through the *IL21R*, inducing the STAT pathway[239]. Activation of this pathway results in expansion of B cells and downstream production of antibodies, class switching and plasma cell differentiation[240,239]. *IL21R* expression has been found to be higher in RA and systemic sclerosis compared to controls[241]. More recently, *IL21R* has been enhanced in other inflammatory conditions such as tendinopathy[238]. *IL21R* has also been found to be upregulated in synovial tissues of people with RA but not osteoarthritis (OA)[241]. There have been studies investigating the efficacy of *IL21R* blockade in the treatment of RA and other autoimmune diseases. Animal models have demonstrated that blockade of the IL-21/*IL21R* pathway was effective in reducing RA disease activity as well as having an inhibitory effect on cytokine production in vitro[242]. Mouse models lacking *IL21R* were found to be unable to develop spontaneous autoimmune disease. Humoral immunity was also comprised in these mice, highlighting the role of IL-21 in antibody production[243]. A recent study demonstrated an increase of *IL21R* on naive and memory B cells in RA in comparison to healthy controls. This was associated with an increase in pSTAT3 levels. The increased *IL21R* was attributed to increased SP1[244].

1.4.5.4 *IL23*

IL23 exists on the Th17 axis along with IL-21 and *IL17A*. *IL23* is a member of the IL-12 family of cytokines[245]. The structure of the cytokine consists of 2 subunits: *IL23* p19, which is exclusive and the IL-12p40 subunit which is shared with *IL12*[246]. Antigen presenting cells, monocytes, macrophages and DCs are the cell types that produce *IL23* the most. When *IL23* binds to its receptor, *IL23R*, it activates the JAK-STAT pathway, specifically JAK2 and subsequent STAT3 and

STAT4[245]. The receptor is made up of *IL23* -R and IL-12R-αß1 complex[246]. This leads to the release of pro-inflammatory cytokines, *IL17A* and RORγt[247]. It is thought the role for *IL23* is maintenance, development and survival of Th17 cells via a positive feedback loop that involves TNFα, IL-6 and IL-1ß[247]. Studies involving the EAE model were crucial in revealing that the IL12-IFNy axis was not responsible for initiation of autoimmunity, but was in fact the IL17-23 axis[248,249]. This cytokine can be found in the synovial fluid and serum of those with RA, and studies have found it to be significantly higher than healthy individuals[250]. The levels of *IL23* in the serum have been shown to correlate with severity of disease activity[246]. The implication of *IL23* in inflammatory arthritis has been evidenced in the CIA model of arthritis. Overexpression leads to development of CIA while reduction is protective of CIA. This study showed that while *IL23* plays a role in the development of disease, once established, *IL23* has less of a role. This was shown by inhibition during disease not reducing disease severity[251]. *IL23* has also been shown to have a role in the production of autoantibodies, hence its role in the early stages of disease[251]. In addition to its role in disease onset, it has been suggested that *IL23* could play a role in disease flare. This is due to successful reduction of disease severity with blockade of *IL23* [251]. Studies have also shown *IL23* to have osteoclastogenic activity, contributing to the bone erosion in RA. While it's role in RA pathogenesis has been demonstrated, pharmacological targeting of *IL23* has been unsuccessful in showing any benefit clinically. Two antibodies have been tested in a stage II clinical trial. One was a monoclonal antibody targeting the *IL23* 12/23 p40 complex, and the other targeting the p19 subunit alone[252].

1.4.5.5 *IFNAR1*

The IFNAR receptor is the receptor for the antiviral cytokines, named interferons. *IFNAR1* and *IFNAR2* make up the single-membrane spanning IFNAR receptor which is ubiquitously expressed[253]. The IFN receptors act by increasing binding of ligands. Once the receptor is activated, intracellular signalling cascades are activated which results in the activation of the STAT pathway[254]. There are three types of interferons in humans, classed as type I, type II and type III. They all signal through the IFNAR receptor, with differing binding affinities[255]. Type I interferons are heavily involved in the regulation of both the

innate and adaptive immune response. Namely, macrophages, NK cells, B cells, T cells and DCs are aided in their differentiation and proliferation by type I interferons[256]. *IFNAR1* has been investigated in various autoimmune conditions. One study found some patients with RA have a higher proportion of interferon response genes compared to other patients[257]. Several studies have indicated the presence of a type 1 interferon signature in RA. One study has shown that one subgroup in RA, with upregulated activity in the innate immune system, complement cascades and fatty acid metabolism[258]. Another study reported *IFNAR1* blockade has been utilised in lyme arthritis[259]. The involvement of IFN signalling in the development of lyme arthritis was also studied using *IFNAR1* -/- mice. Severity of arthritis was reduced in the KO. Many cell types have been found to contribute to the IFN response, including primarily myeloid cells, resident in joint tissues, in addition to fibroblasts and endothelial cells[259].

1.5 Aims

Despite efforts, there is no molecular biomarker currently used at diagnosis to stratify RA patients and ensure they are on the right treatment from the outset. The development of technologies able to interrogate the genome, as well as the growth in biobanks, has made the study of molecular biomarkers for RA more accessible.

While the discovery of a biomarker for MTX response in the treatment of RA is an important and interesting finding, biomarkers require further validation until they can be considered for adoption into a clinical setting. Additionally, research has demonstrated that chromatin conformation reveals insight into gene regulation, therefore there is scope that this MTX CCS could increase understanding about the underlying mechanisms that dictate ability to respond, or not respond to MTX treatment. Moreover, with the EpiSwitch™ pipeline incorporating additional, more informative methods of biomarker discovery, there is the potential of discovering additional CCS with the capacity to further stratify the RA population.

This body of work aimed to:

1) Validate the MTX CCS bioinformatically and in an independent clinical cohort, as well as establish the efficacy to predict response to other csDMARDs.

2) Understand the relationship between CCS and disease pathogenesis.

3) Determine if underlying epigenetic endotypes exist in the early RA population and if chromatin loop candidates exist to predict these.

Chapter 2 Materials and Methods

2.1 Patient Cohorts

2.1.1 Patient Identification - SERA

The Scottish Rheumatoid Arthritis (SERA) cohort is a pan-Scotland inception, longitudinal cohort of patients with early rheumatoid arthritis (RA). Samples were obtained at 6-month intervals, from baseline of treatment. All patients were conventional synthetic disease modifying anti-rheumatic drug (csDMARD) naïve at baseline. Healthy samples came from demographically matched friends or family of enrolled patients. Patients of interest in this study were identified by their response to DMARD therapy. This was done by calculating disease activity using clinical disease activity index (CDAI) and disease activity score 28 (DAS28) measurements at baseline, 6 months and 12 months. These calculations take into consideration swollen (SJC28) and tender joint (TJC28) counts from 28 joints (Figure 2.1). Patient assessment of disease activity (dasVAS) and physician (GlobalVAS) assessment of global health from a visual analogue scale (VAS) of 0-10cm is also included. The closer to 10cm on the scale, the worse the disease activity. Some DAS scores take the inflammatory markers, erythrocyte sedimentation rate (ESR) and c-reactive (CRP) protein into account.

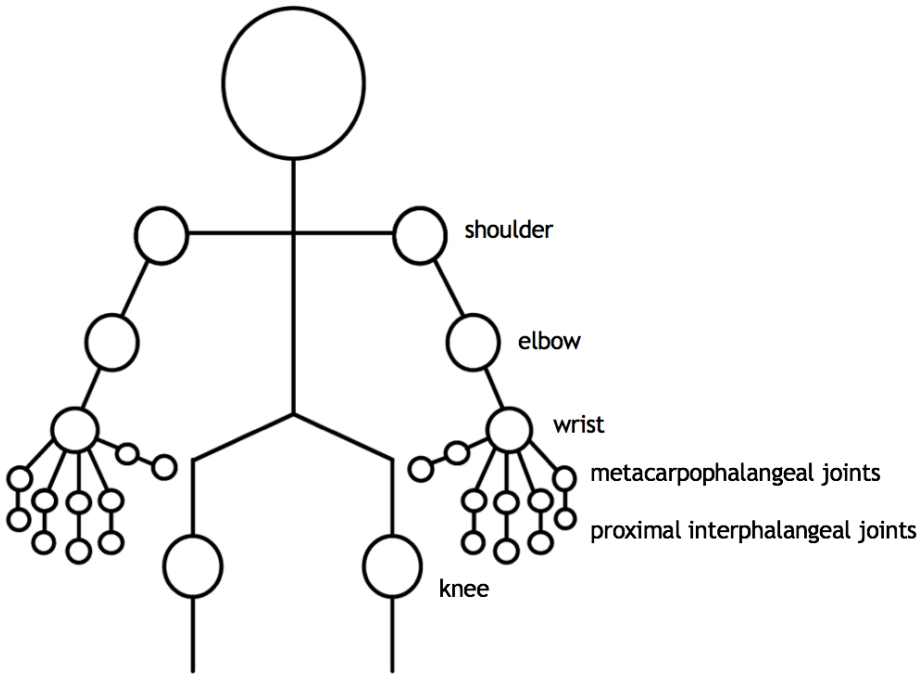


Figure 2.1 Joints included in swollen and tender joint counts

Disease activity calculations used were:

- Clinical disease activity index (CDAI) – (DasVAS/10) + (GlobalVAS/10) + SJC28 + TJC28
- Disease activity score 28 with c-reactive protein (DAS28 CRP) - 0.56*SQRT(TJC28) + 0.28*SQRT(SJC28) + 0.36*ln(CRP+1) + 0.014*GH + 0.96
- Disease activity score 28 with erythrocyte sedimentation rate (DAS28 ESR) 0.56* square root (SQRT)(TJC28) + 0.28* SQRT(SJC28) +0.7*LN(ESR) *1.08+0.16

For the work in this thesis, a combination of patients were chosen: some had reduced disease activity, from high disease activity (HDA) at baseline, to low disease activity (LDA) or remission after 6 months of therapy, and others had minimal, or no reduction in disease activity, representing responders and non-responders respectively. All patients chosen were identified as having HDA at baseline (Table 2.1).

Table 2.1 Categories of Disease Activity Score
CDAI, clinical disease activity index; DAS, disease activity score

CDAI score	DAS score	Disease Activity
≤2.8	≤2.6	Remission
>2.8 to ≤10	>2.6 to ≤3.2	Low
>10 to ≤22	>3.2 to ≤5.1	Moderate
>22	>5.1	High

2.1.2 Patient Identification – TACERA

The Towards A Cure for Early Rheumatoid Arthritis (TACERA) cohort is an early RA cohort that is part of the larger RA-MAP consortium. Like SERA, TACERA is an early RA, longitudinal cohort and patients were DMARD naive at enrollment. The

online knowledge management platform, TranSMART, was used to identify patients that were given monotherapy methotrexate (MTX) at baseline. Samples from this cohort were used for the blinded validation of the MTX signature, originally generated in the SERA cohort, and therefore the only data required at the selection stage was treatment at baseline.

2.1.3 Sample Type - SERA

From the patients identified as per section 2.1.1 in the SERA cohort, frozen buffy coat (BC) samples from baseline, 6-month and 12-month time points were identified and selected. Clinical information was obtained alongside the clinical samples. Of note, this includes age, BMI, smoking status and disease activity measurements.

2.1.4 Sample Type - TACERA

From the patients identified as per section 2.1.2 in the TACERA cohort, peripheral blood mononuclear cells (PBMCs) from csDMARD naïve patients at baseline were identified and selected. Disease activity data were also available, but this was not retrieved until the blinded analysis was complete (See 2.2.1).

2.1.5 Sample retrieval - SERA

Samples were collected on dry ice from the SERA storage facility at Yorkhill Biorepository. On return to the Glasgow Biomedical Research Centre (GBRC) samples were logged in using unique barcodes on each sample tube. Samples were subsequently thawed and then aliquoted into 110µl aliquots and stored at -80°C until required.

2.1.6 Sample retrieval – TACERA

TACERA samples were retrieved from the UK Biocentre on dry ice before shipment to Oxford Biodynamics Plc (OBD) where they were subsequently stored at -80°C until required.

2.1.7 Sample storage

The laboratory information management system (LIMS) was used to identify samples for retrieval from the SERA biobank. Sample barcodes were scanned and recorded on an excel spreadsheet which documented sample location in -80°C freezer. This information was stored in a password protected folder.

2.1.8 Ethical Approval - SERA

Samples were obtained with written consent and under appropriate ethical approval. Ethical approval for the SERA study was obtained on 28/05/2010, under REC approval number 10/S0704/20. A sample access application was submitted to the SERA Access Committee to achieve approval to access the requested samples and associated clinical information on several occasions throughout this study. These applications were approved on 23/08/2017, 07/08/2018, 18/02/2019.

2.1.9 Ethical Approval - TACERA

Samples were obtained with written consent and under appropriate ethical approval. Ethical approval for the TACERA samples was obtained on 02/05/12 under REC approval no 12/LO/0469.

2.2 Chromosome Conformation Capture

Prior to using Chromosome Conformation Capture (3C) on valuable patient samples, the protocol had to be optimised in my own hands using healthy samples. Once optimised, the steps described (2.2.3 - 2.2.7) were performed using RA patient samples. This section describes the original 3C protocol and the methods used to determine if quantitative PCR could be used to capture the MTX chromosome conformation signature (CCS) loci.

2.2.1 PBMC Isolation

Healthy buffy coat donors were provided from the Scottish National Blood Transfusion service (SNBTS). PBMCs were isolated by density gradient centrifugation with Ficoll (GE Healthcare). PBMCs were re-suspended in PBS (Sigma) or cell separation buffer (PBS 1% Fetal Bovine Serum(FBS), 2mM EDTA). Cells were frozen at $2x10^7$ in freeze buffer (10% Dimethyl Sulphoxide (DMSO), 90% fetal calf serum (FCS)) and stored at -80°C for short term storage or in liquid nitrogen for samples getting stored for longer periods.

2.2.2 $CD4^+$ T cell Isolation

PBMCs isolated as described in section 2.2.1, were re-suspended in cell separation buffer and $CD4^+$ T cells were separated by positive selection using magnetic bead separation as described by manufacturer (T cells - Miltenyi Biotec). Briefly, PBMCs were mixed by pipetting with 20µl $CD4^+$ magnetic microbeads in 80µl of cell separation buffer per 10^7 cells for 15 minutes at 4°C. To wash off excess labelling, 10ml of cell separation buffer was added and tube centrifuged at 300g for 10 minutes at room temperature (RT). Supernatant was removed and cells resuspended in 500µl cell separation buffer. Cells were then appropriately labelled and could be passed through a magnetic separation column. A column was placed on the appropriate MidiMACS™ Separator and rinsed with cell separation buffer. The labelled cells were applied to the column, and the column rinsed three times with 3ml of cell separation buffer. The column was removed from the magnet and 5ml of cell separation buffer added. Using a plunger, cells were forced through the column. This, the positive

fraction containing $CD4^+$ T cells, was then available for use in future experiments.

2.2.3 Flow Cytometry Purity Check

For surface staining, 1 x 10^6 cells were resuspended in flow cytometry buffer (PBS with 2% FBS and 5mM EDTA) into 6ml FACs tubes (BD Biosciences). Cells were centrifuged at 400g for 5 minutes before adding CD4-APC antibody (BioLegend). Tubes were incubated at room temperature for 20 minutes in the dark. Cells were fixed in diluted fix buffer (BD Biosciences) and kept at 4°C until they were run on the LSR II flow cytometer. Data was then analysed using FlowJo v10 software.

2.2.4 DNA extraction

PBMCs from healthy donors or patient BC samples were removed from -80°C prior to DNA extraction and thawed at 4°C. This was carried out as per OBD protocol using the Episwitch™ proprietary reagents. A starting volume of 50µl (1 million cells) patient sample was used for each DNA extraction. Briefly, cells were fixed with EpiMix Buffer DE-A (Thermofisher Scientific) and quenched with EpiMix Buffer DE-B. This was followed by cell lysis with 10x EpiMix Buffer DE-C and the nuclei were purified by density cushion centrifugation. Taq1 (Thermofisher Scientific) and T4 DNA ligase (Takara) were used to restrict and ligate the DNA followed by the addition of proteinase K (Roche) to remove any proteins. Incubations with these reagents were carried out on the Veriti thermocycler, see Table 2.2 for thermocycler conditions.

An updated extraction protocol (Protocol 2) was implemented after quantitative PCR was introduced. This was carried out as described above, with the addition of protease inhibitors (Sigma) prior to EpiMix Buffer DE-A treatment and during lysis. During the EpiMix Buffer DE-A step, a non-fixation (NF) control was generated with the addition of water instead of EpiMix Buffer DE-A. See Table 2.3 for cycling conditions associated with the updated protocol. Once extraction was complete, the sample was pelleted using density centrifugation and the pellet

then purified using the QIAamp FFPE tissue kit (Qiagen) as per manufacturer's instructions.

Table 2.2 Thermocycler conditions for 3C – Protocol 1

Step	Temp (°C)	Time
Taq1	65	20 minutes
1x EpiMix Buffer DE-F	65	5 minutes
1x EpiMix Buffer DE-G	37	5 minutes
DNA Ligase	16	10 minutes
Proteinase K	37 94	30 minutes 10 minutes

Table 2.3 Thermocycler conditions for 3C – Protocol 2

Step	Temp (°C)	Time
25x EpiMix Buffer DE-F	65	15 minutes
2x EpiMix Buffer DE-G	37	15 minutes
Taq1	65	20 minutes
1x EpiMix Buffer DE-F	65	5 minutes
1X EpiMix Buffer DE-G	37	5 minutes
DNA ligase	16 80 4	10 minutes 20 minutes hold

2.2.5 DNA Quantification – Picogreen

After DNA extraction and the preparation of 3C libraries, the DNA had to be quantified. The first method of quantification used Picogreen. Here, 20x Tris-EDTA (TE) buffer was diluted with 200x Quant-iT Picrogreen to make a 1 in 10 working solution. Volumes of working solution were dependent on the number of DNA samples being analysed. A 1 in 2 serial dilution of 100μg/ml lambda DNA was created to act as a standard. All standards and samples were diluted 1:100. Next, 100μl of diluted samples were added to a 96-well ELISA plate in triplicate and the 100μl of 1x TE-Picogreen mix was added. The plate was incubated at RT for 5 minutes then read on a Tecan M200 Pro at 480nm.

2.2.6 DNA Quantification – Qubit

The Qubit DNA quantification kit was used as an alternative method of quantification to Picogreen. The Qubit high sensitivity (HS) double-stranded (ds)DNA kit (Thermofisher Scientific) was used. A working solution was prepared by diluting Qubit dsDNA HS reagent 1:200 in Qubit dsDNA HS buffer. For samples, 2μl was added to 198μl of working solution and for standards, 10μl was added to 190 μl of working solution. Standards and samples were briefly vortexed and incubated at RT for 2 minutes before being read on the Qubit 3.0 Flourometer (Thermofisher Scientific). The Qubit dsDNA broad range (BR) kit (protocol as described for HS kit) was implemented if the DNA concentration was out with the range of the HS kit.

2.2.7 Nested PCR

After the DNA quantification, nested PCR was carried out using primers listed in Table 2.4. Primers for nested PCR were designed by OBD using Primer3 software. Primers were stored at -20°C until needed, at which point they were thawed at RT. Samples were normalised to a concentration of 1μg/μl in nuclease free water. A master mix of 16.5μl, nuclease free water (Thermo Fisher Scientific), 4μl of both outer primers and 12.5μl kappa blood mix was made. 37μl of master mix was added to 0.2ml tubes followed by 13μl of diluted template. A non-template control (NTC) was created by adding nuclease free water instead of DNA sample to the mix. Samples were added to the thermocycler for the 1st round. For the 2nd round, master mixes were prepared as before with 24.5μl nuclease free water and inner primers. 45μl of master mix was added to a new set of 0.2ml tubes and 5μl of template (from the 1st round) was added. The tubes were added to the thermocycler for round 2. Cycling conditions are documented in Table 2.5.

Table 2.4 Nested PCR Primers

Primer type	Name	Sequence	Description
Control	MMP13	CAGAAAAGAGTTAAGAGTGTCAGAC	Forward Outer
	MMP1 11	TCCAGGTCCATCAAAAGGAG	Reverse Outer
	MMP1 4	ATTTGTGAAATGGGGAGTGG	Forward Inner
	MMP1 12	AGCTAAGCCAGAAGGGCAAG	Reverse Inner
	MMP1 10	AAGGCTGGAGGTAAACTATACAGG	Forward Outer
	MMP1 9	CAACAGGACACACCTATCAAACC	Reverse Inner
	ER 15	GGAGCATTTGAGGGAGAGAG	Forward Outer
	ER 18	GGTTTTACCAAGACTGCTTGC	Reverse Outer
	ER 16	CGGGGAACTATGGTAAACTCC	Forward Inner
	ER 17	TCACATGCAGCACAGAATACC	Reverse Inner
	RFA17	ACCCAGTCCCCACTCCTATC	Forward Outer
	RFA18	ACTCCCCATAGGCACAAGC	Reverse Outer
	RFA19	TGTGGTGGAGACAAAAATGG	Forward Inner
	RFA20	AGAAGTTGCCAAGGGTGATG	Reverse Inner
	ERTM3	TGTGTGACTCCTTCCTGCAC	Forward Outer
	ERTM4	TTTCAATTTCCCCAGCAGAG	Reverse Outer
	ERTM17	AGTTTGCGTCTGTGCATCTG	Forward Inner
	ERTM18	TGCGTCAATTCCTAGTGTGG	Reverse Inner
MTX CCS loci	RAA65	GAGTCACAGCAGAAGGGTAAG	Forward Outer
	RAA67	ATACAGATGGAGGAGGAGGTAG	Reverse Outer
	RAA66	GGAAGTGCTACACCTTTAAACCA	Forward Inner
	RAA68	CCTCCTCTACACACGACCA	Reverse Inner
	RAA89	CAGTCTGTCACGTGGGTTATT	Forward Outer
	RAA91	CCTTATTCATGTCTGCCCTAAGA	Reverse Outer
	RAA90	TTCTTTCCAGTGGCTGCTTAT	Forward Inner
	RAA92	GCTCTCTGATAGCCAGATGATTC	Reverse Inner
	RAA93	GATGTGGGATGACTCCATCT	Forward Outer
	RAA95	CTGTAAACATCAGGCTCAAAGG	Reverse Outer
	RAA94	GTCTAGTGCATTCAGAGAGTGG	Forward Inner
	RAA96	GACATCCAGTCAGCCTCATTA	Reverse Inner
	RAA98	CAGTAGAAAGGTGCCAGACAT	Forward Outer
	RAA100	CCAAGATCAGAAAGACGCAAAC	Reverse Outer
	RAA97	CGAGGGTTTGAAGTACGAAGA	Forward Inner
	RAA99	TATCCAGGAGGAAGGCTGTA	Reverse Inner
	RAA46	GCCTCCTGCATTCTCTTCTT	Forward Outer
	RAA48	CCCAGCTTTGCTTCATGTATTT	Reverse Outer
	RAA45	CTCACTCTTTCCGGCCTATG	Forward Inner
	RAA47	GGAAGTAGATACCAGCCAAACT	Reverse Inner

Table 2.5 Nested PCR Cycling Conditions

Temp (°C)	Time	Cycles	Round
94	5 minutes	1	1
94 56 72	1 minute 45 seconds 30 seconds	25	1
72	5 minutes	1	1
4		Hold	1
94	5 minutes	1	2
94 62 72	1 minute 45 seconds 30 seconds	25	2
72	5 minutes	1	2
4		Hold	2

2.2.8 qPCR product purification

After DNA extraction of samples intended for qPCR using the method described in 2.2.3 (Protocol 2), the Qiagen FFPE tissue kit was used for purification prior to qPCR. In brief, 1-8 pellets from each sample were suspended in ATL buffer and transferred to DNA LoBind tubes with the addition of 20µl Proteinase K. These were incubated on a heat block for 1 hour at 56°C followed by 1 hour at 90°C. Samples were cooled to RT after which 2µl RNase was added followed by RT incubation for 2 minutes. A master mix of 1:1 AL buffer and 200 proof ethanol was made. 400µl of the AL/ethanol mix was added to the samples which were then transferred to MiniElute columns. 500µl of AW1 buffer from the kit was added to the columns, followed by a 6000g spin in a centrifuge for 1 minute. After flow through was discarded, 500µl of buffer AW2 was added to the column and had a 6000g spin in a centrifuge for 1 minute. For elution, 30µl 1x TE buffer was added to the columns, which were incubated for 5 minutes at RT. This was followed by a 20000g spin for 1 minute. DNA concentration could then be measured by the Qubit dsDNA HS kit, as described in section 2.2.5.

Primers for qPCR were designed using the PrimerQuest tool within Integrated DNA Technologies (IDT). The default primer option for intercalating dyes was selected. This considers primer characteristics such as an optimum melting temperature of 62°C, a GC content of 50%, primer size of 22 nucleotides (nts) and amplicon of 100nts. Once designed, primer specificity was tested using NCBI Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi). All qPCR primers were then tested

experimentally. A master mix was made from 5µl 2x kappa probe force, 0.25µl of each forward and reverse primer (stock 100µM; see Table 2.6), and 0.5µl of 20x PowerSYBR green (Invitrogen). Patient samples were pooled together once normalised to 10ng/µl. These samples, along with NF control (generated as described in 2.2.3), genomic control (PE Biosystems), a loading buffer control (TE) and non-template control (NTC) were prepared. The master mix was plated in a 96-well plate and 4µl of template was added in duplicate to make a 10µl reaction. One CCS loci was tested per 96-well plate, primers shown in Table 2.6. Once prepared, the plates were sealed and centrifuged for 30 seconds. Plates were run on Applied Biosystems StepOne Plus or QuantStudio 7 Flex Real-Time System. The qPCR block set with 6 annealing temperatures of 68°C, 67.5°C, 66.4°C, 64.4°C, 62°C, 60°C to determine the optimal annealing temperature for each primer. A melt curve was also generated per run to identify the presence of only a single product without evidence of primer dimerization. The cycling conditions are shown in Table 2.7.

Table 2.6 qPCR primers

Primer type	Name	Sequence	Description	Melting Temp (ºC)
Control	MMP1 3	CAGAAAAGAGTTAAGAGTGTCAGACC	Forward Outer	
	MMP1 11	TCCAGGTCCATCAAAAGGAG	Reverse Outer	
	MMP1 4	ATTTGTGAAATGGGGAGTGG	Forward Inner	
	MMP1 12	AGCTAAGCCAGAAGGGCAAG	Reverse Inner	
	MMP1 10	AAGGCTGGAGGTAAACTATACAGG	Forward Outer	
MTX CCS loci	IL17A	CCCTCAACATGCAGGGATTA	Sense	62
	IL17A	TCACCCACTTGGATGAGC	Antisense	61
	IL17A	CCCTCAACATGCAGGGATTAC	Sense	63
	IL17A	ACTCACACTCTCACCCACTT	Antisense	63
	IL17A	ACTGCCCTCATAATCCAATCAC	Sense	62
	IL17A	ACTCACACTCTCACCCACTT	Antisense	63
	CXCL13	CTGACATGAGTGATGCGTTT	Sense	61
	CXCL13	ATTTTTATCATCAGATACATAAAATGAGA	Antisense	60
	CXCL13	GCTCTGACATGAGTGATGCG	Sense	63
	CXCL13	ATAGGGAAGTTTTTTTTTTGCTTCA	Antisense	61
	CXCL13	GTGGCTGCTTATATCTCCTACC	Sense	62
	CXCL13	TCAAACCTGATTTTTATCATCAGATACA	Antisense	62
	IL21R	CTGAGGCAGGCAGATCAT	Sense	62
	IL21R	CAGGTGACCTTGTCTCTGG	Antisense	62
	IL21R	CAGTGGCTCACACCTGTAAT	Sense	62
	IL21R	ATAGGGAAGTTTTTTTTTTGCTTCA	Antisense	62
	IL21R	CAGTGGCTCACACCTGTAAT	Sense	62
	IL21R	AGCTCTGGACATCCAGTCA	Antisense	62
	IL23	GGGAGACAGGGTGTCATTC	Sense	61
	IL23	ACGTAAGAACGTAAATGTTTGG	Antisense	61
	IL23	ATAGTGGCATGATCACAGCTC	Sense	62
	IL23	AACGTAAGAACGTAAATGTTTGGG	Antisense	62
	IL23	AGGCTGGAGAATAGTGGCA	Sense	63
	IL23	AGAACGTAAATGTTTGGGTGTTG	Antisense	62
	IFNAR1	GAAGGAGGAAGTGGCTGAG	Sense	61
	IFNAR1	CTCTTCTTGCTCAGGGTGAATA	Antisense	62
	IFNAR1	ACCAGACCGTTGCTGTG	Sense	62
	IFNAR1	CTCTTCTTGCTCAGGGTGAATA	Antisense	62
	IFNAR1	GAAGTGGCTGAGCGACC	Sense	62
	IFNAR1	TTAATGAAATCAAATAAACTCTTCTTGCT	Antisense	62

Table 0 qPCR cycling conditions

Temp (ºC)	Time	Cycles
98	3 minutes	1
95 60-68	10 seconds 20 seconds	44
Melt curve	5 minutes	1
4	Hold	

2.2.9 Gel electrophoresis

The amplified nested PCR products were visualised using gel electrophoresis. A 1.5% agarose gel was prepared with 1x tris-acetate-EDTA (TAE) buffer with 1x Ethidium Bromide (1µg/ml). A 1kb^{+} ladder (Thermofisher Scientific) was used and gels were run until this ladder had migrated adequately through the gel. The gels were imaged with UV light and the image was captured using the Gel Logic 200 imaging system.

2.2.10 Gel Purification

The qPCR amplified products of interest were run on a 1% agarose gel, which was prepared with 1x Ethidum Bromide. The gels were imaged with UV light and once products of interest were confirmed, bands were excised for purification. The excised gel was then put into a labelled 1.5 ml microcentrifuge tube for weighing and images were captured using the Gel Logic 200 imaging system. The excised product of interest was then processed for sequencing. In brief, solubilization and binding buffer (GQ) was added to each 100mg of gel. This was then incubated tor 10 minutes at 15°C to dissolve the gel. Following this, 100µl of isopropanol was added to the sample and mixed. To bind the DNA, the sample was applied the sample to a QIAquick column, and centrifuge at 18,000g for 1 minute. Run-through was discarded and the column was added to a clean tube with 15ul TE buffer to elute the DNA.

After gel purification, samples had to be sent for sequencing. Gel purification as described above did not yield the required concentration of DNA for sequencing. To increase the yield, the optimisation steps included GQ incubation for 15 minutes with shaking every 2 minutes, and incubation with elution buffer (10 mM Tris-Cl, pH 8.5) for 3 minutes. Additionally, DNA was eluted into heated buffer (37^{o}C water bath).

2.2.11 LabChip GX

As an alternative to gel electrophoresis as described above (2.2.8, 2.2.9), the LabChiP GX microfluidic system was used to visualise PCR products in high

throughput at a later stage in the study. The 1K reagent kit was used as per manufacturer's instructions. In brief, the LabChip was washed and reagents filled in the appropriate wells. Subsequently, a 96-well plate was prepared with samples in triplicate. The LabChip and the plate were loaded onto the LabChip GX Touch Nucleic Acid Analyzer, which was run for 3 hours, changing LabChip reagents after 1.5 hours. The samples were analysed using the LabChip GX software. Product band sizes were observed to confirm presence or absence of loop of interest.

2.2.12 Tubeseq

After qPCR products were purified, they had to be sent for sequencing to confirm that it was our product of interest. qPCR products were sent to Eurofins genomic sequencing to confirm the qPCR product of interest. The Tubeseq service was used. Samples were prepared at 1ng/µl for the 150-300bp products. The total volume of sample was 17µl, made up of 15µl of DNA sample at appropriate concentration, and 2µl primer at 10pmol/µl. Samples were prepared in 1.5ml tubes and labelled with unique Tubeseq barcodes for identification. Samples were transferred at RT to the Eurofins sequencing facility. Sequencing results were emailed several days after samples arrived at the facility. It should be noted that the first sequencing run was unsuccessful, and optimisation steps were required, see section 2.2.12 and 2.2.13 below.

2.2.13 Cloning

The first set of qPCR samples that were sent for sequencing were not successfully sequenced as the sample was of poor quality. As such, cloning of the qPCR product was introduced to increase quality. In brief, the chosen qPCR product was inserted into the TOPO 2.1 vector (Thermofisher) and incubated for 5 minutes at RT. This was then transformed into OneShot cells and plated on agar plates coated with kanamycin (50µg/µl). These were incubated overnight at 37°C. White clones were chosen and placed in liquid culture in a shaking incubator (200rpm) overnight. The plasmid was then isolated using the purelink miniprep kit as per manufacturer's instructions. The plasmid was then analysed for inserts by restriction digest with EcoR1. Inserts were identified by running an ethidium bromide gel (see section

2.2.8) and looking for a product 300 bp in length. This is based on the plasmid size of 3.9kb, EcoR1 restriction size of ~3kb.

2.2.14 PolyA tailing

Cloning alone was unsuccessful in achieving product of interest during restriction digest, and optimisation of sequencing preparation was carried out. A polyA tail was added to the qPCR product with the aim of generating a more stable product for future steps in the cloning protocol. Briefly, a master mix was created by adding, 2µl of 5x GoTaq reaction buffer, 2µl of 1mM dATP, 1µl GoTaq flexi DNA polymerase and 0.6µl of 25mM $MgCl_2$. 2µl of purified blunt-ended DNA fragment was added and nuclease free water was added to bring the final volume to 10µl. This was incubated at 70°C for 15-30 minutes in a water bath. PolyA tailing alone was unsuccessful, and success of generating a restriction product involved optimisation of the ligation ratio for the desired concentration of 1.5ng of insert.

2.3 Validation of MTX CCS in TACERA Cohort

Data was analysed with Ewan Hunter and Christina Koutsothanasi, OBD. 3C libraries were created and nested PCR was carried out as described in 2.2.3 and 2.2.6 on 70 PBMC samples chosen from the TACERA cohort (2.1.2). Briefly, two machine learning algorithms were employed to test the ability of the MTX CCS to stratify R and NR to MTX. Both models, namely XGBoost (https://xgboost.readthedocs.io/en/latest/) and LightGBM (https://lightgbm.readthedocs.io/en/latest/), utilise a gradient boosting decision tree algorithm and were used via R studio. 47 samples were unblinded (R vs NR status revealed) for classification; 23 samples remained blinded during analysis. R and NR status was determined by disease activity (CDAI) at 6 months after treatment.

2.4 *In-silico* data analysis of MTX CCS epigenomic environment

2.4.1 DeepBlue Data retrieval

Online datasets were utilised for these investigations. Namely the DeepBlue Epigenomic Data Server and the Promotor-Capture HiC (PCHiC) dataset, generated by Javierre *et al*[260] . To analyse the data, a combination of Microsoft Excel and packages within R and R studio were used. From the DeepBlue Epigenomic Server, various datasets were downloaded dependent on the epigenetic feature of interest. Specifically, data from Chromatin immunoprecipitation followed by sequencing (ChIP-Seq) experiments identifying H3K27ac, H3K4me3, H3K4me1, H3K27me3, H3K36me3 and K3K9me3, in addition to data from Bisulphite–Seq and DNase-Seq to identify methylation marks and DNase hypersensitive sites (DHSs), respectively.

Table 2.8 DeepBlue coordinates

Genome	Gene	Chromosome	Site a	Site b
hg38	*IL17A*	6	52161697-52172165	52184632-52187067
	CXCL13	4	77510412-77512093	77602625-77605433
	IL21R	16	27356311-27357534	27449257-27451508
	IL23	12	56345719-56347275	56361069-56361825
	IFNAR1	21	33324378-33325411	33373955-33376515
hg19	*IL17A*	6	52026495-52036963	52049430-52051865
	CXCL13	4	78431566 -78433247	78523779-78526587
	IL21R	16	27367632-27368855	27460578-27462829
	IL23	12	56739503- 56741059	56754853-56755609
	IFNAR1	21	34696683-34697716	34746261-34748821

2.4.2 DeepBlue data analysis

Once downloaded, the files were processed in R to identify if the marks were present in the regions of interest (MTX CCS loci). Dependent on the dataset downloaded, either hg38 or hg19 coordinates (Table 2.8) were used in the analysis script (Appendix). Regions of the genome 500 kb upstream of the first anchor site and 500 kb downstream of the second anchor site of chromosome loops were also investigated. Outputs from R were then quantified to understand enrichment of different epigenetic marks at each chromatin loop site.

2.5 Discovery Microarray

2.5.1 Microarray Set-up

For this analysis 54 buffy layer samples from the SERA cohort were used. Here, 18 healthy samples were used as a pooled standard on the array. 4x180k, custom Agilent microarrays were designed by OBD and run at their facility. OBD proprietary EpiSwitch™ pattern recognition algorithm was used to identify high probability chromatin folding interactions in combination with findings from Walsh *et al*[209] were used to generate a list of probes that were functionally relevant in RA.

Each probe was present in quadruplicate on the EpiSwitch™ microarray. The Agilent protocol for enzymatic labelling was followed. In brief, the standard EpiSwitch™ extraction as described previously (2.2.3) was used to generate the

3C library for each sample. Subsequently, the DNA concentration of each sample was determined using the absorbable nanoquant plate on the Tecan Infinate M200 Pro. 800ng DNA was used per sample. A pool of the healthy samples was generated by determining DNA concentration as before and adding 6.2ng of each together. An ethanol precipitation step was then conducted to clean the sample before beginning the sample labelling. The Agilent DNA enzymatic labelling kit (Agilent p/n 5190-0449) was used to label the DNA library. This kit uses random primers and exo-Klenow fragments to label the DNA with fluorescently labelled nucleotides using cyanine 3-dUTP and cyanine 5-dUTP dyes. Samples were spun for 1 min at 6000g in a centrifuge. 5µl of random primers were then added and the samples were incubated at 95°C in a thermocycler for 3 minutes. Samples were again spun at 6000g for 1 min in a thermocycler. A master mix of Cy3 and Cy5 was prepared by mixing nuclease free water, 5xbuffer, 10xdNTP, Cy3 or Cy5 and Exo-Klenow fragment. 19µl of master mix was added to each reaction tube, giving a total of 50µl. Samples were then incubated at 37°C for 2 hours, 65°C for 10 mins then held at 4°C. The hybridisation master mix was then prepared by mixing cot-1 DNA, Agilent 10x Blocking reagent and 2x Hi-RPM buffer. The master mix was incubated at 95°C for 3 minutes, then 37°C for 30 minutes. After incubation, samples were spun at 6000g for 1 minute in a centrifuge.

2.5.2 Microarray processing and feature extraction

100µl of hybridisation sample (section 2.5.1) was dispensed onto a clean gasket slide in the Agilent SureHub chamber base. The assembled chamber slide was placed in the rotator rack in the hybridisation oven at 65°C, 20 rotations per minute (rpm) and left for 24 hours. Then, the slide staining dishes, rack and bars were washed thoroughly with milli-Q water to remove any contaminated material. The slide rack and bar were then added to the slide staining dish, which was filled with 100% acetonitrile. The magnetic stir plate was set to a speed of 4 and washed for 5 mins at RT. The step was repeated and then the plate was dried in a fume hood. To wash the array slides, the first 2 staining dishes were filled with Oligo aCGH wash buffer 1 at RT and placed on magnetic stir plate. The pre-warmed glass dish filled with water and containing slide staining dish 3 was also placed on the magnetic stir plate. Staining dish 3 was filled with Oligo aCGH wash buffer 2, which had been warmed to 37°C. A 4th staining dish with acetonitrile was placed in a fume food with a magnetic stir bar

added and placed on magnetic stir plate. A 5th staining dish was filled in the fume hood with hybridisation and drying solution, and was place on magnetic stir plate with magnetic bar. The hybridisation chamber was disassembled and the array slide placed in staining dish 1 and this was repeated for further slides. Slides were then transferred to staining dish 2 and stirred on setting 4 for 5 minutes. Slides were then transferred to dish 3 for 1 min, and dish 4 for 10 seconds and dish 5 for 30 seconds. Slides were removed with barcode facing upwards with a slide cover was placed on top. Slides then were immediately scanned using the SureScan DA model.

Features were then extracted using feature extraction software and images were extracted as .tif. A QC report was carried out to ensure each extraction was completed successfully.

2.5.3 Microarray analysis - Limma

Once the array was completed, feature extraction data was downloaded from the raw OBD server. Data was analysed using several packages within R studio, namely, Limma and RankProd 2.0. Appropriate target files were generated for each analysis run. The use of the common reference healthy control sample allowed comparison of the loop expression across all RA samples. Target and raw data files were read into the R package (Appendix). Briefly, agilent control probes were removed first followed by probes that had a saturation signal above 65525. The Limma background correction and the data was then normalised within arrays using the locally weighted polynomial regression (Loess) method. A log matrix of $\log_2$ ratios of fluorescence intensities was generated from the normalised data. Since duplicate probes were used on the array, a matrix was generated from the mean, median and cv values that could be taken forward in the analysis. The log median matrix was used for the analysis. A design matrix, followed by a contrast matrix were generated and a linear regression model was then fitted to the data based on the design. Statistics were then computed using the empirical bayes method (ebayes). A table of probes (loops) could then be extracted that had differential abundance between samples. Extra filtration could be implemented if desired, such as a specified number of loops to be output, and filtering on adjusted P.value (FDR correction) and abundance scores (AS), such as adj.P.Val ≤ 0.05 and AS $-1.1\leq$ or ≥ 1.1.

2.5.4 Microarray analysis - RankProd

Data was analysed with assistance from Ewan Hunter and Christina Koutsothanasi, OBD. After exploration of the data using Limma, it was decided a more stringent method of analysis should be implemented to differentiate the 3 responder groups. For this, the RankProd package 2.0 was used to subsequently analyse the data. The Rank Product (RP) is a statistical technique which is used to find differentially expressed marks from molecular profiling studies. The RankProd package utilises this technique. The RP and Rank Sum (RS) are non-parametric tests which can determine up or downregulated variables in repeat experiments. The P value for RP has strict bounds and calculated in a computationally fast manner. For this analysis, firstly data was normalised (Loess) and an expression matrix produced. Data were then filtered on adjusted P value and abundance scores (AS); loops with an adjusted P value ≤0.05 and of -1.1≤ or ≥1.1 were carried forward for further analysis.

2.5.5 Microarray Analysis – Searchlight

Data was also analysed using Searchlight (https://github.com/Searchlight2/Searchlight2), an automated, platform for the analysis and visualization of RNAseq data, which was adapted for our microarray dataset. Data was analysed by John Cole, University of Glasgow. Searchlight provided an alternative way to analyse our data from the microarray experiment, and provided a streamlined, expedited way to facilitate deep exploration of the data that could not be achieved with RankProd in the same time frame. In summation, the three types of analysis performed were:

- Expression – how much of a loop was present in a sample
- Differential expression – how did the loop abundance differ between 2 groups
- Signature analysis – did groups of differently abundant loops generate a signature that would allude to a predictive biomarker

The normalised expression matrix data was used for this analysis, and was generated as described above (2.5.3). To generate differential expression signatures, numpy was used to generate mean expression values and differential expression of loop abundance. Comparisons were: R vs NR, NR vs IR and R vs IR

at each time point. To determine signatures, each loop was classified by its starting signature based on the differential expression from the comparisons described. Expression values for each loop was converted to a z-score which allowed metagene expression for all samples. 2 expression metagenes were correlated to each other using the Spearman Correlation Coefficient to find signatures with similar expression profiles.

2.5.6 Microarray Biological Interpretation - Bedtools

To begin biological exploration of array data the Bedtools programme was used. This was implemented via the terminal to analyse protein coding loci in proximity to loops of interest, found via Limma and RankProd analysis pipelines. The Bedtools *closest* function was used to identify the closest three protein coding loci to each loop of interest. Once a list of loci was produced, these were put through the online tool, Hugo Gene Nomenclature Committee, to filter and ensure all information was captured. This list was then entered into the online platform, GeneAnalytics to identify functional enrichment of genes and other genomic features. GeneAnalytics aims to identify potential associations of gene sets with pathways, compounds and Gene Ontology (GO) terms (biological process and molecular function). The results are ranked by relevance to the analysed gene set.

2.5.7 Microarray Biological Interpretation - STRING

Gene lists of interest generated by GeneAnalytics were analysed further using the Search Tool for Retrieval of Interacting Genes/Proteins (STRING) (https://string-db.org) version 11.0, a database consisting of over 9 million known and predicted protein sequences. Gene lists corresponding to various analyses were entered into the online STRING platform and interaction networks were generated. Network nodes within the string represent proteins and the edges indicate functional associations between proteins. Proteins that are grouped based only on shared homology are excluded. The PPI enrichment value identified if the network had significantly more enrichment than expected. The interaction scores are given from zero to 1 and are based on the confidence that

the interaction/association is true. Other information such as enriched biological processes were extracted as .csv files, in addition to downloading the networks as images (Appendix).

2.5.8 Microarray Biological Interpretation – Cytoscape

Based on additional functionality, Cytoscape version 3.7.2 (https://cytoscape.org) was used to conduct further network analysis of protein networks generated using STRING. Cytoscape represents genes or molecular marks as nodes and edges represent interactions between them. The network .csv file taken from STRING was loaded into the Cytoscape software. The network analysis tool was implemented to identify the network nodes with the most connected edges. The most connected nodes were identified as those with the most directed edges in the network.

2.5.9 Microarray Biological Interpretation - IGV

The Integrated Genome Viewer (IGV) (https://software.broadinstitute.org/software/igv/) was used for the exploration and visualisation of loops of interest, in addition to other genomic and epigenomic features. Files of interest were prepared in '.bed' format and loaded into IGV version 2.4.14. Files are represented as 'tracks' on the viewer. Files included in the analysis included the Janssen expression quantitative trait loci (eQTL) files, and loop anchor sites. Images could be saved from analyses.

2.5.10 EpiSwitch™ Data Portal

The longitudinal data from the RankProd analysis was uploaded to the EpiSwitch™ data portal (https://episwitch3dgenomicsportal.com), an interactive interface to allow for easy manipulation and visualisation of the 3D genome data. The portal incorporates Bedtools functionality, described above (2.5.6) as well as IGV visualisation software that has also been described above (2.5.9). From the portal, data could be downloaded and images saved for future use and analysis.

2.6 Statistical Analysis

Statistical analysis not already described above was conducted in GraphPad Prism 6 software. Tests were chosen based on the distribution of the data and the desired comparisons to be made. Figure legends detail the exact statistical test used on each data set. In this study 0.05 was considered significant, with * $P<0.05$, ** $P<0.01$, ***$P<0.001$, *** $P<0.0001$.

Chapter 3 Validation and Further Characterisation of Methotrexate Chromosome Conformation Signature and Optimisation of Detection Method

3.1 Introduction

Previous work from our lab, in collaboration with Oxford BioDynamics Plc (OBD), produced a chromosome conformation signature (CCS) with promising capacity to differentiate responders (R) and non-responders (NR) to methotrexate (MTX) in an early rheumatoid arthritis (RA) population[219]. The discovery and validation populations both came from the Scottish Early Rheumatoid Arthritis (SERA) cohort. This signature was developed to be exclusive to RA patients, and was not in the same conformation in healthy samples. While molecular biomarker investigations can show promise in the discovery and preliminary validation stages, studies have shown subsequent validation can produce less efficacious results[261]. Therefore, before clinical implementation, it is fundamental that a biomarker is validated, proving the efficacy and ensuring it is a true representation of a heterogeneous disease population[262]. RA is a well-characterised heterogeneous population, and it must be established if the MTX CCS can predict response in other early RA cohorts, and identify any potential confounding factors that may impact the predictive ability of the biomarker[262]. Validation in a completely independent cohort that is demographically matched to SERA would be both interesting and clinically important.

While the results from Carini *et al*[219] demonstrate the potential of a MTX biomarker, it is not only MTX that is given at baseline of RA treatment. Some patients cannot tolerate the therapy and suffer side effects such as nausea and hepatotoxicity, or in the case of around 50%, will not respond clinically to MTX[263,264,265,266]. Other conventional synthetic disease modifying anti-rheumatic drugs (csDMARDs), primarily hydroxychloroquine (HCQ) and sulphasalazine (SSZ) exist as alternative first-line therapies. Despite investigations to find biomarkers of treatment response to these csDMARDs, there has yet to be a clinically implementable finding and often studies investigate HCQ and SSZ in combination with MTX[267,268]. The study by Kremers *et al.*, identified HCQ and MTX as having better retention than other treatments. However, this was predicted using survival analysis techniques, taking into consideration comorbidities and disease

characteristics such as duration of disease, instead of employing molecular biology methodologies[268]. Molecular biomarker studies for HCQ alone are limited. One study investigating SSZ, as part of triple csDMARD therapy, suggested gene variants in *NAT2* and *ABCG2* were associated with a limited response to SSZ[112]. This study has yet to be validated. Predictors of response to csDMARD treatment as a whole have been attempted. One recent study investigated the ability of both molecular and synovial signatures to predict response to csDMARDs. They identified cellular synovial and molecular signatures that had the potential to predict disease progression and treatment response. This study offers the potential for a blood based pan-DMARD predictor that would be of clinical benefit[212]. However, this study incorporated low-throughput techniques that would not be advantageous in a clinical setting. Therefore, it would be valuable to know if the MTX CCS biomarker has the capacity to predict response to csDMARD treatment, regardless of which monotherapy or combination therapy is assigned.

The work for the MTX CCS study was carried out at OBD where the EpiSwitch™ proprietary technology exists. The work in this thesis required the establishment of this 3C propriety protocol at the University of Glasgow. This chapter details the steps taken to ensure efficient establishment of this technique, which included exploring transition to a higher throughput method of chromosomal loop detection. The nested PCR method used in the MTX CCS discovery study could be considered low throughput. Moreover, quantification capabilities are limited with gel electrophoresis outputs and there is difficulty in determining any weak 3C signals[269]. Implementing a quantitative PCR (qPCR) method would not only be of benefit in our study, but in future clinical use. It would be higher throughput and offer the opportunity of multiplexing[270]. qPCR has been successfully used in the study of the mouse HoxB1 loci[271]. The process of qPCR implementation in the work in this thesis followed the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines for primer design to ensure accuracy and robust results moving forward. These guidelines state the *in-silico* and wet lab steps required to create publishable-qPCR results[272]. As such, this chapter explores the investigation into qPCR as a method of loop detection, adhering to the MIQE guidelines.

The aim of the work in this chapter was to further characterise the predictive ability of the MTX CCS and identify the optimal way to investigate the chromatin architecture of samples throughout this study. To achieve this, the aims were:

1) Validate the MTX CCS both bioinformatically and experimentally using an independent clinical RA cohort

2) Set up 3C protocol independently and determine the optimal method for loop detection through exploration of different PCR methodologies

3) Assess whether MTX CCS is stable after treatment and if it can accurately predict response to csDMARD treatment as a whole

3.2 Results

3.2.1 Bioinformatic validation of MTX CCS

Validating the prior MTX CCS signature *in-silico* by independently analysing the data was an important first step in this study. The initial step to achieve this was to use a Limma-based analysis to interrogate the quality of the microarray-generated data (Figure 3.1). Limma is a package that facilitates the analysis of gene expression arising from microarray or RNA-Seq experiments. Limma utilises linear models to identify differential expression. In the case of data in our study, Limma was used to identify abundance changes of loops between healthy and RA samples. The starting data file was the intensity of each spot on the array, which had been extracted by the Agilent Feature Extraction Software. The red-green density plots were used to visualise the signal distribution across the arrays. The density plot before normalisation indicated that the signals had an expected distribution and there were no outliers (Figure 3.1A). Moreover, the dye intensity of both red and green dyes were similar, indicating the absence of dye bias. The 'within array' normalisation step was successful by bringing the signal distributions closer together (Figure 3.1B). Loess normalisation was used, which is a type of Generalised Additive Model (GAM). MA plots were then used to understand the relationship between the red-green intensity log ratio (M) and average intensity of a spot on the array (A). Figures 3.1C and Figure 3.1D highlight the successful normalisation, by the flattening of the line around the M value of 0. A boxplot was then produced to illustrate the distribution of M values across all 8 arrays (Figure 3.1E). The normalisation brought the medians close together and the range of M values can still be visualised. Most of the M values are distributed around 0. This analysis supported the concept that the data was of good quality and suitable for further interpretation.

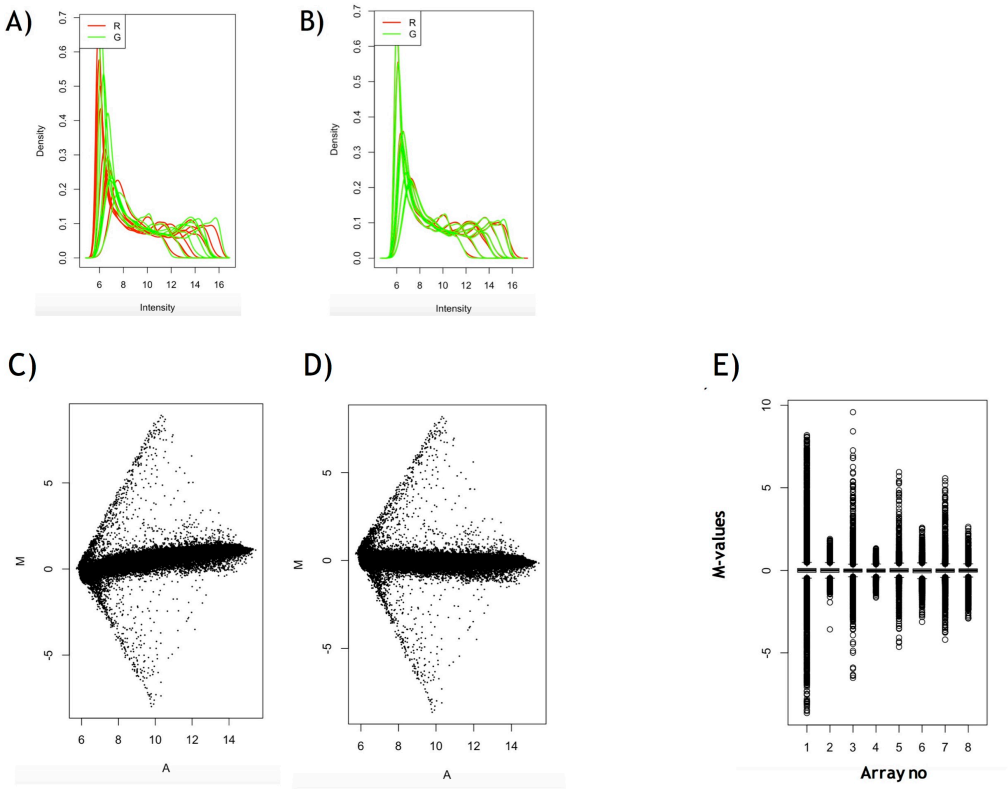


Figure 3.1 Quality Control Assessment of MTX CCS Microarray Data
Series of plots demonstrating quality of raw microarray data and the influence of normalisation. 8 dual-colour arrays in total. Array 1-4 compared R and NR, array 5+6 compared HC and R, array 7+8 compared HC and NR. 13,322 EpiSwitch™ sites across 123 loci were analysed. A) Red-green density histogram before normalisation. B) Red-green density histogram after 'within array' locally weighted polynomial regression (Loess) normalisation. C) MA plot before normalisation. D) MA plot after 'within array' Loess normalisation. E) Boxplot illustrating M value (log-ratios) distribution after Loess normalisation. Plots created using Limma package on R studio.
A, mean average; HC, healthy control; M, log ratio; R, responder; NR, non-responder

After using multiple plots to visualise that the MTX CCS data was of good quality and normalisation procedures were effective, it was important to find out if the predictive loops would be replicated using the original nested PCR dataset (Figure 3.2). Limma was used to apply a linear model to the data to identify loops that could differentiate responders (R), non-responders (NR) and healthy controls (HC). The results shown for the MTX CCS genomic regions indicate the potential for stratification. Differences of loop abundance was used as a measure of stratification potential with a positive value associated with the condition on the left of the contrast model, and a negative value associated with the condition on the right side of the model (Figure 3.2A). The first output from the Limma contrast model, contrasting NR and R, illustrated that *IL17A* and *CXCL13* had a positive fold change. In contrast, *IL21R*, *IL23* and *IFNAR1* had negative fold change values. This demonstrated the association of *IL17A* and *CXCL13* with NR and *IL21R*, *IL23* and *IFNAR1* with R as per the MTX CCS. Next, a

classification model was used to test the predictive value of the signature. A Random Forest method was used to test the data that was taken from the nested PCR results from 55 RA patients, collected at the time of the study by Carini *et al*[219]. The model successfully predicted 26 R, and 22 NR correctly (Figure 3.2B). This resulted in a true positive rate of 0.96 and 0.79 for R and NR, respectively which gave an overall accuracy of 87%.

A)

MTX CCS gene	EpiSwitch nomenclature	FC	Association
IL17A	IL17A_3_1_RR	1	NR
CXCL13	CXCL13_1_3_RR	1	NR
IL21R	IL21R_5_2_RR	-1	R
IL23	IL23_4_5_FR	-1	R
IFNAR1	IFNAR1_2_4_RR	-1	R

B)

	Classified	
Observed response	Responder	Non-responder
Responder	26	1
Non-responder	6	22

C)

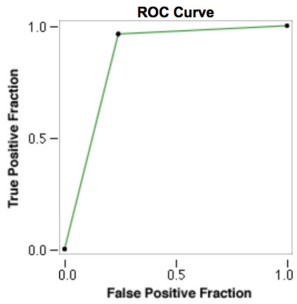


Figure 3.2 Statistical Validation of MTX CCS
Limma linear model results and validation using PCR data from 55 RA patients using 5-loop MTX CCS. A) Limma linear model results of contrast model between R and NR. B) Binary classifier model conducted in Weka using Random Forest classification. C) ROC curve illustrating relationship between sensitivity and specificity. ROC curve generated using web-based calculator which utilises JROCFIT program.

3.2.2 Technical optimisation of 3C in peripheral blood from healthy donors

The first step in establishing the EpiSwitch™ 3C assay in house was to use a range of primary cells from human donors. It should be noted that the original MTX CCS was generated in peripheral blood mononuclear cells (PBMCs) from RA patients. However, given the precious nature of clinical samples, for the 3C technical optimisation it was not deemed appropriate to use RA patient samples. Thus, PBMCs were isolated from healthy buffy coat (BC) donors. Flow cytometry analysis was performed to check cell purity prior to further analysis (Figure 3.3). 70.2% of PBMCs were lymphocytes. Analysis revealed that 56.5% of PBMCs were $CD4^+$ T cells, and after isolation by magnetic bead separation, 92.5% were $CD4^+$ T

cells. These values confirmed that it was a normal PBMC population that was isolated.

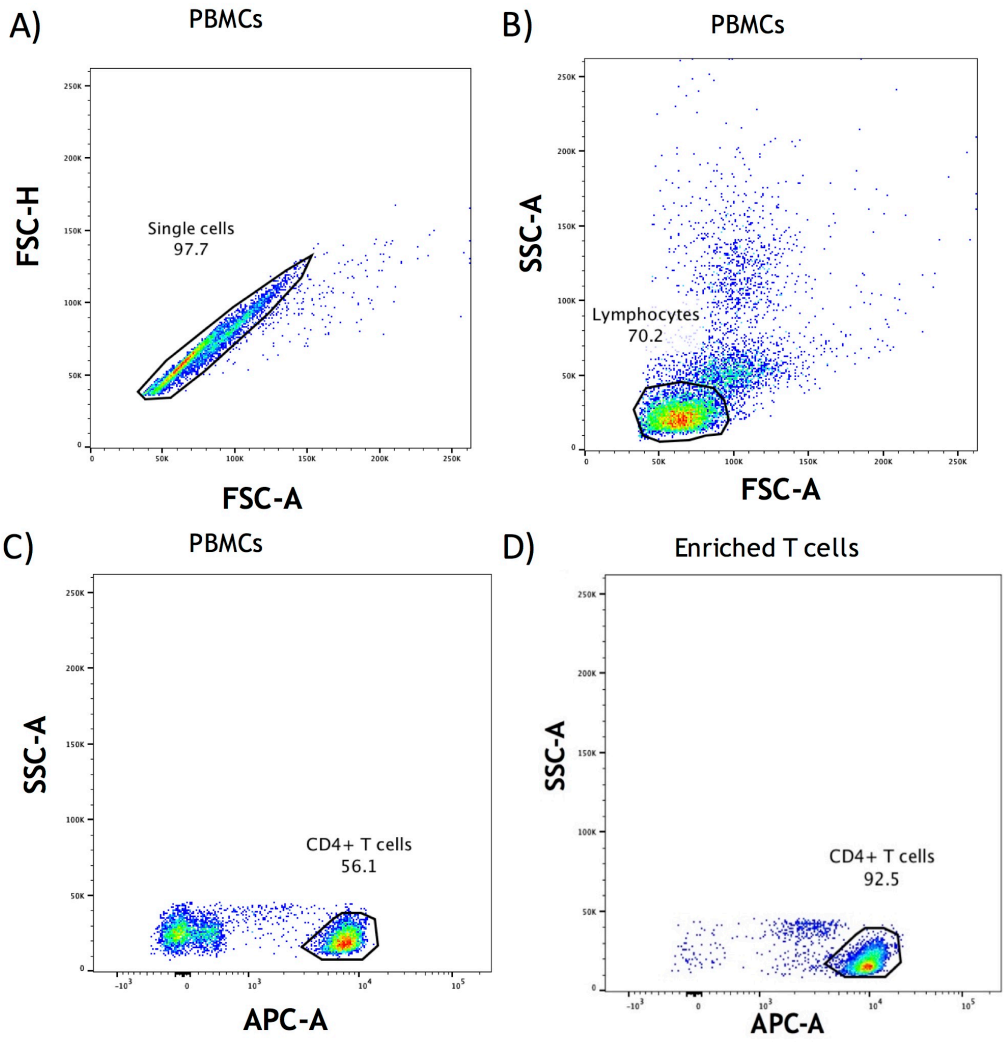


Figure 3.3 PBMC Isolation Purity Check
Purity of the isolation was examined using flow cytometry. Representative flourescence-activated cell sorting (FACS) plots of PBMCs and $CD4^+$ T cells from healthy BC donors. PBMCs and T cells stained with CD4-APC antibody. A) Gate demonstrating defined population of single cells within PBMCs, percentage of gate is displayed within plot. B) Percentage of lymphocytes in total PBMCs. C) Percentage of $CD4^+$ T cells in single cells. D) Percentage of $CD4^+$ T cells after $CD4^+$ magnetic separation.

Once PBMCs (and purified $CD4^+$ T cells) were successfully isolated, 3C DNA extraction was carried out (See Section 2.2). After extraction, the DNA library concentration had to be measured to confirm successful isolation of DNA, and to determine a reference for the normalisation in later protocol stages. The Quant-iT PicoGreen DNA quantification assay was initially used to determine DNA concentration in PBMCs, as this was an optimised methodology used at OBD and

suitable for our sample type. This method revealed samples contained more than 20ng/μl of DNA (Figure 3.4).

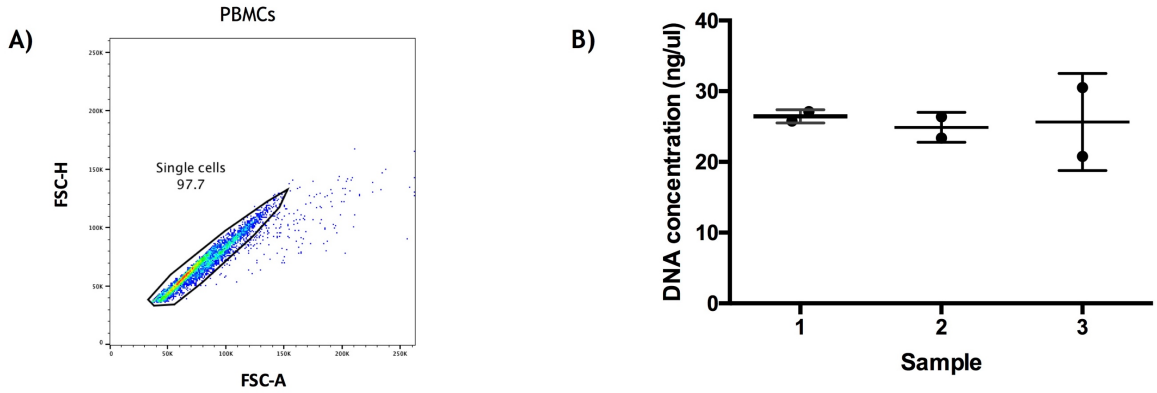


Figure 3.4 3C DNA Library Isolation Check
Confirmation of successful extraction of DNA library from PBMCs. A) Representative FACS plot of PBMCs used for DNA library preparation. B) DNA concentration of first DNA

During the course of the optimisation, it was anticipated that fewer samples would be processed at one time than generally occurs at OBD, and thus, the Quant-iT PicoGreen assay (tailored for large sample number) was not optimal. The Qubit High Sensitivity (HS) double-stranded DNA (dsDNA) kit was chosen as a suitable alternative quantification method, based on sample number requirements, speed of protocol and rapid reading capabilities. To confirm that this was an appropriate alternative quantification method, samples were run in parallel using both methods, enabling the comparison of the techniques to determine if that resulted in consistent in DNA concentrations. Surprisingly, the resulting DNA concentrations calculated from the two methods were not comparable, producing different concentrations in the same sample (Figure 3.5). Consequently, this experiment was repeated to determine if the methods continued to produce different results. Each experiment produced different concentrations per sample, with the Quant-iT PicoGreen assay consistently measuring higher concentrations of DNA compared to the Qubit. Experiment 1 demonstrated a significant difference between the two methods. Additionally, there was a significant difference between the concentrations measured by Quant-iT PicoGreen between experiment 1 and experiment 2. Combining the results from all experiments (Figure 3.5B), there was a significant difference between the 2 methods. The PicoGreen method showed a much larger variation

in DNA concentration values with a standard deviation of 10.98 compared to 2.312 for the Qubit. The maximum DNA concentration measured by the PicoGreen method was 30.53ng/μl and the maximum measured by the Qubit HS dsDNA assay was 10.30ng/μl. Based on the more consistent DNA concentration measurements, and the concentrations considered within normal range for this type of library preparation, the Qubit dsDNA HS assay was selected as the most robust method to move forward with.

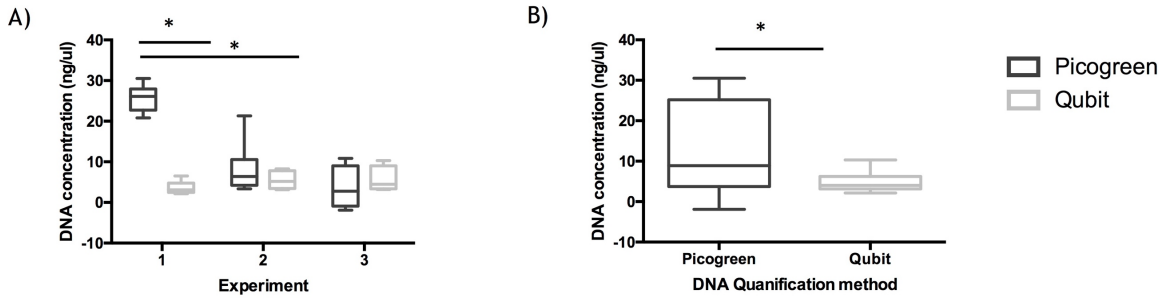


Figure 3.5 Comparison of DNA Quantification Methods
Comparison of Quant-iT PicoGreen and Qubit dsDNA HS assay. 3C DNA libraries were extracted from PBMCs and both methods were used to measure the DNA concentration of each sample. A) DNA concentrations (ng/ul) calculated via the 2 independent methods are plotted for comparison, data across 3 experimental repeats. Experiment 1 and 2, n=7, experiment 3, n=4. Non-parametric T-test used to compare methods in one experiment. Wilcoxon test to compare between experiments. B) Combined DNA concentrations (ng/ul) from 3 experimental repeats, calculated via 2 independent methods - plotted for method comparison (N=18). Mann Whitney T test to compare methods. * P< 0.05. Data is shown as box and whisker plot showing the median and range.

The 3C assay preparation continued with the optimisation of nested PCR and ultimate visualisation of the ligated DNA on an agarose gel, which is the surrogate for the presence of a loop in a sample. Establishment of a reliable detection method of DNA concentration (Qubit) enabled normalisation of sample DNA concentration prior to nested PCR. The first step in the PCR process was to test the variety of control primers that can aid the 3C assay (Table 3.1) to confirm if the 3C could be replicated at our site.

Table 3.1 Control Primers
List of control primers tested for the 3C protocol.
+ = positive control, - = negative control, PCR = controlling for PCR reaction

Primer type	Name	Description
Control	MMP1 3	+
	MMP1 11	+
	MMP1 4	+
	MMP1 12	+
	MMP1 10	+
	MMP1 9	+
	ER 15	PCR
	ER 18	PCR
	ER 16	PCR
	ER 17	PCR
	RFA17	+
	RFA18	+
	RFA19	+
	RFA20	+
	ERTM3	-
	ERTM4	-
	ERTM17	-
	ERTM18	-

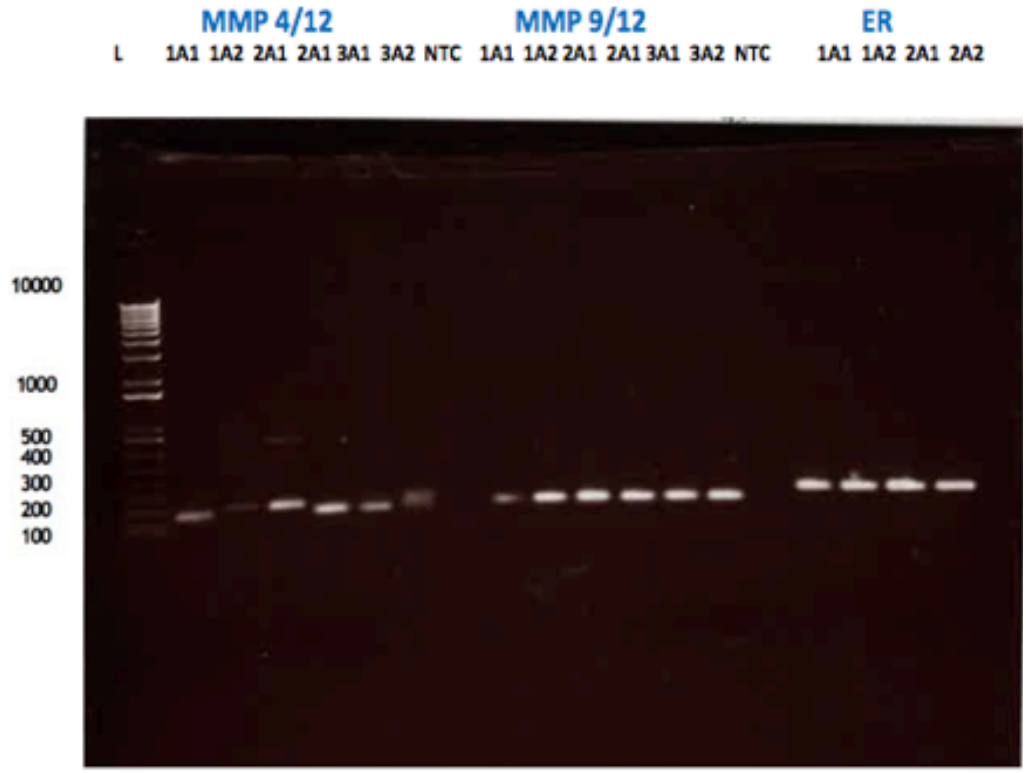
Table 3.2 Gel electrophoresis interpretation
Interpretation of gel data when samples are loaded on gel in triplicate. Sample names hypothetical.

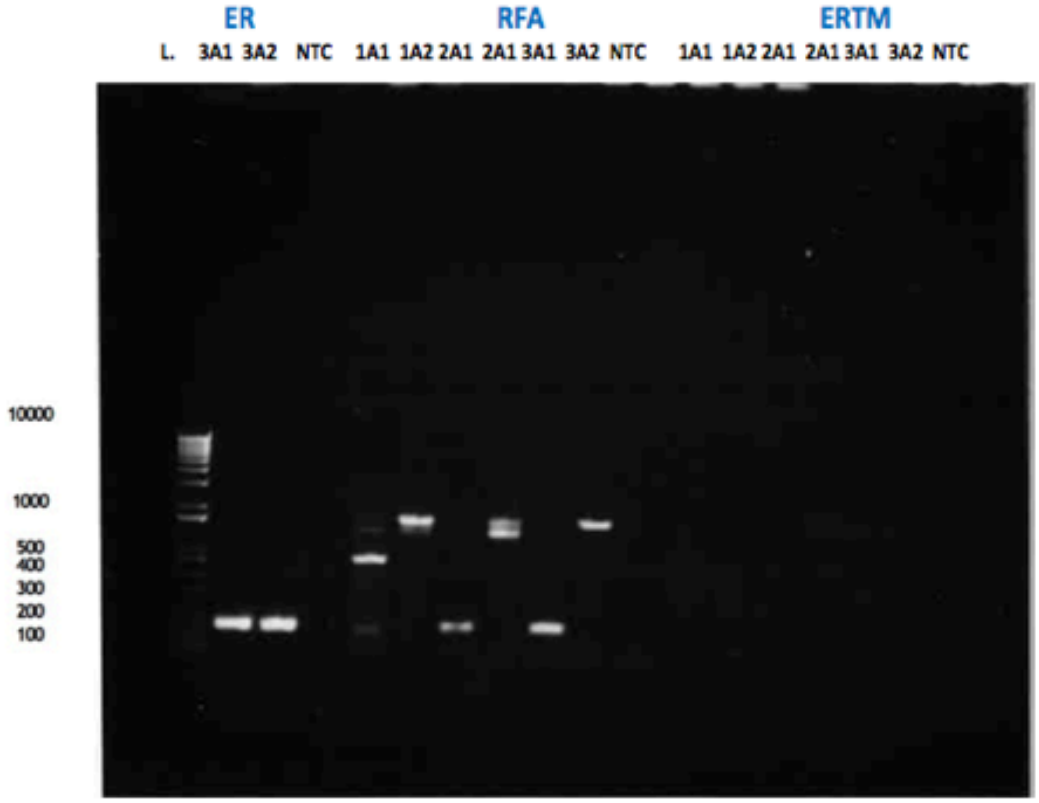
Sample	Number of bands across triplicate samples	Binary score	Loop prediction
1A	0	0	No loop
1B	1	0	No loop
2A	2	1	Loop
2B	3	1	Loop

DNA loops from PBMCs (Figure 3.6A) and $CD4^+$ T cells (Figure 3.6B) were used to test the control primers. Three primers were positive 3C controls, one was a PCR control and 1 was a negative control. 6 PBMC samples had bands present for the control primers MMP1 4/12, MMP1 9/12, ER 16/17, and RFA 17/19 at the expected size of 281bp or 556bp, 185bp, 246bp and 252bp, respectively

confirming successful 3C and loops present at these loop sites. It should be noted, however, that there was a level of variation in the RFA control. Across the 6 samples, some samples had the expected 252bp whilst other had a band that indicated a larger size product. The $CD4^+$ T cells also had bands present at a variation of sizes using primers MMP1 4/12, MMP1 9/12 and RFA 17/19. These results suggested possible incomplete digestion of chromatin resulting in multiple ligation products. The ER primer consistently showed clear bands in each sample. This illustrated the PCR protocol was successfully executed. The ERTM negative control also worked along with the non-template control (NTC).

A)


MMP 4/12
MMP 9/12
ER
L 1A1 1A2 2A1 2A1 3A1 3A2 NTC 1A1 1A2 2A1 2A1 3A1 3A2 NTC 1A1 1A2 2A1 2A2
10000
1000
500
400
300
200
100


ER
RFA
ERTM
L. 3A1 3A2 NTC 1A1 1A2 2A1 2A1 3A1 3A2 NTC 1A1 1A2 2A1 2A1 3A1 3A2 NTC
10000
1000
500
400
300
200
100

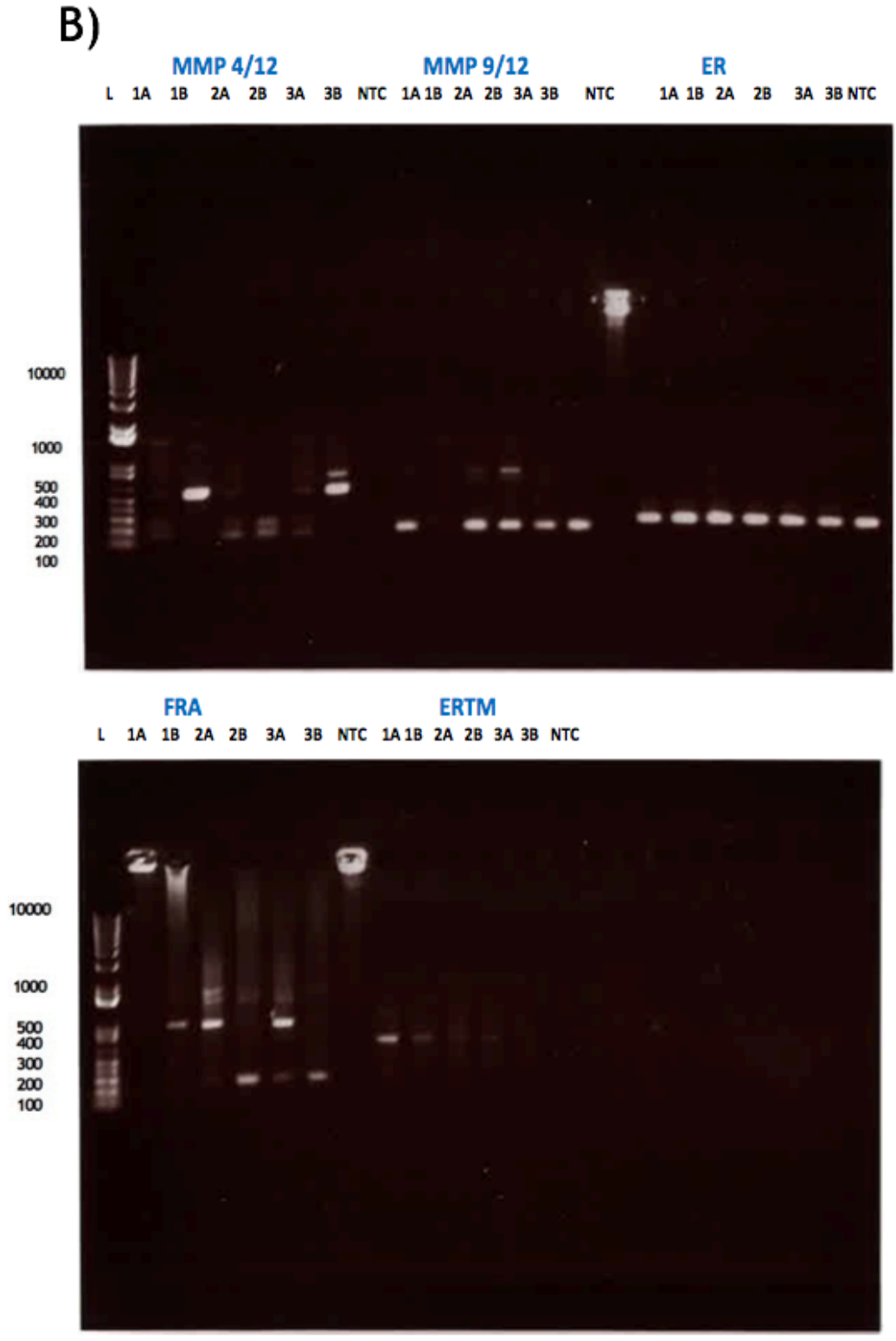


Figure 3.6 Gel Electrophoresis of Control Primers
Gel electrophoresis of nested PCR using DNA extracted from PBMCs and $CD4^+$ T cells testing 5 3C controls. 3C DNA libraries were amplified using nested PCR and then loaded onto a 1.5% gel, N=3. L= $1kb^+$ ladder, 7 μl DNA ladder and 15 μl sample loaded. Samples loaded in 1 well each in duplicate. MMP1 4/12 and MMP1 9/12 = 3C controls, ER = PCR control, RFA = positive control and ERTM = negative control. Non-template control (NTC) used for each primer. A) DNA was extracted from PBMCs. B) DNA was extracted from $CD4^+$ T cells

While the results presented in Figure 3.6 illustrated using the PCR method was successful in some samples for determining DNA loop presence, some consider it non-quantitative. Therefore, a semi-quantitative approach was attempted. After initial gel electrophoresis, MMP1 4/12 and MMP1 9/12 were considered the most robust 3C controls to use based on the most consistent presence of bands at the expected size. Additionally, after the initial gel electrophoresis, samples were loaded onto the gels in triplicate in line with the protocol used at OBD (Table 3.2). When carried out in triplicate, if a sample presents with 1 band or lower it can be stated that no loop is present at the locus of interest. However, if 2 or 3 bands are present at the predicted size, it can be stated that a loop is present at that locus. The samples for the semi-quantitative method were prepared by creating a 1:2 serial dilution of samples. When carrying out this method it would be expected that 3C copy number would decrease as the DNA concentration decreased.

The semi-quantitative method was tested with control primers and RA primers (Figure 3.7 and Figure 3.8). There was not a clear concentration dependent effect on loops using the control MMP 4/12 primer set. The neat sample had only 1 set of bands at expected size, however diluted samples of 0.25ng/ul and lower had multiple bands in some samples. Both samples used had instances with multiple bands. The images shown in Figure 3.8A reveal that only 1 patient, at 1 DNA concentration had bands present for the *IL17A* loop. The images in Figure 3.7B (*CXCL13*) demonstrate no bands were present. Due to the lack of bands in A + B, it was difficult to determine the effect of the varying DNA concentrations on 3C copy number. In contrast, Figure 3.8C (*IL21R*), D (*IL23*) and E (*IFNAR1*) highlight the semi-quantitative method well. Particularly in 3.8C, it was evident that the 3C copy number decreased gradually with decreasing DNA concentration in sample 1A1. This is also presented in 3.8D with sample 1A1. It is evident from this figure that the *IL21R*, *IL23* and *IFNAR1* loci had several loops in multiple samples. Overall, considering the unreliable semi-quantitative results in 3C controls, and the limited reliability in MTX CCS genes, the semi-quantitative method was not taken forward. However, based on the other data collected, it was confirmed that the 3C assay could be successfully conducted in my hands.

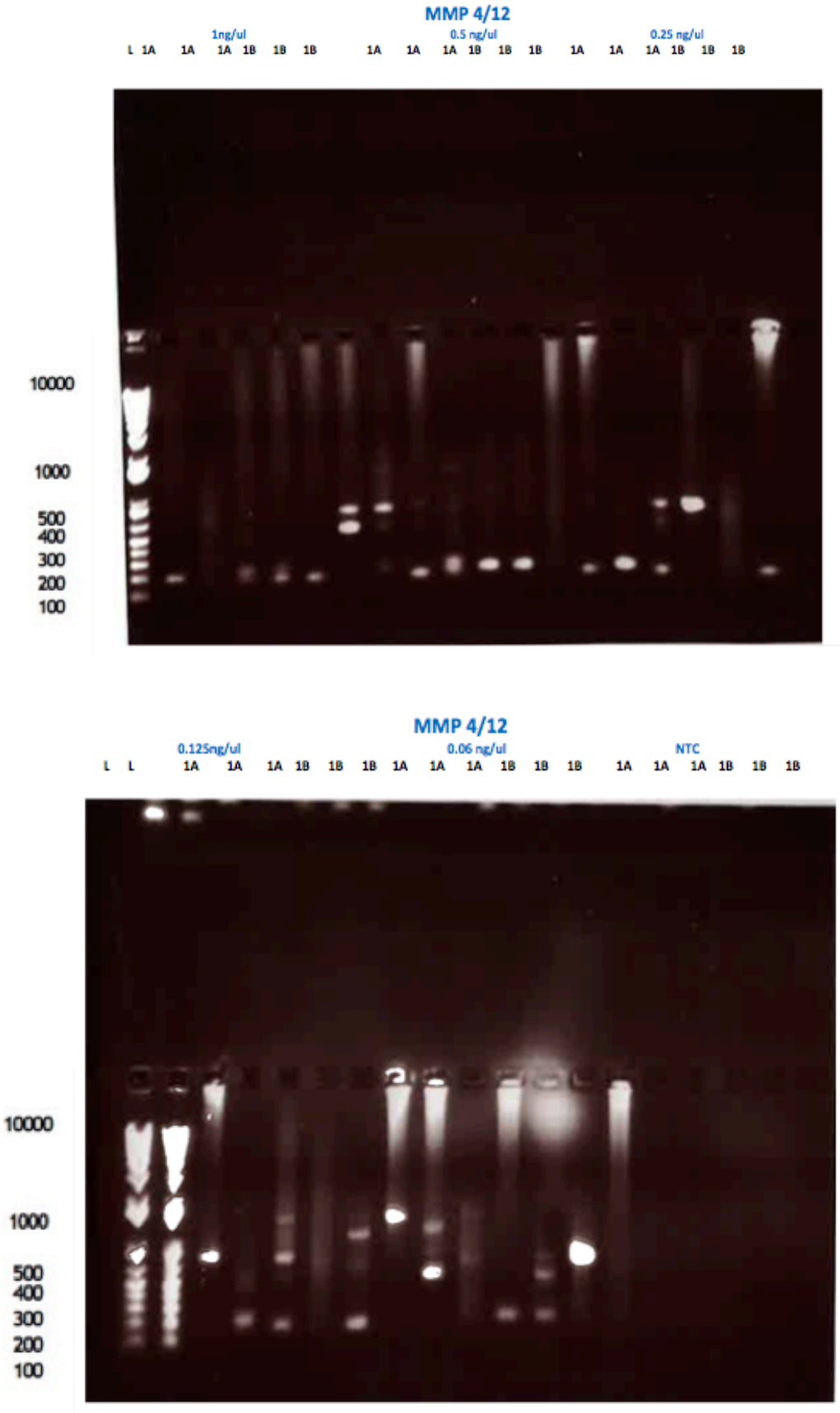
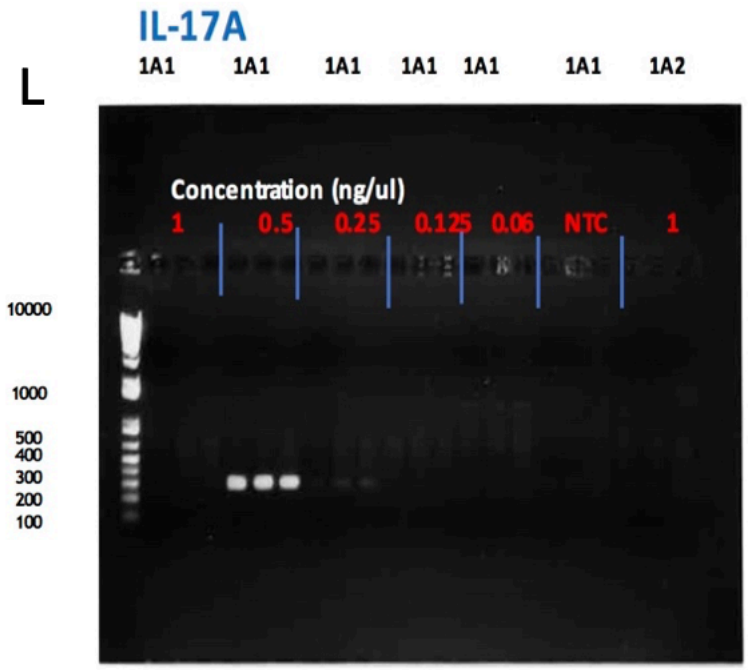
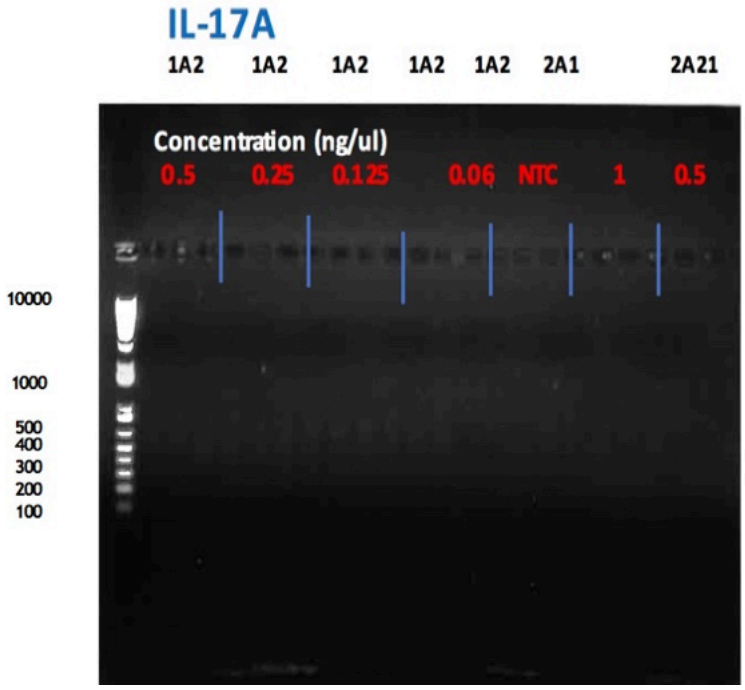


Figure 3.7 Gel Electrophoresis of 3C control loop using semi-quantitative method DNA was extracted from PBMCs of healthy donors, N=2, labelled A+B. 3C DNA libraries were amplified using nested PCR and then loaded onto a 1.5% gel. L= 1kb^{+} ladder, 7 μl DNA ladder and 13 μl of sample loaded in triplicate. Various concentrations of DNA

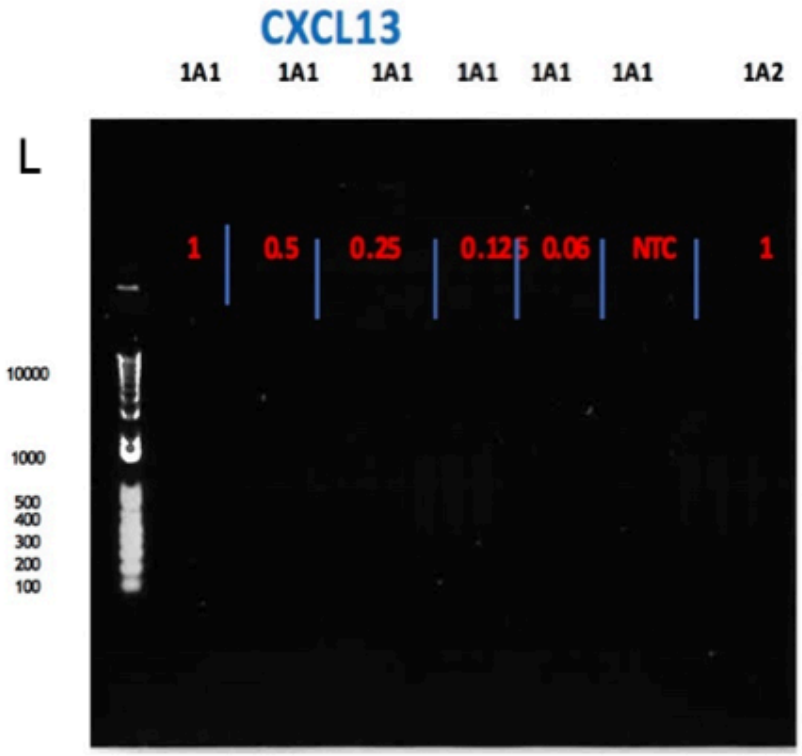
used: 1 ng/μl, 0.5 ng/μl, 0.25 ng/μl, 0.125 ng/μl, 0.06 ng/μl. Non-template control (NTC) also used.

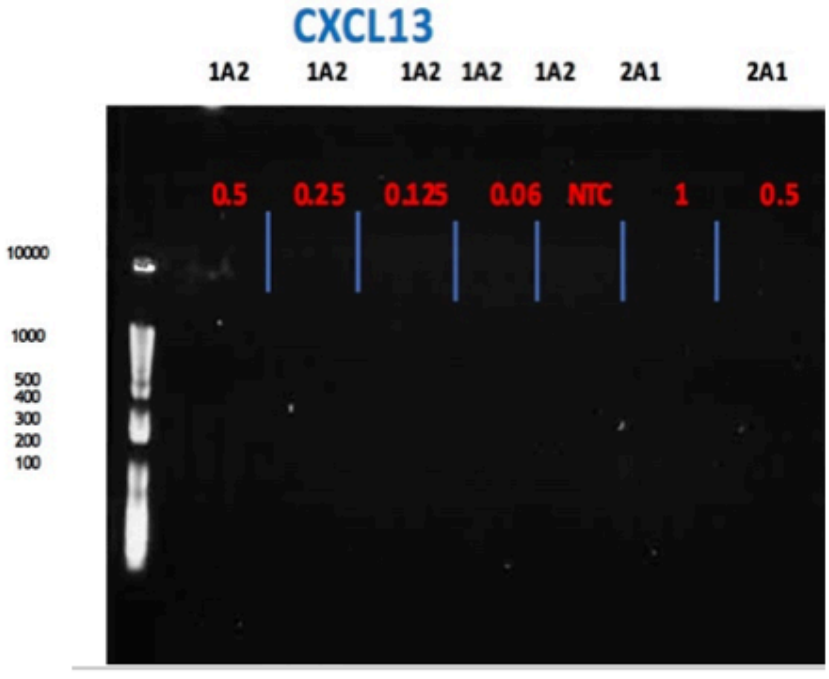
A)



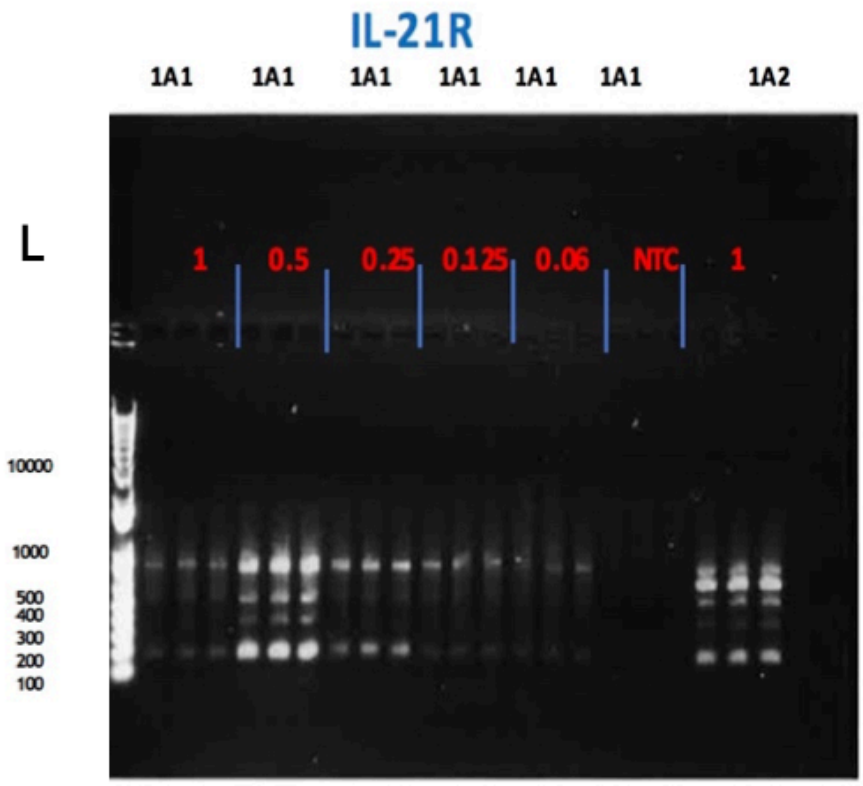


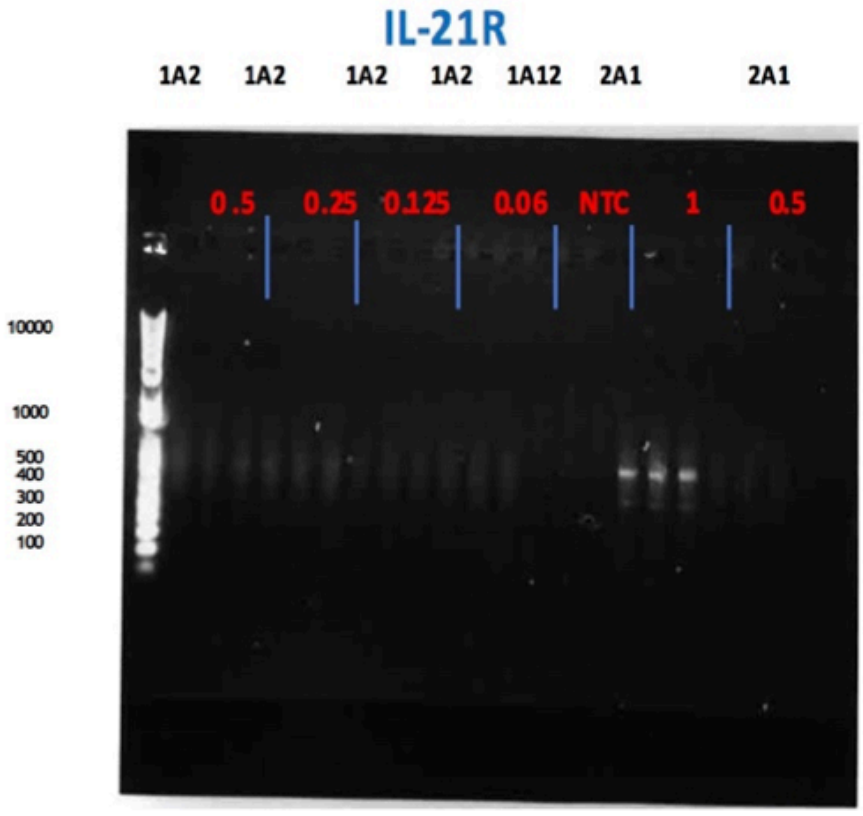
B)


CXCL13
1A1 1A1 1A1 1A1 1A1 1A1 1A2
L
1 0.5 0.25 0.125 0.06 NTC 1
10000
1000
500
400
300
200
100

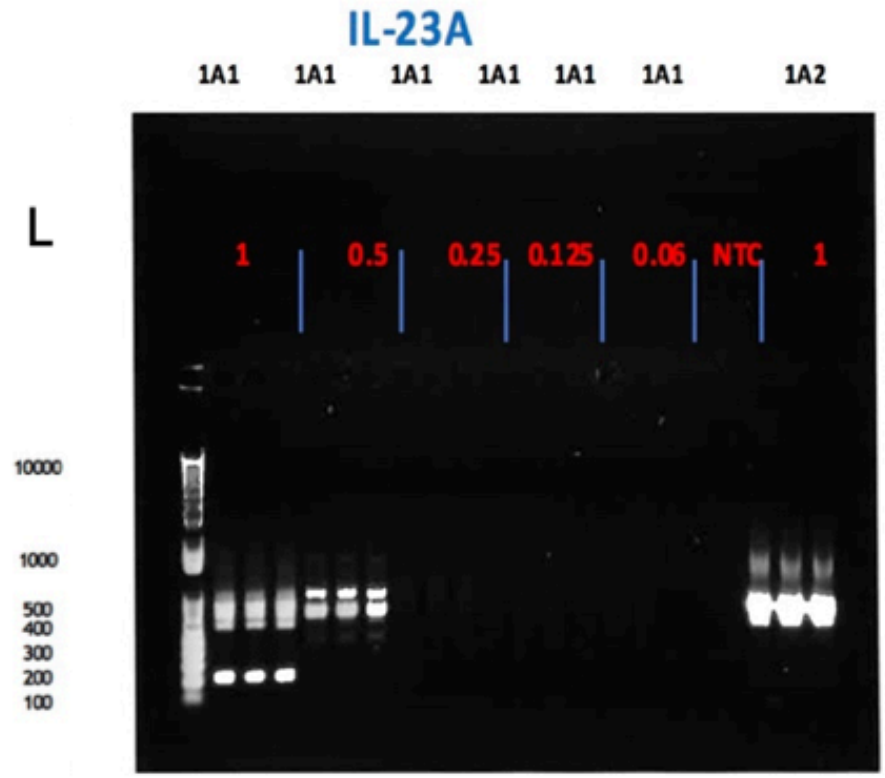

CXCL13
1A2 1A2 1A2 1A2 1A2 2A1 2A1
0.5 0.25 0.125 0.06 NTC 1 0.5
10000
1000
500
400
300
200
100

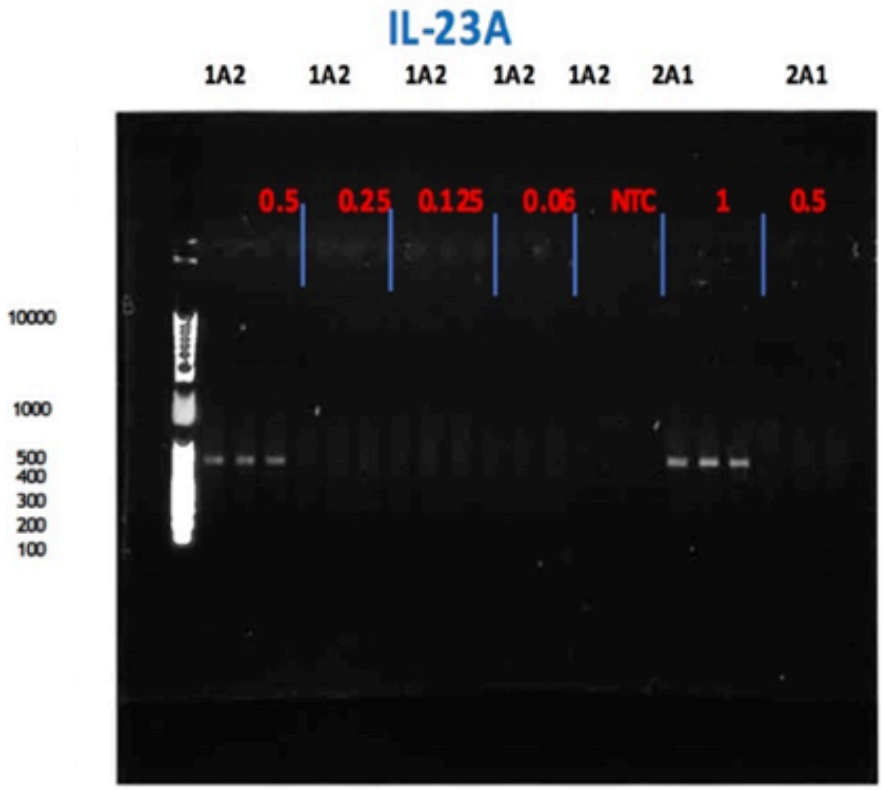
C)


IL-21R
1A1
1A1
1A1
1A1
1A1
1A1
1A2
L
1
0.5
0.25
0.125
0.06
NTC
1
10000
1000
500
400
300
200
100

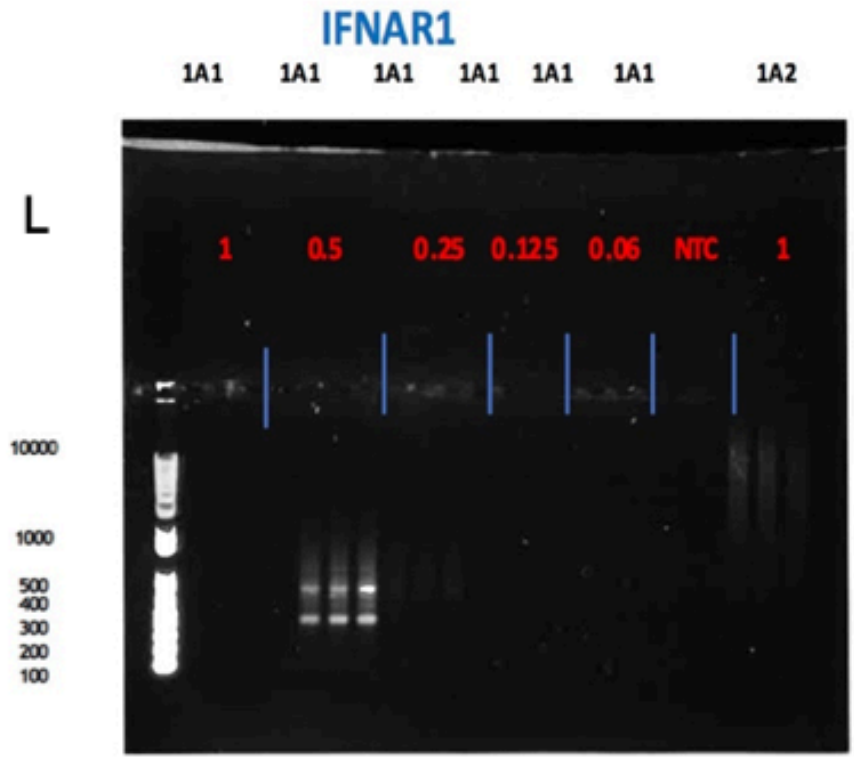

IL-21R
1A2
1A2
1A2
1A2
1A12
2A1
2A1
0.5
0.25
0.125
0.06
NTC
1
0.5
10000
1000
500
400
300
200
100

D)


IL-23A
1A1
1A1
1A1
1A1
1A1
1A1
1A2
L
1
0.5
0.25
0.125
0.06
NTC
1
10000
1000
500
400
300
200
100


IL-23A
1A2
1A2
1A2
1A2
1A2
2A1
2A1
0.5
0.25
0.125
0.06
NTC
1
0.5
10000
1000
500
400
300
200
100

E)


IFNAR1
1A1 1A1 1A1 1A1 1A1 1A1 1A2
L
1 0.5 0.25 0.125 0.06 NTC 1
10000
1000
500
400
300
200
100

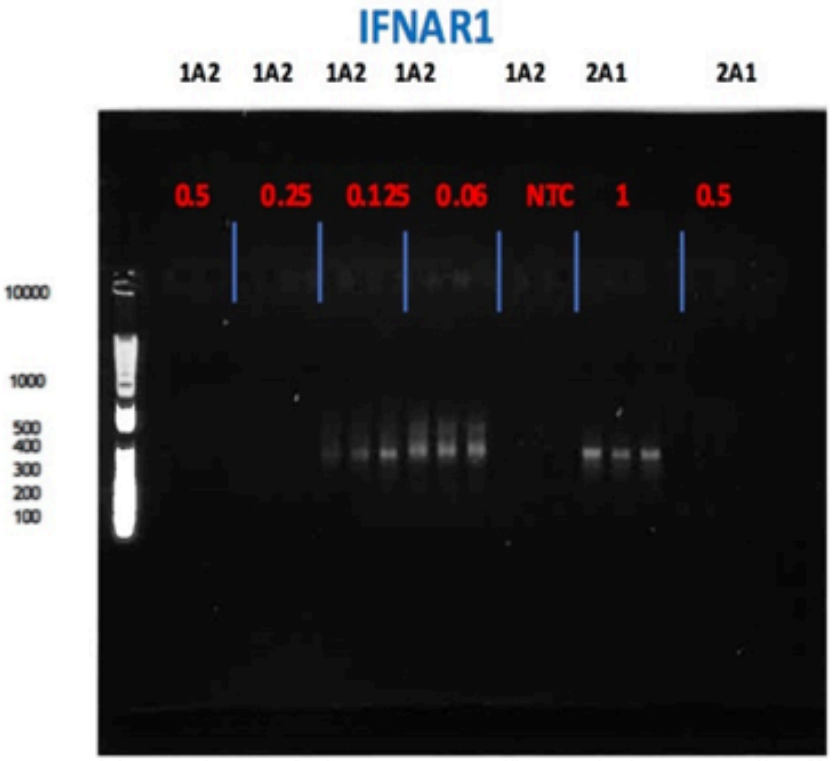
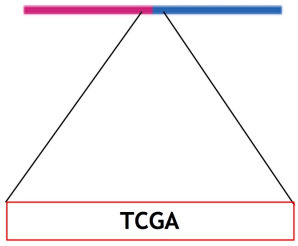

IFNAR1
1A2 1A2 1A2 1A2 1A2 2A1 2A1
0.5 0.25 0.125 0.06 NTC 1 0.5
10000
1000
500
400
300
200
100

Figure 3.8 Gel Electrophoresis of MTX CCS Loops using Semi-quantitative Method
DNA was extracted from PBMCs of healthy donors, N=3, labelledA1, A2 + 2A. 3C DNA libraries were amplified using nested PCR and then loaded onto a 1.5% gel. L= 1kb^{+} ladder, 7 μl DNA ladder and 13 μl of sample loaded in triplicate. Various concentrations of DNA used: 1 ng/μl, 0.5 ng/μl, 0.25 ng/μl, 0.125 ng/μl, 0.06 ng/μl. Non-template control (NTC) also used. A) RA1 (*IL17A*). B) RA2(*CXCL13*). C) RA3 (*IL21R*). D) RA4(*IL23*). E) RA5(*IFNAR1*). Patient sample annotation indicated above gel image, DNA library concentration shown below gel.

3.2.3 Establishing use of quantitative PCR for MTX CCS

Having verified that it was possible to run the original 3C method independently, but clearly demonstrating that the semi-quantitate method failed to produce robust results, I considered introducing a higher-throughput, more informative method of observing chromosomal loops. Moreover, OBD were in the process of transitioning to this methodology and therefore this work aligned with the direction of travel for EpiSwitch^{TM} technology. A more informative method would be one which could successfully quantify the loops within a patient sample. A higher-throughput method would not only be of benefit in the short-term of the study, but also in the long-term if the signature was to be implemented clinically. Quantitative PCR (qPCR) would fill these criteria, however, the development and refinement of this process for chromosomal loops is complex and required a systematic approach. Various steps were carried out to determine if qPCR primers were suitable for detection of chromosomal loops of interest, and determine the optimal annealing temperature to use for the primers. This was carried out in accordance with the MIQE guidelines. Primers were designed to capture the 3C ligated DNA product which centred around a 4 base TCGA sequence (Figure 3.9A). Firstly, primers were designed (Section 2.2.7) and tested for the 3C control gene, MMP1 (Figure 3.10). For the process of determining the optimal primer annealing temperature, 3 primer versions for each loop of the MTX CCS were tested.

A)



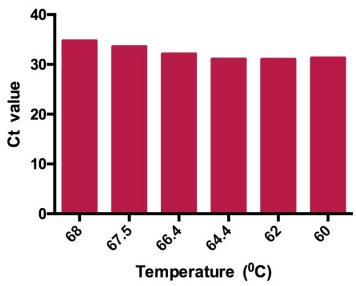
B)

Primer Version	Gene	Expected Amplicon size (bp)
A		
	IL17A	84
	CXCL13	135
	IL21R	110
	IL23	138
	IFNAR1	117
B		
	IL17A	94
	CXCL13	104
	IL21R	119
	IL23	81
	IFNAR1	117
C		
	IL17A	121
	CXCL13	116
	IL21R	119
	IL23	85
	IFNAR1	149

Figure 3.9 Primer Design for qPCR
A) Ligated DNA region in which qPCR primers are designed around, full sequences found in Appendix. B) Expected sizes of qPCR products for all primers designed and tested

The chromosomal loop, from pooled RA patient samples, was amplified at 6 annealing temperatures and gave cycle threshold (Ct) values ranging from 31 to 34 (Figure 3.10A). Melt curve analysis showed amplification of a single product (Figure 3.10B). This could be considered a successful qPCR amplification. The high Ct values for MMP1 may reflect weak ligation and the results were a possible indication that other loops from the MTX CCS may also amplify at a higher Ct value than expected for other qPCR reactions.

A)



B)

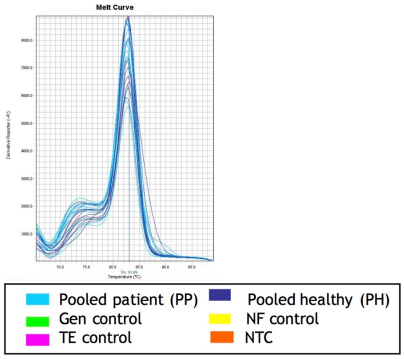
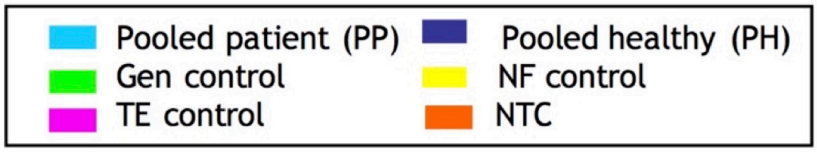


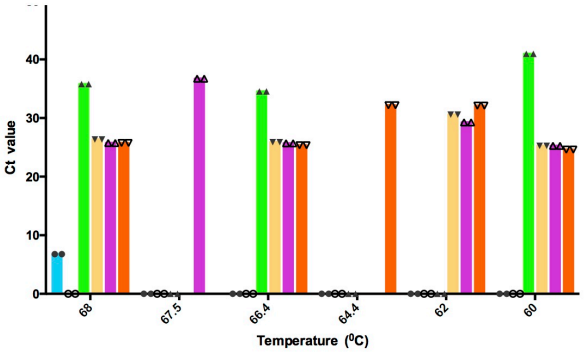
Figure 3.10 Quantitative PCR with 3C control
Quantitative PCR using pooled RA patient samples with MMP1 primers. A) Representative plot from 1 qPCR experiment of Ct values at 6 annealing temperatures with pooled RA patient samples, (N=8). B) Representative melt curve of from 1 qPCR experiment that was repeated twice with pooled patient samples (N=8): NF, non-fixtion, NTC, not-template control; PH, pooled healthy, PP, pooled patient; TE, Tris-EDTA; NF= non-

fixation - control generated by using nuclease-free water to fix the DNA, instead of formaldehyde.

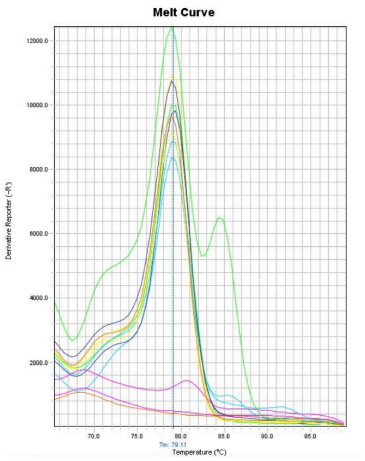
Once it was confirmed qPCR could be successful for amplifying chromosomal loops of interest, the next step was to determine the optimal annealing temperature of qPCR primers for all 5 loci in the MTX signature. This single step process used a temperature gradient to identify the optimum annealing temperature and overall primer efficiency. Amplification of only the product of interest (loop in pooled patient sample), without any product present in the negative controls (pooled healthy sample, Gen control, NF control, TE control, NTC) was the aim. The first primer set for *IL17A* amplified the product of interest at a Ct value of 9, at an annealing temperature of 68°C, alongside amplification of negative controls at higher Ct values (Figure 3.11A). Primer set B for *IL17A* could amplify the product of interest at all 6 annealing temperatures, however, there was also better amplification of negative controls at all 6 annealing temperatures (Figure 3.11B). Primer set 3 was also able to amplify the product of interest at all 6 temperatures at Ct values ranging from 27 to 40. At 67.5°C only the product of interest was amplified (Figure 3.11G). The high Ct values of primer set B and C suggested presence of off-target products. Melt curve analysis was conducted alongside to confirm primer specificity. Many melt curves for *IL17A* primers, had multiple peaks (Figure 3.11B,E,H) indicating the presence of off-target amplification products. The qPCR products were then purified using gel electrophoresis, which could also be used to check the amplified product was of the expected size. All 3 primer sets had products at the expected size (representative gel images can be found in Appendix). There was also the presence of multiple bands for all primers sets in pooled patient (PP) and pooled healthy (PH) samples at lower annealing temperatures. Primer set A and B did not meet requirements, however primer set C met requirements at 67.5°C only. As pooled heathy controls were amplified at neighbouring annealing temperatures, these results should be interpreted with caution.


Pooled patient (PP)
Pooled healthy (PH)
Gen control
NF control
TE control
NTC

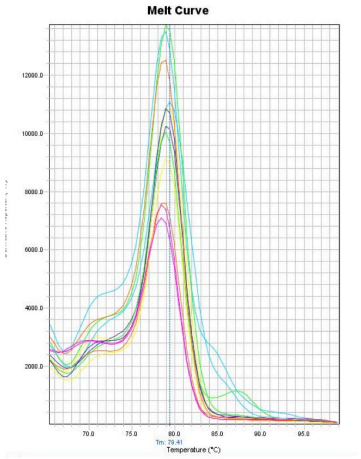
A)


Ct value
Temperature (°C)
68
67.5
66.4
64.4
62
60

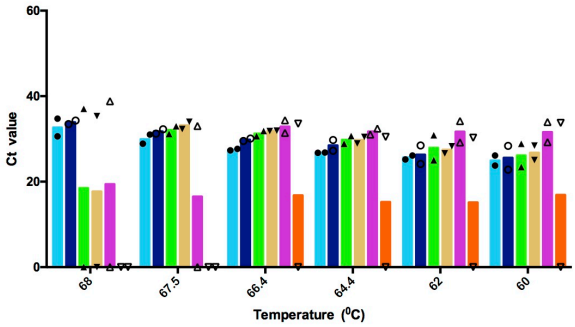
B)


Melt Curve
Temperature (°C)

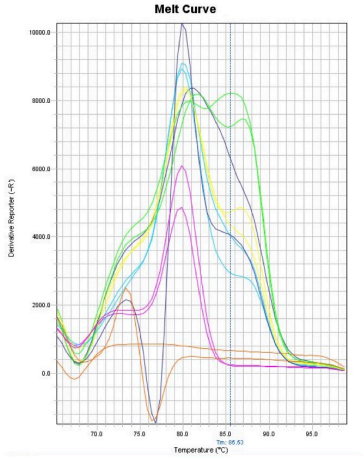
C)


Melt Curve
Temperature (°C)

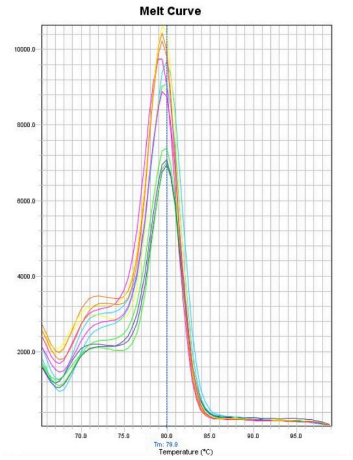
D)


Ct value
Temperature (°C)
68
67.5
66.4
64.4
62
60

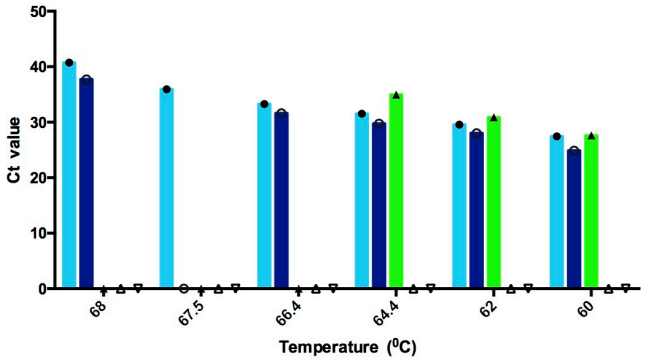
E)


Melt Curve
Temperature (°C)

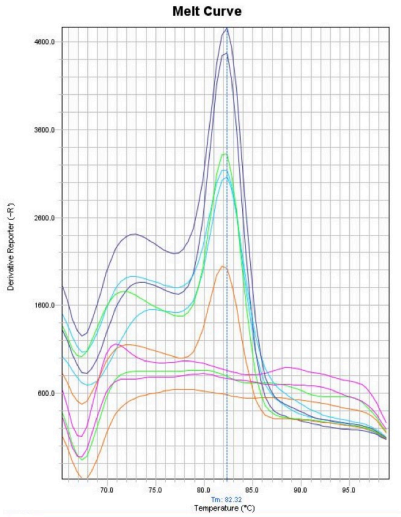
F)


Melt Curve
Temperature (°C)

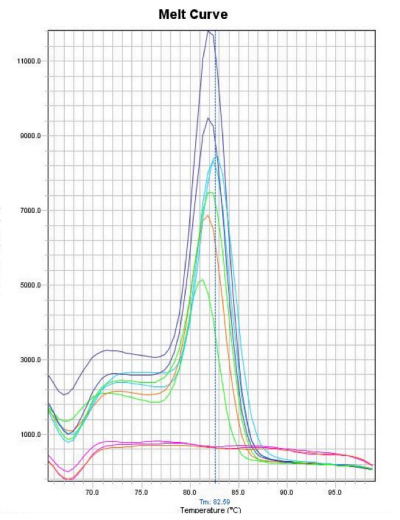
G)



H)



I)

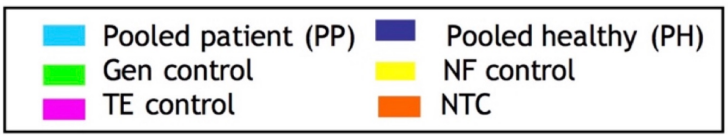


J)

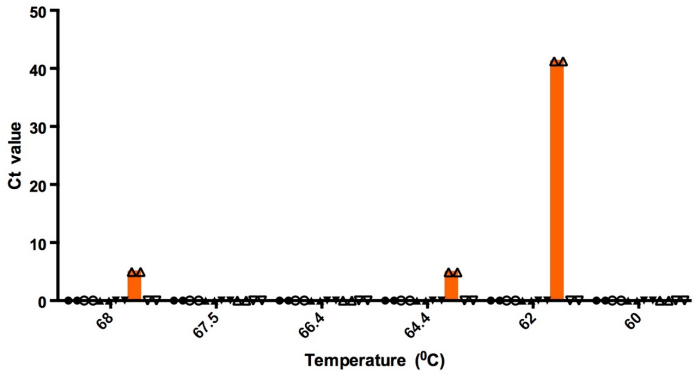
Primer version	Temp (⁰C)	PP		PH		Gen		NF		TE		NTC	
A	68	1	1	0	1	0	1	0	1	0	0	0	0
	67.5	1	1	0	1	0	1	0	1	0	0	0	0
	66.4	1	1	0	1	0	1	0	1	0	0	0	1
	64.4	1	1	1	1	2	1	0	1	0	0	0	1
	62	2	1	2	1	2	1	0	1	0	0	0	1
	60	2	1	2	1	2	1	0	1	0	0	0	1
B	68	1	1	1	1	0	0	1	0	0	1	0	0
	67.5	1	1	1	1	1	1	1	0	0	1	0	1
	66.4	1	1	1	1	1	1	1	1	1	1	0	1
	64.4	1	1	1	1	2	1	1	1	1	1	0	1
	62	1	1	1	1	1	1	0	1	0	1	0	1
	60	1	1	1	1	1	1	1	1	1	1	1	1
C	68	1		1		1		0		1		1	
	67.5	0		0		0		0		0		1	
	66.4	1		1		0		0		1		1	
	64.4	1		1		1		0		1		2	
	62	2		2		1		0		1		1	
	60	3		3		1		0		1		1	

Figure 3.11 Temperature Gradient Quantitative PCR with *IL17A* primers
3C DNA libraries were loaded in duplicate on 96-well reaction plate. qPCR analysis of 3 versions of *IL17A* primers, Ct values shown as individual samples, and mean of 2. A) Ct values of version A primers, N=2. B) Representative melt curve at 68°C. C) Representative melt curve at 66.4°C. D) Ct values of version B primers, N=2. E) Representative melt curve at 62°C. F) Representative melt curve at 64.4°C. G) Ct values from version C primers, N=1 H) Representative melt curve at 68°C. I) Representative melt curve at 64.4°C. J) Summary of gel electrophoresis results from 3 primer versions. Numbers highlighted in red indicate one of the bands are at the expected size of 84bp for version A primers (2 technical repeats), 94bp for version B primers (2 technical repeats) and 104bp for version C primers.
NF, non-fixation, NTC, not-template control; PH, pooled healthy, PP, pooled patient; TE, Tris-EDTA

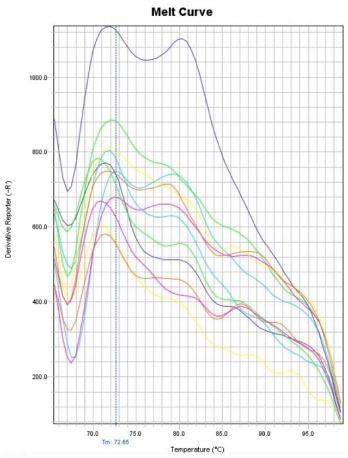
The first primer set tested for *CXCL13* was unable to amplify the product of interest. At 68°C, 64.4°C, and 62°C the NTC control was amplified. Melt curves from 68°C (Figure 3.12B) and 66.4°C (Figure 3.12C) highlighted the presence of multiple products. The second primer set was also unable to amplify the product of interest (Figure 3.12D). At 62°C and 60°C there was amplification of the genomic control and TE control respectively. The melt curve from 64.4°C (Figure 3.12F) suggested there was a single product. The melt curve from 60°C (Figure 3.12E) also showed evidence that the product of interest was amplified. The product of interest was amplified with primer set C, at annealing temperatures of 64.4°C, 62°C and 60°C at Ct values of 33, 31 and 32, respectively. At each of those temperatures there was also amplification of negative controls. The melt curve at 60°C indicated that there was only a single product of interest amplified (Figure 3.12I). Gel electrophoresis of qPCR products for primer set B were reflective of amplification and melt curve results. A selection of products of interest at the expected size of 116 and 135bp were present on a gel for primer set A and C. No functional primer set was found for the *CXCL13* chromosomal loop.


Pooled patient (PP)
Pooled healthy (PH)
Gen control
NF control
TE control
NTC

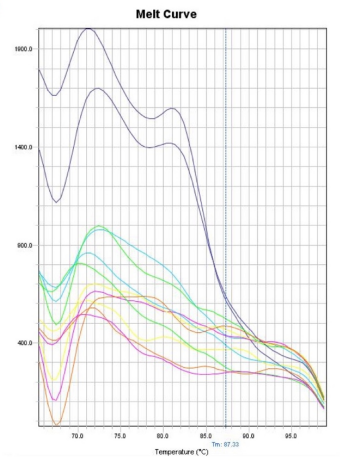
A)


Ct value
50
40
30
20
10
0
68
67.5
66.4
64.4
62
60
Temperature (⁰C)

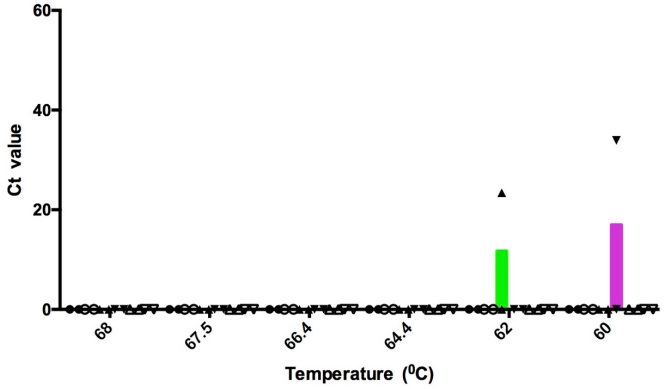
B)


Melt Curve
Temperature (°C)

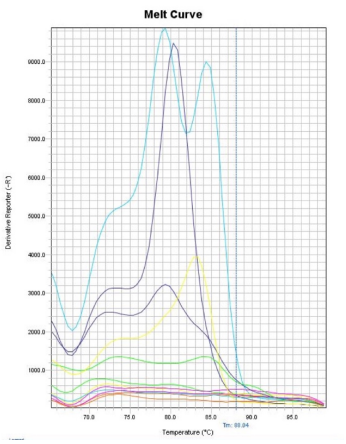
C)


Melt Curve
Temperature (°C)

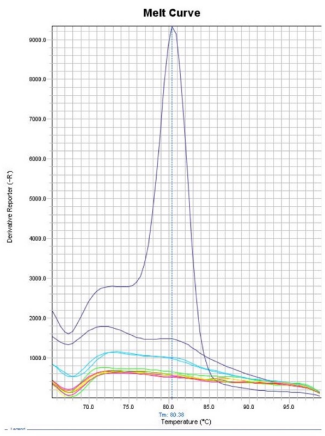
D)


Ct value
60
40
20
0
68
67.5
66.4
64.4
62
60
Temperature (⁰C)

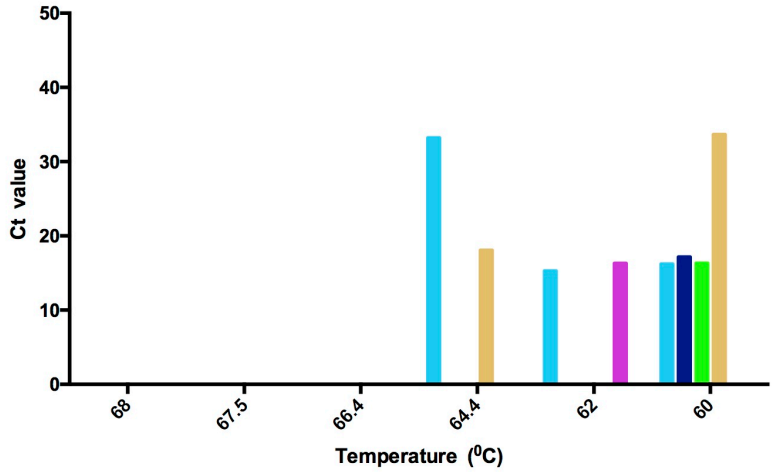
E)


Melt Curve
Temperature (°C)

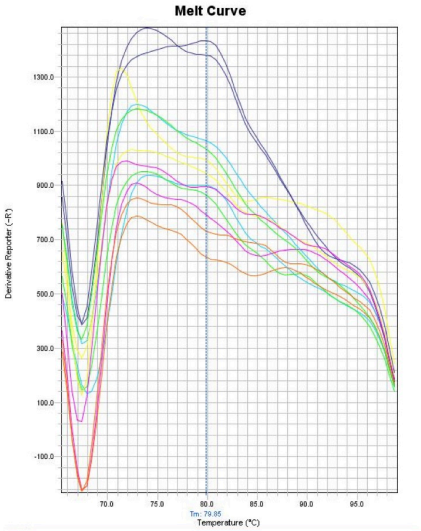
F)


Melt Curve
Temperature (°C)

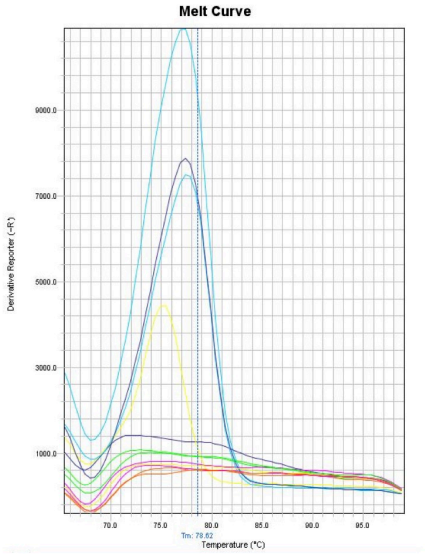
G)



H)



I)

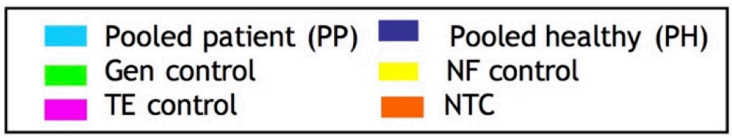


J)

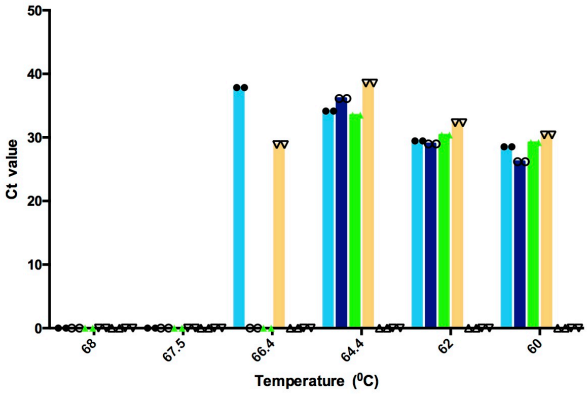
Primer version	Temp (⁰C)	PP		PH		Gen		NF		TE		NTC	
A	68	0	1	0	1	0	1	0	1	0	0	0	0
	67.5	0	1	0	1	0	1	0	1	0	0	0	1
	66.4	0	1	0	1	0	1	0	1	0	0	0	0
	64.4	0	0	0	0	0	0	0	0	0	0	0	0
	62	0	0	0	0	0	0	0	0	0	0	0	0
	60	0	0	0	0	0	0	0	0	0	0	0	0
B	68	0	0	0	0	0	0	0	0	0	0	0	0
	67.5	0	0	0	0	0	0	0	0	0	0	0	0
	66.4	0	0	0	0	0	0	0	0	0	0	0	0
	64.4	0	0	0	0	0	0	0	0	0	0	0	0
	62	0	0	0	1	0	0	0	0	0	0	0	0
	60	0	1	0	1	0	1	0	1	0	0	0	0
C	68	0	0	0`	0	0	0	0	0	0	0	0	0
	67.5	0	0	0	0	0	0	0	0	0	0	0	0
	66.4	0	0	0	0	0	0	0	0	0	0	0	0
	64.4	1	1	1	1	0	0	0	1	1	0	1	1
	62	1	1	1	0	0	0	1	0	0	0	1	0
	60	1	1	1	1	0	0	0	1	0	1	0	0

Figure 3.12 Temperature Gradient Quantitative PCR with *CXCL13* Primers
3C DNA libraries were loaded in duplicate on 96-well reaction plate. qPCR analysis of 3 versions of *CXCL13* primers, Ct values shown as individual samples, and mean. A) Ct values of version A primers, N=2. B) Representative melt curve at 68°C. C) Representative melt curve at 66.4°C. D) Ct values of version B primers, N=2. E) Representative melt curve at 60°C. F) Representative melt curve at 64.4°C. G) Ct values form version C primers, N=1. H) Representative melt curve at 68°C. I) Representative melt curve at 60°C. J) Summary of gel electrophoresis results from 3 primer versions. Numbers highlighted in red indicate one of the bands are at the expected size of 135bp for version A primers, 104bp for version B primers and 116bp for version C primers, 2 technical repeats for all primer sets.
NF, non-fixation, NTC, not-template control; PH, pooled healthy, PP, pooled patient; TE, Tris-EDTA

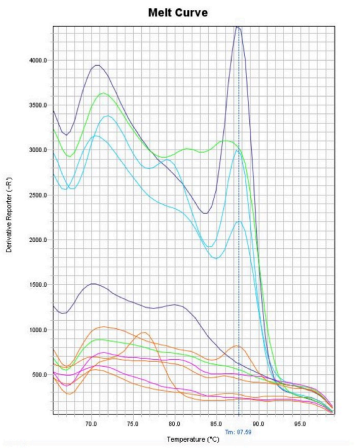
IL21R primers could amplify the product of interest, but only alongside amplification of negative controls. Temperatures of 66.4°C to 60°C for primer set A had amplification of the pooled patient (PP) sample (Figure 3.13A). This was accompanied by amplification of negative controls. Melt curves from primer set A had multiple peaks indicating presence of off-target products (Figure 3.13B,C). Primer set B successfully amplified the product of interest at all 6 annealing temperatures (Figure 3.13D). Negative controls were also amplified at all 6 temperatures. Melt curve from 64.4°C indicated peaks for only PP and PH. Primer set C also had amplification of PP at all annealing temperatures, as well as amplification of negative controls (Figure 3.13G). Gel electrophoresis of the qPCR results highlighted the presence of multiple bands at all temperatures in PP and PH samples (Figure 3.13J). Primer set A did not have any qPCR products at the expected size, however primer set B and C had bands at expected size. Primer set B and C met primer requirements at some temperatures. However, taken together, there is not enough evidence that any *IL21R* primer set could successfully translate to a qPCR platform.


Pooled patient (PP)
Pooled healthy (PH)
Gen control
NF control
TE control
NTC

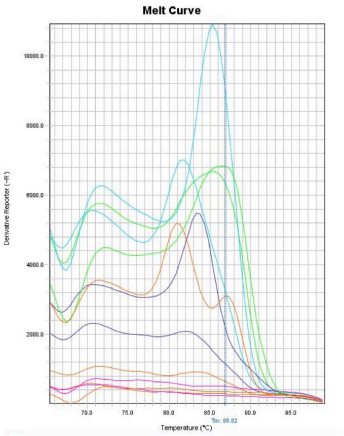
A)


Ct value
Temperature (⁰C)
68
67.5
66.4
64.4
62
60

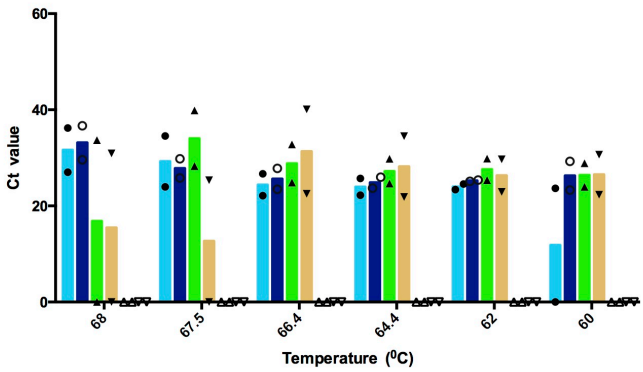
B)


Melt Curve
Temperature (°C)

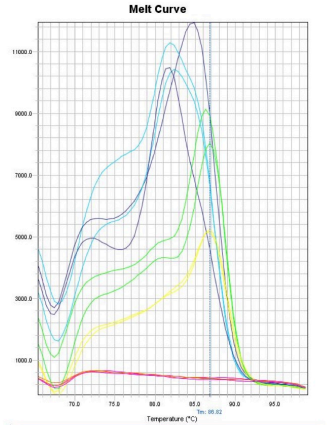
C)


Melt Curve
Temperature (°C)

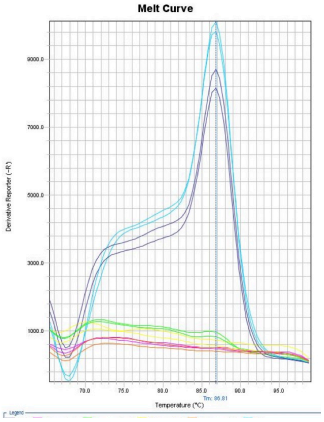
D)


Ct value
Temperature (⁰C)
68
67.5
66.4
64.4
62
60

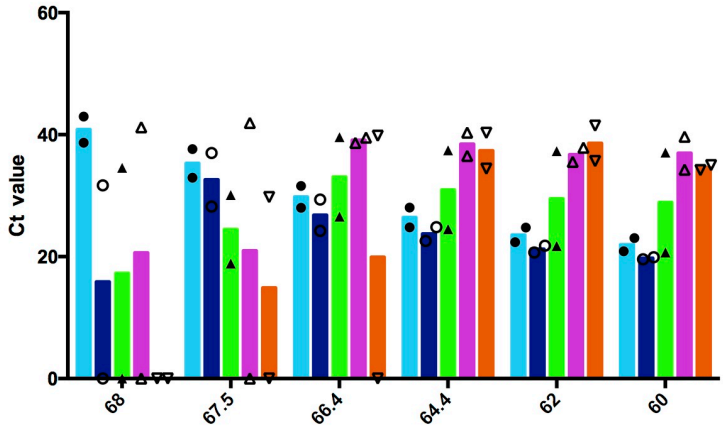
E)


Melt Curve
Temperature (°C)

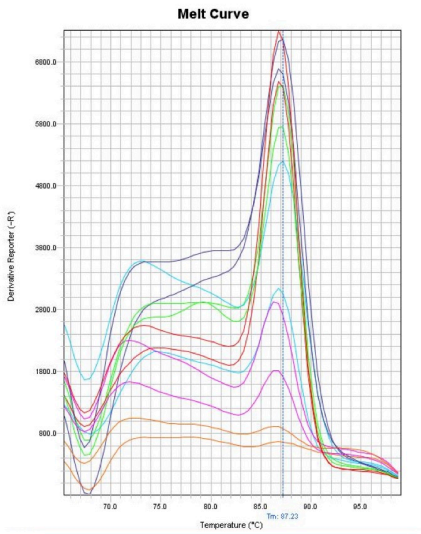
F)


Melt Curve
Temperature (°C)

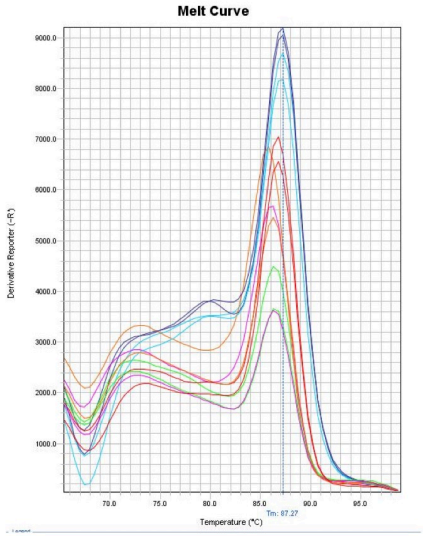
G)



H)



I)

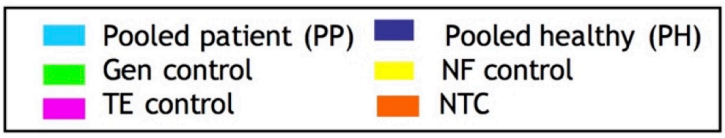


J)

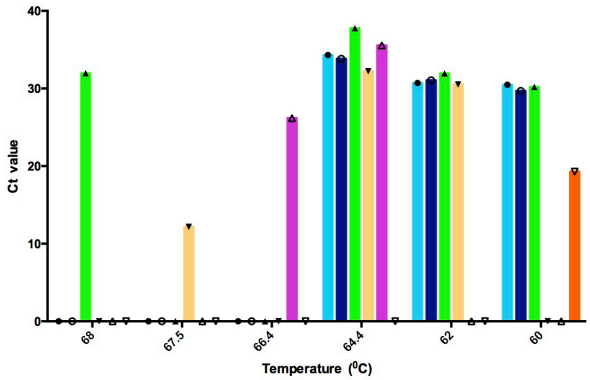
Primer version	Temp (⁰C)	PP		PH		Gen		NF		TE		NTC	
A	68 67.5 66.4 64.4 62 60	0 0 0	0 0 0 1 1 2	0 0 0	0 0 0 2 3 2	0 0 0	0 0 0 0 2 2	0 0 0	0 0 0 1 2 2	0 0 0	0 0 0 0 0 0	0 0 0	0 0 0 0 0 0
B	68 67.5 66.4 64.4 62 60	0 0 1 2 2 2	1 1 0 1 2 2	0 0 1 2 2 2	1 1 0 1 3 0	1 1 1 2 2 2	0 0 0 3 3 2	1 1 1 0 0 0	0 0 0 2 3 2	0 0 0 0 0 1	0 0 0 0 0 0	0 0 0 1 1 1	0 0 0 0 0 0
C	68 67.5 66.4 64.4 62 60	1 3 0 3 3 3	0 1 2 3 3 2	3 3 1 3 3 3	0 1 2 3 3 2	2 2 2 0 0 0	0 0 0 1 1 1			0 0 0 0 0 1	0 1 0 1 1 1	1 0 1 1 1 1	1 1 0 1 1 1

Figure 3.13 Temperature Gradient Quantitative PCR with *IL21R* Primers
3C DNA libraries were loaded in duplicate on 96-well reaction plate. qPCR analysis of 3 versions of *IL21R* primers, Ct values shown as individual samples, and mean. A) Ct values of version A primers, N=2. B) Representative melt curve at 66.4°C. C) Representative melt curve at 64.4°C. D) Ct values of version B primers, N=2. E) Representative melt curve at 64.4°C. F) Representative melt curve at 68°C. G) Ct values form version C primers, N=2. H) Representative melt curve at 68°C. I) Representative melt curve at 64.4°C. J) Summary of gel electrophoresis results from 3 primer versions. Numbers highlighted in red indicate one of the bands are at the expected size of 110bp for version A primers, 119bp for version B primers and 119bp for version C primers 2 technical repeats for all primer sets.
NF, non-fixation, NTC, not-template control; PH, pooled healthy, PP, pooled patient; TE, Tris-EDTA

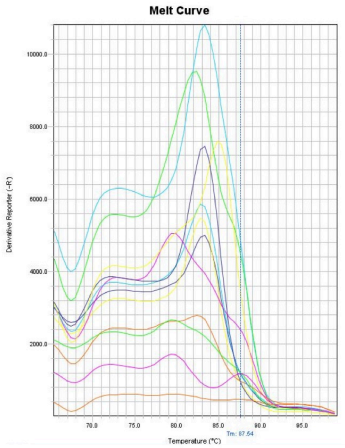
Like *IL21R*, *IL23* primer sets achieved amplification of the patient product of interest but not without amplification of control primers. Primer set A saw amplification at temperatures 64.4°C, 62°C and 60°C. Amplification of negative controls also occurred (Figure 3.14A). Primer set B also had amplification of the product of interest and negative controls at these temperatures, as well as 66.4°C (Figure 3.14D). There was amplification of only the product of interest at annealing temperature of 66.4°C using primer set C (Figure 3.14G). Amplification also occurred at the 3 lower annealing temperatures along with negative controls. The melt curves for primer set A and B indicate the presence of non-specific products (Figure 3.14B,C,E,F). At 62°C using Primer set C, the melt curves appeared to have less evidence of non-specific products (Figure 3.14I). Primer set A produced no products at the expected size when gel electrophoresis of qPCR products was conducted (Figure 3.14J). At temperatures from 64.4°C to 60°C primers set B and C had bands at the expected sizes. Primer C met requirements at 66.4°C but again, annealing temperatures 2°C below had amplification of negative controls and this that temperature should be considered further.


Pooled patient (PP)
Pooled healthy (PH)
Gen control
NF control
TE control
NTC

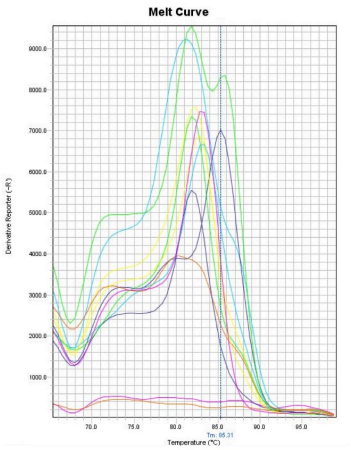
A)


Ct value
Temperature (°C)
68
67.5
66.4
64.4
62
60

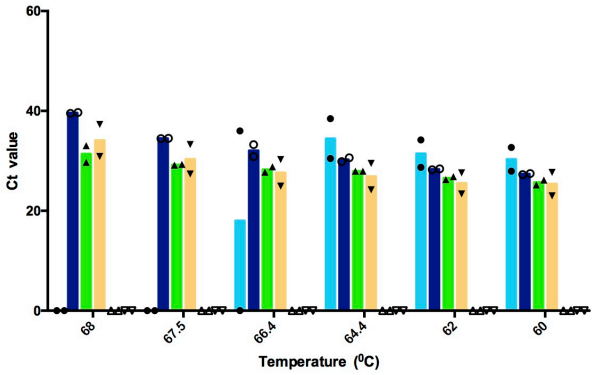
B)


Melt Curve
Temperature (°C)

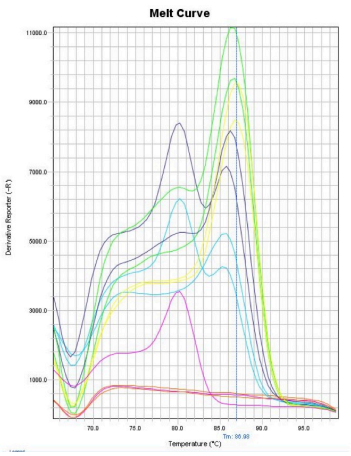
C)


Melt Curve
Temperature (°C)

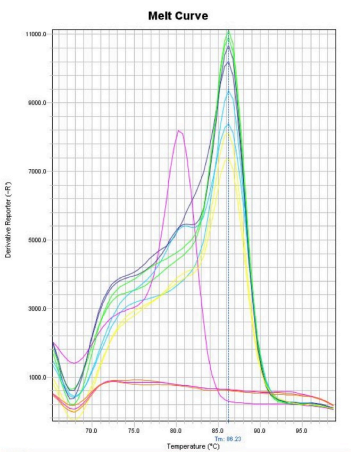
D)


Ct value
Temperature (°C)
68
67.5
66.4
64.4
62
60

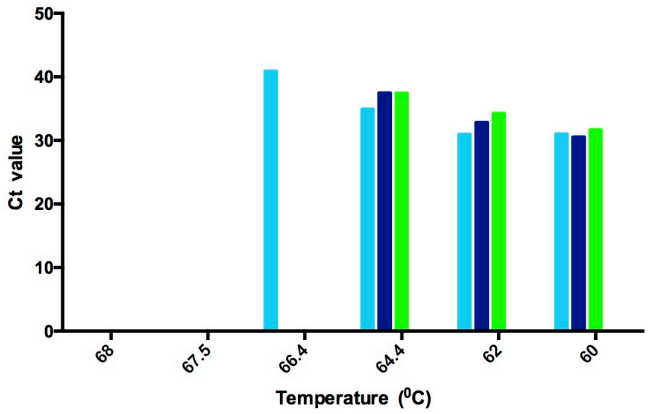
E)


Melt Curve
Temperature (°C)

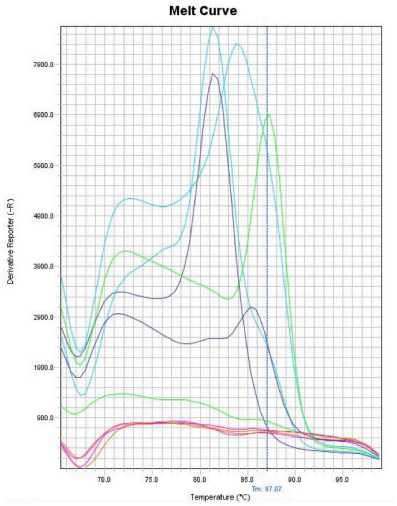
F)


Melt Curve
Temperature (°C)

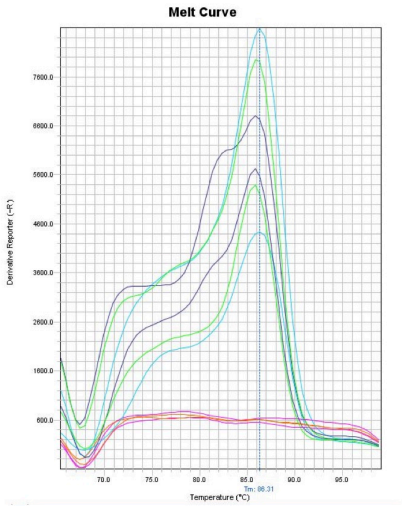
G)



H)



I)

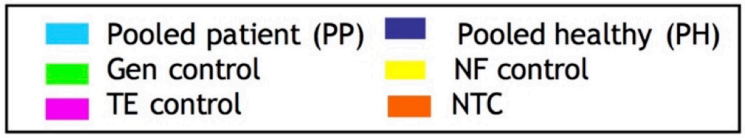


J)

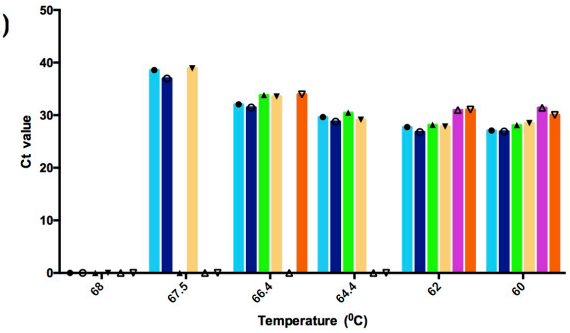
Primer version	Temp (⁰C)	PP		PH		Gen		NF		TE		NTC	
A	68	0		0		0		0		0		0	
	67.5	0		0		0		0		0		0	
	66.4	0		0		0		0		0		0	
	64.4	0		1		0		0		0		0	
	62	1		2		2		0		0		0	
	60	2		0		2		0		0		0	
B	68	0	0	0	1	2	2	2	1	0	0	0	0
	67.5	1	2	2	2	1	2	0	0	0	0	0	1
	66.4	0	2	0	2	1	0	1	0	0	1	0	1
	64.4	2	1	3	1	2	2	1	2	0	0	0	1
	62	2	2	2	2	2	2	0	1	0	1	0	1
	60	1	1	1	1	1	1	1	1	0	0	0	0
C	68	0		0		0				0		1	
	67.5	0		0		0				0		1	
	66.4	0		0		0				0		1	
	64.4	1		0		0				0		1	
	62	1		1		1				0		1	
	60	2		2		2				0		1	

Figure 3.14 Temperature Gradient Quantitative PCR with *IL23* Primers
3C DNA libraries were loaded in duplicate on 96-well reaction plate. qPCR analysis of 3 versions of *IL23* primers, Ct values shown as individual samples, and mean. A) Ct values of version A primers, N=2. B) Representative melt curve at 62°C. C) Representative melt curve at 60°C. D) Ct values of version B primers, N=2. E) representative melt curve at 66.4°C. F) Representative melt curve at 62°C. G) Ct values form version C primers, N=1. H) representative melt curve at 64.4°C, I) Representative melt curve at 60°C. J) summary of gel electrophoresis results from 3 primer versions. Numbers highlighted in red indicate one of the bands are at the expected size of 80bp for version A primers, 81bp for version B primers and 85bp (2 technical repeats) for version C primers.
NF, non-fixtion, NTC, not-template control; PH, pooled healthy, PP, pooled patient; TE, Tris-EDTA

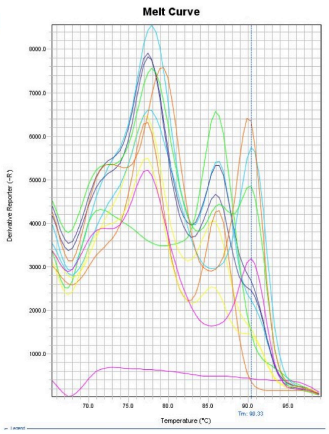
qPCR using primer sets A, B and C for the *IFNAR1* loop resulted in amplification of the product of interest, but not without amplification of negative controls. There was amplification of the patient product with annealing temperatures ranging from 67.5°C to 60°C using primer set A (Figure 3.15A). There was also amplification of negative controls at these temperatures. Primer set B achieved amplification of PP at temperatures from 66.4°C to 60°C (Figure 3.15D). Using primer set C, the product of interest was not amplified without negative controls (Figure 3.15G). Melt curves using primers set A produced varying results. At annealing temperature 66.4°C the melt curve shows evidence of multiple products (Figure 3.15B), but at 62°C the melt curve suggests only presence of a single product (Figure 3.15C). Melt curves using primer set B and C suggest presence of non-specific products (Figure 3.15E,F,H,I). There were bands at the expected sizes using all 3 primer sets, as well as multiple bands at all temperatures in many samples reflecting the melt curve results (Figure 3.15J). Primer set B and C met some requirements at some temperatures, but it was not robust.


Pooled patient (PP)
Pooled healthy (PH)
Gen control
NF control
TE control
NTC

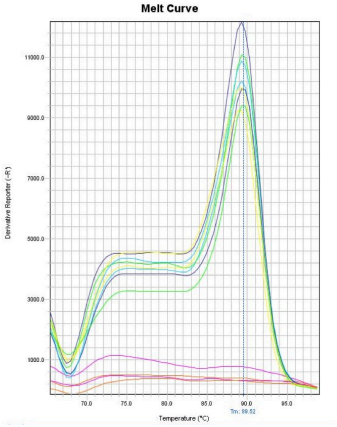
A)


Ct value
Temperature (⁰C)
68
67.5
66.4
64.4
62
60

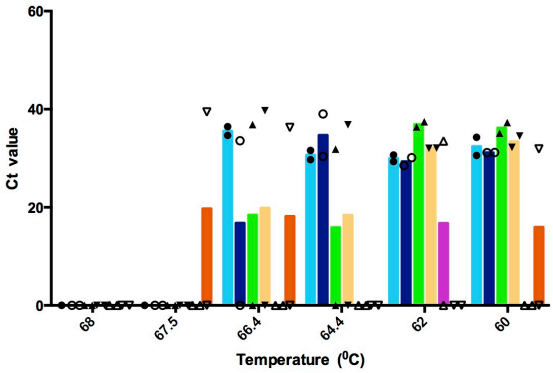
B)


Melt Curve
Temperature (°C)

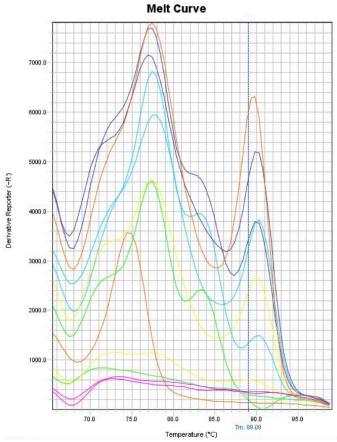
C)


Melt Curve
Temperature (°C)

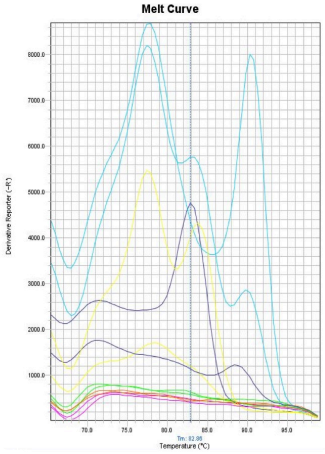
D)


Ct value
Temperature (⁰C)
68
67.5
66.4
64.4
62
60

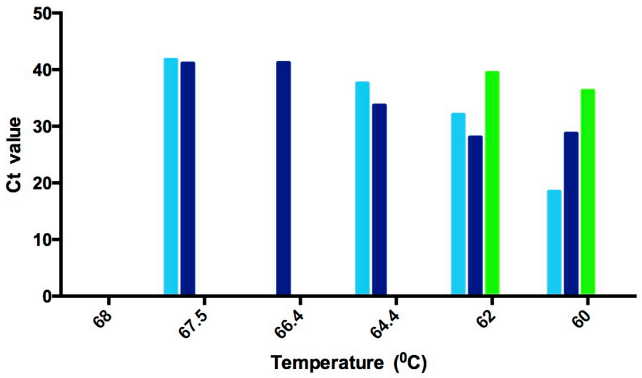
E)


Melt Curve
Temperature (°C)

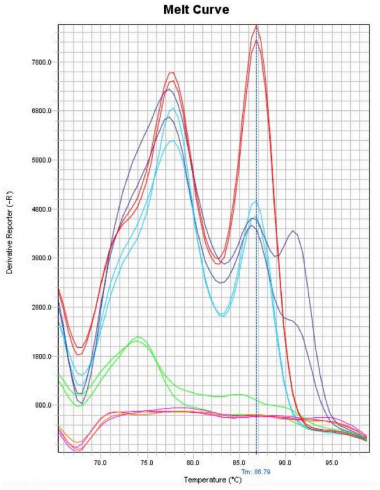
F)


Melt Curve
Temperature (°C)

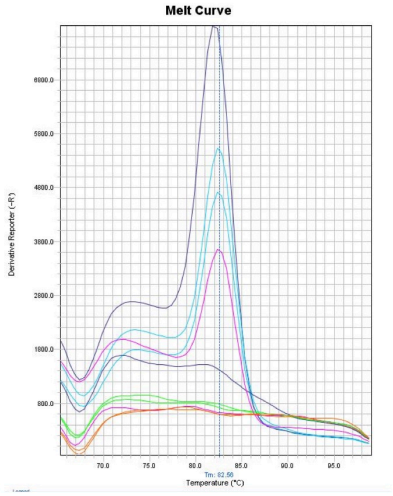
G)



H)



I)



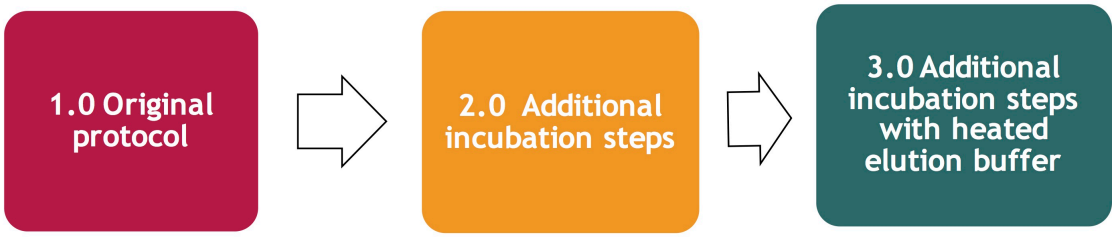
J)

Primer version	Temp (⁰C)	PP		PH		Gen		NF		TE		NTC	
A	68	0		0		0		0		0		0	
	67.5	0		0		0		0		0		0	
	66.4	1		2		2		0		0		0	
	64.4	2		3		2		0		0		0	
	62	2		3		2		0		0		0	
	60	2		3		2		0		0		0	
B	68	0	0	0	0	0	0	0	0	0	0	0	0
	67.5	0	0	0	0	0	0	0	0	0	0	0	0
	66.4	0	0	0	0	0	0	0	0	0	0	0	0
	64.4	2	1	1	0	1	0	0	1	0	0	0	0
	62	2	2	1	2	1	1	0	1	0	0	0	0
	60	0	1	0	1	0	1	0	1	0	0	0	0
C	68	0		0		0				0		1	
	67.5	0		0		0				0		0	
	66.4	0		0		0				0		1	
	64.4	1		2		0				0		1	
	62	1		1		2				0		1	
	60	3		2		0				0		1	

Figure 3.15 Temperature gradient quantitative PCR with *IFNAR1* primers
3C DNA libraries were loaded in duplicate on 96-well reaction plate. qPCR analysis of 3 versions of *IFNAR1* primers, Ct values shown as individual samples, and mean. A) Ct values of version A primers, N=1. B) Representative melt curve at 66.4°C. C) Representative melt curve at 62°C. D) Ct values of version B primers, N=2. E) Representative melt curve at 66.4°C. F) Representative melt curve at 64.4°C. G) Ct values from version C primers, N=1. H) Representative melt curve at 62°C. I) representative melt curve at 67.5°C. J) Summary of gel electrophoresis results from 3 primer versions. Numbers highlighted in red indicate one of the bands are at the expected size of 117bp for version A primers, 117bp for version B primers and 149bp for version C primers.
NF, non-fixtion, NTC, not-template control; PH, pooled healthy, PP, pooled patient; TE, Tris-EDTA

Overall, the qPCR protocol up to this point revealed some potentially useful primer candidates from successful amplification of the product of interest without parallel amplification of negative controls. Some annealing temperatures were emerging successful, but this could not be robustly replicated in multiple experiments. Despite the results not being conclusive, the gel products had to be purified before sending to Eurofins Genomics for sequencing (See section 2.2.11). This would be another measure to confirm amplification of the product of interest was successful, as the sequenced product could be checked against the known sequence (Appendix). The original gel purification protocol (as described in 2.2.9) was attempted but yielded little DNA, with concentrations ranging from 0.25ng/μl to 0.88ng/μl (Figure 3.16B). These concentrations were not sufficient for sequencing protocols, indicating optimisation of the gel purification process was needed. The first step to improve DNA yield was to add extra incubation steps at the buffer GQ stage and the elution stage of the protocol. This decreased DNA yield further to the lowest concentration of 0.106ng/μl (Figure 3.16B). The second optimisation step combined extra incubation stages (Section 2.2.9) with elution into heated elution buffer. This successfully increased yield to concentrations suitable for future sequencing. The minimum concentration from this attempt was 1.07 ng/μl with a maximum of 2.8 ng/μl (Figure 3.16B). All protocols had significantly different DNA concentrations, with protocol 1.0 and 3.0 also having significantly different yields. We proceeded with protocol 3.0 based on the yield of DNA produced.

A)



B)

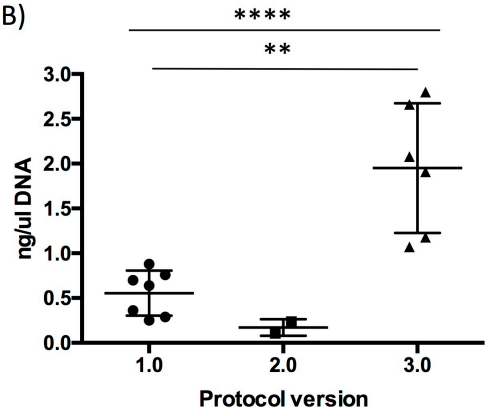
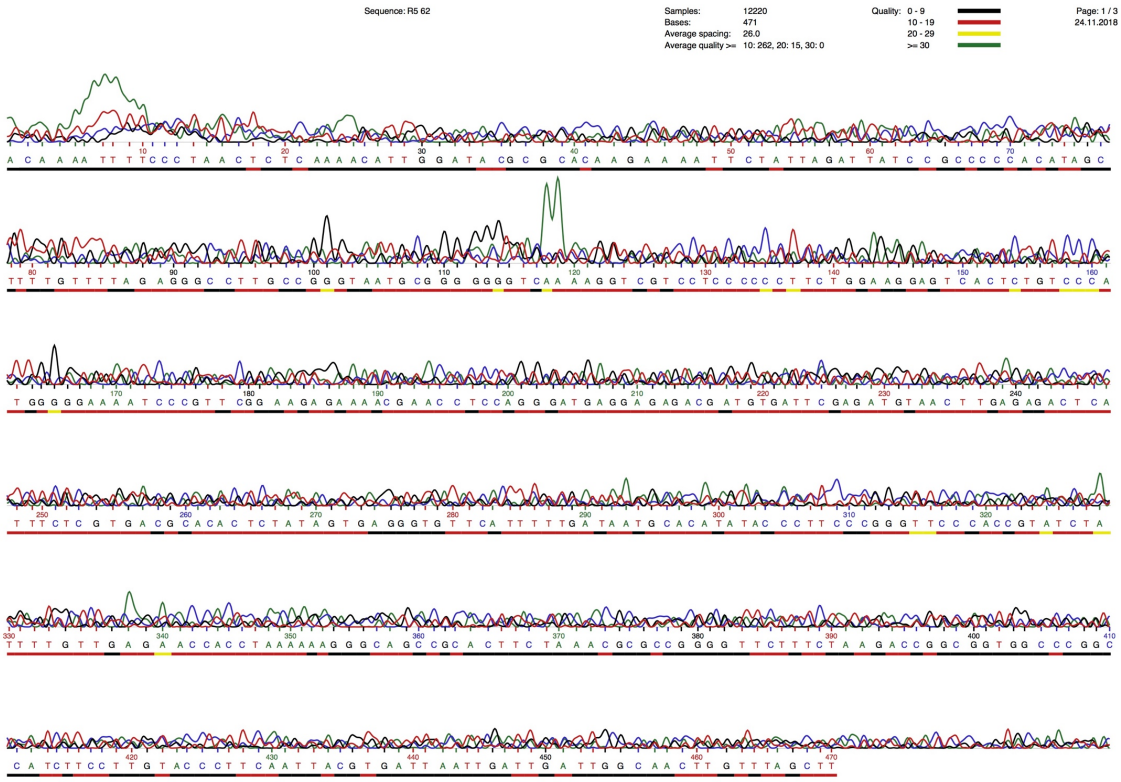


Figure 3.16 Optimisation of Gel Purification
Optimisation of gel purification to increase DNA yield. A) Schematic representing optimisation steps used. Protocol 1.0: running qPCR products on 1.5% agarose gel, gel fragments of interest excised and dissolved using buffer GQ from QIAquick gel purification kit. Incubation with buffer GQ for 10 mins at 50°C. Incubation with elution buffer (1xTE) for 1 minute at RT. Protocol 2.0: As protocol 1.0 with buffer GQ incubation for 15 minutes with shaking every 2 minutes, incubation with elution buffer for 3 minutes. Protocol 3.0: as protocol 2.0, with heated (37°C water bath) elution buffer. B) Concentration of DNA ng/μl using each protocol version, protocol 1.0 N=7, protocol 2.0 N=2, protocol 3.0 N=6. 2 tailed Mann Whitney test used to compare DNA yield between protocols, Kruskal-Wallis test comparing variation between 3 groups as a whole. ** P< 0.01, **** P <0.0001.

When the gel purification protocol was optimised to yield enough DNA for sequencing, the first sample to be sent was a product amplified by *IFNAR1* Version B primers at an annealing temperature of 62°C. This sequencing run proved unsuccessful. Images provided by Eurofins revealed the DNA was poor quality, as illustrated by the large proportion of black underneath each sequence row, representing DNA of 0-9% quality (Figure 3.17A). The alignment with the desired *IFNAR1* loop sequence was 6.91% (Figure 3.17B), although as the sequencing was such low quality, this could not be interpreted with any certainty. The known sequence, based on DNA digestion and re-ligation, primer design, predicted amplicon size (Figure 3.9 and Appendix) facilitated the ability to determine alignment. The decision was made to implement cloning of the gel

purified qPCR product as a way to improve the quality of the sequencing product (Figure 3.17C). Plasmids produced from cloning steps would be expected to be of high quality and ideal for sequencing. After the cloning of the qPCR product into the Topo 2.1 cloning vector, a restriction digest using the EcoR1 enzyme using a standard protocol (Section 2.2.12), was conducted to confirm presence of the qPCR product. The first attempt based on the original cloning protocol (section 2.2.12) did not yield the PCR product at expected size (Figure 3.17D). Version 2.0 of the protocol added an polyA tail to the qPCR product with the aim of increasing product stability (Section 2.2.13). This also did not yield a product at the expected size (Figure 3.17E). Version 3.0 of the protocol combined a polyA tail addition with an optimised ratio of reagents for ligation. This was hoped to increase the chances of the product being successfully inserted into the plasmid and creating a positive clone. After restriction digest, this version of the protocol yielded a product at a size around 300bp, which was in the range to be expected. This product, which came from *IL21R* at 64.4°C, was sent for sequencing. This product was chosen as results appeared more robust than the *IFNAR1* product used previously. Images from Eurofins were provided to illustrate sequencing quality. The quality of the sequencing was better than the first sample sent, highlighted by the large proportion of green, representing DNA of >30% quality. (Figure 3.17G). However, alignment with the desired sequence was only 6.96%.

A)



B)

Alignment of Sequence_1: [RA5 620C B sequence unclipped.xdna] with Sequence_2: [IFNAR1 sequence .xdna]

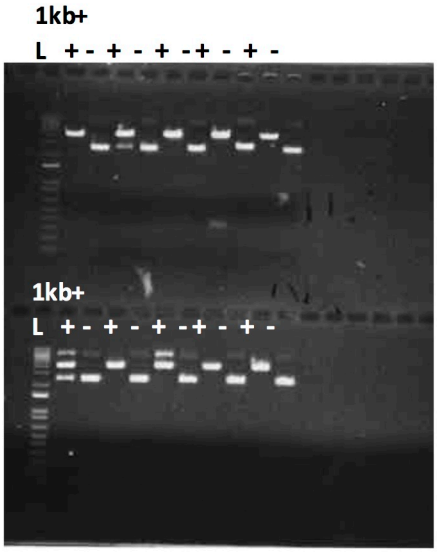
Similarity : 108/1562 (6.91 %)

```
Seq_1   1      ------------------------------------------------g-ttatctaa--g  10
                                                              |  ||
Seq_2   1141   agGgGtGcAgtGagGtttgtCgAgCgTGgtcCggGtatGtcAgccataTtCgACTGGgGt  1200

Seq_1   11     gA-TT-CTa-Cca--GtaTagatcCttactTtCacCtTTgtAtatcGatatctagtAgcc  65
               |  |  |   |    ||  ||    ||           | | |||    |  |
Seq_2   1201   GtttACCGgCcATgggTTCaGcatattgAcgatGGtgCgctgtttcgtcCCCcgGgGCGa  1260

Seq_1   66     cggctgCACcACccttAGgcCgActcccgcgCGgTaTtaCCaTagaCtTCatagagTttT  125
                  |     |  |  | |      ||     | ||     |     ||  |    ||||
Seq_2   1261   tCcCcaatAtAGTCAaactcgGGGCgAcggTTTTTTTTTTTTTTTTTTTTTTTTTTTTT  1320

Seq_1   126    TatgatctcgacacgtCaatCatTtcttagaagttacAgActctCagttgttAgtattgt  185
               | |  | |        |   |  ||  |          ||  |  ||    |  ||||
Seq_2   1321   TTTTTTTTTTgTTTTTTTTTTgttgatGTCCAgaCAGGaaAccGAggAATacTTaTtTa  1380

Seq_1   186    gttatctaCCTggTgTcCCATCgAtattctgCTGcaAagatcgcGGtaatgGcAcAAcga  245
                           |          |  | | ||  || ||       |  ||
Seq_2   1381   aAatCAGCAGcTcgAGttgCtAatCtGATccCGccCAtGaatgCTGtgGaCTGaGCacTT  1440

Seq_1   246    aCTTTTCACTaaacgaacttGGGTCcTCcCaCtttacTTgatgcgtgAaATtaTGaCaGG  305
                 ||  ||     |           |||      |  | |    | |  | |
Seq_2   1441   CtCattgcCTCgtAgtcAgactAcTgGGAGGcttgCgAaGTggGggATacTATTTATTTC  1500

Seq_1   306    gcAtCccGccgAcAtGGcGATaCgtAttatattCgcCgTcataAacatCttTttTtcCat  365
               |   | |  |       |  |  | |       |  |  |  |  |  |  ||  || | |
Seq_2   1501   gtgACGcCtCCTtCGagGcaaTcATTagcCTCttTctctAttctaggttgtTaGttgcct  1560

Seq_1   366    CatatatAAtataatat  382
                 ||
Seq_2   1561   ca---------------  1562
```

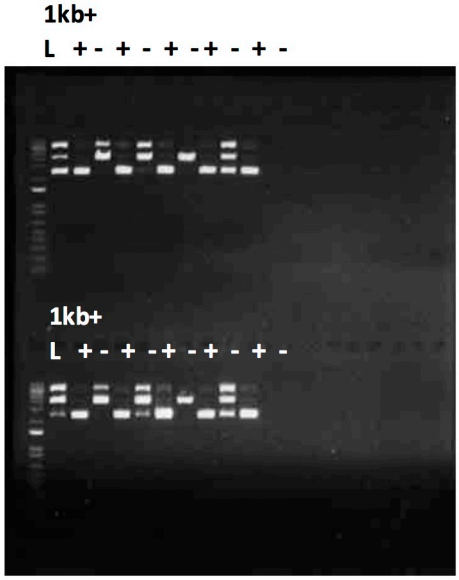
c)


1.0 Original cloning protocol
2.0 Addition of polyA tail
3.0 Addition of purified polyA tail Optimised ratio for ligation

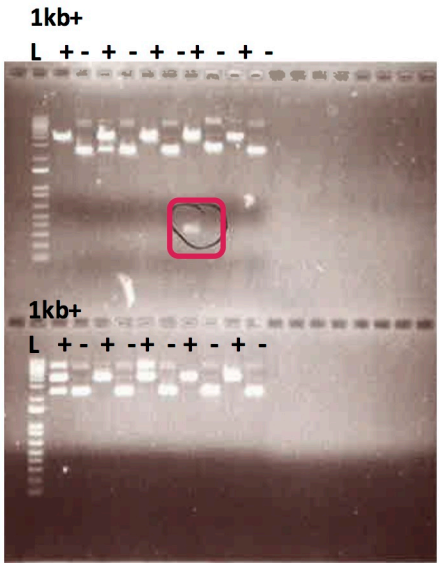
D)



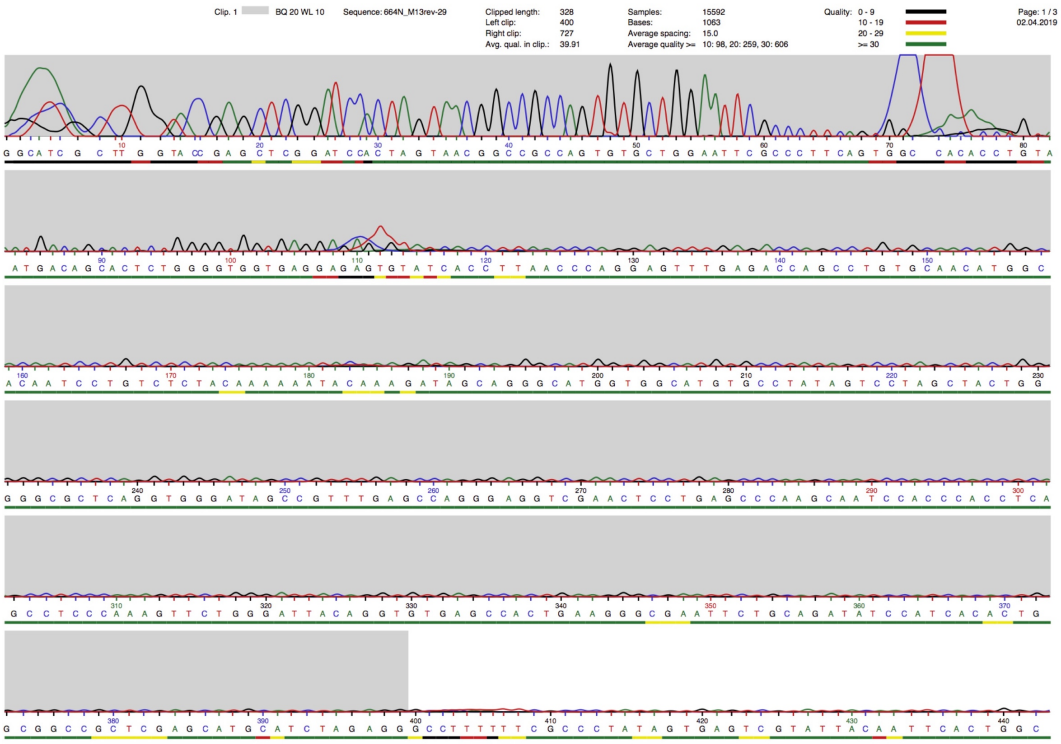
E)



F)



G)



H)

```
Similarity : 25/359 (6.96 %)

Seq_1   1      gttatctaCgggGgtaGTtcGTataTagatcCttactTtCacCtTTgtAtatcGatatct  60
Seq_2   1      ------------------------------------------------------------  0

Seq_1   61     agtAgcccggctgCACcACccttAGgcCgActcccgcgCGgTaTtaCCaTagaCtTCata  120
Seq_2   1      ------------------------------------------------------------  0

Seq_1   121    gagTttTTatTgatctcgacacgtCaatCatTtcttaTgaagttacAgActctCagttgt  180
Seq_2   1      ------------------------------------------------------------  0

Seq_1   181    tAgtattgtgttatctacC-gTgCctA-TcGTCCTAaCCCagaACttactcATGtCCAAc  238
                              | | |  |  |  ||  |  |       |
Seq_2   1      --------------C-AGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGCA  45

Seq_1   239    aagAAtAcctCtTcTCaTagtgCTggatGgTATgatGgGAcAGtTtcCtGaGCAGGATac  298
                  |  |  |   |||     |     |
Seq_2   46     GGCAGATCATGAGGTCAGGAGTTCGAGCCCTGGACCCCAGGCCAGCTAATGAGGCTGACT  105

Seq_1   299    GGagaGTCcAAGcCATtCctaAacatCttTttTtcCatCatatatAAtatacccacccta  358
               |||      |  |||
Seq_2   106    GGATGTCCAGAGCT----------------------------------------------  119

Seq_1   359    t  359

Seq_2   120    -  119
```

Figure 3.17 Optimisation of Sequencing of Quantitative PCR Products
Optimisation of qPCR product cloning to increase sequencing quality. Eurofins Tubeseq service was used for sequencing. A) Sequencing of product from *IFNAR1* primer version B, at annealing temperature of 62°C, image provided by Eurofins. B) Alignment of sequenced product with expected *IFNAR1* loop sequence, image provided by Eurofins. C) Schematic representing optimisation of the cloning method to generate plasmids containing loop products. Protocol 1.0: Topo 2.1 cloning vector used. Plasmid isolated using purelink miniprep kit. Restriction digest with EcoR1. Protocol 2.0: as protocol 1.0 with addition of PolyA tail to qPCR product before insertion to cloning plasmid. Protocol

3.0: as protocol 2.0 with ligation ratio optimised for 1.5ug DNA insert. D) Gel image of restriction digestion of cloned qPCR product using protocol 1.0. E) Gel image of restriction digestion of cloned qPCR product using protocol 2.0. F) Gel image of restriction digestion of cloned qPCR product using protocol 3.0. G) Sequencing of product from *IL21R* primer version C, at annealing temperature of 66.4°C, image provided by Eurofins. H) Alignment of *IL21R* sequenced product with *IL21R* expected loop sequence, image provided by Eurofins.

In summation, reviewing the collection of data from the qPCR optimisation (Figure 3.10 to Figure 3.17), I concluded that it was not possible for me to translate the nested PCR-based method previously used to evaluate our chromosomal loops of interest into a qPCR-based method. Whilst some primers showed some evidence that they could successfully amplify the loop of interest in the absence of negative control amplification at some temperatures, sequencing revealed they were not the expected loop product. After discussions with OBD, after they acquired additional propriety information about the translatability of the 5 markers to the qPCR platform, it was agreed that it was not appropriate to invest more time into the translation of these particular loops. Notably, since that point, the platform at OBD has been optimised and markers are now more easily translated from the nested to the qPCR platform.

Overall, based on minimal evidence that loops from the MTX signature could successfully be detected with qPCR following MIQE guidelines, the decision was taken to revert to nested PCR to amplify our 3C libraries. Based on this decision, an alternative method of gel visualisation was sourced. The decision to use the Lab Chip GX provided results quicker than a traditional gel electrophoresis protocol and results that were easily interpreted, highlighted by a clear band in combination with clear peaks at the expected size (Figure 3.18). These features allow high-throughput, more informative analysis than the original gel electrophoresis protocol, which is advantageous for a biomarker.

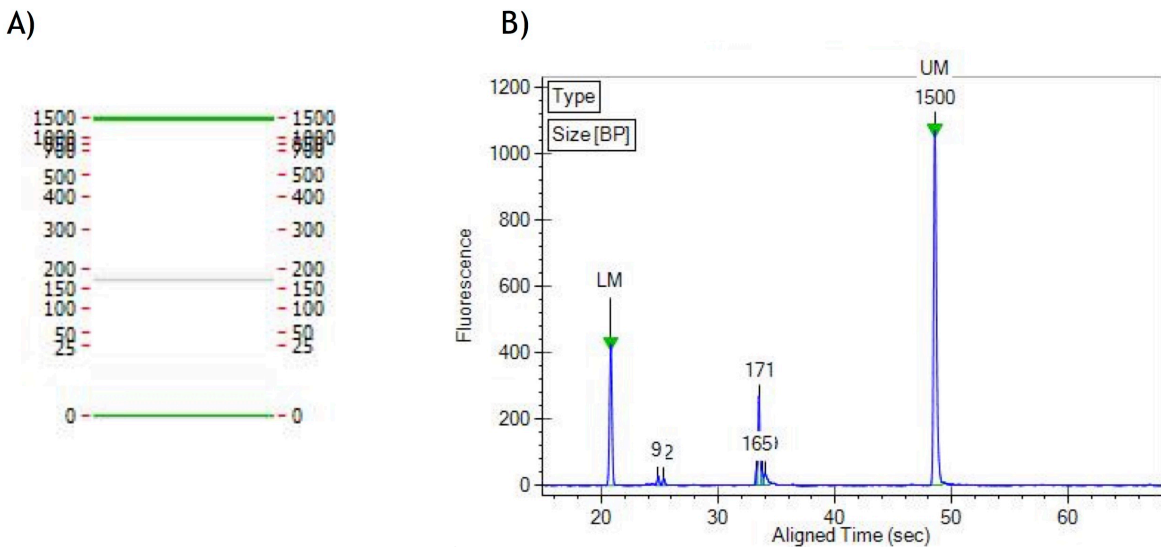


Figure 3.18 LabChip GX platform
Representative images from LabChipGX Touch imaging platform of *IL23* loci. 3C from SSZ treated RA patient. A) 1kb^{+} DNA ladder highlighting a band at 171bp. B) Peaks highlighting the lower (LM) and upper (UM) of the 1kb+ ladder and band at 171bp.

3.2.4 Understanding the stability of the MTX CCS

Having determined that it was not possible to translate the assay into a qPCR-based platform, I decided to go back to the nested PCR-based assay and evaluate the stability of the MTX CCS signature over time. To achieve this, baseline and 6 month samples from the original SERA cohort (R and NR groups used to discover the MTX signature) were used. It is important to note that patients in the R group at 6 months had reduced clinical disease activity index (CDAI) scores compared to baseline. All R had CDAI of less than 7.7, meaning all had low disease activity (LDA) or were in remission (Figure 3.19). NR had little or no improvement in CDAI scores by 6 months. All samples were investigated to find the presence or absence of each of the 5 loops of the MTX CCS at 6 months. Across all 5 loci, there were several samples that had a loop at both time points (labelled stable), a number that had no loop at baseline and one at 6 months, and another group that had a loop at baseline, but no loop at 6 months. These three categories of loop dynamics were present in both R and NR. The genomic locations with the most stable loops were *IL17A* NR, with 21 patients (72.4%) having loops present at both time points (Figure 3.19C). The least stable loop was *IL21R* from R, with 3 patients (10.7%) having loops at 6 months who had them at baseline (Figure 3.19E). 13.8% of NR had the CXCL13 loop still present at

6 months (Figure 3.19D). 35.7% and 46.4% of R had *IL23* (Figure 3.19F) and *IFNAR1* loops present at 6 months, respectively (Figure 3.19G).

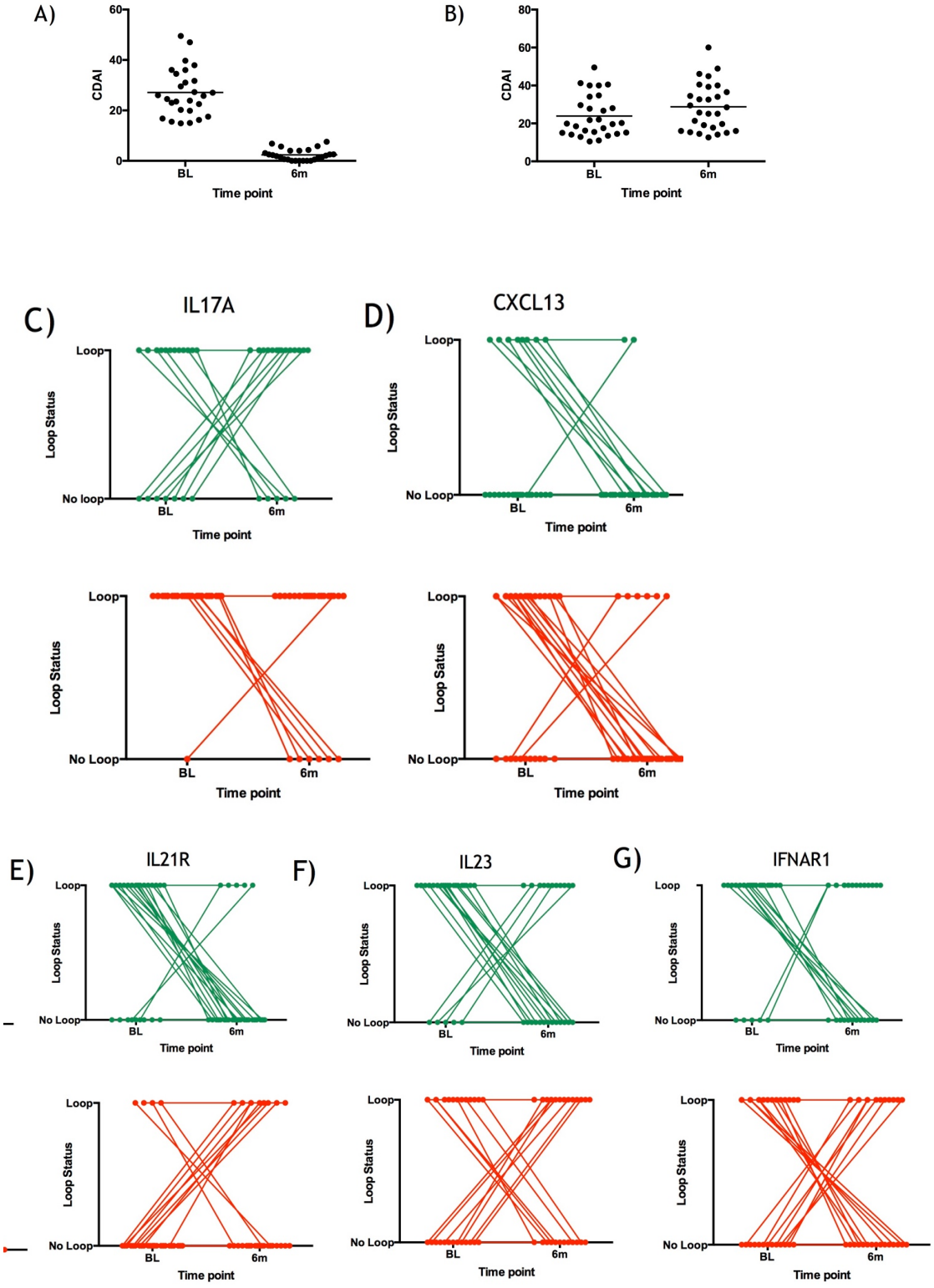


Figure 3.19 MTX CCS stability at 6 months
Analysis of presence of MTX CCS loops at 6 months in MTX R and NR from SERA cohort. A) CDAI of R at baseline (BL) and 6 months (6m), N=28. B) CDAI of NR at BL and 6m, N=29. C) Loop status of *IL17A* loop at BL and 6m. D) Loop status of *CXCL13* loop at BL and 6m. E) Loop status of *IL21R* loop at BL and 6m. F) Loop status of *IL23* loop at BL and 6m. G) Loop status of *IFNAR1* loop at BL and 6m. R shown in green and NR shown in red.

6m, 6 months; BL, baseline; CDAI, clinical disease activity index; m, month; MTX, methotrexate; NR, non-responder; R, responder; SERA, Scottish Early Rheumatoid Arthritis Cohort

Statistical tests were carried out on this data to gain further insight into the stability of the MTX CCS. The Boschloo independence tests were employed to measure the change in the signature at both time points for both groups, as well as measure the ability to stratify the two groups at each time point. This test was chosen based on the sample number available. The score of 0.48 for *IL17A* showed that is the only loop in R to not significantly change between baseline and 6 months, i.e., partly stable. In NR, *IL17A*, *CXCL13* and *IL21R* significantly changed between time points (Figure 3.20A). The score of 1 for *IL23* suggested the loop does not change between time points and suggests stability of that loop in NR. With a score of 0.69 in *IFNAR1*, the loop in NR could be considered partly stable. The test of how the CCS could stratify between R and NR to MTX at both time points confirmed, at baseline only, all loci of the signature could successfully stratify, with significant scores of 0.1 and below for all gene loops (Figure 3.20B). This was expected as it was based on the original MTX CCS study data and illustrates the 5 loci required for the signature. At 6 months, the score of 1 for *IL17A* suggests the loss of significant stratification ability to differentiate MTX R and NR for that loop. Other scores ranging from 0.19 to 0.76 implies limited stratification ability for the 4 remaining CCS loops at 6 months. Ultimately the data showed that the signature is not stable and would only be successful if used at baseline.

A)

Gene	R-BL vs R-6m	Gene	NR-BL vs NR-6m
IL17A	0.47982255	IL17A	0.07552019
CXCL13	0.03852677	CXCL13	0.003585556
IL21R	0.00036069	IL21R	0.184399036
IL23	0.02968988	IL23	1
IFNAR1	0.05777856	IFNAR1	0.688879761

B)

Gene	R-BL vs NR-BL	Gene	R-6m vs NR-6m
IL17A	0.00863913	IL17A	1
CXCL13	0.10479109	CXCL13	0.4088694
IL21R	0.00024335	IL21R	0.191031
IL23	0.0350002	IL23	0.7630735
IFNAR1	0.0350002	IFNAR1	0.5900533

Figure 3.20 Statistical analysis for identification of stable markers and markers with stable stratification ability
Analysis conducted by Ewan Hunter (OBD). Boschloo independence test carried out using binary data from MTX 6m stability analysis at baseline (BL) and 6 months (6m). Green for significant, red for insignificant and black for limited significance. A) Test analysing difference between time points, significant ≤0.4. B) Test analysing difference stratification ability of R and at BL and 6m, significant ≤ 0.1.
BL, baseline; CDAI, clinical disease activity index; m, month; NR, non-responder; R, responder

With the results indicating a largely instable CCS, an initial investigation was carried out using clinical scores to determine whether disease severity played a role in this variability between time points. As such, the correlation between CDAI (Figure 3.21) and DAS28 (Figure 3.22) scores with number of stable loops was examined. The results showed no clear correlation between chromatin stability and disease activity state. The numbers of stable loops are similar between R and NR. Furthermore, individual patient scores are widely distributed within each stable loop category, from as low as CDAI 10 to 49.5 (Figure 3.21C). This showed that loop architecture was not influenced by disease activity.

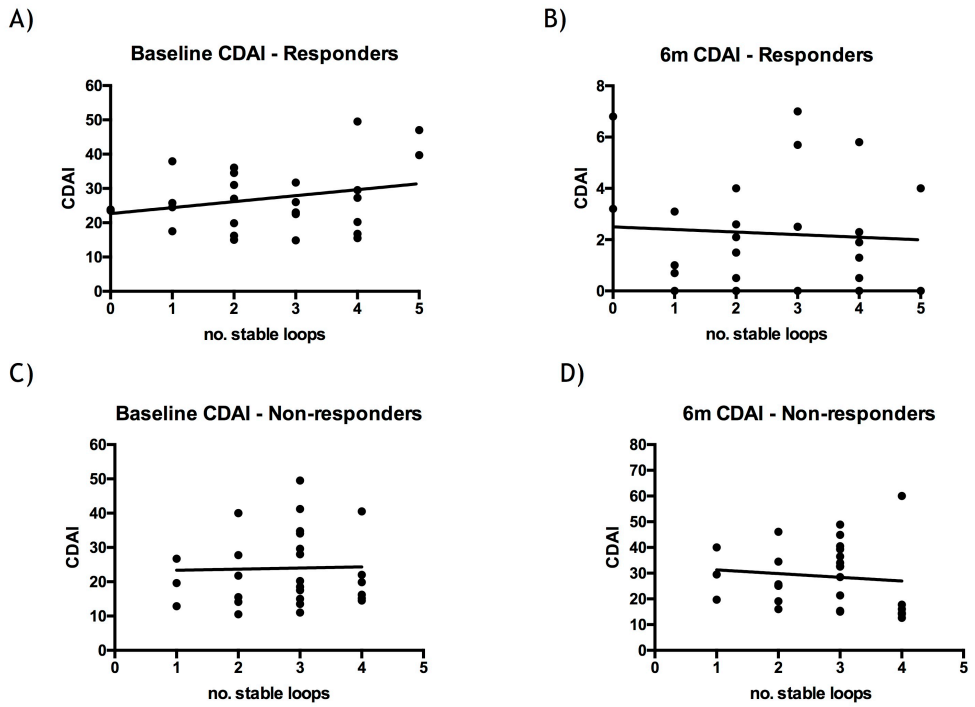


Figure 3.21 CDAI Correlation with Loop Stability
The CDAI scores for each patient sample at BL and 6m was plotted against the no of stable loops. A) baseline CDAI in R. B) 6m CDAI in R. C) baseline CDAI in NR. D) 6m CDAI in NR. The linear regression of the data was plotted. R, N=27, NR, N=28.
6m, 6 months; CDAI, clinical disease activity index

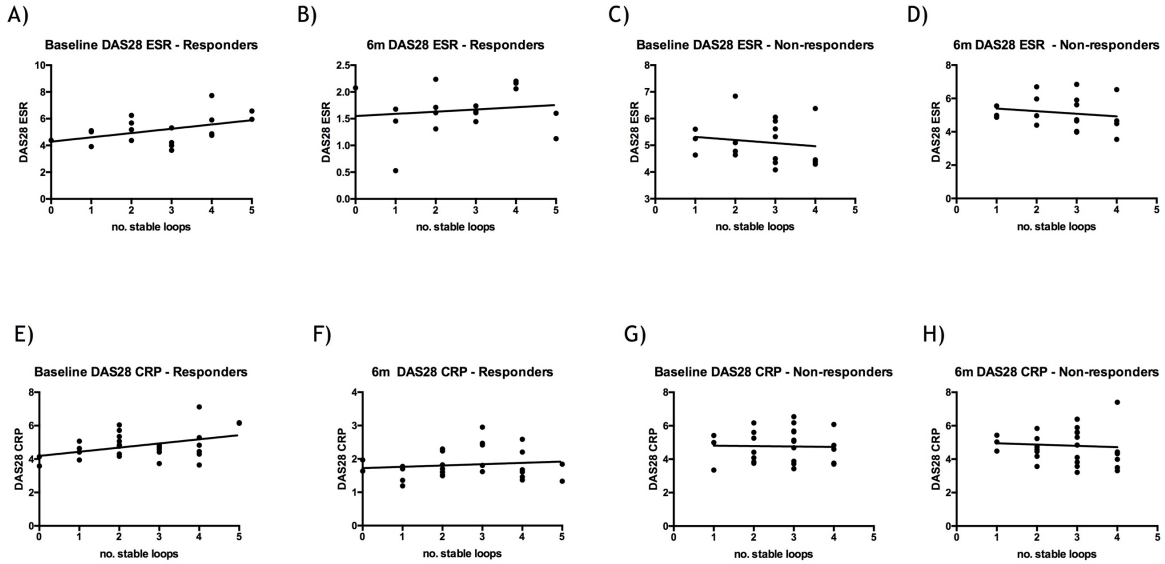


Figure 3.22 DAS28 correlation with loop stability
The DAS28 scores for each patient sample at BL and 6m was plotted against the no. of stable loops. A) Baseline DAS28 ESR in R, N=18. B) 6m DAS28 ESR in R N=17. C) Baseline DAS28 ESR in NR, N=18. D) 6m DAS28 ESR in NR, N=18. E) Baseline DAS28 CRP in R, N=27. F) 6m DAS28 CRP in R, N= 27. G) baseline DAS28 CRP in NR, N=27. H) 6m DAS28 CRP in NR, N=28. The linear regression of the data was plotted.
CDAI, clinical disease activity index; CRP, C-reactive protein;
DAS, disease activity score; ESR, erythrocyte sedimentation rate

3.2.5 Ability of the MTX CCS to predict response csDMARD treatment

While a biomarker for MTX is beneficial, there would be value in identifying if the MTX CCS biomarker could predict response to baseline treatment, regardless of which csDMARD or csDMARDs were given to a patient. We wanted to assess whether the MTX CCS 5-loop signature for MTX would be applicable for determining response to patients treated with a combination of csDMARD treatment.

To evaluate the suitability of this, the disease activity at baseline, 6 months and 12 months was assessed in the SERA cohort. Patients were assigned a R or NR status based on the disease activity scores at 6 months and 12 months (Figure 3.23). To differ from the cohort used previously, patients were identified that had been treated with HCQ or SSZ, which may have been in addition to MTX. 35 patients treated with HCQ or SSZ, with or without MTX, were selected that had strong R or NR status (Table 3.3). A strong status was defined as a R remaining in remission or low disease activity at 12-months, or a NR with a high disease activity score at that time point. Patient baseline demographics revealed a similar profile between groups (Table 3.4 and 3.5). The presence/absence of MTX CCS loops was assessed in patients from these groups.

Table 3.3 Chosen SERA Patient Treatment Assignment
csDMARDs received in first 12 months of treatment
HCQ, hydroxychloroquine; MTX, methotrexate; SSZ, sulphazine

Treatment	No of patients
HCQ monotherapy	5
SSZ monotherapy	10
HCQ and MTX	13
SSZ and MTX	7

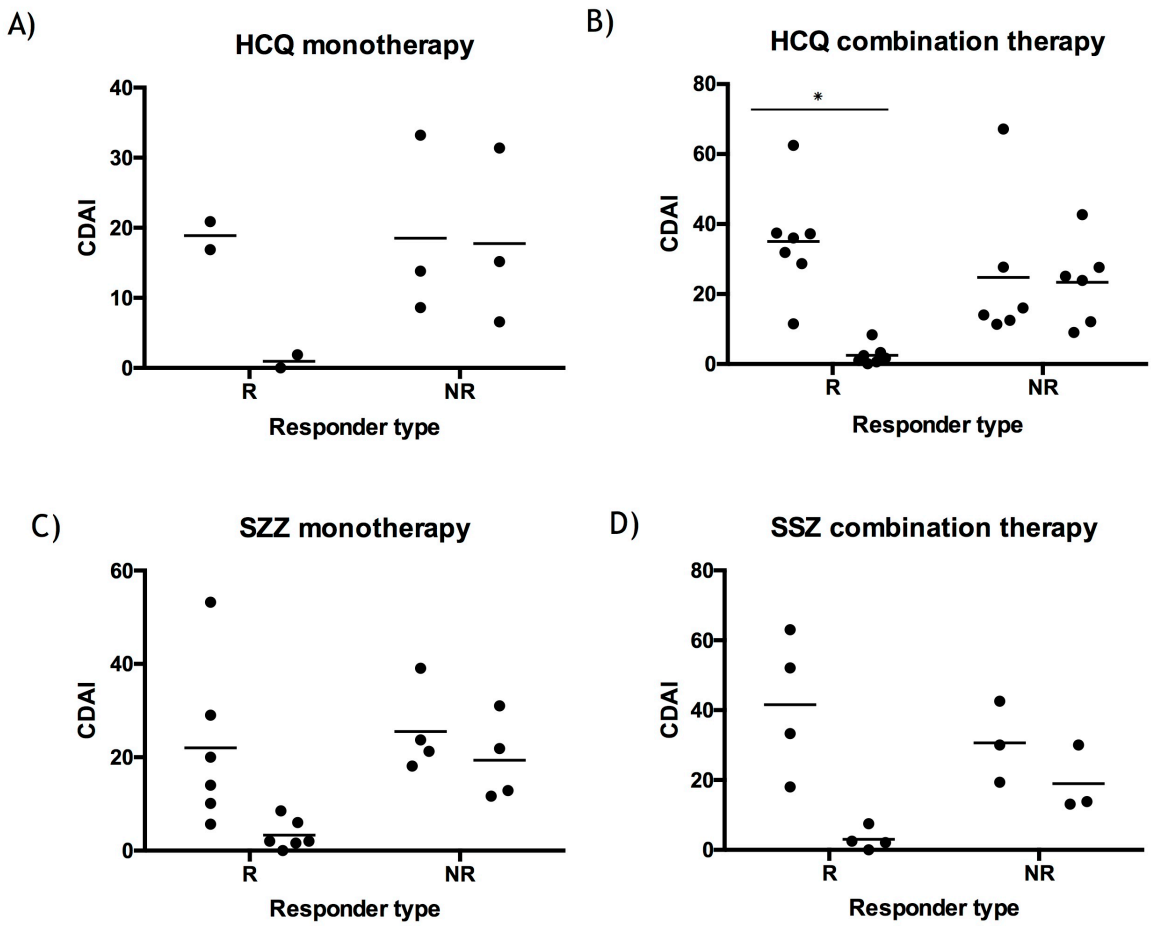


Figure 3.23 Disease activity in responders and non-responders to HCQ and SSZ
CDAI scores at baseline and 6 months (shown on left and right within each responder type, respectively) in SERA patients treated with HCQ or SSZ with or without MTX. CDAI represented as individual scores. A) R (N=9) and NR (N=12) to HCQ. B) R (N=11) and NR (N=8) to SSZ. T test to compare CDAI between time points, *P< 0.05.
CDAI, clinical disease activity index; HCQ, hydroxychloroquine; NR, non-responder; R, responder; SSZ, sulphasalzine

Table 3.4 HCQ +/- MTX-treated Patient Demographics

	Responders		Non-responders	
	Monotherapy	Combination therapy	Monotherapy	Combination therapy
Number	2	7	3	6
Age (years)	47 ± 5	70.1 ± 6.9	57.8 ± 3.1	55.7 ± 10.8
Sex no (%) male	2 ± 100	3 (42.9)	2 (50)	2 (33.3)
Caucasian no (%)	2 (100)	7 (100)	4 (100)	6 (100)
BMI kg/m^2	27.0 (7.8)	29.4 ± 3.9	27.4 ± 3.8	28.3 ± 4.4
Alcohol excess (%)	0 (0)	0 (0)	0 (0)	0 (0)
Current smoker no (%)	0 (0)	1 (14.3)	2 (50)	1 (16.6)
Previous smoker no (%)	0 (0)	2 (16.6)	1 (25)	1 (22.2)
RF (IU/ml)	NA	211.3 ± 473.3	112.3 ± 137.7	15.5 ± 4.5
CCP(U/ml)	225.5 ± 223.5	50.3 ± 86.6	51.7 ± 50.3	220 ± 150.0
CDAI	18.9 ± 2	35.0 ± 14.0	17.0 ± 9.5	24.8 ± 19.7
DAS28 CRP	4.55 ± 0.55	5.3 ± 1.2	3.9 ± 0.9	4.8 ± 1.3

Table 3.5 SSZ+/- MTX-Treated Patient Demographics

	Responders		Non-responders	
	Monotherapy	Combination therapy	Monotherapy	Combination therapy
Number	6	4	4	3
Age (years)	69.8 ± 19.9	67.8 ± 5.4	58.3 ± 6.18	60 ± 5.9
Sex no (%) male	3 (50)	3 (75)	4 (100)	3 (66.7)
Caucasian no (%)	6 (100)	4 (100)	4 (100)	3 (100)
BMI kg/m^2	27.4 ± 5.3	30.2 ± 4.7	48.5 ± 3.5	27.4 ± 4.9
Alcohol excess (%)	3 (50)	2 (50)	0 (0)	0 (0)
Current smoker no (%)	1 (16.7)	0 (0)	1 (25)	2 (66.7)
Previous smoker no (%)	3 (50)	2 (50)	1 (25)	1 (33.3)
RF (IU/ml)	357.7 ± 471.1	43.3 ± 40.3	234.9 ± 311.2	NA
CCP(U/ml)	374.3 ± 251.8	1.4 ± 0.43	1.5 ± 0.5	36.5 ± 51.2
CDAI	22 ± 15.8	41.6 ± 17.2	25.6 ± 8.1	30.7 ± 9.5
DAS28 CRP	4.4 ± 1.6	6.0 ± 1.3	4.6 ± 0.9	5.5 ± 0.8

3C analysis was used to examine the presence of loops belonging to the MTX signature in HCQ and SSZ (with or without MTX) treated patients at baseline. Patients had a variety of loop combinations, however there was no clear difference between R or NR in most treatment subgroups (Figure 3.24). Monotherapy HCQ was the only treatment subgroup that showed a differentiation in signature loci between R and NR (Figure 3.24A). However, this

was also the smallest treatment group, and not large enough for robust interpretation. Loops were only present at *IL17A*, *CXCL13* and *IL23* loci. There were no loops detected at *IL21R* and *IFNAR1* loci in any response group which meant these loci always clustered together within the heat map. The MTX CCS NR signature was observed 5 times (Figure 3.24B, C), but was only once observed in a patient with a clinical NR status (Figure 3.24C). The HCQ combination therapy group had most variation in loop presence of all subgroups, but no clear differentiation between R and NR could be observed. All subgroups were very low in number, therefore solid conclusions cannot be drawn from the data.

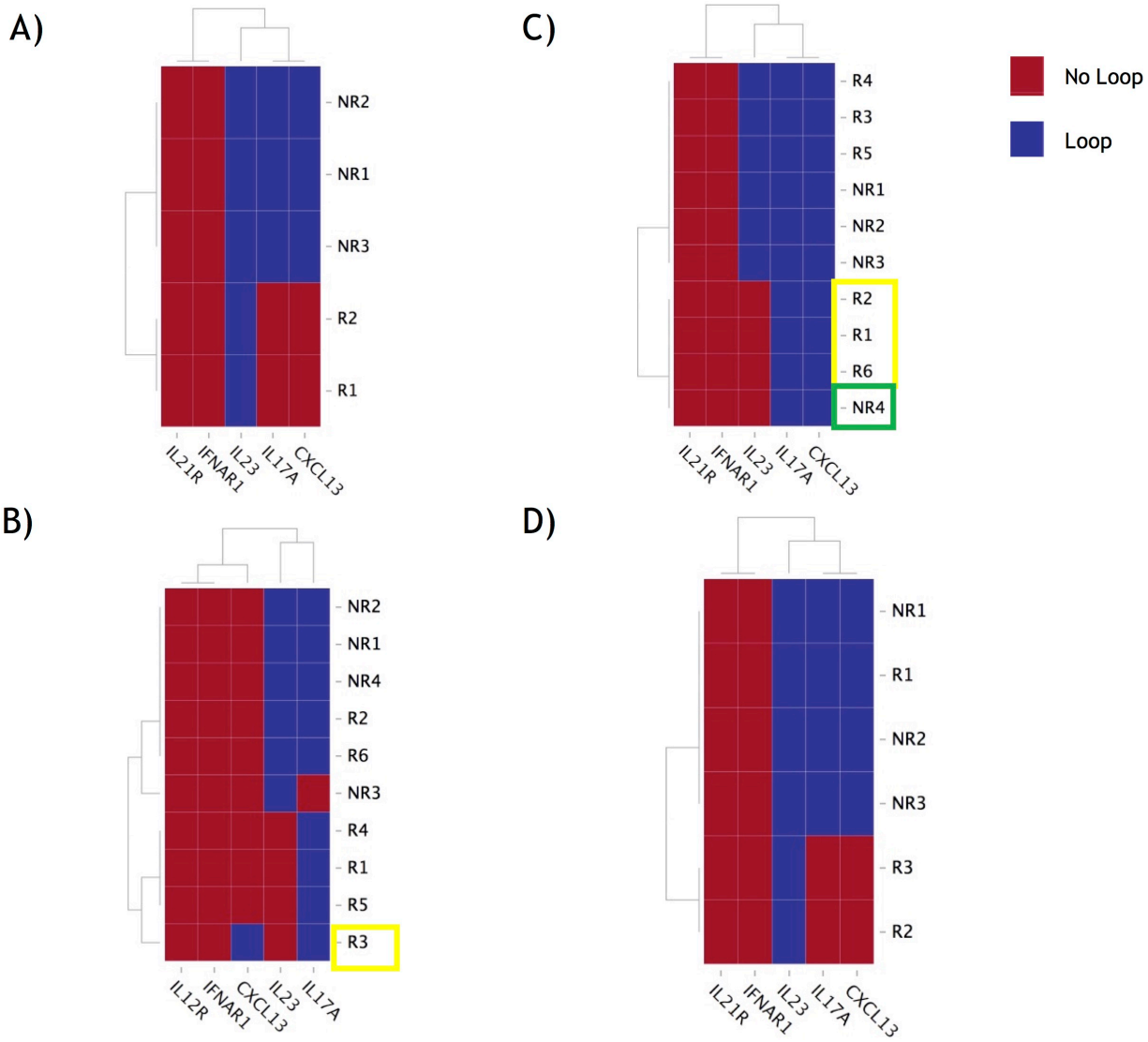


Figure 3.24 Heat map of MTX CCS loop presence in R and NR
3C carried out on baseline samples from patients treated with HCQ or SSZ, with or without MTX. Heat map indicating binary presence or absence of 5 MTX CCS loci loops. A) Patients treated with monotherapy HCQ. B) Patients treated with combination HCQ + MTX. C) Patients treated with monotherapy SSZ. D) Patients treated with combination SSZ +MTX. Blue = loop, Red = no loop. Yellow box indicates NR signature in a clinical response, green box indicates NR signature in clinical NR
NR, non-responder; R, responder

Based on the MTX signature, HCQ and SSZ +/- MTX treated patients were assigned a predicted response. The correct response status was only assigned in 1 patient (Figure 3.25B). This patient received monotherapy. Several patients, also monotherapy SSZ-treated, had a NR signature present, but were clinically responders to therapy. The remaining patients had a combination of loops out-with the MTX signature conformation and therefore their response could not be predicted. These values were used to employ ROC analysis to determine specificity and sensitivity of the signature in this group. The ROC curve for HCQ monotherapy treated patients had a sensitivity of 50% and a specificity of 33.3% (Figure 3.26A) and combination treated ROC curve had a sensitivity of 33.3% and a specificity of 50%. ROC analysis of monotherapy SSZ treated patients revealed a sensitivity of 0% and a specificity of 25% (Figure 3.26D). Combination treated SSZ patients had sensitivity of 0% and specificity of 100%. The highest accuracy value was 50%, which confirms that there is no predictive potential for the MTX CCS for response to csDMARD treatment.

A)

	Predicted response		
Observed response	Responder	Non-responder	Undefined
Responder	0	0	2
Non-responder	0	0	3

B)

	Predicted response		
Observed response	Responder	Non-responder	Undefined
Responder	0	1	5
Non-responder	0	0	4

E)

	Predicted response		
Observed response	Responder	Non-responder	Undefined
Responder	0	3	3
Non-responder	0	1	3

F)

	Predicted response		
Observed response	Responder	Non-responder	Undefined
Responder	0	0	3
Non-responder	0	0	3

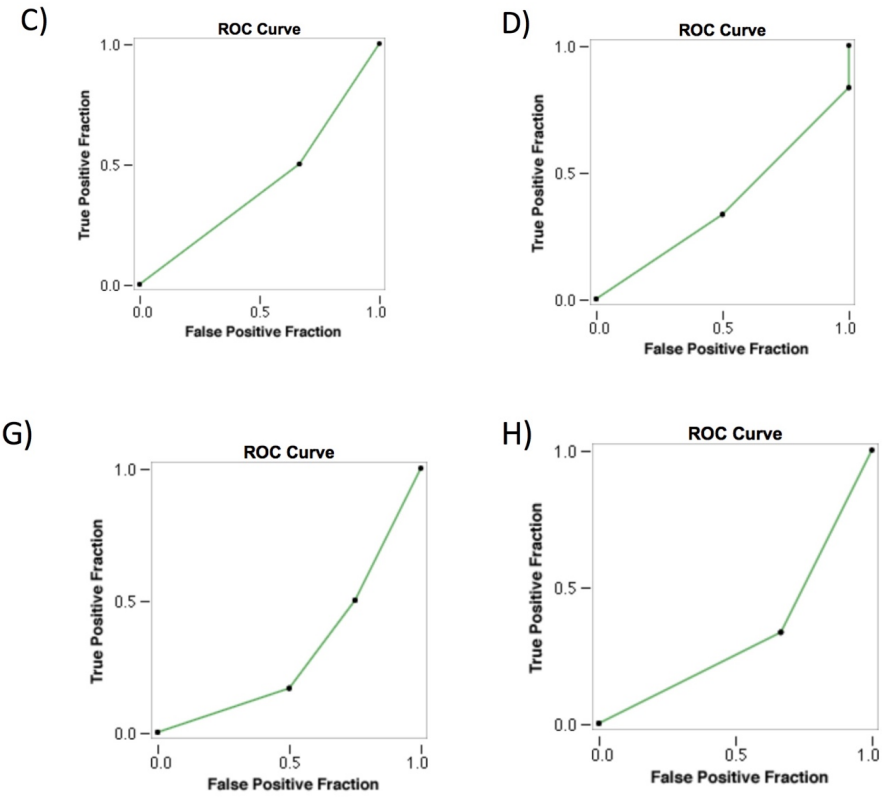


Figure 3.25 Binary classifier and ROC curve analysis of MTX signature for csDMARD response prediction
Assessment of the predictive ability of the MTX CCS to stratify HCQ and SSZ +/- MTX treated patients. Classifier of observed and predicted response status for A) Monotherapy HCQ and B) SSZ + MTX. C) Monotherapy SSZ. D) SSZ + MTX. E) ROC curve for monotherapy HCQ. F) ROC curve for HCQ + SSZ. G) ROC curve for monotherapy SSZ. H) ROX curve for SSZ and MTX. ROC curves generated using web-based calculator which utilises JROCFIT program.

To understand why the CCS could successfully predict response in only 1 patient, and other patients had a mostly undefined signature, demographic and disease activity was measured in each category (Figure 3.26). The patients were split into 4 categories: 1) responders who were predicted to be non-responders (R-NR, 2) non-responder predicted to be non-responder (NR-NR), 3) responders who could not be given a prediction response (R-un-defined (UD) and 4) non-responders that could not be given a prediction (NR-UD). The NR that had the NR

signature had an age within the range of ages in the other 3 categories of patients. Unlike the other 3 categories, the successfully predicted NR was neither a current smoker, or had an ex-smoker status. In the other categories, there was at least 1 current smoker and ex-smoker. The alcohol intake in the NR was within the range of the other categories. When considering the baseline disease activity of the NR, it was within the CDAI and DAS28 CRP upper range with the other categories. The CDAI and DAS28 CRP of the NR were 39.1 ad 5.7 respectively. The other interesting group to consider was the 4 R who had the NR signature. While the prediction was wrong, the MTX CCS was present. This R group had the largest age range from a minimum of 28 to a maximum of 88. The number of current smokers was the same as the R-UD group and ex-smoker number half of the UD groups. Alcohol intake, CDAI and DAS28 CRP scores were in similar range with the UD groups. As already discussed, the sample number for this part of the stud were very low and higher numbers in every subgroup would be needed to draw the right conclusions.

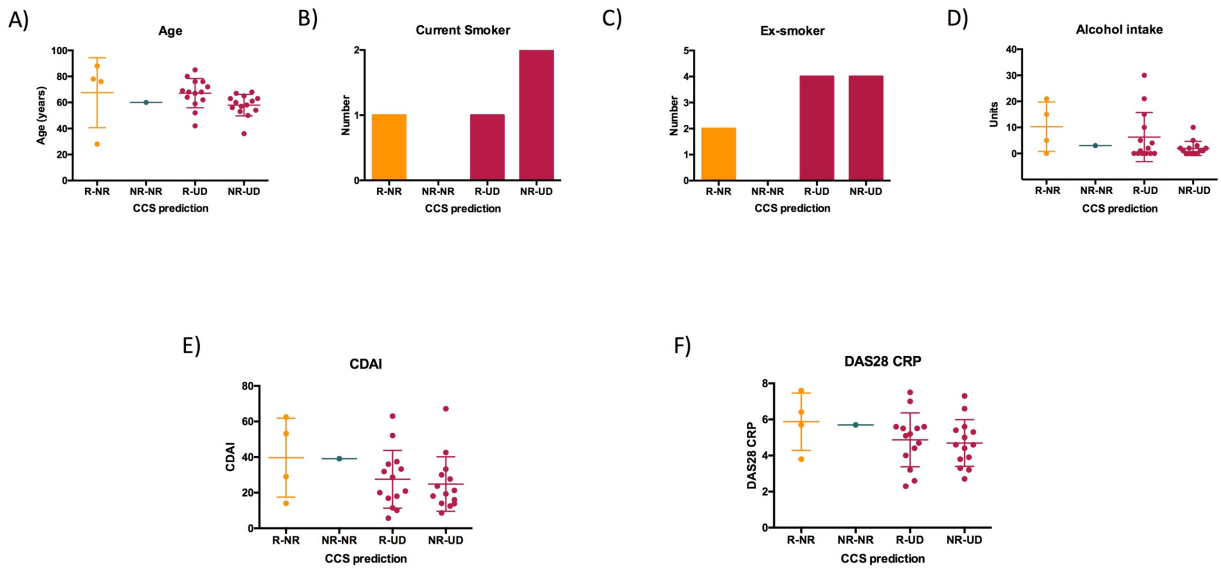


Figure 3.26 Demography and disease activity relationship with csDMARDs response prediction
BL demographic factors and disease activity was measured in all HCQ/SSZ +/- MTX-treated patients. A) Age. B) Number of patients who currently smoke. C) Number of patients who smoked previously. D) Units of alcohol consumed weekly. E) CDAI at BL. F) DAS28CRP at BL.
CDAI, clinical disease activity index; CRP, C-reactive protein; NR, non-responder; R, responder; UD, undefined

As well as exploring the relationship between demographic factors and the ability of the MTX CCS to predict response correctly, it was important to

understand if mono or combination therapy had an influence. Most patients with the presence of the MTX NR signature were SSZ monotherapy treated. There was 1 patient in this group that was treated with a combination of HCQ and MTX (Figure 2.27). The R and NR group that had an UD signature had a combination of the 4 treatment options of monotherapy HCQ or SSZ, or combination therapy with MTX. 35.7% of R with the UD signature had most patients treated with the HCQ and MTX combination. 26.6% had combination SSZ and MTX, followed by 21.4% with monotherapy SSZ and 14.3% with HCQ. 31% of NR with UD signature were monotherapy HCQ treated. The remaining 3 combinations made up 23% each of this group. R with UD signature had over 60% using combination therapy, however the R with NR signature have 75% on monotherapy. Overall, most patients that had loci in a conformation from the MTX CCS were monotherapy treated. This could be expected as the signature was developed in a monotherapy cohort. However, numbers are too low to draw a solid conclusion.

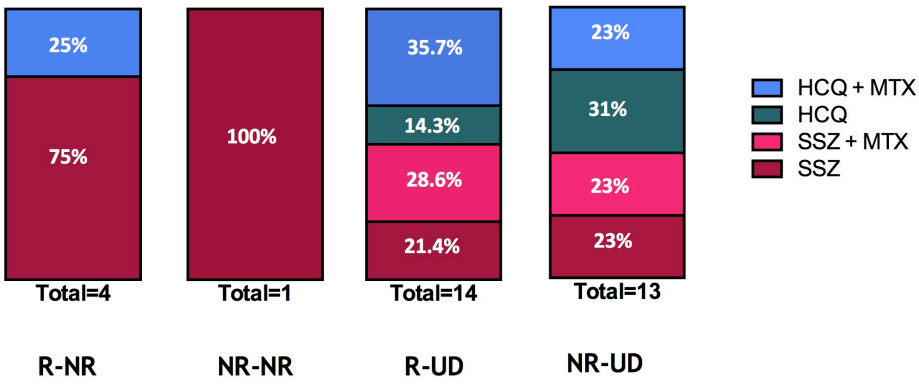


Figure 3.27 Relationship of mono and combination therapy with csDMARD response prediction
BL treatment was recorded and split into 4 groups: HCQ monotherapy, SSZ monotherapy, HCQ+MTX combination therapy and SSZ+MTX combination therapy. R predicted to be NR = R-NR, NR correctly predicted as NR=NR-NR, R without prediction= R-UD, NR without prediction = NR-UD
HCQ, hydroxychloroquine; MTX, methotrexate; NR, non-responder; R, responder; SSZ, sulphsalazine; UD, undefined

3.2.6 Validation of MTX CCS in new clinical cohort

With a robust PCR protocol established for assessing the 3C epigenome in RA patients, the decision was made to validate the MTX CCS in another RA cohort. A group of early RA samples from the Towards A Cure for Early Rheumatoid Arthritis (TACERA) cohort were selected based on the assignment of

monotherapy MTX at baseline. The TACERA cohort replicates the SERA cohort based on collection of samples from early RA patients who were treatment naïve at baseline. Peripheral blood samples were retrieved from these patients. At selection for our study, patient characteristic information was not retrieved as it was to be conducted blinded. Nested PCR was carried out at OBD and statistical analysis was used to understand if the MTX CCS could predict MTX response in an independent clinical cohort. 2 models were used to assess the predictive ability and plots can be used to visualise the loops in the CCS with the best predictive potential (Figure 3.28). These models have been developed since the discovery of the MTX CCS and are sophisticated, boosted machine learning models. Compared to the Weka model used to develop the original MTX CCS, the newer models are better at classification. The other advantage is the ability to see use these models directly in R Studio, as opposed to using externally with the Weka model. Weka could be used within R, but functionality is limited. Dilutions of primers to detect each loop were also used to help assess the influence of each loop on the model. The plot for one of the models, named XGBoost, shows SHapley Additive exPlanations (SHAP) scores which best represent results from tree-based algorithms. The plot shown highlights that *IL17A* provided the most stratification potential to the model with a score of 0.705 (Figure 3.28). Many patient samples with a low abundance of this loop had good prediction of a producing a negative in the model, and another group of patient samples with a high abundance of this loop contributed to a positive predictive score. Overall, this contributes to the highest ranking of this loop in the model. *IL21R* did not add any value to the model with all dilutions of the sample producing a score of zero. The plot indicates that most dilutions over 2-fold did not add anything to the model with scores of zero.

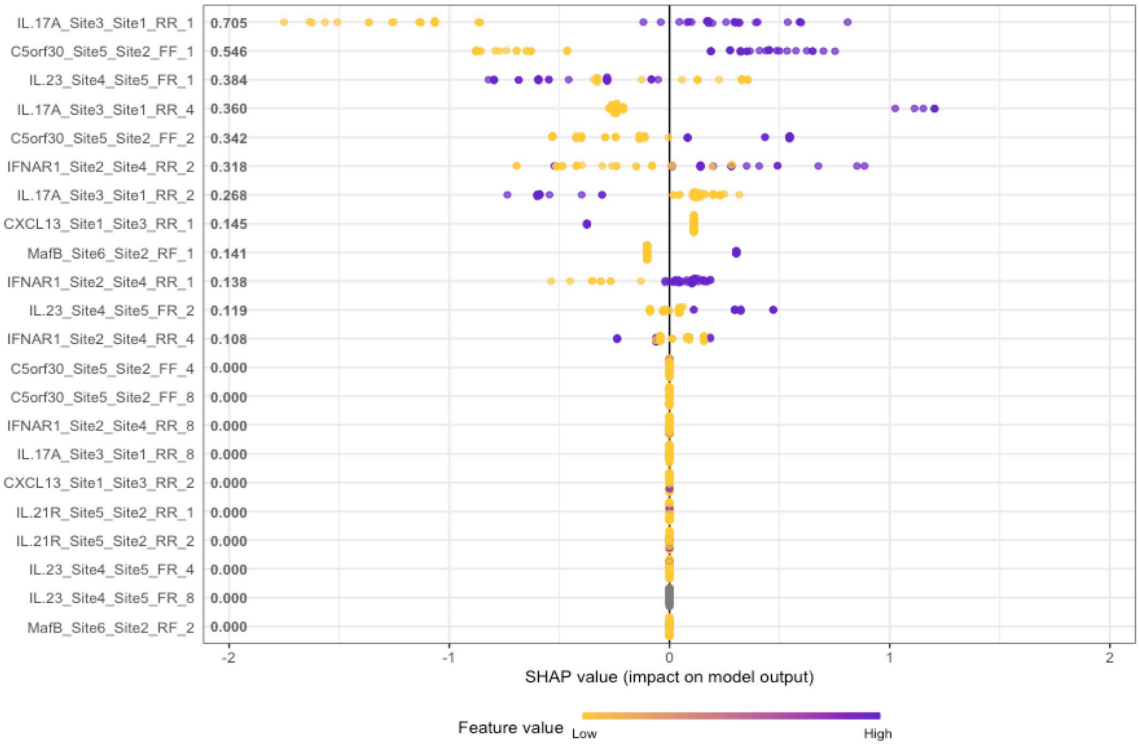


Figure 3.28 Identification of MTX CCS Loops that add value to Model in New RA Cohort
Data from XGBoost training model to plot contribution of each MTX CCS to new prediction model using the SHapley Additive exPlanations (SHAP) values. Each dot represents patient sample. Positive and negative values on x-axis indicate whether the loop is associated with a positive or negative prediction. Colour indicates whether loop abundance is high or low. Loops are ranked from top to bottom (from best to worst), based on the contribution to model. EpiSwitch™ loop detecting primers diluted used neat, or diluted 2, 4 or 8 fold; '_1', neat; '_2', 2-fold; '_4', 4 fold; '_8' 8 fold.

2 training models were built using 2 methods, XGBoost and Light GBM (Figure 3.29). For the purposes of prediction, these patients had to be unblinded. The first model, XGBoost, identified 20 R successfully and 15 NR successfully (Figure 3.29A). 9 NR were incorrectly identified as R, and 3 R wrongly assigned a NR status. With the Light GBM model, 12 R were correctly identified and 20 NR successfully identified (Figure 3.29B). 11 NR were predicted to be a R and 6 R predicted to be NR. The XGBoost model revealed a sensitivity of 0.833 and specificity of 0.69 (Figure 3.29C). The Light GBM model revealed a sensitivity of 0.62 and specificity of 0.67 (Figure 3.29D). The XGboost model had 74.5% [95% CI (0.59, 0.86)] accuracy and the light GBM an accuracy of 63.8% [95% CI (0.49, 0.77)].

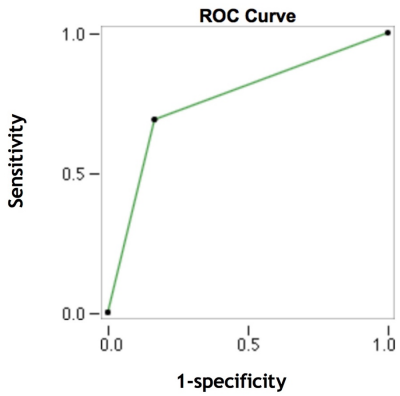
A)

	Predicted response	
Observed response	Responder	Non-responder
Responder	20	3
Non-responder	9	15

B)

	Predicted response	
Observed response	Responder	Non-responder
Responder	12	6
Non-responder	11	18

C)



D)

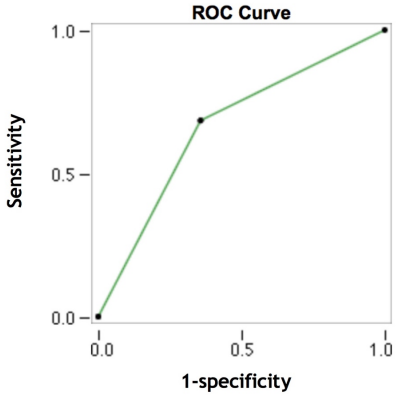


Figure 3.29 Binary classifier and ROC curve analysis of Training Models of MTX CCS for MTX Response prediction
Assessment of the predictive ability of the training models on 47 RA patients. Classifier of observed and predicted response status for A) XGBoost model and B) Light GBM model C) ROC curve for XGBoost. D) ROC curve for LightGBM. ROC curves generated using web-based calculator which utilises JROCFIT program.

Once the training sets had been generated, they were tested on 23 blinded patients (Figure 3.30). The XGBoost identified 6 R successfully and 8 NR successfully (Figure 3.30A). 5 NR were incorrectly identified as R, and 4 R wrongly assigned a NR status. With the Light GBM model, 7 R were correctly identified and 8 NR successfully identified (Figure 3.30B). 4 NR were predicted to be R and 4 R predicted to be NR. The XGBoost model revealed a sensitivity of 0.62 and specificity of 0.6 (Figure 3.30C). The Light GBM model revealed a sensitivity of 0.67 and specificity of 0.64 (Figure 3.30D). The XGBoost model had 65.2% [95% CI (0.39, 0.8)] accuracy and the Light GBM an accuracy of 63.8% [95% CI (0.49, 0.77)]. Overall, I can't conclusively say that the signature was successfully validated in another cohort. There is some evidence that some loci hold their predictive capacity, but ultimately, the signature was not validated. Based on the recognition of the models used and their predictive power, evidence suggests that the signature should be improved, as the sophistication of the models used would have likely extracted the results of stratification if it was possible in this group.

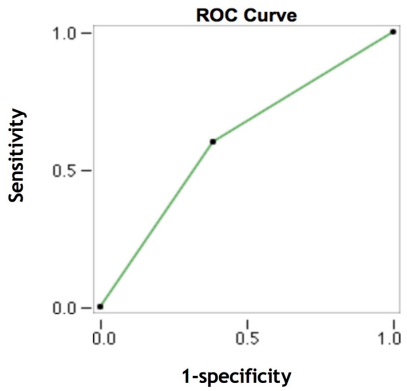
A)

	Predicted response	
Observed response	Responder	Non-responder
Responder	6	4
Non-responder	5	8

B)

	Predicted response	
Observed response	Responder	Non-responder
Responder	7	4
Non-responder	4	8

C)



D)

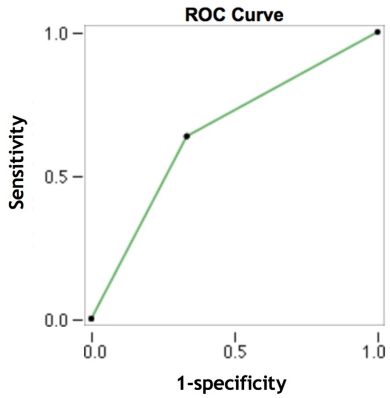


Figure 3.30 Binary classifier and ROC curve analysis of testing models of MTX CCS for MTX Response Prediction

Assessment of the predictive ability of the testing models on 23 blinded RA patients. Classifier of observed and predicted response status for A) XGBoost model and B) Light GBM model C) ROC curve for XGBoost. D) ROC curve for LightGBM. ROC curves generated using web-based calculator which utilisies JROCFIT program.

3.3 Discussion

The discovery that chromosome conformation signatures had the ability to successfully predict MTX response in treatment naive RA patients at baseline was a promising finding, but validation was needed. The work in this chapter explores the validation of this biomarker using bioinformatics techniques and experimentally using an independent clinical cohort. It also explores the transition of the 3C technique from OBD to the lab in house and the attempt to transition to alternative protocols. Further measurements to define the stratification potential by measuring stability of the MTX CCS at 6 months, and ability to predict response to treatment, regardless of csDMARD(s) assigned at baseline was also explored.

The first step in the *in-silico* validation process was to assess the quality of the data. It is recognised that pre-processing and appropriate normalisation steps are crucial to produce robust results that lead to reliable biological interpretation[273]. Limma was chosen as the software to validate the results due to its features to analyse data quality and normalisation, and for its linear modelling potential. Multiple studies have demonstrated that dye bias exists in dual colour microarray experiments, so this was one of the first steps to assess quality in our dataset[274]. This dye bias can contribute to the inaccurate interpretation of the expression intensities of the material being measured. There is debate in the literature about the methods that should be used to combat dye bias. Some researchers believe that dye swaps should be used, however, others believe that it is not essential, and the latter was the approach in this study[274,275]. The red-green dye densities shown did not highlight any arrays with a dye bias and the distribution of both colours was comparable (Figure 3.1A). MA plots were also used to visualise the red-green intensity log ratio (M) and mean signal intensities of each of spot on the array (A). With our dataset, it was assumed that most loops captured on the array were unlikely to change, meaning the plots should have most spots centred around the middle 0 line. The results in Figure 3.1C show a curved line, highlighting that pre-normalisation, there were more spots than expected with a positive or negative fold change representing loop abundance. There was also substantial variability at the low intensities, shown by the V shape on the left side to the middle of the

plot. This shape of plot has been shown in the literature[273] and suggests a low intensity bias for large fold changes. Despite the indication that there may be some bias at certain intensities for large fold changes, this analysis highlights that the most loops do not change between samples, but that there were enough to warrant further investigation.

After this first visualisation step of the array parameters, normalisation of the data was required. Loess normalisation is considered an appropriate normalisation approach for datasets where most measured elements are unlikely to change. Loess normalisation successfully reduced variability between the dyes in each array, indicated by more uniform histogram distributions (Figure 3.1B). In the MA plots (Figure 3.1D), it is evident that Loess normalisation reduces the curvature of the middle of the plot, highlighting the reduction in the number of loops with differential fold change. The last QC assessment shown is the boxplot (Figure 3.1E). This represents normalisation of signal intensities across 8 the arrays, and similar to other plots, indicated normalisation was successful and there were no problematic arrays that should be removed from the analysis.

The second stage of the validation process was to confirm the stratification potential of the MTX CCS loci. Analysis confirmed the potential for stratification using the MTX CCS loops. Contrast models in Limma, contrasting NR and R revealed the association of *IL17A* and *CXCL13* with NR and *IL21R*, *IL23* and INFAR1 with R, as is present in the signature (Figure 3.2A). Next, the 5 loci had to be tested on a set of patient samples. Using a Random Forest classification method, the model revealed an accuracy of 87%, shown visually with a ROC curve, which can be considered in alignment with the original model. Overall these tests indicated that the signature could be validated and there was a basis to explore this signature further throughout this study.

The first stage in establishing that I could carry out the 3C independently was to obtain samples to work with. These samples were generated from healthy BC donors. A purity check was carried out to ensure successful isolation of the intended cell type (Figure 3.3). These checks illustrated that the isolation was effective. The population of $CD4^+$ T cells were within the expected range of 20-60% of the total PBMC population[276].

The generation of the first 3C libraries using these cells was successful, as demonstrated by the Quant-iT PicoGreen method of DNA quantification (Figure 3.4). After measuring the first samples using the PicoGreen assay, it was considered that there may be a higher-throughput method of quantification for the 3C libraries, which would be more suited to the number of samples in our study. The Qubit platform has been shown to be a reliable method of choice to calculate DNA concentration when carrying out next-generation sequencing (NGS) or microarray methods[277]. Compared to other DNA quantification methods, the Qubit is rapid, precise and sensitive[278]. As such, this was the alternative method to PicoGreen that was chosen. A direct comparison was carried out between PicoGreen and the Qubit dsDNA HS assay (Figure 3.5). The results revealed a statistically significant difference in the yield of DNA measured. This experiment was repeated and there continued to be a disparity in results between the two quantification methods, with the Qubit consistently calculating lower concentrations. Exploration of these methods has also been shown in the literature[279]. Like the results described, the Qubit measured lower DNA yields than an alternative method, specifically, the Taqman RNAse assay. This could be considered surprising as the Qubit should capture all dsDNA. However, since the Qubit only measures dsDNA and not any single stranded (ssDNA) or other contaminants which may be present in the sample, that could be a contributor to the disparity in this example. Interrogation of the data showed that the Qubit results were the most consistent over the 3 experiments. These results, coupled with the reliability of this assay in the literature meant the Qubit was the method chosen to continue the study with. Moreover, the Qubit system has various kits that can be tailored to the yield of DNA expected from a sample, which could be useful in the future[278].

Another essential check that had to be carried out was nested PCR and gel electrophoresis using the control primers for EpiSwitch™ 3C before future work could commence. Figure 3.6 illustrates these results and highlights that the control primers perform as expected. MMP1 4/12 and MMP1 9/12 are positive 3C controls as there is a stable copy number in normal blood. Both controls were consistently present at the expected size. The primers designed to capture the loops in the MTX signature were also tested in healthy BC before being used on

patient samples (Figure 3.7 and Figure 3.8). This employed a semi-quantitative method, which was used in the seminal 3C paper by Dekker *et al*. The analysis of gel electrophoresis from healthy donors displayed a wide variation between samples. The higher copy number of loops in the *IL21R* and *IL23* loci compared to the other 3 loci was interesting. It could suggest that these genes are more active genes in the general population or could mean the alternative if the looping of the gene causes inactivation. Until the function of the loops within the signature is investigated, it cannot be fully understood why there are a higher number of loops at these loci in healthy donors over *IL17A* and *CXCL13*. There is the likelihood that absence of loops in these loci could mean that *IL17A* and *CXCL13* are more associated with disease phenotype and pathogenesis compared to the other genes in the signature. Marwa *et al* have suggested a polymorphism in *IL17A* has been associated with response to MTX[280]. As our signature has the predictive power in only RA patients, it is not unsurprising that the healthy donors used for the 3C assays thus far would have a low number of individuals presenting with loops in our genomic regions of interest. It should be noted that the results shown are from a limited number of healthy donors therefore it is difficult to draw conclusions from this data. In summary, the results presented thus far demonstrate that I was able to successfully execute the 3C assay and analysis using patient samples from the SERA biobank could commence.

In my hands, I considered whether we could transition to a high-throughput method of biomarker detection. Work in this chapter details the attempt to optimise a qPCR assay for this purpose. This aligned with the work being carried out at OBD, as they were translating other signatures to the higher throughput qPCR platform. qPCR has been described widely in the literature, but it is recognised that there are challenges in achieving a robust assay for 3C templates. A high DNA template concentration and primer-dimers can result in non-specific fluorescence[281]. As such, it was important that the MIQE guidelines were followed which state the minimum requirements for publication of PCR results[272]. qPCR MMP control primers were obtained from OBD and tested in RA patient samples (Figure 3.10). The results for qPCR with the MMP primer indicated that generally there is a low level of 3C template, indicated by the high Ct value. This has been consistently shown with work since carried out at

OBD where the MMP positive control primer has a low Ct value indicating very low DNA template. (Figure 3.31). This gave an indication of what could be expected from the MTX CCS loci.

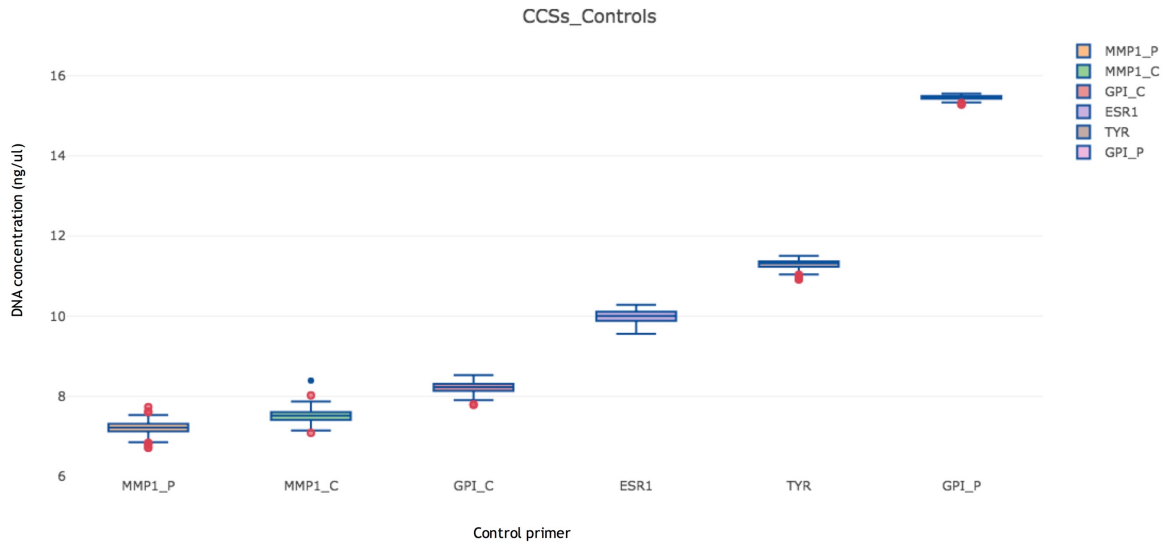


Figure 3.31 qPCR Positive Control Analysis
Concentration of DNA of 3C products using positive 3C controls from recent qPCR microarray data generated at OBD. Data is shown as box and whisker plot showing the median and standard deviation, N=1.

One of the first stages in the optimisation process was primer design. 3C regions are non-germline and low complexity which makes them prone to hairpin loops and primer dimers, therefore making primer design challenging. A low complexity sequence will reduce the primers discriminatory power on the sequence of interest and result in nonspecific binding[282,283]. This challenge is reflected in the results shown in this chapter. To be taken forward, primers had to demonstrate evidence of efficiency and specificity. Efficiency was measured by temperature gradient qPCR, with the aim of identifying the optimal temperature that a single product could be amplified. Specificity is shown by the existence of a single product, which can be evidenced through the melt curve and gel electrophoresis analysis. Three separate primer sets were tested for all 5 loci from the MTX signature, and results indicated that most primer sets were marginally improved on the set before, i.e. A better than B, and B better than C. This can be seen from primer set A having the most amplification of negative controls and primer set C, designed last, having the least amplification of negative controls. However, this was not the case for *IL21R*. Moreover, it was

evident that primers for *CXCL13* were not efficient or specific with any set (Figure 3.12). The three primer sets tested were unable to amplify the patient product of interest at most annealing temperatures. Any occurrence of the loop of interest was accompanied by amplification of negative controls. In general, the lower the annealing temperature, the lower the Ct of the patient loop of interest. However, it is at temperatures of 66.4°C and above that appear to be most optimal to amplify only the patient loop of interest without negative controls. Based on analysis of the Ct values, primer set C for *IL17A* and *IL23* appeared good candidates for 3C qPCR (Figure 3.11 and Figure 3.14). However, when gel electrophoresis was carried out on the qPCR products, a band could not be found at the temperatures that resulted in single product amplification. Manual gel loading could be considered a reason for this. Additionally, gel and melt curve analysis taken together suggest the presence of non-specific amplification. The mean Ct values for *IL21R* and *IFNAR1* suggest that no primer set was suitable for amplifying the patient loop of interest. However, limited individual experiments provided evidence that the primers may be good candidates for 3C qPCR. 68°C was an optimal temperature for primer set B and C to amplify the patient *IL21R* loop of interest without amplification of negative controls. For primer set B, the qPCR product of interest did not appear at the expected size, however it did appear at the expected size for primer set C. The optimal annealing temperatures for amplification of only the *IFNAR1* loop were 66.4°C and 64.4°C. This only occurred with primer set B, and this was in conjunction with a gel band at the expected size. Based on the limited primer success there were several candidates that were chosen for sequencing: *IL21R* version C and *IFNAR1* version B.

Sequencing was used to confirm if the amplified qPCR product was correct. The known sequence of the ligated 3C product with the characteristic TCGA sequence in the centre was used as the reference. This, combined with the predicted product size based on the specific primer design facilitated this prediction. In order to send a qPCR product for sequencing, the band from gel electrophoresis had to be purified and the yield of DNA established. As documented in Figure 3.16, this process required optimisation as the original protocol yielded only 0.88ng/μl which was not sufficient for sequencing protocols. The first attempt to improve the yield added additional incubation

steps to the protocol. The kit used for this purification is based on the centrifugal filtration method. This method has been shown to have a normal recovery rate of between 60%[284]. This was chosen as the first factor to modify, based on evidence from the literature that states increased incubation temperature or incubation time could improve yield in cross-linked DNA[285]. This proved unsuccessful, yielding less DNA than the first attempt. A possible cause could be degradation of the DNA. Alternatively, since the sizes of DNA fragments are not considerably large, this alteration of the protocol may not have had much influence. The third attempt to increase the DNA yield comprised of additional incubation steps in addition to elution in heated buffer with the aim that this would allow more DNA to be released from the membrane. This optimisation step resulted in a statistically significant increase in DNA yield from the first attempt (Figure 3.16) and provided sufficient DNA concentrations to facilitate sequencing steps.

As discussed, the first product sent for sequencing was a product from an *IFNAR1* reaction; annealing temperature 62°C. This resulted in sequencing that was low quality and had a very low alignment with the known *IFNAR1* loop sequence based on only a small number bases of the known sequence aligning with the sequencing product (Figure 3.17). This meant optimisation was required to ensure a good quality DNA product was sent for sequencing. Cloning of the gel-purified qPCR product was chosen as a method of increasing DNA quality. This would allow amplification of only the single qPCR product. After cloning, a restriction digest was conducted to cleave the cloned qPCR product. The results from the first cloning attempt did not yield the product with restriction digest (Figure 3.17D). This revealed that optimisation of the cloning protocol would be necessary. The addition of a PolyA tail to the qPCR product was not enough, and an optimised ligation ratio to facilitate increased chance of clone insert was needed. With the protocol optimised, another qPCR product was sent for sequencing. Despite the quality improvement on the first sequencing attempt, the alignment with the loop sequence was too low. The resulting high quality product, with similarly low alignment to the first sequencing attempt which suggested the primers were not amplifying the correct product.

Combined, the data generated suggested that the translation of this assay to a qPCR-based platform was not easily achievable. This could be due to a multitude of reasons, but it is conceivable that the qPCR assay could be less sensitive due to low ligation frequency in 3C loops in our signature of interest. As mentioned above, 3C ligated sequences produce low-complexity templates, and it has been documented that in some circumstances, low complexity sequences are excluded from primer design processes[286]. Furthermore, these low-complexity templates have been explored and shown to contribute to the formation of pathological ribonucleoprotein assemblies[287]. Taken together, this evidence illustrates the challenge of designing qPCR primers for these loop regions and why it is plausible that optimisation would be ineffective. However, with the addition of a more easily interpreted, high-throughput gel electrophoresis technology (Figure 3.18), I was confident that I could move forward with nested PCR.

We also wanted to understand if the biomarker was stable after 6 months on treatment. The literature describes chromosome conformation as both stable and dynamic[185], so it was important to establish chromatin dynamics in the context of RA. Establishing the time frame within which this biomarker can successfully stratify R and NR to MTX is of great importance. It has been recognised that the timing of biomarker detection is critical and that plasticity of the epigenome is a complex factor to consider in such studies[288,289]. If the biomarker is to measure disease progression, biomarkers that fluctuate with disease progression is desirable, as demonstrated by Selaas *et al*[290]. They found that *IL-6* and *VEGF-A* could be promising candidate disease biomarkers due to their reduction over the disease course. However, as the MTX CCS has the aim of establishing treatment response, and MTX can be given at any time throughout the RA treatment regimen, stability would be considered beneficial. Results illustrated that the signature was not stable in the majority of the patients. All 5 loops from the signature, had variability between time points in both R and NR. The loop dynamics came in the form of loops being lost and gained over the 6 months on treatment. It is highly plausible that MTX treatment can influence the structural epigenome. Studies of the transmembrane receptor tyrosine kinase, HER2, have illustrated the influence of treatment on the epigenetic landscape. In two breast cancer cell lines, there was genome-wide reprogramming of HER2

binding sites after treatment with the growth factor EGF[291]. Molecular changes caused by treatment have also been shown in RA studies. A study by Tasaki *et al* investigated the levels of multiple serum proteome signatures associated with RA, such as serum CRP and ESR over time. They found that treatment with Infliximab and Tocilizumab reduced proteome signatures, as did MTX to a lesser extent[292]. While this study did not monitor the molecular features of the MTX signature, it does highlight the ability of MTX to modify other markers and highlights that loops would be likely to alter conformation as a result of treatment. The Boschloo test was used to statistically test the degree of change in the signature loops between baseline and 6 months in each responder group. Only *IL23* was considered statistically stable as the confirmation in most patients remained the same at both time points. Boschloo independence tests were also used to determine the predictive ability of differentiating R and NR to MTX at 6 months (Figure 3.20B). The tests clearly confirmed the ability of the signature to differentiate between R and NR at baseline, as expected, but by 6 months this capability is lost. It must be considered how the underlying disease has changed in this time, and may be the reason why the signature no longer has predicative capacity.

To try and understand if there was a relationship between the stability of the biomarker, and demographic factors, correlations were carried out between CDAI and DAS28 scores and the number of stable loops (Figure 3.21 and 3.22). There was a minimal positive correlation between CDAI at baseline and the number of stable loops. Those with a lower CDAI at baseline have marginally lower chance of having more stable loops. This relationship is also observed with DAS28 ESR and CRP scores at baseline in R (Figure 3.22A and E). In all NR, there is the indication that there is a negative correlation that exists between increased disease activity and number of stable loops. However, collectively, the results suggest there is no relationship between disease activity and the number of loops in the signature that are present at both time points. Several R had all 5 loops in the MTX CCS at baseline and 6 months, whereas there were no NR with all 5 loops present at both time points. This may allude to a stable epigenetic set point that facilities a good response to MTX. However, the distribution of stable loop number is otherwise similar between these 2 groups. Due to the lack of interim time point to test the presence of the biomarker loops in all patients

we are unable to conclude if the chromatin loops change immediately after treatment, or at some point in between. Many studies in the field of oncology have illustrated the ability of drug treatment to alter chromatin architecture. Using histone modification enzymes, Gerrard *et al* illustrated that structural epigenome changes occurred within days of drug treatment[293]. Based on those findings, it would be reasonable to consider an earlier time point after MTX administration to test the presence of MTX CCS loops.

Another important question to explore for the MTX CCS was the ability to use it to predict response to csDMARD treatment as a whole. The results demonstrated that the signature was not able to predict response to treatment, whether it be monotherapy or combination therapy. The results shown in Figure 3.24 illustrated that there was no difference in signature loci conformation between R and NR. Only monotherapy HCQ showed a clear differentiation between R and NR based on loop confirmation. However, 5 patients is not enough to draw a robust conclusion. Moreover, in the HCQ +/- MTX R group, there was a marked age difference between patients receiving monotherapy and those receiving combination therapy. Specifically, monotherapy patients had an average age of 47, and combination therapy patients, an average of age 70. Due to the low numbers of patients already in each responder, and therapy, subgroup, unfortunately, the groups could not be stratified further to account for age. This is something that should be considered in the future, and enough samples should be obtained to ensure subgroups are comparable in age, or numbers are sufficient to stratify by age group. In contrast to the MTX signature that groups *IL17A* and *CXCL13* together, and *IL21R*, *IL23* and *IFNAR1* together, all heat maps appear to group *IL23* with *IL17A* and *CXCL13*. However, a new signature involving these new groupings would not be possible, as this grouping is the same in R and NR.

Based on the binary results, a predicted response type was assigned to each patient. The results shown in Figure 3.25 show that most predictions could not be made and were labelled undefined (UD). No patient sample had an R signature, but 5 patients in total had an NR signature. However, of those, only 1 patient was a clinical NR. The ROC analysis conducted from this data highlighted the specificity and sensitivity of the signature in these patients. All ROC analyses

had poor sensitivity and specificity results. The predictive ability of the MTX CCS for these combination of therapies is poor, at no more than 50% accuracy. This is greatly reduced from the predicative capacity of the CCS for MTX R and NR of 87% and 90%. The findings of the monotherapy HCQ group highlight, on a small scale, the importance of a large patient group for biomarker testing. Recent CCS discovery studies have used 74, 59 and 116 patients in their cohorts to define their signature[216,218,217]. The MTX CCS was systematically developed with a group of treatment naive samples that were given MTX monotherapy. Due to this systematic approach, it is not surprising there is little capacity for the same signature to stratify patient treated with a combination of other csDMARDs.

In general, there was no obvious demographic or clinical characteristic which defined why some patients had the MTX CCS present, and others didn't (Figure 3.26). The successfully predicted NR had no history of smoking, while patients in the other 3 groups had a combination of current and ex-smokers. It could be disputed that as smoking is well recognised influence on the RA epigenome, that this may play a role. However, in the patient groups used to identify the original MTX CCS, there were current and previous smokers. The same numbers of current and previous smokers were present in from the HCQ and SSZ treated patient groups, as well as no smoking history. Therefore, it is unlikely that smoking plays a significant role in the efficacy of the biomarker. The successfully predicted NR was in the middle of the age range of all other groups, suggesting this is not an influential factor either. Baseline disease activity measurements were in the upper range of values. CDAI and DAS28 CRP for this patient were close to the maximum values of the response groups with an undefined signature. The other group of responders that had a signature, although unsuccessfully predicted, also had patients with high baseline disease activity. This may suggest that the presence of the MTX CCS is more likely in those with worse disease at baseline.

When investigating the potential influence of treatment option on the ability of the MTX CCS to successfully predict response, results in Figure 2.27 suggest treatment has little influence. Observing these results from the perspective of treatment option on response, suggests monotherapy or combination therapy has around equal chance of being successful. The clear majority of patients with the

presence of a MTX signature were monotherapy treated. However, there was 1 combination DMARD treated patient. R and NR with UD signatures have patients representing all 4 treatment options. R with UD signatures have marginally more combination treated patients, however, non-responders have marginally more monotherapy treated patients. However, with the MTX CCS loci in the 'correct' conformation in only 5 patients, the influence of therapy is difficult to comprehend fully. Looking at this data with a view of understanding if DMARD treatment option had an influence on response, once again, overall monotherapy and combination therapy both appear to work just as well as each other in this patients group. This work is very limited by patient number and if this was to be explored further, a much larger patient group would be needed. This data confirms that it is not a csDMARD biomarker that was found, but a biomarker specifically for MTX response and highlights the need to explore other options for a CCS for baseline csDMARD treatment. The 5 genetic loci for the MTX CCS began with a list of RA associated loci, so it is plausible to consider that a pan-DMARD CCS could be generated from this list. Similar questions of CCS extending to other therapy areas have been discussed, but not fully explored in the literature. Work by Rousseau *et al*, identified a CCS could be used to classify a subtype of leukaemia. They acknowledged that it was likely other genes would have to be consulted to find a signature able to identify other subtypes[294]. Similarly, in another study by Rousseau, they found evidence that distinct cellular states in macrophages had distinct chromatin conformations[295]. This means with further investigation, different cellular states, captured by 3C, may shed more light on the MTX CCS and biological consequences.

Overall, while the data has been interpreted with a lot of caution based on limitations described, these findings suggest that the patient cohorts used to test the stratification potential of the signature were comparable, and the difference in sensitivity and specificity measures were unlikely to be caused by patient demographics or disease activity measures. This experiment was very limited by available patient samples and the fact that many patients were on combination therapy makes it difficult to interpret where the clinical response is arising from and therefore the meaning of the stratification using the chromatin signature. Based on the data gathered, it is evident that the MTX CCS is not

capable of stratifying patients who are on a different csDMARDs and a new signature would be required to stratify RA patients at baseline.

Another clinically important question that had to be asked was if the MTX CCS could also be validated in an independent clinical cohort. The results from this investigation indicated that the CCS could not differentiate R and NR to MTX as well as in the SERA cohort. Despite both experiments using different models (Weka vs XGBoost and Light GBM), evidence has shown that the newer models used to assess the signature in the TACERA cohort are statistically superior, and if the signature could differentiate R and NR, these models would have most likely identified this. Only 3 of the 5 loops from the MTX CCS were shown to have albeit limited, stratification potential with the newer models. A dilution series was used as a semi-quantitative method that was described previously (Figure 3.7, 3.8). Only the neat primer cocktail and the 2-fold dilution produced results that added any value to the predictive model. It must be considered that the sample collection processes of the SERA samples and TACERA samples were different and that could impact the ability of the signature to perform as well. It is known that nucleic acids are susceptible to oxidative damage after blood collection[296]. One study compared blood extraction protocol for whole blood gene expression profiling experiments using mRNA. They revealed that there were substantial differences in the transcriptomic profiles of PBMCs that had gone through three different blood collection processes[297]. However, it is largely understood that DNA is more stable than RNA and therefore, for our signature, differences in blood collection should not have as much of an effect. Interestingly, the 3 loci with promising predictive potential were *IL17A*, *IL23* and *IFNAR1*. The results from the qPCR optimisation revealed *IL21R* and to a lesser extent, *IFNAR1*, as better candidates than other loci for an effective qPCR assay, while *CXCL13* showed little evidence of success. This may be due to the advantage of the updated extraction protocol (Protocol 2, Section 2.2.3) that has been implemented by OBD, supporting the idea that the extraction protocol favours loops in some genes over others. This could be attributed to the variation of copy number of each library that is evidenced with the 3C control primers (Figure 3.31). In contrast, these 3 loops suggesting some of their predictive potential was retained across cohorts may suggest that the regulation of these three genes is more important in RA pathogenesis, or in response to

MTX. The *IL17A*/*IL23* axis and its role in driving chronicity in autoimmunity has been widely explored[248]. Moreover, this axis is heavily implicated in PsA. Based on that evidence, it could be possible that the CCS may hold some predictive potential in patients with PsA. However, as has been suggested by our data, heterogeneity and protocol variances may play a large role in the stratification potential of the biomarker and therefore, a tailored biomarker for PsA would most likely need to be specifically found. The role of this axis in response to MTX has been investigated in murine models of psoriasis[298]. One human study with etanercept plus MTX revealed higher levels of *IL17A* and *IL23* in PBMCs in psoriasis patients[299]. The combination therapy significantly reduced cytokine levels and the addition of MTX improved therapeutic response. Another study in humans, with RA, explored the effect of anti-TNFα on levels of *IL17A* and *IL23* among other pro-inflammatory cytokines. At the beginning of the study there were increased level of cytokines in the sera of patients. After 24 weeks of treatment this was reduced[300]. Loops form in *IL17A* and *IL23* loci in NR and R, respectively, therefore it is important to establish the biological impact of loop formation to understand gene expression and it's relation to drug response. This will be explored later in this thesis. However, there are multiple studies that have explored the involvement of *IFNAR1* and *CXCL13* in RA pathogenesis, and therefore it can't be said with certainty that the biomarker translates better for some loci than other due to underlying pathogenesis.

Of the 2 models tested, the XGBoost model had better capacity to differentiate R and NR. The sensitivity was much higher than Light GBM in the training set and comparable with Light GBM in the testing cohort (Figure 3.29 and 3.30). The literature supports use of both models and each have their own advantages and disadvantages. While Light GBM can find results faster in some cases, in some experiments, while slower, XGBoost has been shown to find the more accurate answer[301]. Moreover, compared to Light GBM, XGboost has been found to require less training time to produce an accurate model. While not impacted by our dataset, studies have shown LightGBM to be superior for large datasets due to memory limitations for XGBoost and this should be considered going forward.

One of the limitations of this part of the study was the difference in blood collection protocols and difference in DNA extraction method. Both differences

could contribute to the reduced predictive capacity of the signature and reduction of statistically significant loops. Furthermore, a blinded cohort of 23 is not large, and a larger cohort may reveal more about the predictive potential of the CCS in this cohort. However, as the aim would be to have this work at the individual level, it must be considered if a larger cohort size would be of relevance.

The selection process for sample collection for this validation was based only on csDMARD assignment at baseline, as it was the important to carry out the testing blinded. Therefore, disease activity information was only retrieved once the testing had been carried out. This meant that the heterogeneity that is present in RA could not be fully interrogated and shed light on the clinical information which may impact the stratification ability. Moreover, of the work documented in this thesis, this validation experiment was the most recently conducted and therefore there was not sufficient time to investigate correlations between demographic factors and predictive ability of the signature as was done with the cohort for testing broad csDMARD stratification potential. With more time, this would be useful to understand if differences in demographics or disease activity contributed to the decreased sensitivity and specificity of the signature in this cohort. Furthermore, additional work is planned to test the TACERA samples with additional markers that have been found through other studies (which will be described in Chapter 5) to identify the best stratification marker.

A significant caveat to the work described in this chapter from an exploratory perspective is the heterogeneity of PBMCs. Studies have successfully illustrated the differences in the epigenome between subsets of cells within PBMCs[209,302] and therefore it is important to consider that the CCS will likely be structurally different in the cell subsets. The aim would be to identify the loops that are relevant in one state but not in other states. For successful assessment of the subsets within PBMCs, they would have to be isolated before freezing. The biobank samples used in this study did not have cell subsets prepared in this way and therefore this question could not be easily addressed. Despite this, EpiSwitch™ is a technique developed to work in a mixed population of cells and is sensitive enough to detect loops even if they were present in only in a certain subtype. Future work examining CCS in patient groups would benefit from bio

banking cell subsets separately in order to fully understand the heterogeneity of each cell type.

3.4 Conclusion

In summation, this chapter details the complexity of investigating the structural epigenome in RA. It was ascertained that assessment of the MTX CCS should be carried out with nested PCR methodology and the Qubit dsDNA HS assay and LabChip GX technology could facilitate a high-throughput protocol for the number of samples normally processed. It was also established that the MTX CCS was not stable, but largely dynamic. Additionally, we ascertained that some loci of the MTX signature had some predictive capacity in an independent clinical cohort, but the full signature had limited capacity. Further work should be done to find a more powerful signature suitable to a range of sample collection protocols. Furthermore, results confirmed the CCS was specific to MTX response prediction and not suitable to simply predict response regardless of baseline treatment.

Further work is warranted to understand the consequence of MTX CCS loop formation on the underlying RA cellular biology. Moreover, ways to interrogate the data in cell subtypes should be explored.

Chapter 4 Exploration of the Possible Functional Implications of Methotrexate Chromosome Confirmation Signature on Underlying Cellular Biology

4.1 Introduction

The work in Chapter 3 explored the stratification potential of the methotrexate (MTX) chromosome conformation signature (CCS). Following on from these findings, it was important to begin to understand the potential biological consequences of loop formation in MTX responders (R) and non-responders (NR). It is known that chromosome looping has a role in the regulation of transcription, by bringing regulatory regions such as enhancers and promotors together[288]. As epigenetic features of the genome are subject to changes through exposure to medication, understanding the epigenetic landscape in CCS genes pre-treatment may help to understand the differences in R and NR to MTX [288]. The work by Carini *et al* highlighted that expression quantitative trait loci (eQTL), identified from work by Walsh *et al*, were present at R loop sites but absent at NR loop sites, which provides a foundation for the suggestion of regulatory differences between response types[209,219].

Some epigenetic features that could help elucidate the function of the MTX CCS loops are methylation, histone modifications, DNase hypersensitive sites (DHSs), and their relationship with transcription factors (TF). When hypermethylation of a promotor occurs, this results in gene repression due to the inability of TF to bind to the promoter region. However, hypomethylation, largely characterised in cancer, often occurs in heterochromatin and can aid in upregulation of gene expression[303]. Hyper and hypomethylation have been previously explored in RA T cells and monocytes in the context of MTX treatment. Andres *et al* found that RA patients had global hypomethylation before initiation of MTX treatment, and treatment appeared to reverse this. However, this study did not incorporate the influence of the chromatin conformation, which may impact the methylation effects[304]. Multiple histone modifications have been characterised, with some associated with increased transcription, and others with transcriptional repression. The work in this chapter focuses on 6 histone modifications, chosen based on availability of data and their representation of various transcriptional consequences. H3K4me1, H3K4me3, H3K27ac and H3K36me3 represent

transcriptional activation marks and H3K27me3 and H3K9me3 represent transcriptional repression marks. Furthermore, DNase hypersensitivity can provide an additional layer to the understanding of the 3D epigenetic landscape. DNase I is a DNA sequence non-specific endonuclease. This enzyme at open chromatin sites and collocated with transcriptionally active genes. As such, these areas are termed DHSs[305,306]. These sites are often located at transcriptionally active genes and are susceptible to multiple regulatory elements.

As research into the genome, and epigenome, has grown over the last decade, there are multiple online databases where data can be downloaded and analysed for other research interests. Numerous datasets from peripheral blood monocular cells (PBMCs), and derivatives are available. Two such databases are the encyclopedia of DNA elements (ENCODE), and the BLUEPRINT epigenomics consortium, both of which have previously yielded informative results in Rheumatology[307,308,309]. However, utilising this information for an extension of an epigenetic biomarker has yet to be done. It has been recognised that combining this data is ultimately what will result in a clinically useful biomarker[310].

Based on the loci within the MTX signature, multiple cell types could be explored to provide insight into the potential biological impact of loop formation. However, to ensure a comprehensive analysis of multiple epigenetic features in the context of this study, PBMCs, $CD4^+$T cells and $CD14^+CD16^-$ monocytes were chosen as a focus.

The aim of this chapter was to elucidate potential regulatory differences between R and NR, which may indicate the relationship between the MTX CCS and underlying pathogenesis. To accomplish this, the aims were:

1) Utilise publicly available online datasets from healthy cells to extract information about epigenetic regulatory features present at the MTX loop sites in PBMCs, $CD4^+$T cells and $CD14^+CD16^-$ monocytes

2) Where possible, use RA datasets to build on findings from healthy data.

4.2 Results

4.2.1 Identification of publicly available data available to explore 3D epigenetic environment at MTX CCS loop sites

The first step in exploring the 3D epigenomic landscape was to identify the data that was publicly available. Through a search of the literature, the DeepBlue Epigenomic Data Server (https://deepblue.mpi-inf.mpg.de) was found. This database collates the findings from several large-scale epigenome studies. Numerous studies are included, namely the BLUEPRINT Epigenome, ChIP-Atlas, ENCODE and Roadmap Epigenomics (Figure 4.1A). Once this data source was found, it was important to establish the experiments that would yield data that was of interest in our study. As such, the techniques, biosources and epigenetic marks from the database were explored, and biologically relevant sources quantified. Many experiments were available for interrogation. Analysis of the experimental techniques included in the database highlighted the epigenetic motifs that had the most interest from researchers. There were significantly more ChIP-Seq experiments than any other methodology, with 31,922 experiments included in the database (Figure 4.1B). As such, this technique was removed and a second graph plotted to clearly identify the distribution of other techniques (Figure 4.1C). The next technique with the most data available was DNAse-Seq, used to identify DHSs throughout the genome, with 3,855 experiments available. The next stage in the process was to ascertain the biosources available. Blood, with 3,817, was the biosource with the most experiments accessible (Figure 4.1D). Multiple experiments from cells types of interest were found, namely PBMCs and derivatives including $CD14^{+}CD16^{-}$ classical monocytes and $CD4^{+}$ T cells with 46,287 and 1,171 experiments available, respectively. Lastly, the epigenetic marks of interest were quantified (Figure 4.1E). In line with the techniques available, histone modifications were the highest epigenetic mark represented, with 3,635 experiments available measuring H3K27ac, closely followed by DNase with 3,652 experiments available.

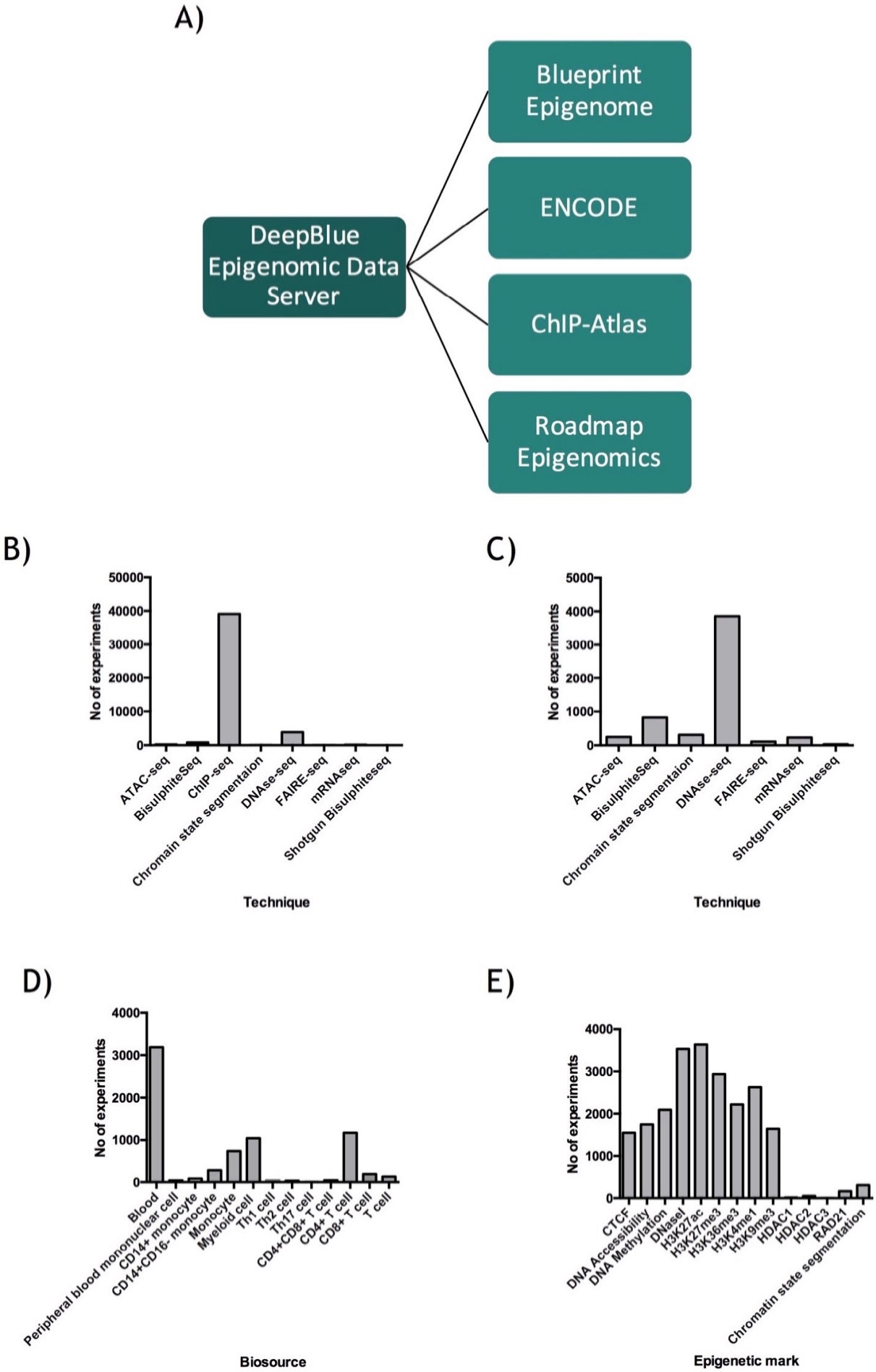

A)
Blueprint Epigenome
ENCODE
DeepBlue Epigenomic Data Server
ChIP-Atlas
Roadmap Epigenomics
B)
No of experiments
50000
40000
30000
20000
10000
0
ATAC-seq
BisulphiteSeq
ChIP-seq
Chromain state segmentaion
DNAse-seq
FAIRE-seq
mRNAseq
Shotgun Bisulphiteseq
Technique
C)
No of experiments
5000
4000
3000
2000
1000
0
ATAC-seq
BisulphiteSeq
Chromain state segmentaion
DNAse-seq
FAIRE-seq
mRNAseq
Shotgun Bisulphiteseq
Technique
D)
No of experiments
4000
3000
2000
1000
0
Blood
Peripheral blood mononuclear cell
CD14+ monocyte
CD14+CD16- monocyte
Monocyte
Myeloid cell
Th1 cell
Th2 cell
Th17 cell
CD4+CD8+ T cell
CD4+ T cell
CD8+ T cell
T cell
Biosource
E)
No of experiments
4000
3000
2000
1000
0
CTCF
DNA Accessibility
DNA Methylation
DNaseI
H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K9me3
HDAC1
HDAC2
HDAC3
RAD21
Chromatin state segmentation
Epigenetic mark

Figure 4.1 Quantification of DeepBlue Epigenomic Server Data
Identification and quantification of epigenetic data available for our study. A) Schematic of relevant studies collated within DeepBlue Epigenomic Data Server. B) Quantification of techniques from server. C) Quantification of techniques, with ChIP-Seq excluded. C) Quantification of Biosources from server. D) Quantification of Epigenetic marks available from server. Data includes experiments using both Hg38 and Hg19 genomes.
CTCF, CCCTC-binding factor

Once appropriate biosources and epigenetic marks were identified, the next step was to extract the relevant information for the MTX CCS sites. A custom script (Appendix) was used to pull out any epigenetic marks that were found in and around the signature anchor sites (Figure 4.2). As well as the coordinates of the loop anchor sites (Figure 4.2A,B), coordinates 500 kilobases (kb) up stream of anchor point a and downstream of anchor point b (Figure 4.2C,D) were included in the analysis, as epigenetic marks from that distance can be brought into close proximity with the formation of a chromosomal loop. Data from between the loop anchor sites was also captured (Figure 4.2E), which could include up to 91,723 base pairs (bp) of the genome (Table 1).

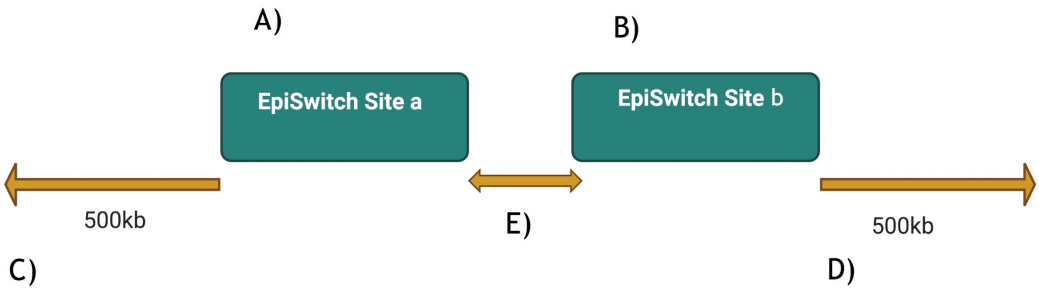


Figure 40 Sites Included in *In-silico* Analysis Methodology
Schematic illustrating sites included in custom script to map epigenetic marks to CCS loop sites. Site a = first EpiSwitch™ anchor point in CCS loop, site b= second EpiSwitch™ anchor point in CCS loop. A) EpiSwitch Site a = *IL17A* (Site 3), *CXCL13* (Site 1), *IL21R* Site 5), *IL23* (Site 4), *IFNAR1* (Site 2). B) EpiSwitch site b = *IL17A* (Site1), *CXCL13* (Site 3), *IL21R* (Site 2), *IL23* (Site 5), *IFNAR1* (Site 4). C) 500kb upstream of EpiSwitch Site a. D) 500kb downstream of EpiSwitch Site b. E) Distance between site a and site b. Coordinates used to capture epigenetic marks at EpiSwitch sites, the distance 500kb upstream and downstream and in-between (See Appendix).

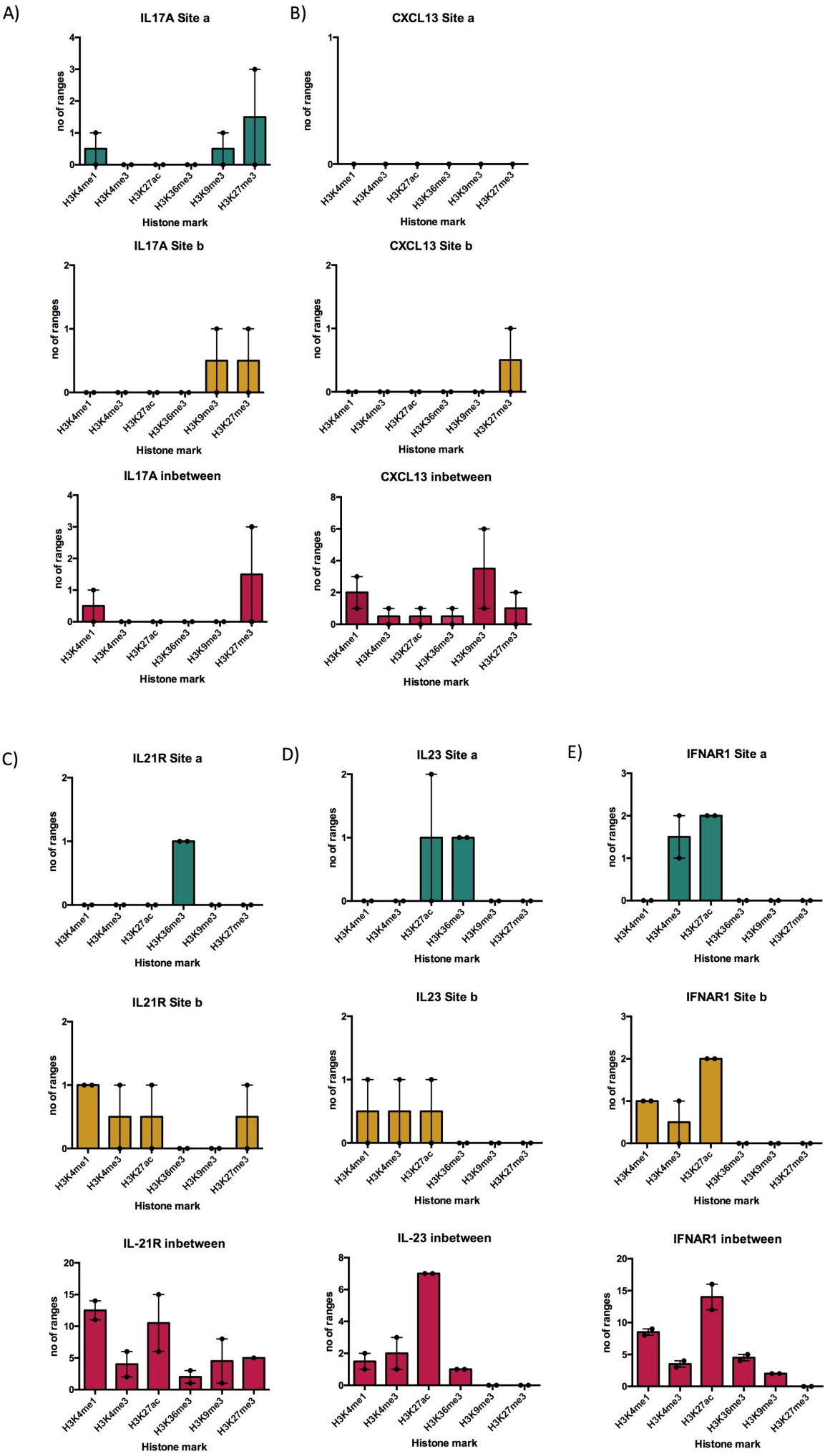
Table 4.1 Genomic Distances Between CCS Anchor Sites

Gene	EpiSwitch site a	Length of site (bp)	EpiSwitch site b	Length of site (bp)	Distance between sites (bp)
IL17A	3	10,468	1	2,435	12,467
CXCL13	1	1,681	3	2,808	90,532
IL21R	5	1,223	2	2,251	91,723
IL23	4	1,556	5	756	13,794
IFNAR1	2	1,033	4	2,560	48,544

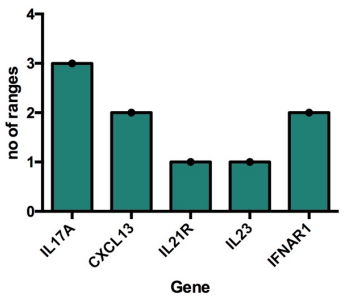
4.2.2 Exploration of epigenetic landscape in PBMCs

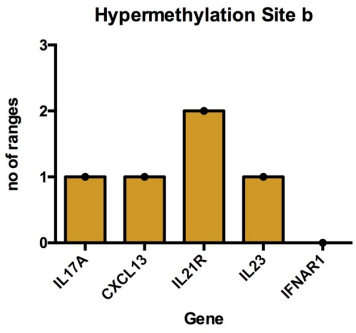
In the first instance, as the biomarker was found in the PBMC population, broad analysis of epigenetic marks in PBMCs was explored (Figure 4.3). Firstly, ranges of the genome within which histone peaks were found were mapped around MTX CCS loop anchor sites. All loop sites had the presence of at least one histone modification. At the *IL17A* site there was the presence of H3K4me1 at site a and H3K9me3 and H3K27me3 at site b (Figure 4.3A). At the *CXCL13* site there was the presence of only H3K27me3 (Figure 4.3B). There were H3K4me1, H3K4me3, H3K27ac, H3K36me3 and H3K27me3 ranges at the IL-12R loop sites (Figure 4.3C). Both *IL23* and *IFNAR1* had H3K4me1, H3K4me3 and H3K27ac histone peaks at loop sites, with a maximum of 2 ranges quantified (Figure 4.3D,E).Between loop anchor sites, the histone profile was similar to that of anchor sites at the *IL17A* locus and *IL23* locus. However, at the other 3 signature loci, histones with both enhancer and repression transcriptional consequences were captured. Methylation was also explored in PBMCs. At all loop sites hypermethylation was present at the minimum of one anchor site. Most ranges, maximum of 3, were present at site a for *IL17A*. At site b, *IL21R* had most (2) hypermethylation ranges (Figure 4.3F). Hypomethylation was less present than hypermethylation, with only ranges present at *IL21R* and *IFNAR1* loop sites (Figure 4.3G). Between anchor sites, *IL21R* has the most hyper and hypomethylation ranges (20 and 24, respectively), with *CXCL13* having a similar number of hypermethylation ranges (23). DHSs represent accessible areas of the genome, and as such, these sites

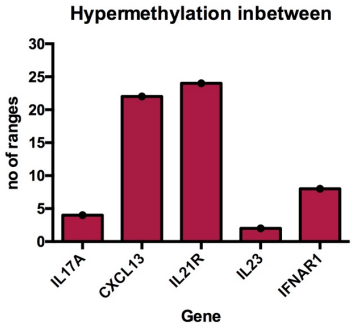
were mapped at loop regions and could be layered above other epigenetic marks to understand how likely the other modifications may impact gene expression. *IFNAR1* had the most ranges, with an average of 7.5 at site a (Figure 4.3H). *IL21R* had the most ranges at site b, and between anchor sites, with an average of 5 and 20, respectively. *CXCL13* was the only loop site not to have any DNase I sites present. Overall, the data suggest that there is the potential for a more repressive transcriptional environment at *IL17A* and *CXCL13* loci, than at *IL21R*, *IL23* and *IFNAR1* loci. Moreover, based on the DNAse I sites mapped, *IL21R* may be the loci with the potential to be most accessible and could aid in enhancing transcriptional in combination with the other epigenetic marks.


A)
IL17A Site a
no of ranges
H3K4me1
H3K4me3
H3K27ac
H3K36me3
H3K9me3
H3K27me3
Histone mark
B)
CXCL13 Site a
IL17A Site b
CXCL13 Site b
IL17A inbetween
CXCL13 inbetween
C)
IL21R Site a
IL21R Site b
IL-21R inbetween
D)
IL23 Site a
IL23 Site b
IL-23 inbetween
E)
IFNAR1 Site a
IFNAR1 Site b
IFNAR1 inbetween

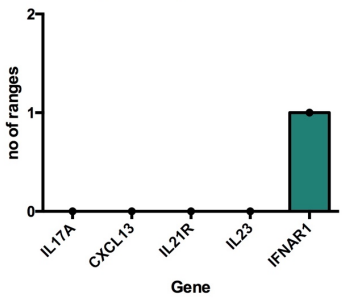
E)

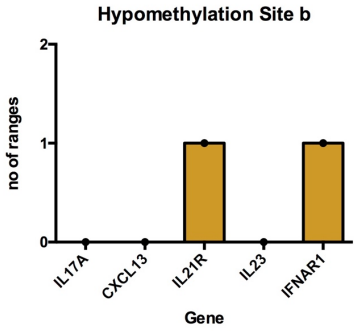

Hypermethylation Site a
no of ranges
0
1
2
3
4
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

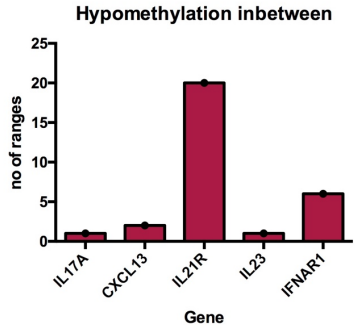

Hypermethylation Site b
no of ranges
0
1
2
3
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene


Hypermethylation inbetween
no of ranges
0
5
10
15
20
25
30
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

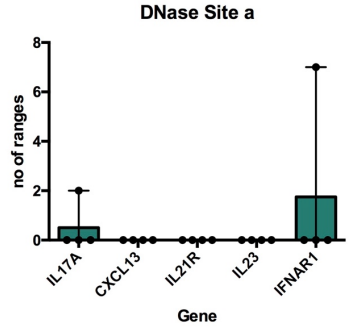
F)

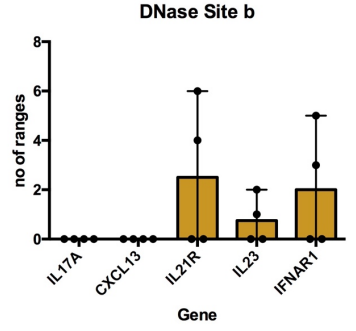

Hypomethylation Site a
no of ranges
0
1
2
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene


Hypomethylation Site b
no of ranges
0
1
2
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene


Hypomethylation inbetween
no of ranges
0
5
10
15
20
25
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

G)


DNase Site a
no of ranges
0
2
4
6
8
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene


DNase Site b
no of ranges
0
2
4
6
8
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

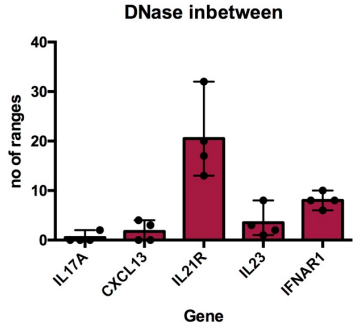

DNase inbetween
no of ranges
0
10
20
30
40
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

Figure 4.3 Mapping of Epigenetic Marks at MTX CCS Loop Anchor Sites in PBMCs
Mapping of histones, methylation and DHSs in PBMCs from healthy samples, extracted from DeepBlue Epigenomic Server. Data from hg38 genome. Histone data from ChIP-Seq technique (N=2), methylation data from Bisulphite-Seq (N=1) and DNase data from DNase-Seq (N=1). Ranges = regions of the genome within which peaks were recorded from ChIP-seq, Bisulphite-seq or Dnase-seq experiments. Site a = first EpiSwitch™ anchor point in CCS loop, site b= second EpiSwitch™ anchor point in CCS loop. In-between = stretch of DNA between CCS sites. Coordinates (See Appendix). A) Histones from *IL17A*. B) Histones from *CXCL13*. C) Histones from *IL21R*. D) Histones from *IL23*. E) Histones from *IFNAR1*. F) Hypermethylation at EpiSwitch™ sites for all MTX CCS loops. G) Hypomethylation at EpiSwitch™ sites for all MTX CCS loops. H) DHSs at MTX CCS loop sites. Data is presented as mean with range. 1-way ANOVA with Tukey's multiple comparisons. *p<0.05, **p< 0.01, ***p < 0.005, ***p<0.0001

Taking the association of regions of DNA with an epigenetic mark further, the Integrated Genome Browser (IGV) was used to visualise how the epigenetic marks may overlap at the loop anchor sites. The *IL21R* site is shown as a representative of this visualisation in PBMCs (Figure 4.4). Data representing eQTLs were also overlaid in this data. The *IL21R* loop site 5 can be visualised overlapping with H3K36me3. At *IL21R* loop site 2, there is overlap between multiple eQTLs, H3K4me1, H3K27me3, and multiple DNase I sites. These data further suggest that enhanced gene transcription at this locus is possible. Moreover, the anchor sites coming together would bring the DNase I sites closer to H3K36me3, potentially making this a highly active transcription site.

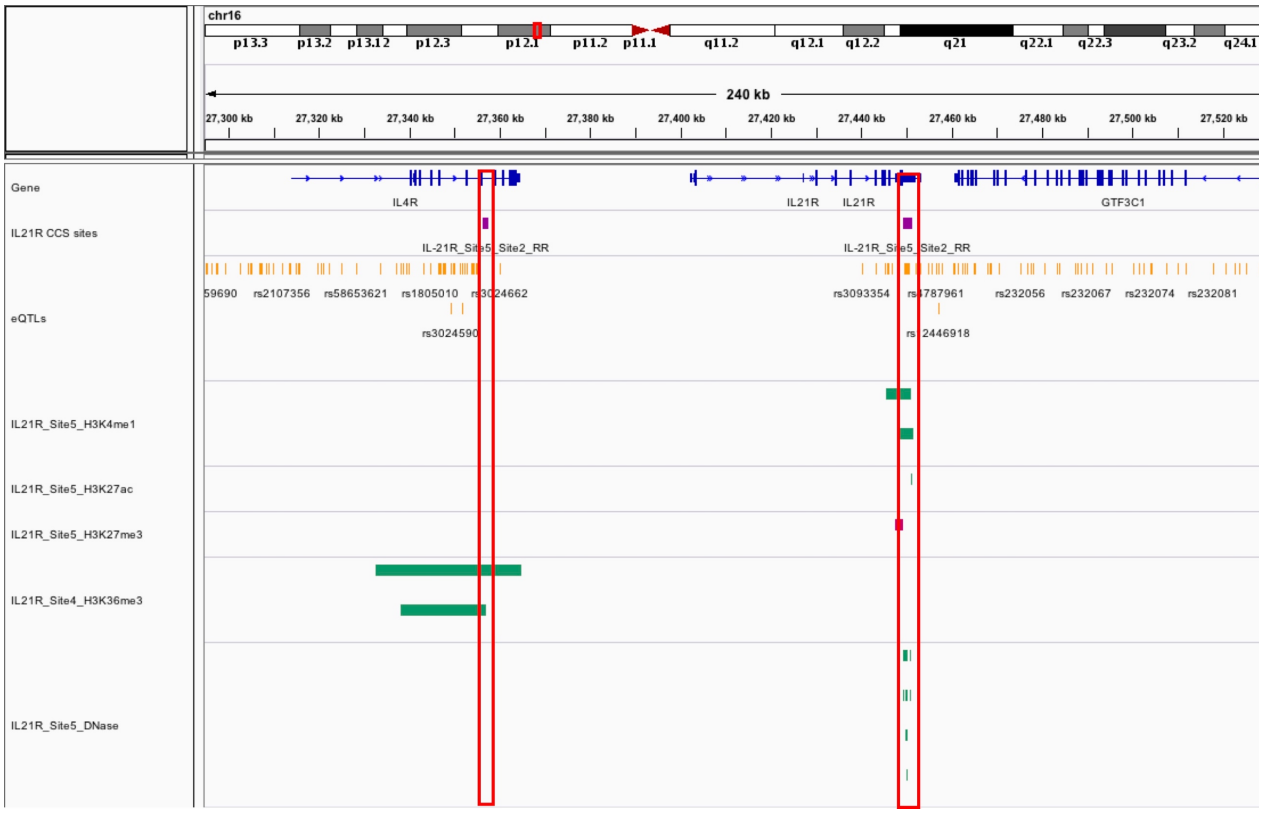


Figure 4.4 Environment Surrounding IL21R MTX CCS Loop Anchor Sites in PBMCs
Images generated in IGV. Representative image of IL21R MTX CCS sites. hg38 genome used. Histone and DNase I data from DeepBlue Epigenomic server. eQTL data from Walsh *et al*. [209]. CCS sites boxed in red encompassing epigenetic marks that lie within that region. Tracks shown in collapsed format. Track marks in green represent epigenetic marks that are likely to increase transcription, and track marks in red representative of epigenetic marks with potential repressive transcriptional impact. Site 5 is CCS site on left, Site 2 is CCS site on right.
CCS, chromosome conformation signature; eQTLs, expression quantitative trait loci; kb, kilobase; RR, reverse-reverse orientation

Following on from the association analysis of epigenetic marks at the loop anchor sites, the number of histones were quantified 500kb up and downstream of the loop anchor sites (Figure 4.5). At 500kb upstream, *IL17A* appears most distinct from the other 4 loci in the signature (Figure 4.5A). This is based on the higher number of H3K27me3 histone marks and low number of other histone marks. All other signature loci have many ranges of H3K4me1 and H3K27ac 500kb upstream. 500kb downstream from the CCS sites, the distribution of histones appears different. All loci have the presence of H3K4me1, H3K36me3 and H3K27ac (Figure 4.5B). *IL17A* has the highest presence of H3K4me1, *IL21R* the highest for H3K36me3 and *CXCL13*, *IL23* and *IFNAR1* highest for H3K27ac. There were low numbers of H3K9me3 and H3K27me3 captured across all loci. Upstream, this data somewhat reflects the data at the anchor sites at the *IL17A* loci, as it suggests a largely repressive environment. Yet downstream, there is

little differentiation between *IL17A* and *CXCL13* from *IL21R*, *IL23* and INFAR1, as there was at the anchor sites themselves. This indicates the importance of considering the whole stretch of DNA to provide a more thorough picture of the potential effect on transcription with the formation of a loop.

Moving on, using the online datasets offered the opportunity to explore cell types within the PBMC population and shed further light on regulatory differences between loci. As such, $CD4^+$ T cells were analysed next.

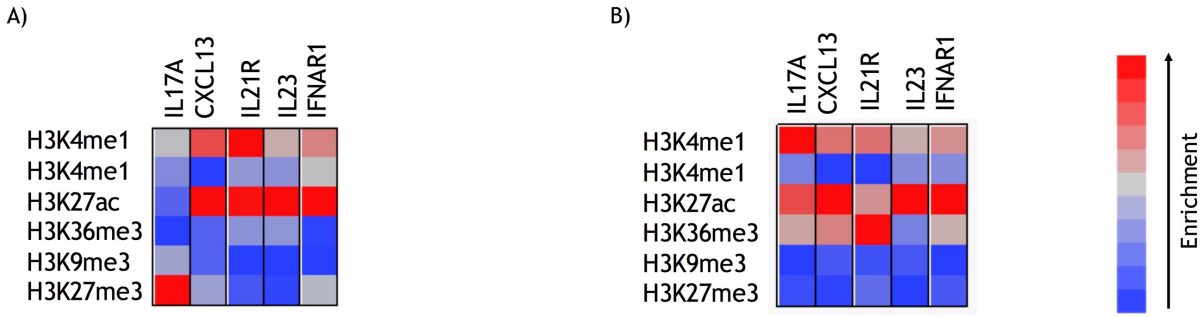
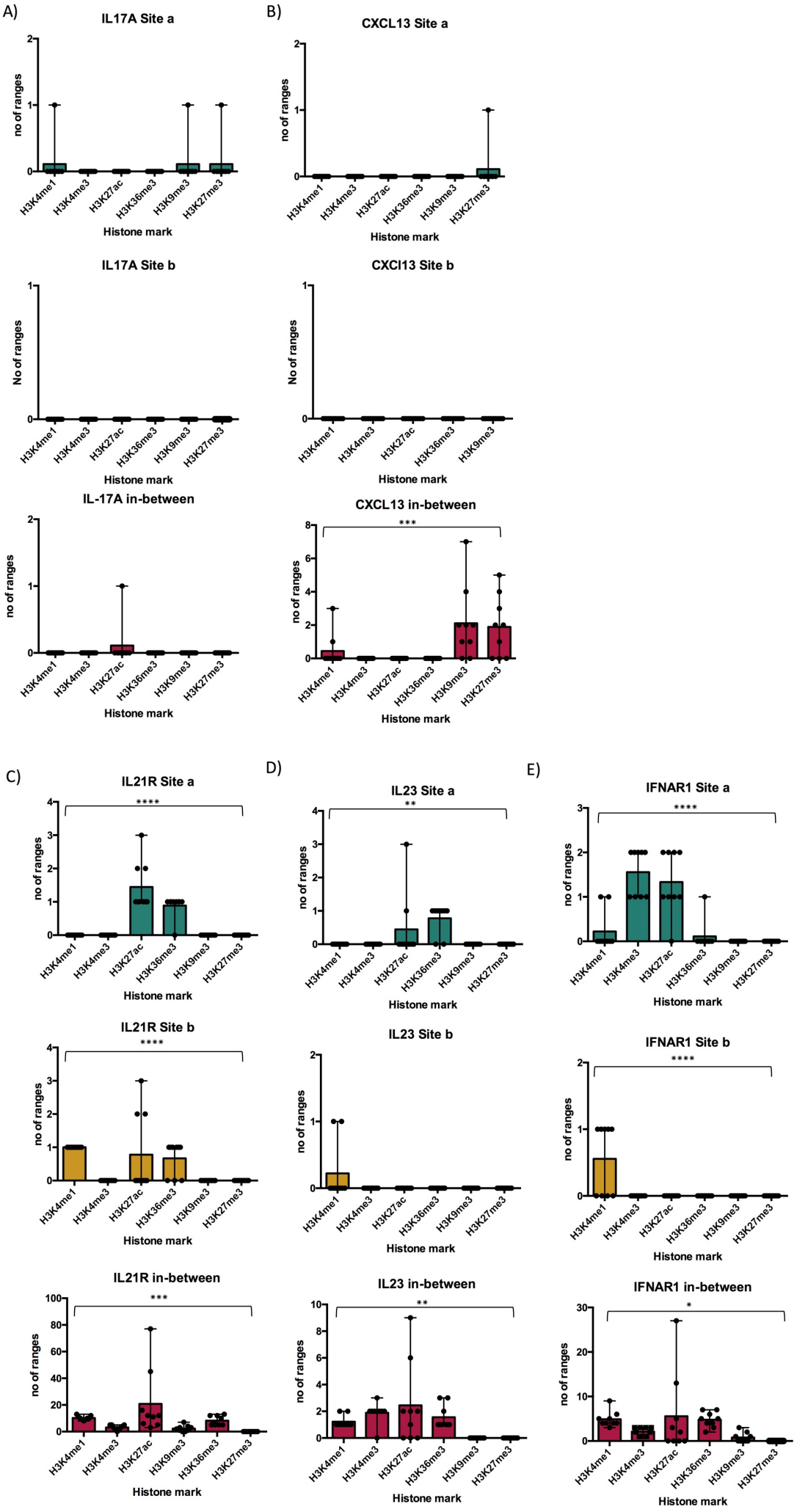


Figure 4.5 Histone Enrichment Up and Downstream of MTX CCS Loop sites in PBMCs Plots generated in JMP. Heat map representing number of ranges within which histones 500kb up and downstream of MTX CCS sites in PBMCs were mapped. Data from DeepBlue Epigenomic Server. Number of ranges were unique for each locus and each histone mark (See Appendix). A) Enrichment of histones mark 500kb upstream of MTX CCS site a. B) Enrichment of histone marks 500kb downstream of MTX CCS site b. Coordinates used to capture epigenetic marks at EpiSwitch sites, the distance 500kb upstream and downstream and in-between (See Appendix).

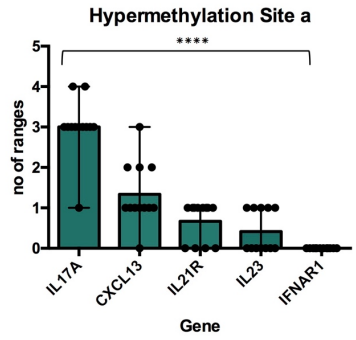
4.2.3 Exploration of epigenetic landscape in $CD4^+$ T cells

$CD4^+$ T cells were chosen as one of the cell types relevant for our loci of interest. These were chosen as $CD4^+$ T cells are known producers of *IL17A*, *CXCL13*, and express *IL21R* and *IFNAR1*[222,231,238]. As before, histone and methylation enrichment were mapped, in addition to DHSs (Figure 4.6). Site b for both *IL17A* and *CXCL13* loci had no histone marks recorded. H3K27me3 was only recorded at *IL17A* and *CXCL13* site a and between *CXCL13* and *IFNAR1* anchor sites (Figure 4.6A,B,E). The *CXCL13* site also had the presence of H3K27ac captured (Figure 4.6B). H3K4me1 marks were present at *IL17A*, *IL21R*, *IL23* and *IFNAR1* sites. H3K27ac had more ranges than the other histone marks and these were found at the *IL21R* site. Between anchor sites *IL21R* had considerably more, with one experiment finding 79 H3K27ac peaks. H3K36me3 marks were also present at *IL21R* and *IFNAR1* sites. There were statistically significant differences between

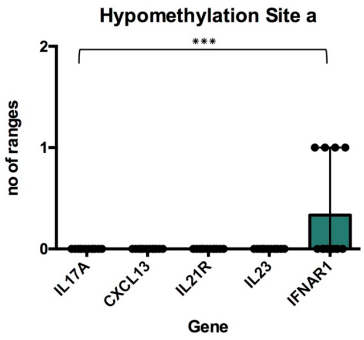
the histones across the *IL21R* and *IFNAR1* loci, and site a and in-between on the *IL23* loci. At the CXCL13 in-between site, there was also a statistically significant different between the histones. Overall, the most statically significant differences appeared to be at site a or b of the loci. Similarly, when quantifying DNase ranges, there was statistical significant differences between the loci at site a and b, but not in-between. Ranges with hypermethylation and hypomethylation were recorded at both sites in at least 1 loci. Hypermethylation was present in more abundance than hypomethylation (Figure 4.6E,F). *IL17A* site a had the most hypermethylation recorded with 4 ranges in 2 experiments. Other genes had similar levels of hypermethylation, between 1 and 3 ranges. Hypomethylation marks were only found at *IFNAR1* and *IL21R* at site a and b respectively. Hypomethylation was more represented in between anchor points, but only at *IL21R*, *IL23* and *IFNAR1* loci only. At anchor sites, *IL17A*, *CXCL13* and *IFNAR1* sites had DHSs present (Figure 4.6G). *IL17A* had a large variation of ranges between samples, from 1 to 6. *CXCL13* had DNase I marks only at 1 anchor point (Figure 4.6H). Yet, between sites, *IL21R* is clearly the locus with the most DNase I sites, and no sites are recorded at the *IL17A* loci. This data suggests that a stable gene repression state may exist in $CD4^+$ T cells at the *IL17A* locus and an enhanced transcription state may be possible at the *IL21R* site. Interestingly, this would be on contrast to the known expression of *IL17A* in T cells in the inflammatory setting.

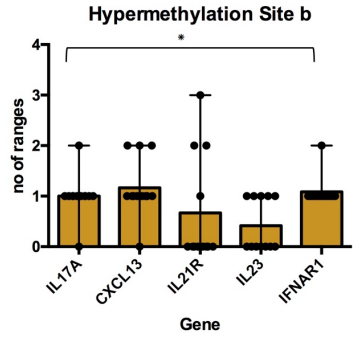

A)
IL17A Site a
no of ranges
H3K4me1
H3K4me3
H3K27ac
H3K36me3
H3K9me3
H3K27me3
Histone mark
B)
CXCL13 Site a
IL17A Site b
No of ranges
CXCl13 Site b
IL-17A in-between
CXCL13 in-between
***
C)
IL21R Site a
****
IL21R Site b
IL21R in-between
D)
IL23 Site a
**
IL23 Site b
IL23 in-between
E)
IFNAR1 Site a
IFNAR1 Site b
IFNAR1 in-between
*

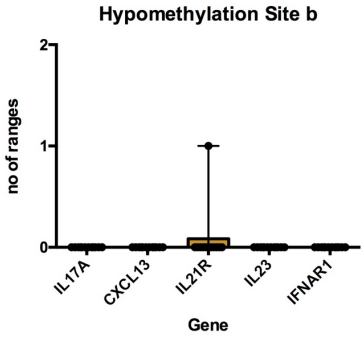
E)

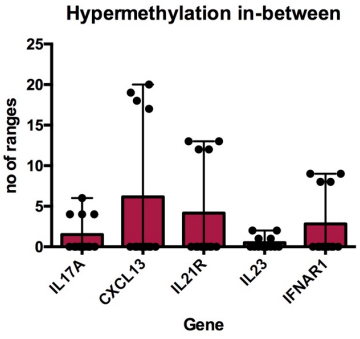

Hypermethylation Site a
****
no of ranges
0
1
2
3
4
5
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

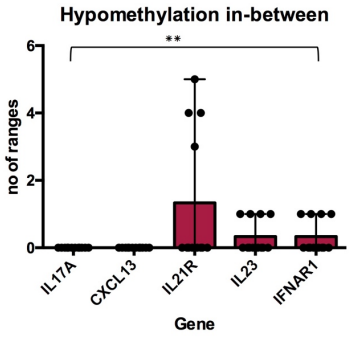
F)


Hypomethylation Site a
***
no of ranges
0
1
2
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

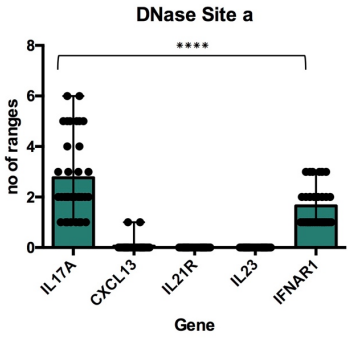

Hypermethylation Site b
*
no of ranges
0
1
2
3
4
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

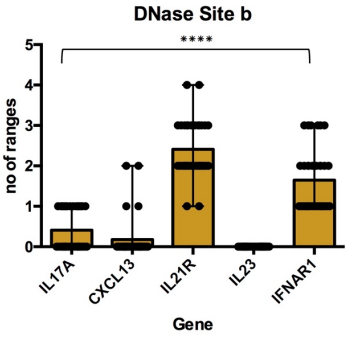

Hypomethylation Site b
no of ranges
0
1
2
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene


Hypermethylation in-between
no of ranges
0
5
10
15
20
25
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene


Hypomethylation in-between
**
no of ranges
0
2
4
6
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

G)


DNase Site a
****
no of ranges
0
2
4
6
8
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene


DNase Site b
****
no of ranges
0
1
2
3
4
5
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

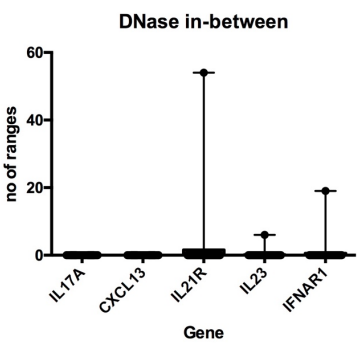

DNase in-between
no of ranges
0
20
40
60
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

Figure 4.6 Mapping of Epigenetic Marks at MTX CCS Loop Anchor Sites in CD4$^+$ T cells Mapping of histones, methylation and DHSs in CD4$^+$ T cells from healthy samples, extracted from DeepBlue Epigenomic Server. Data from hg38 and hg19 genome. Histone data from ChIP-Seq technique (N=9), methylation data from Bisulphite-Seq (N=34) and DNase data from DNase-Seq (N=12). Ranges = regions of the genome within which peaks were recorded from ChIP-seq, Bisulphite-seq or Dnase-seq experiments. Site a = first EpiSwitch^{TM} anchor point in CCS loop, site b= second EpiSwitch^{TM} anchor point in CCS loop. In-between = stretch of DNA between CCS sites. Coordinates used to capture epigenetic marks at EpiSwitch sites, the distance 500 kb upstream and downstream and in-between (See Appendix). A) Histones from *IL17A*. B) Histones from *CXCL13,* site, in-between,. C) Histones from *IL21R*. D) Histones from *IL23*. E) Histones from *IFNAR1*. F) Hypermethylation at EpiSwitch sites for CCS loops. G) Hypomethylation at EpiSwitch sites for CCS loops. H) DHSs MTX CCS loop sites. Data is presented as mean with range. 1-way ANOVA with Tukey's multiple comparisons. *p<0.05, **p< 0.01, ***p < 0.005, ***p<0.0001

As with PBMCs, IGV software was used to image the relationship between epigenetic marks. IGV demonstrated the location of histone marks at *IL17A* loop anchor site 3 (Figure 4.7). The site is dominated by the repressive marks H3K9me3 and H3K27me3. However, overall there is no overlap of histone marks at the same genomic locations. This suggests that both enhancing/repressive marks could have an influence on the gene transcription as they are unlikely to be in direct contact with each other, and there are no marks at anchor site b to come into contact with, with the formation of a loop.

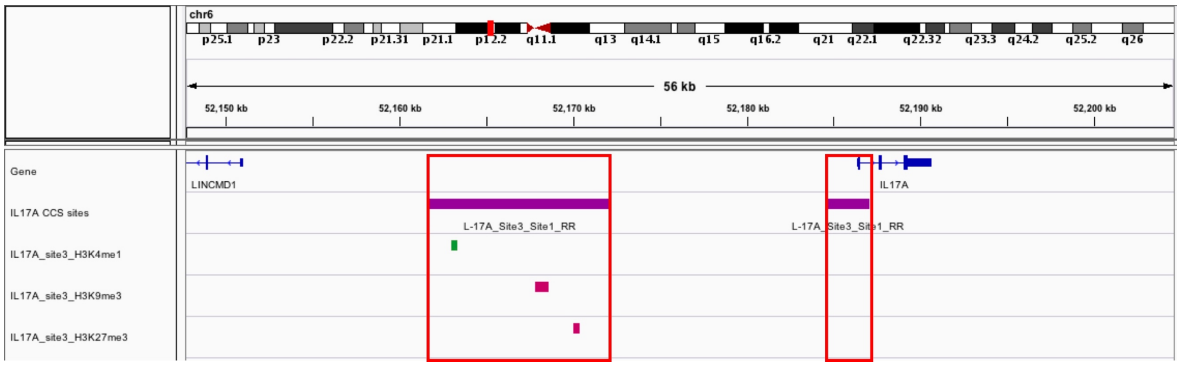


Figure 4.7 Environment Surrounding IL17A MTX CCS Loop Anchor Sites in CD4$^+$ T Cells
Image generated in IGV. Representative image of IL17A MTX CCS sites in CD4$^+$ T cells. Histone data from DeepBlue Epigenomic server. CCS sites boxed in red encompassing histone marks that lie within that region. Tracks shown in collapsed format. Track marks in green represent epigenetic marks that have a positive effect on transcription, and track marks in red are representative of a potential repressive transcriptional histone modifications. Site 3 is CCS site on left and Site 1 is CCS site on right.
CCS, chromosome conformation signature; kb, kilobase; RR, reverse-reverse orientation

At the anchor sites, there was little evidence which indicated that the regulatory environment differed between loci that are known to be expressed by CD4$^+$T cells, and those that are not. Therefore, at this stage of the analysis, I chose to focus on genes known to be expressed by CD4$^+$ T cells, thus the presence of histones 500kb up and downstream of more biologically relevant CCS sites in CD4$^+$T cells was determined (Figure 4.8). At the region 500kb upstream, *IL17A* appears more distinct from *CXCL13*, *IL21R* and *IFNAR1* (Figure 4.8A). *IL17A* had most H3K27me3 ranges, followed by H3K9me3, suggestive of an environment that could supress gene expression. Conversely, *CXCL13*, *IL21R* and *IFNAR1* have most ranges of H3K4me1 and H3K27ac, with less of the other histone marks. At 500kb downstream of the CCS anchor sites, *IL21R* appeared most distinct with high enrichment of H3K36me3 only (Figure 4.8B). *IL17A* most H3K4me1 ranges, with many ranges of H3K27ac also quantified. *CXCL13* and *IFNAR1* appear to share similar enrichment of H3K36me3 ranges with *IL21R*. The number of repressive histone marks quantified at *IL17A* build on earlier evidence which suggests a highly repressive environment which extends upstream. However, the data downstream suggest a largely gene enhancing environment which, if coming into contact with the region upstream, may create something of a more poised, non-enhanced outcome on gene transcription.

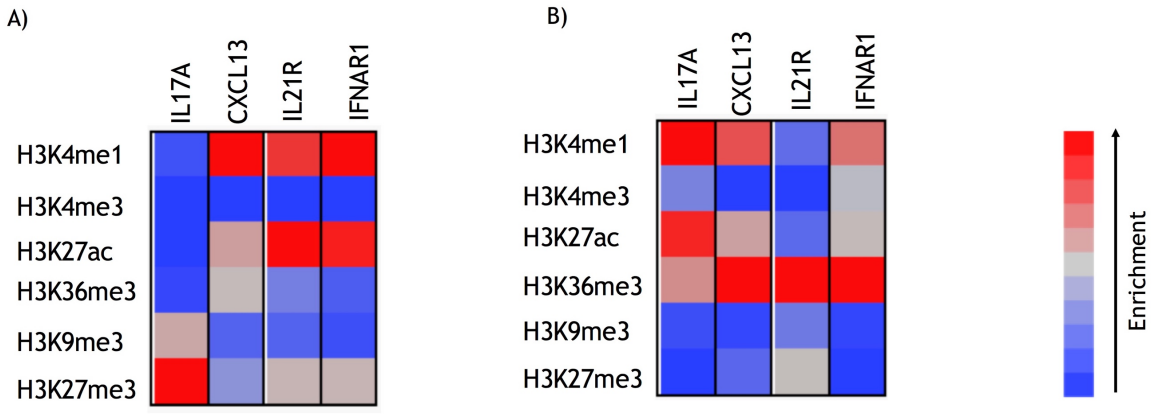
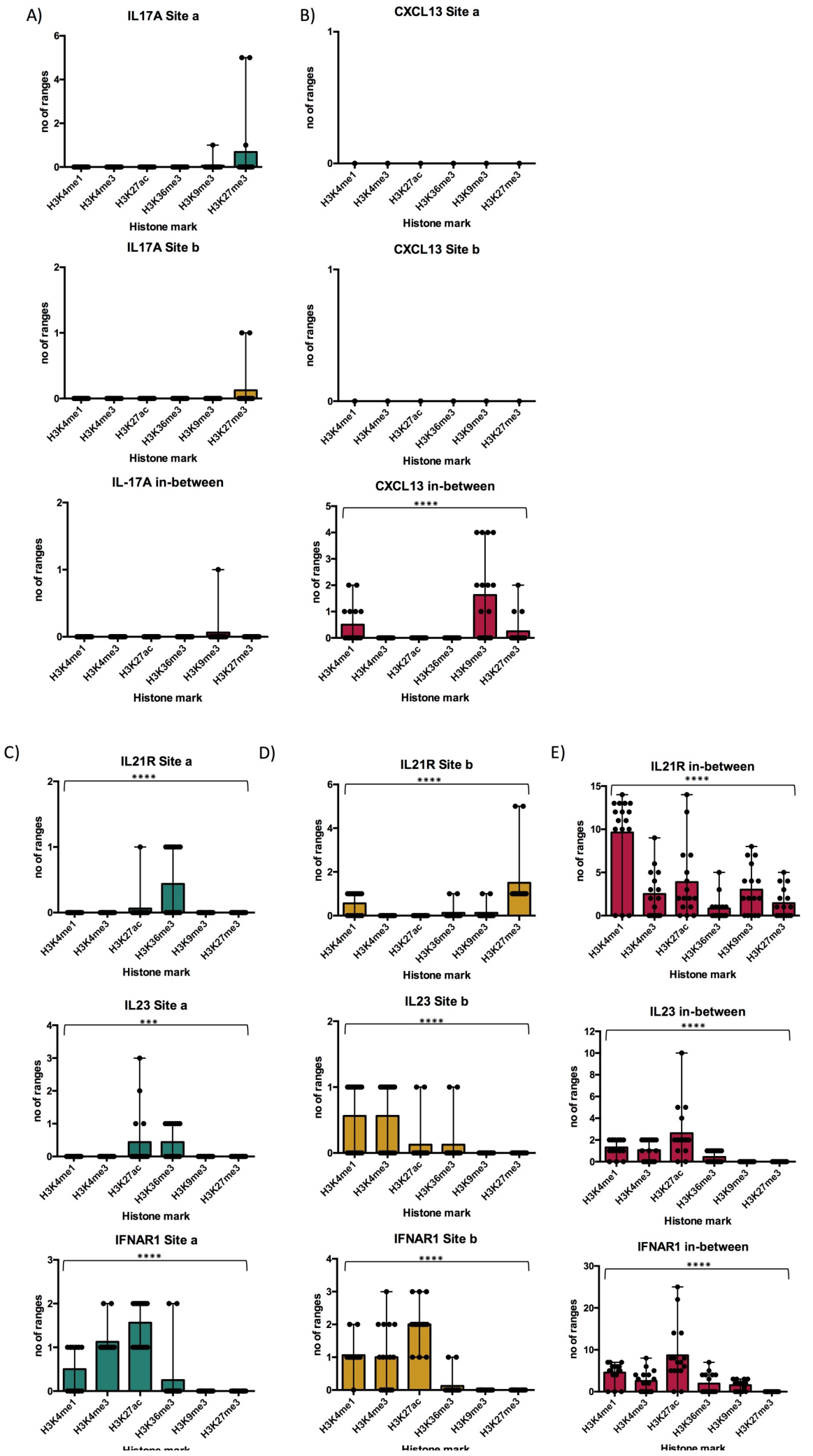


Figure 4.8 Histone Enrichment Up and Downstream of MTX CCS Loop Sites in CD4$^+$ T cells
Plots generated in JMP. Heat map representing enrichment of histones 500kb up and downstream of biologically relevant MTX CCS sites in CD4$^+$ T cells. Data from DeepBlue Epigenomic Server. Enrichment values were unique to each locus and each histone mark (See Appendix). A) Enrichment of histone marks 500kb upstream of MTX CCS site a. B) Enrichment of histone marks 500kb downstream of MTX CCS site b. Coordinates (See Appendix).

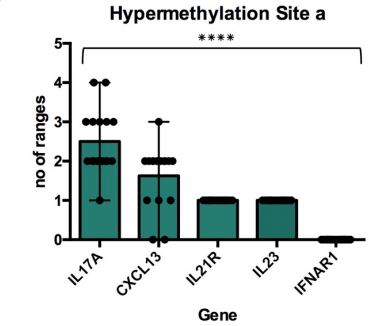
4.2.4 Exploration of epigenetic landscape in CD14$^+$ CD16$^-$ Monocytes

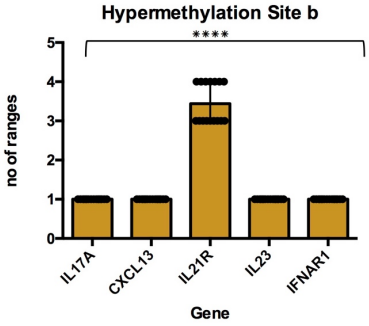
CD14$^+$CD16$^-$ monocytes were determined another appropriate cell type to allude to the potential functional impact of loop formation. This was decided due to the known expression of *CXCL13* and *IL23*, as well as *IFNAR1*$^{311\text{-}313}$. As before, histone marks, methylation and DHSs were mapped at all CCS sites (Figure 4.9). Both *IL17A* and *CXCL13* sites had only few histones ranges associated, with H3K27me3 and H3K9me3 being the dominant marks, with 4 samples with 4 ranges of H3K9me3 in the stretch between *CXCL13* anchor sites (Figure 4.9A,B). Across it's anchor sites, the *IL21R* locus had ranges with all histone marks apart from H3K4me3 (Figure 4.9C). Between anchor sites, ranges with all histone marks were found. *IL23* had the presence of H3K27ac and H3K36me3 at both loop anchor points, as well as H3K4me1 and H3K4me3 at site b (Figure 4.9D). This distribution was also found in the region in-between. At the *IFNAR1* site, both anchor points had ranges of H3K4me1, H3K4me3, H3K27ac and H3K36me3 (Figure 4.9E). H3K4me3 and H3K27ac had the most variation in ranges, between 1 and 3. At all points across the *IL21R*, *IL23* and *IFNAR1* loci measure, there was a statistically significant difference between the number of histone ranges. Hypermethylation ranges were found at all site b anchor points. (Figure 4.9F).

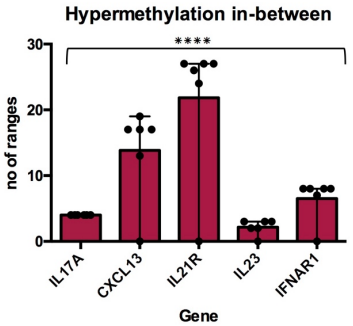
Hypomethylation was only found in the anchor regions of *IL17A*, *IL21R* and *IFNAR1* (Figure 4.9G). Ranges with DNase I were not found at the *IL17A* or *CXCL13* loop anchor sites. However, they were present at *IL23* and *IFNAR1* sites (Figure 4.9G). *IFNAR1* site b had the most hypersensitive ranges, with a maximum of 9 and average of 5 between experiments. This observation was also similar in the region between *IFNAR1* anchor sites, with many more DNase I ranges than the other loci. There was a statistically significant difference between the loci at all methylation sites and DNase sites measured. Taken together, these data suggest that there is a similar regulatory environment at *IL17A* and *CXCL13*, which differs to *IL21R*, *IL23* and *IFNAR1*; with epigenetic marks indicating is a repressive environment, and an environment that would facilitate gene expression, respectively.


A)
IL17A Site a
B)
CXCL13 Site a
IL17A Site b
CXCL13 Site b
IL-17A in-between
CXCL13 in-between
C)
IL21R Site a
D)
IL21R Site b
E)
IL21R in-between
IL23 Site a
IL23 Site b
IL23 in-between
IFNAR1 Site a
IFNAR1 Site b
IFNAR1 in-between
no of ranges
Histone mark
H3K4me1
H3K4me3
H3K27ac
H3K36me3
H3K9me3
H3K27me3
****
***

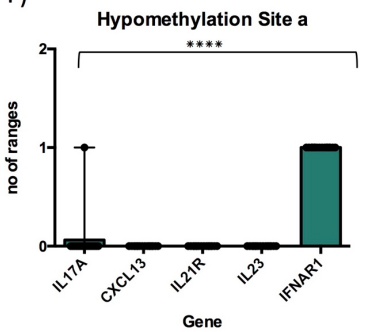
E)

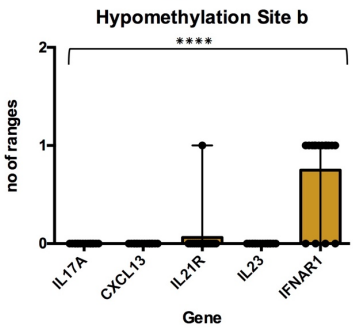

Hypermethylation Site a
****
no of ranges
0
1
2
3
4
5
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

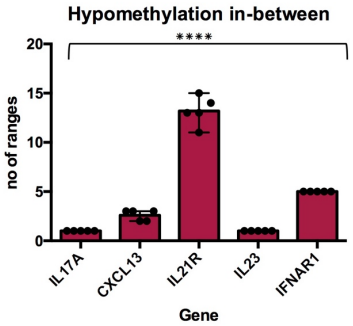

Hypermethylation Site b
****
no of ranges
0
1
2
3
4
5
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene


Hypermethylation in-between
****
no of ranges
0
10
20
30
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

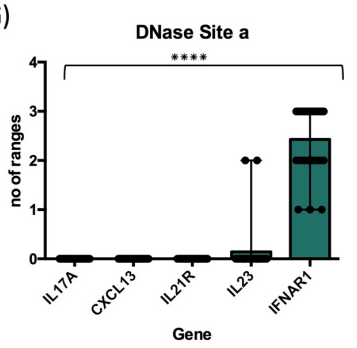
F)

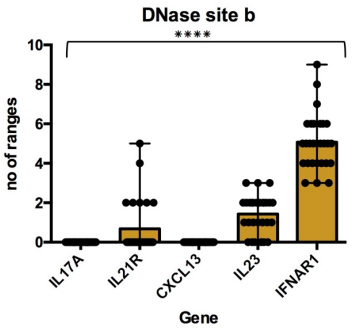

Hypomethylation Site a
****
no of ranges
0
1
2
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene


Hypomethylation Site b
****
no of ranges
0
1
2
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene


Hypomethylation in-between
****
no of ranges
0
5
10
15
20
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

G)


DNase Site a
****
no of ranges
0
1
2
3
4
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene


DNase site b
****
no of ranges
0
2
4
6
8
10
IL17A
IL21R
CXCL13
IL23
IFNAR1
Gene

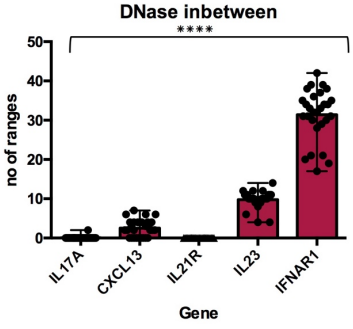

DNase inbetween
****
no of ranges
0
10
20
30
40
50
IL17A
CXCL13
IL21R
IL23
IFNAR1
Gene

Figure 4.9 Mapping of Epigenetic Marks at MTX CCS Loop Anchor Sites in $CD14^{+}CD16^{-}$ monocytes
Mapping of histones, methylation and DHSs in $CD14^{+}CD16^{-}$ from healthy samples, extracted from DeepBlue Epigenomic Server. Data from hg38 and hg19 genome. Histone data from ChIP-Seq technique, methylation data from Bisulphite-Seq and DNase data from DNase-Seq. Ranges = regions of the genome within which peaks were recorded from ChIP-seq, Bisulphite-seq or Dnase-seq experiments. Site a = first EpiSwitch™ anchor point in CCS loop, site b= second EpiSwitch™ anchor point in CCS loop. In-between = stretch of DNA between CCS anchor sites. Coordinates used to capture epigenetic marks at EpiSwitch sites, the distance 500kb upstream and downstream and in-between (See Appendix). A) Histones from *IL17A*. B) Histones from *CXCL13*. C) Histones at *IL21R*. D) Histones from *IL-23*. E) Histones from *IFNAR1*. F) Hypermethylation CCS sites. G) Hypomethylation at CCS sites. H) DNaseI hypersensitive sites at CCS loop sites. Data is presented with range. 1-way ANOVA with Tukey's multiple comparisons. ***$p < 0.005$, ***$p<0.0001$

Once again, IGV was used to better visualise the epigenetic landscape at the CCS sites. As multiple histone marks were associated with the loop sites of *IFNAR1*, this was a useful region to visualise. There is enrichment of enhancer associated histones (Figure 4.10A). Each loop anchor point clearly overlaps with multiple histone marks. These marks also overlap with DHSs. The transcription associated histone, H3K36me3, is also evident at this loop site. The DeepBlue epigenomic server also provided data on transcription factors at this site of interest. IGV demonstrated that DHSs at the CCS anchor points intersect with several TF (Figure 4.10B). *IFNAR1* loop site 2 is branched by CCCTC-binding factor (CTCF) and Signal transducer and activator of transcription 1 (STAT1). The *IFNAR1* loop site 4 overlaps with CTCF, interferon regulatory factor (IRF) and STAT1. As demonstrated previously, multiple eQTLs are present at this site and as such, overlap with the epigenetic marks described. Despite represented separately, it is evident that TF would also be present at the site of histone marks. Taken together, this data suggests that there is likely to be enhanced gene expression with the formation of a loop at this locus.

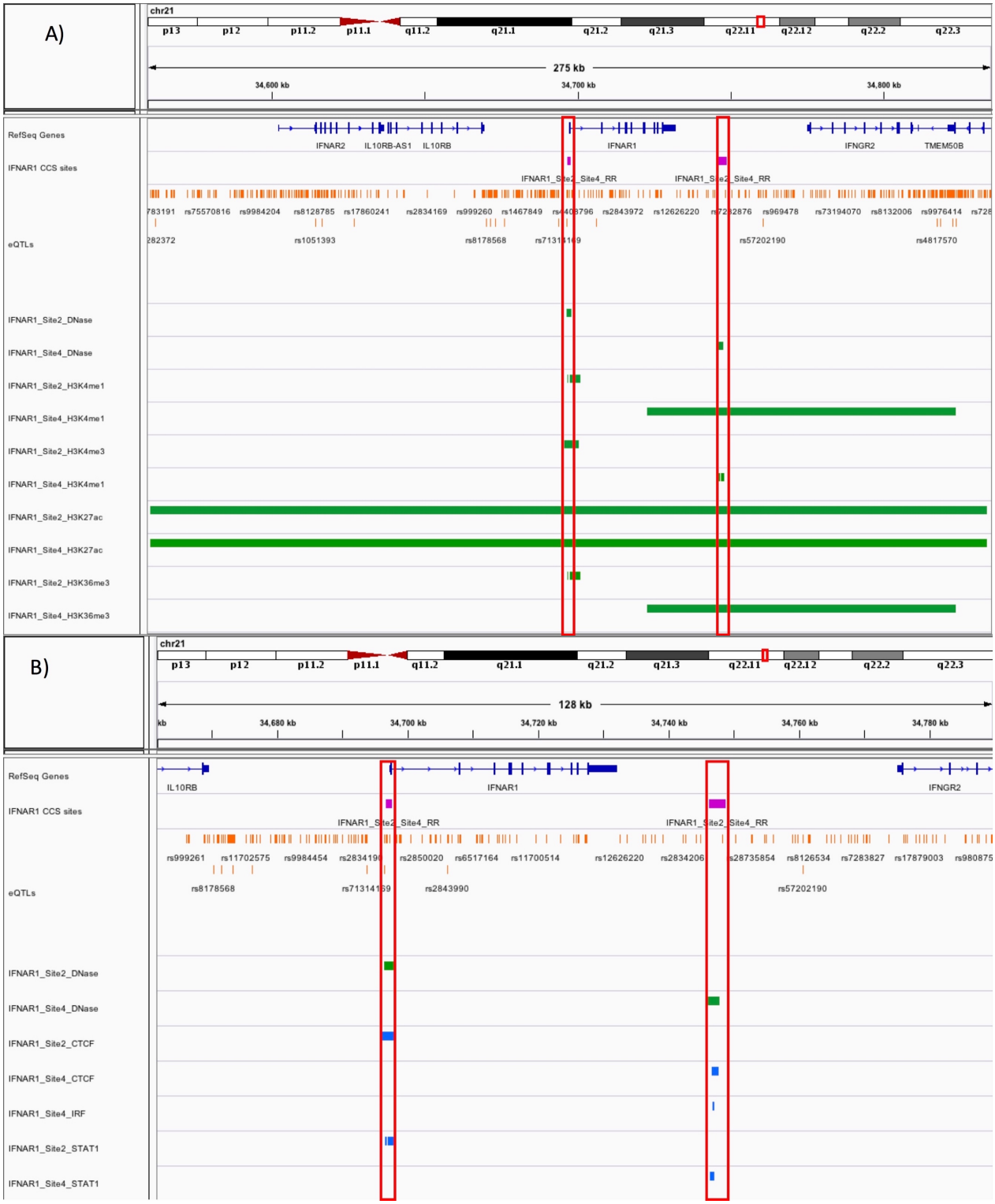


Figure 4.10 Environment Surrounding *IFNAR1* MTX CCS Loop Anchor Sites in CD14^{+}CD16^{-} monocytes

IGV used to generate image of area surrounding *IFNAR1* MTX CCS sites in CD14^{+}CD16^{-} monocytes. CCS sites boxed in red encompassing epigenetic marks that lie within that region. Histone and DNase I data from DeepBlue Epigenomic server. eQTL data from Walsh *et al.*,2016. Tracks shown in collapsed format. Track marks in green represent epigenetic marks that have a positive effect on transcription, and track marks in red are representative of potential negative transcriptional epigenetic marks. A) Histones at *IFNAR1* CCS sites. B) DHSs and transcription factors at *IFNAR1* CCS sites.
CCS, chromosome conformation signature; CTCF, CCCTC-binding factor; eQTLs, expression quantitative trait loci; IRF, interferon-regulatory factors;

kb, kilobase; RR, reverse-reverse orientation; STAT1, 1 Signal transducer and activator of transcription 1

Again, presence of histone marks 500kb up and downstream of sites in monocytes were mapped (Figure 4.11). As with $CD4^+$ T cells, I chose the most biologically relevant loci to focus analysis on at this stage. Upstream, in the region of *CXCL13*, H3K4me1 was the mark with most ranges (Figure 4.11A). H3K27ac was the most enriched mark in the region of *IL23* and *IFNAR1*. The lowest number of H3K27me3 ranges were found at the *CXCL13* region, closely followed by *IL23*. H3K36me3 was the histone with least ranges at the *IFNAR1* region. Enrichment of histones 500kb downstream was similar to upstream (Figure 4.11B). Broadly, the histone profile for all 3 loci appear similar and suggest an enhanced gene expression environment which would be in line with what we know about their expression in monocytes.

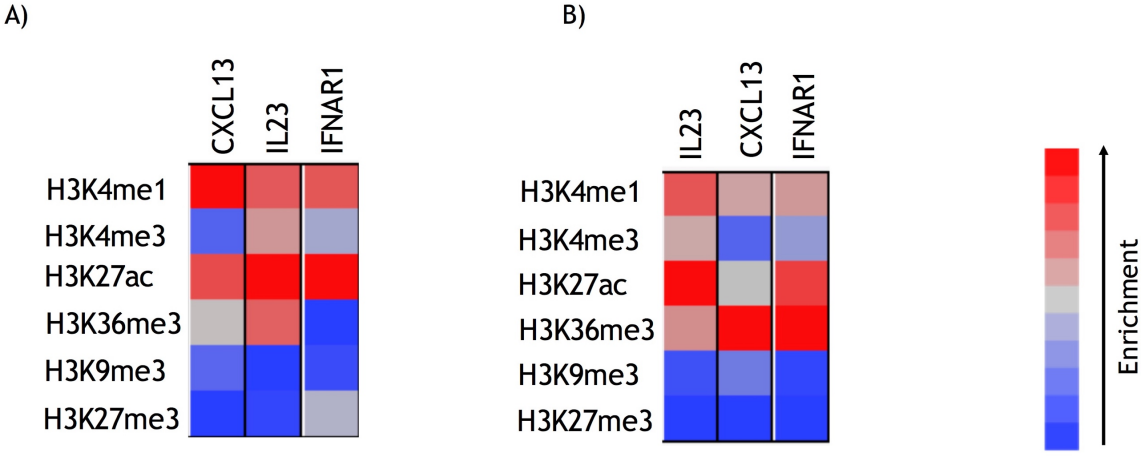
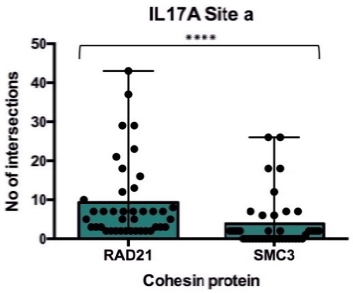


Figure 4.11 Histone Enrichment Up and Downstream of MTX CCS Loop Sites in $CD14^+CD16^-$monocytes
Plots generated in JMP. Heat map representing enrichment of histones 500kb up and downstream of biologically relevant MTX CCS sites in $CD14^+CD16^-$monocytes. Data from DeepBlue Epigenomic Server. Enrichment values unique to each histone modification (See Appendix). A) Enrichment of histone marks 500kb upstream of MTX CCS site a. B) Enrichment of histone marks 500kb downstream of MTX CCS site b. Coordinates used to capture epigenetic marks at EpiSwitch sites, the distance 500kb upstream and downstream and in-between (See Appendix)

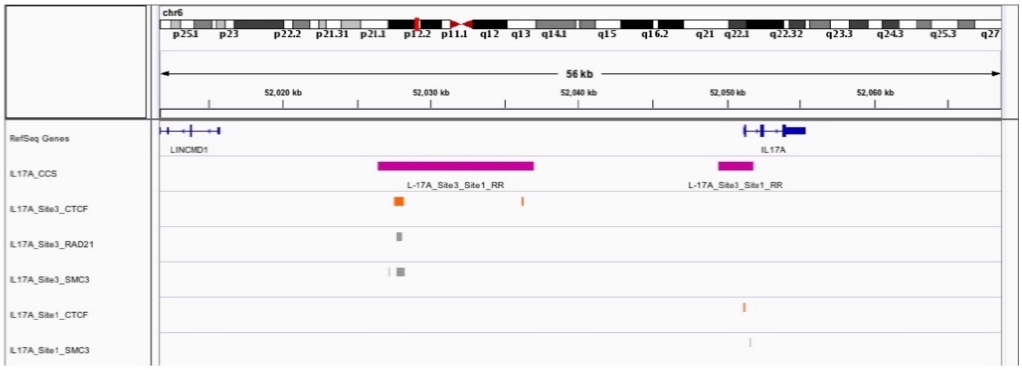
4.2.5 Exploration of markers of chromatin stability

As explored in earlier work detailed in this thesis (Chapter 3, Section 3.2.4), data suggest the MTX CCS biomarker cannot be considered stable. There are known markers of stability that can be found throughout the genome, therefore it was decided that it would be of interest to identify if these markers could be found at loop anchor sites. Once only known for their function in holding sister chromatids together, cohesin proteins, and co-localisation to CTCF, have now been implicated in loop stability and gene regulation. As such, cohesin protein and CTCF interaction was quantified at MTX CCS loop sites (Figure 4.12). To determine interaction, a function of the Bedtools program in R software was used. This program determines overlap between genomic elements and provides an output of interactions, which can then be quantified. Only data from the GM12878 cell line was available in the server, which is representative of B cells. Whilst not a focus in this chapter, B cells are very much a biologically relevant cell type and suitable for this part of the study. Based on available data, RAD21 and SMC3 proteins (which are part of the cohesin complex) were quantified. At site a, the *IL17A* loop had a maximum of 43 interactions between RAD21 and CTCF, and at site b, a maximum of 26 interactions between SMC3 and CTCF (Figure 4.12A). The difference between the number of these proteins was statistically significant. These interactions were visualised in IGV software (Figure 4.12B). This visualisation helps visualise that a loop would be required to form, to bring the cohesin complex into contact with the *IL17A* gene. At the *IL21R* loop site a, there were 3 interactions between CTCF and SMC3 (Figure 4.12C). At the *IFNAR1* loop site a, there were a maximum of 10 interactions between SMC3 and CTCF, and maximum of 27 with SMC3 and CTCF at site b (Figure 4.12E). The difference between the number of these proteins was statistically significant. Interactions were visualised with IGV (Figure 4.12E,F). These data indicate that there is the potential for a stable loop formation at these loci, but since data of all cohesin proteins was not available, these data do not provide a full picture.

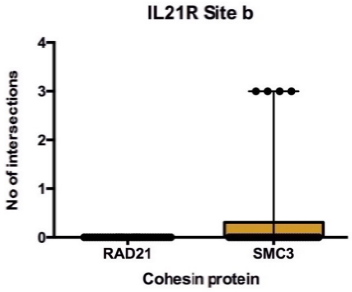
A)


IL17A Site a
****
No of intersections
50
40
30
20
10
0
RAD21
SMC3
Cohesin protein

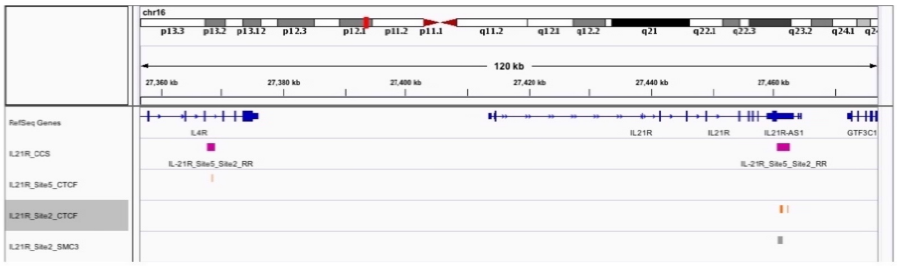
B)


chr6
56 kb
52,020 kb
52,030 kb
52,040 kb
52,050 kb
52,060 kb
RefSeq Genes
LINCMD1
IL17A
IL17A_CCS
L-17A_Site3_Site1_RR
L-17A_Site3_Site1_RR
IL17A_Site3_CTCF
IL17A_Site3_RAD21
IL17A_Site3_SMC3
IL17A_Site1_CTCF
IL17A_Site1_SMC3

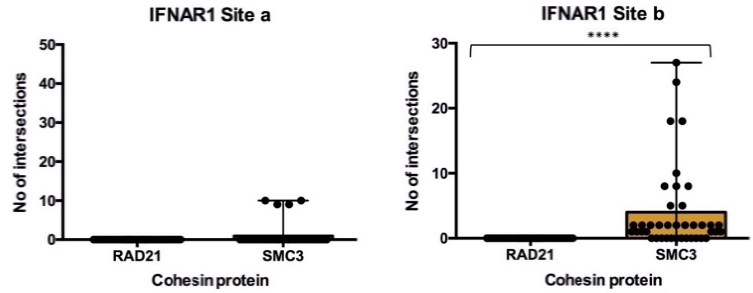
C)


IL21R Site b
No of intersections
4
3
2
1
0
RAD21
SMC3
Cohesin protein

D)


chr16
120 kb
27,360 kb
27,380 kb
27,400 kb
27,420 kb
27,440 kb
27,460 kb
RefSeq Genes
IL4R
IL21R
IL21R
IL21R-AS1
GTF3C1
IL21R_CCS
IL-21R_Site5_Site2_RR
IL-21R_Site5_Site2_RR
IL21R_Site5_CTCF
IL21R_Site2_CTCF
IL21R_Site2_SMC3

E)


IFNAR1 Site a
No of intersections
50
40
30
20
10
0
RAD21
SMC3
Cohesin protein
IFNAR1 Site b
****
No of intersections
30
20
10
0
RAD21
SMC3
Cohesin protein

F)

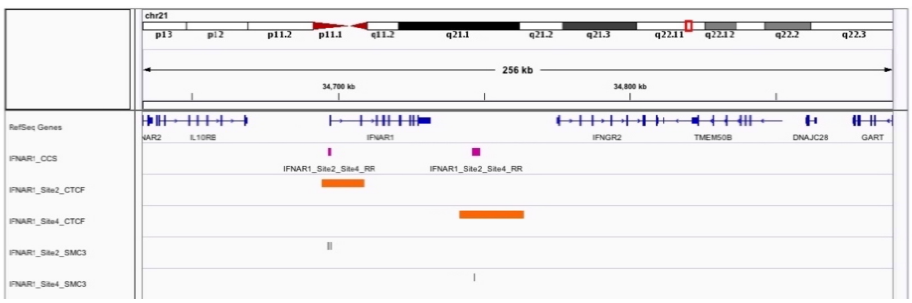
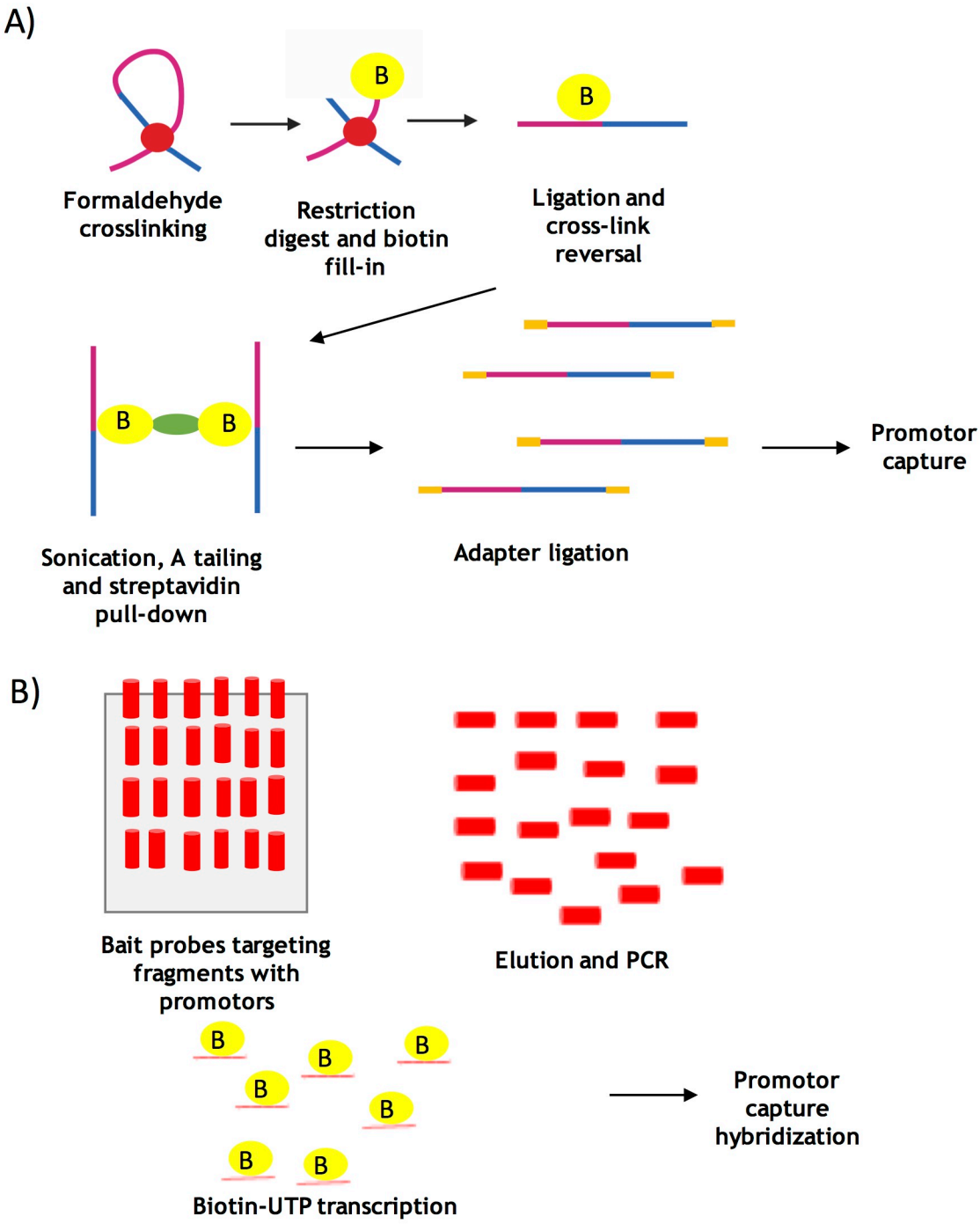

chr21
256 kb
34,700 kb
34,800 kb
RefSeq Genes
IL10RB
IFNAR1
IFNGR2
TMEM50B
DNAJC28
GART
IFNAR1_CCS
IFNAR1_Site2_Site4_RR
IFNAR1_Site2_Site4_RR
IFNAR1_Site2_CTCF
IFNAR1_Site4_CTCF
IFNAR1_Site2_SMC3
IFNAR1_Site4_SMC3

Figure 4.12 CTCF and Cohesin Protein Overlap at CCS MTX Loop Anchor Sites
Quantification of interactions between CTCF sites and cohesin proteins from the GM12878 cell line, extracted from DeepBlue Epigenomic Server. Data from hg19 genome. Number of samples variable for each loci. Site a = first EpiSwitch™ anchor point in CCS loop, site b= second EpiSwitch™ anchor point in CCS loop. Coordinates used to capture epigenetic marks at EpiSwitch sites, the distance 500kb upstream and downstream and in-between (See Appendix). A) Intersections from IL17A, site a – RAD21, N=40, SMC3, N=21. B) IGV visualisation of intersections from IL17A. C) Intersections from IL21R, site 6 – SMC3, N=4. D) IGV visualisation of intersections from IL21R. E) Intersections from IFNAR1, site a – SMC3, N=4, site b – SMC3, N=28. F) IGV visualisation of intersections from IFNAR1. Data is presented as mean ± SD. Mann whitney test. ****p<0.0001
CCS, chromosome conformation signature; CTCF, CCCTC-binding factor; kb, kilobase; RR, reverse-reverse orientation; SMC3, structural maintenance of chromosomes protein 3

4.2.6 Exploration of promotor sites at MTX CCS sites

Another technique that we thought could be informative and enable interpretation of the interactions was Promotor Capture HiC (PCHiC). This technique aims to capture loops from the genome that occur at the promotor site. Javierre *et al* conducted a study in 17 human primary blood cell types to determine the relationship between 3D architecture and gene regulation using promotor capture HiC [260]. Data from the MTX CCS loop sites was extracted from their dataset and significant interactions were quantified (Figure 4.13). *IL17A* understandably had significant interactions in the lymphoid compartment. *IL21R*, *IL23* and *IFNAR1* had significant interactions in myeloid and lymphoid cells. There were no peaks at the *CXCL13* site. Most significant interactions for *IL17A* were in non–activated (na) CD4 cells. All lymphoid cells had 3 significant interactions in the *IL21R* loop region. naCD4 cells also had the most significant interactions at the *IL23* loop site, and (total B) tB cells had the most significant interactions for *IFNAR1* loop site. This data shows that at the MTX signature loci, there is the potential for a loop to cause the activation of the gene, and suggests this is most likely in lymphoid cells. As with all data explored to this point, it was found in healthy cell populations, therefore finding an inflammatory cell type of comparison would be of use.


A)
B
B
B
B
B
Formaldehyde crosslinking
Restriction digest and biotin fill-in
Ligation and cross-link reversal
Promotor capture
Sonication, A tailing and streptavidin pull-down
Adapter ligation
B)
Bait probes targeting fragments with promotors
Elution and PCR
B
B
B
B
B
B
Biotin-UTP transcription
Promotor capture hybridization

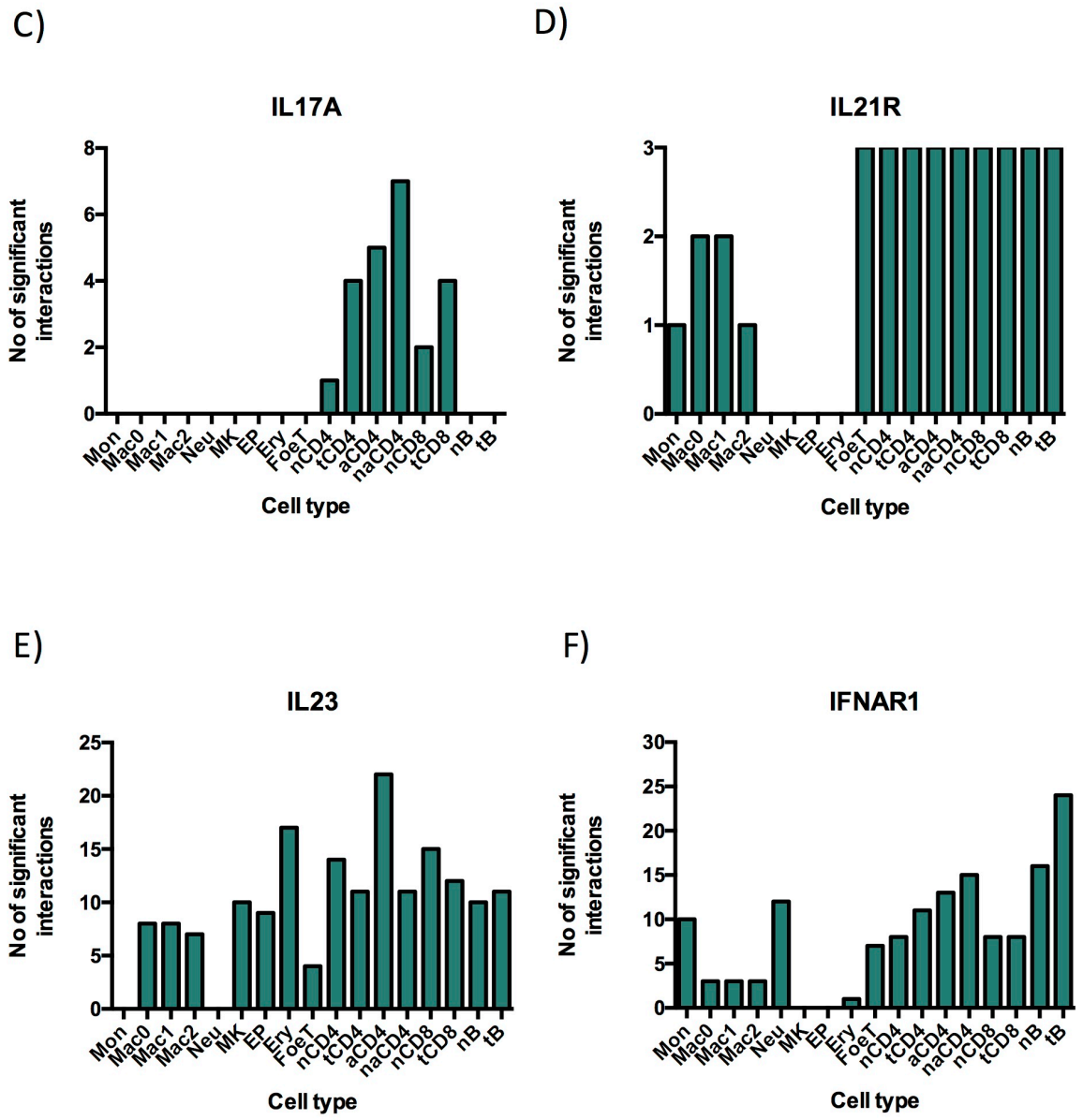
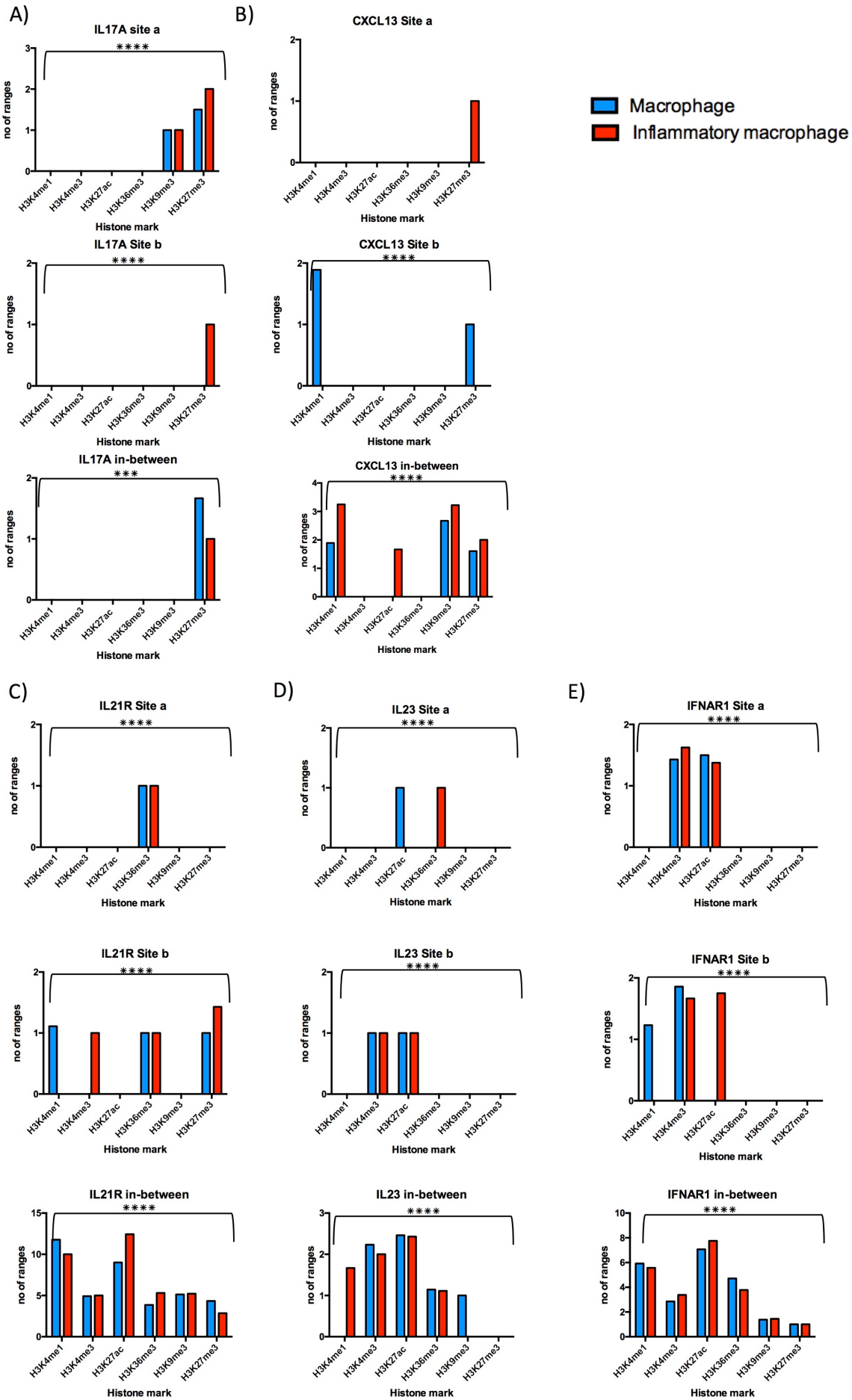


Figure 4.13 Mapping of Promotor Loops at MTX CCS Loop Anchor Sites
Mapping of number of significant interactions defined by CHiCAGO score >5 (Appendix). Data from Promotor Capture HiC (PCHiC) study by Javierre *et al.*, 2016. A) Part 1 of PCHiC protocol, generation of libraries for capture. Briefly, DNA is formaldehyde crosslinked, restriction digested and labelled with biotin. This is followed by ligation. Next, sonication and streptavidin pull down is used to shear ligated fragments and enrich the HiC library for ligation products, alongside A-tailing to repair the sheared ends. Adapters are then ligated to the ends of each fragment ahead of sequencing. B) Part 2 of PCHiC protocol, bait capture. C) Quantification of significant interactions at IL17A loop site. D) Quantification of significant interactions at IL21R loop site. E) Quantification of significant interactions at IL23 loop site. F) Quantification of significant interactions at IFNAR1 loop site. Monocytes (Mon), Naïve Macrophages (M0), Type 1 Macrophages (M1), Type 2 Macrophages (M2), Neutrophils (Neutrophils), Megacaryocytes (MK), Erythrocytes (Ery), Fetal Thymus FeT), Naïve $CD4^+$ T cells (NCD4), Total $CD4^+$ T cells (tCD4), non-activated $CD4^+$ T cells (naCD4), activated total $CD4^+$ T cells, naïve $CD8^+$ T cells (nCD8) total $CD8^+$ T cells (tCD8), naïve B cells (nB), total B cells (tB).

4.2.7 Exploration of epigenetic landscape in Inflammatory Macrophages

As most of the data was obtained from healthy cells, data from inflammatory samples was sourced as a way of understanding the epigenome in the inflammatory environment more clearly. Data from inflammatory macrophages was obtained from the DeepBlue server. These were derived from healthy primary cells cultured with beta glucan to induce an inflammatory phenotype. Data from macrophages was also mapped to allow comparison of the inflammatory and non-inflammatory state and allude the influence of an inflammatory environment (Figure 4.14). In both cell types, only inhibitory histone marks were present at the *IL17A* locus (Figure 4.14A). At the *CXCL13* locus, only H3K4me3 and H3K36me3 marks were not present, across both sites and in-between (Figure 14.4B). Across both anchor sites, all 6 histone marks were found at the *IL21R*, *IL23* and *IFNAR1* loci in both cell types (Figure 4.14C,D,E). As before, methylation marks were also mapped. Across all 5 loci, hypermethylation was more present than hypomethylation (Figure 4.14F,G). At anchor site a, *IL17A* had the most hypermethylation ranges, at site b, *IL21R*, and in-between sites, *CXCL13* has the highest number of ranges, closely followed by *IL21R*. Hypomethylation ranges were only found at the IFNAR1 locus at anchor site a. *IL17A* had no hypomethylation marks at any site. Ranges with DNase I sites were present at *IL17A* and *IFNAR1* loci at anchor site a, *IL21R*, *IL23* and *IFNAR1* at anchor site b, and across all loci in between sites (Figure 4.14H). In general, macrophages and inflammatory macrophages had similar epigenetic landscapes at the 5 CCS loci. DNase I was the only epigenetic mark to indicate any difference between macrophages and inflammatory macrophages, however this profile differed between the anchor sites and the region in-between. Specifically, at anchor sites of *IFNAR1*, macrophages had more DNase I ranges than inflammatory macrophages. Yet, in between anchor sites, inflammatory macrophages had more ranges. Therefore, taking this region as one, the difference is minimal and conclusions cannot be drawn. Taken together, this data broadly reflects earlier data from PBMCs, $CD4^+$ T cells and monocytes, and shows that repressive epigenetic marks are more present at *IL17A* and *CXCL13* sites, and more enhancing epigenetic marks at the other CCS loci. However, as this dataset was gathered to understand the influence of an inflammatory

setting and overall, macrophage and inflammatory macrophage as largely comparable this did not further that understanding. However, it is important to note that the length of activation time of macrophage in an *in-vitro* setting is considerably different than the chronic activation that is present in RA. Ultimately these data highlight that to ascertain the difference in epigenetic marks between inflammatory and non-inflammatory environments, more datasets need to be used. It was considered that data from disease and healthy samples may provide that some of that insight.


A)
IL17A site a
****
no of ranges
H3K4me1
H3K4me3
H3K27ac
H3K36me3
H3K9me3
H3K27me3
Histone mark
IL17A Site b
****
IL17A in-between
***
B)
CXCL13 Site a
CXCL13 Site b
****
CXCL13 in-between
****
Macrophage
Inflammatory macrophage
C)
IL21R Site a
****
IL21R Site b
****
IL21R in-between
****
D)
IL23 Site a
****
IL23 Site b
****
IL23 in-between
****
E)
IFNAR1 Site a
****
IFNAR1 Site b
****
IFNAR1 in-between
****

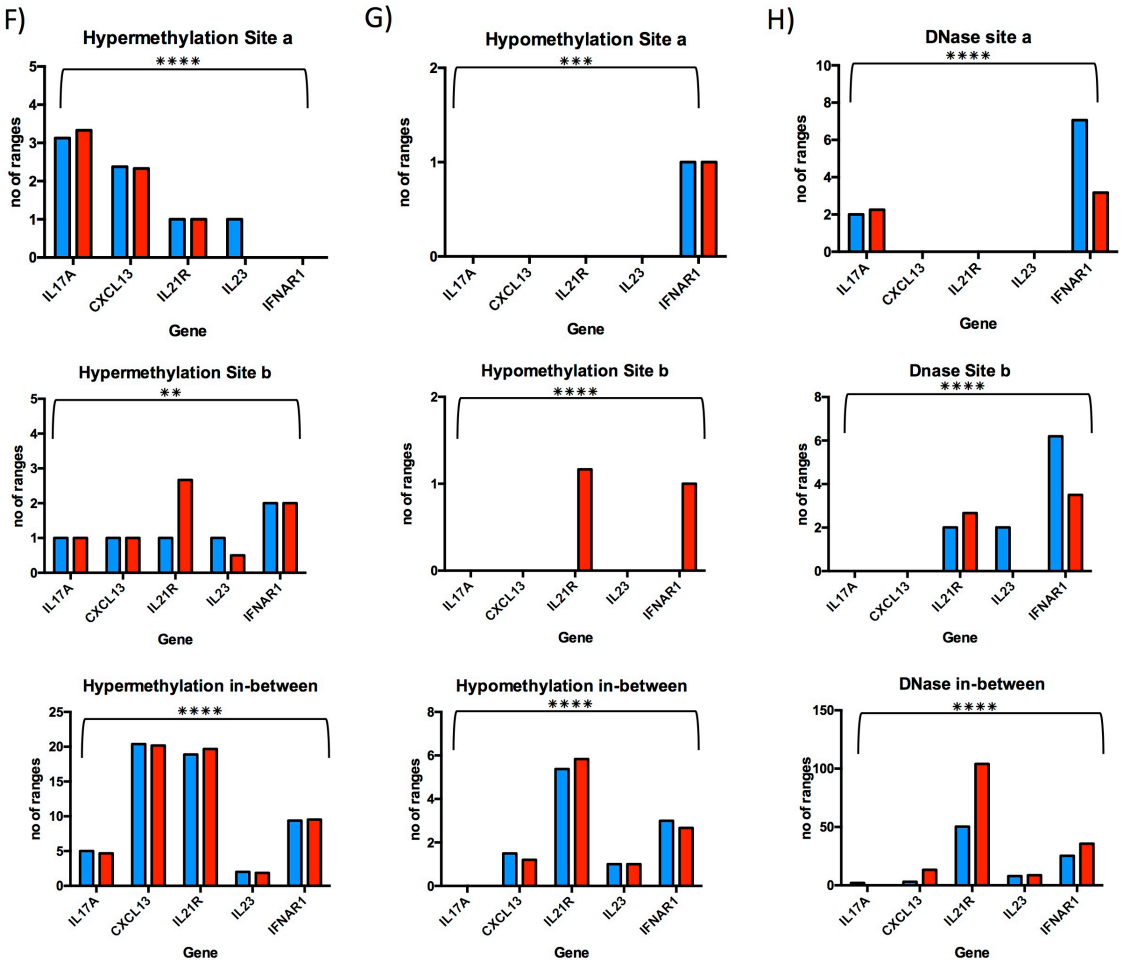


Figure 4.14 Mapping of Epigenetic Marks at MTX CCS Loop Anchor Sites in Inflammatory Macrophages

Mapping of histones and methylation in inflammatory macrophages from healthy samples, extracted from DeepBlue Epigenomic Server. Data from hg38 genome. Histone data from ChIP-Seq technique (N=14), methylation data from Bisulphite-Seq (N=14) and DNase data from DNase-Seq (N=14). Ranges = regions of the genome within which peaks were recorded from ChIP-seq, Bisulphite-seq or Dnase-seq experiments. Site a = first EpiSwitch™ anchor point in CCS loop, site b= second EpiSwitch™ anchor point in CCS loop. In-between = stretch of DNA between CCS anchor sites. Coordinates used to capture epigenetic marks at EpiSwitch sites, the distance 500kb upstream and downstream and in-between (See Appendix). A) Histones from *IL17A*. B) Histones from *CXCL13*. C) Histones from *IL-21R*. D) Histones from *IL-23*. E) Histones from *IFNAR1*. F) Hypermethylation at EpiSwitch™ sites. G) Hypomethylation at EpiSwitch™ sites. Data is presented as mean. H) DNaseI hypersensitive sites at CCS loop sites. 2-way ANOVA with Tukey's multiple comparisons. **p< 0.01, ***p < 0.005, ***p<0.0001

4.2.8 Exploration of epigenetic landscape in RA $CD14^{+}CD16^{-}$ Monocytes

While data collected in healthy subjects could provide some insight, data from RA samples offered the opportunity to improve understanding of the epigenetic landscape and the potential functional consequences in disease. Other lab colleagues, namely John Cole and Cecilia Ansalone, conducted a ChIP-Seq experiment on RA $CD14^{+}$ monocytes to understand the H3K4me3 profile in those cells. From the data, the same approach employed with the DeepBlue data was used to extract the MTX CCS relevant information. All patients were comparable in age, but there was a large variation on clinical disease activity index (CDAI) at the time the sample was taken (Table 4.2). Broadly, at sites of interest, RA and HC samples had similar H3K4me3 profiles (Figure 4.15). At Site a, *IFNAR1* was the only locus to have any H3K4me3 peaks, both RA and HC had 1 peak present (Figure 4.15A). At Site b CCS loci, *IL21R*, *IL23* and *IFNAR1* had H3K4me3 peaks, with a maximum of 3 recorded in 1 RA patient (Figure 4.15B). In-between CCS sites, only *IL17A* (RA + HC) and *CXCL13* (HC) had no peaks (Figure 4.15C). The *IL21R* site had most peaks, with a maximum of 16 recorded in 1 RA patient. These in-between data highlight the largest difference between RA and HC samples; at the *IL21R* site, RA samples have more peaks, but at the *IFNAR1* site, HC have more peaks. Considering H3K4me3 is associated with enhancer activity, these data replicate earlier findings, that loop sites associated with R have regulatory features that could enhance transcription.

Table 4.2 Characteristics of RA Patients used for ChIP-Seq
Demographic and clinical information of 9 RA patients at the time peripheral blood samples were taken.

Annotation	Sex	Age	CDAI
1	F	58	33.5
2	F	64	15.5
3	F	78	20
4	F	75	6
5	F	69	10
6	F	80	20
7	F	53	18
10	F	72	35
11	F	61	27

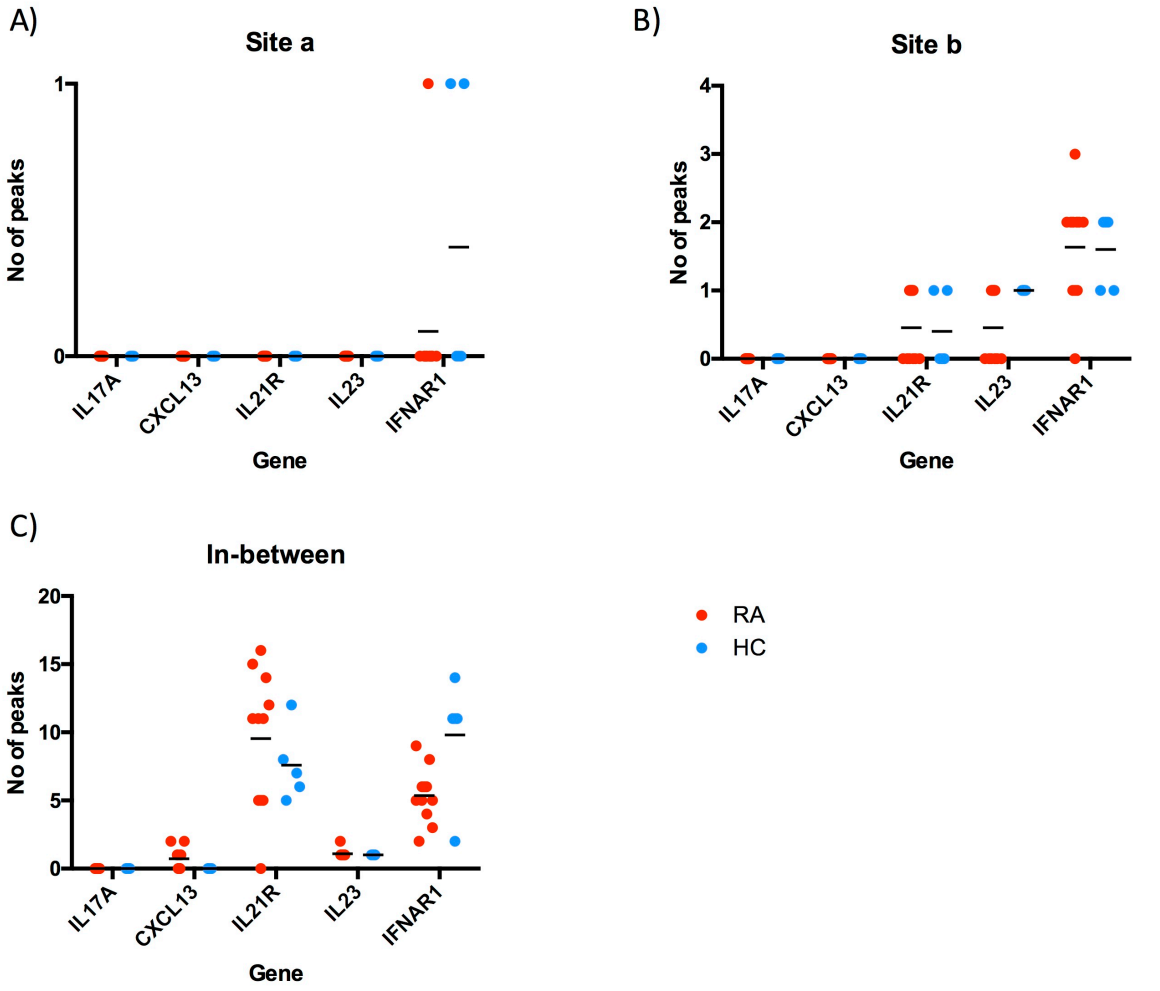


Figure 4.15 Mapping of H3K4me3 at MTX CCS Loop Anchor Sites in RA CD14^{+} Monocytes
Data from ChIP-Seq experiment in CD14^{+}CD16^{-} monocytes from RA peripheral blood, N=9 and peripheral blood from HC, N=5. A) Number of peaks of H3K4me3 at Site a at all CCS loop sites. B) Number of peaks of H3K4me3 at Site b at all CCS loop sites. C) Number of peaks of H3K4me3 in-between all CCS loop sites. Coordinates used to capture epigenetic

marks at EpiSwitch sites, the distance 500kb upstream and downstream and in-between (See appendix). Data is presented as mean.

4.3 Discussion

The work in this chapter aimed at identifying possible functional implications of the loop formation in the MTX CCS genes. Previous data indicated there may be a functional difference between R and NR loops, and epigenetic marks were used to investigate this further. Publicly available datasets were utilised to identify these epigenetic marks at the sites of interest. In the search for appropriate datasets to use, the DeepBlue Epigenomic Data Server was identified. Within this server were hundreds of datasets that could be mined for relevant information in this study. Included in the server were data on histone modifications, DNA methylation, DNA accessibility and markers of chromatin stability. Taken together the work explored in this chapter suggests the loops in NR may be more inhibitory for gene expression, and the R loops may be causing enhanced gene expression.

The datasets available within the DeepBlue Epigenomic server reflect the literature. Most experiments were from blood, a part of which (PBMC) would be relevant in our work. The availability of data from various cell types also offered the opportunity to breakdown the potential impact of loops forming in each cell type. ChIP-Seq experiments were the most represented in the DeepBlue database. This is expected based on the discovery of ChIP-Seq in 2007, providing 12 years to gather data using this methodology. Furthermore, ChIP-Seq is a relatively low complexity analysis and offers the ability to increase sensitivity by increasing sequencing depth[314]. DNase-Seq was another technique with abundant experiments available. This may be based on this technique being the hallmark for the identification of epigenetic modifications of the genome, and many other techniques have been adapted from that[315]. DNase-Seq is a versatile technique that can identify open chromatin, leading to identification of many regulatory features from enhancers and promotor regions to silencer regions. Moreover, it can often be applied to any cell type and applied genome-wide[316]. Other techniques, including ATAC-Seq, are in their relative infancy and therefore it was not expected that there would be an abundance of data for these techniques[317]. Data for multiple histone modifications also had thousands of experiments. Less data was available for cohesin proteins, but there was a minimal set of data that could be applied to this study. Ultimately, there was enough data to interrogate regions of interest for the MTX CCS.

With the datasets identified, data specific for our study could be extracted. We were interested in quantifying the epigenetic marks at the EpiSwitch™ anchor sites, including regions up to 500kb up and downstream. This was important as loops have been shown to range in size, comprising a large genomic area and bringing large stretches of DNA into close-proximity. Mamberti and Cardaso have shown loops to range from 30-90kb in size[318] and loops in a study by Zhao *et al* were a median of 16kb in size[319]. Some other studies have suggested a region of ~500kb to 2mb could be possible for a loop, and could even reach over 7mb[141,320]. Notably, this variation in sizes will depend of sequencing depth used, and larger loops may be less regulatory dense. Moreover, the EpiSwitch™ algorithm identifies loops, which may differ in size, but that are reproducibly detected. Loop formation can have a variety of consequences, dependent on the other epigenetic features in the 3D genomic area. Including the larger region in our analysis provided a more detailed, informative picture of the possible functional consequences of loop formation. This also extends on previous studies that have chosen to include regions 5kb up and downstream of genes of interest[321].

With the availability of data on epigenetic marks with differing regulatory consequences, the potential functional implications of loop formation could be explored. Firstly, data was studied from experiments using healthy PBMC samples. *IL17A* and *CXCL13* loop sites (associated with NR) had predominantly the presence of known inhibitory histone modifications, H3K9me3 and H3K27me3. *IL21R*, *IL23* and *IFNAR1* loop sites (associated with R) had the majority known to be associated with increased transcription (Figure 4.3). The disparity between activation and repression marks were not as clear with methylation. At the EpiSwitch™ loop anchor sites, hypermethylation was marginally more present than hypomethylation, suggesting the potential for transcriptional inhibition. DHSs at loop anchor points were also measured. The *IFNAR1* loop was the most accessible, based on DHSs, while *CXCL13* had no presence of any hypersensitive sites. This suggests the *CXCL13* locus may be less accessible for other epigenetic modifications to have a regulatory impact. Mapping the data between the anchor sites showed the disparities at the anchor sites were not replicated, which makes the potential functional impact of loop

formation less clear. IGV was used to visualise how these epigenetic marks may interact with each other (Figure 4.4). The representative image of the *IL21R* EpiSwitch™ site illustrates that DHSs interact with several histone marks. The overlap of enhancer marks H3K27ac and H3K4me1 suggest the site is an active enhancer site. However, these marks also overlap with the known repressor mark, H2K27me3, suggesting the potential for a more poised state of this gene. Identifying transcription factor binding sites would be helpful to further understand the enhancer potential of this loop and others in the signature[322]. The enrichment of histone marks up and downstream of the site also alluded to the potential function of signature loops (Figure 4.5). Results revealed the *IL17A* locus to be most distinct based on enrichment. Similar to results at the EpiSwitch™ anchor points, data suggests that the *IL17A* loop may be inhibitory in nature. This may result in the downregulation of other proinflammatory cytokines. Secukinumab, an *IL17A* inhibitor has been shown to be effective in the reduction of RA disease activity[323]. Based on the efficacy of this therapy, it may suggest if *IL17A* activity in RA is already reduced, MTX or other pharmacological interventions may be less effective.

The data from $CD4^+$ T cells appeared to replicate the findings from PBMCs (Figure 4.6). *IL17A* and *CXCL13* had mostly inhibitory histone marks while *IL21R*, *IL23* and *IFNAR1* regions were absent of inhibitory histone marks. Similarly, hypermethylation was highest in *IL17A* at site a, and highest in *IL21R* at site b, as with PBMCs. Again, hypomethylation levels were lower than hypermethylation. As with PBMCs, *IFNAR1* and *IL21R* were the only loci to have ranges of hypomethylation in their region. There were many more DHSs recorded in $CD4^+$ T cells than in PBMCs. Most loci, at both loop anchor sites, and in-between anchor sites, had considerably large variation in the number of DHSs recorded between experiments, which serves as a reminder that the results should be interpreted with care. IGV visualisation suggests *IL17A* is in a poised state, based on the co-localisation of histone modifications with opposite regulatory functions (Figure 4.7)[324]. Considering the histone enrichment 500kb upstream, $CD4^+$ T cells are comparable to PBMCs and illustrate *IL17A* is the most distinct gene (Figure 4.8). The enrichment of H3K27me3 continues the suggestion that *IL17A* loop may be inhibitory in nature. *IL21R* and *IFNAR1*, are almost identical in enrichment of all histones measured. This splitting of enrichment, grouping *IL21R* and *IFNAR1*

together follows the hypothesis that R and NR loops have functional differences. The profile associated with particular loops is lost when monitoring the enrichment of histones 500kb downstream of loop anchor sites. In this region, *IL21R* appears most distinct, with high enrichment of only H3K36me3. Compared to upstream, the enrichment of histones at the region around *IL17A* implies an environment likely to enhance transcription. Overall, this suggests the formation of a loop would create more of a poised environment.

$CD14^+CD16^-$ monocyte data suggests histone enrichment at loop sites to be similar to previous data from PBMCs and $CD4^+$ T cells (Figure 4.9). There were no histone marks mapped to *CXCL13* sites and no inhibitory histone marks mapped to *IL23* and *IFNAR1* anchor sites. At *IFNAR1* anchor site b, there was a substantial degree of variation between ranges recorded between samples, particularly at H3K4me3 and H3K27ac, which highlights the importance of interpreting the data with caution. As with previous data, there were more ranges of hypermethylation than hypomethylation recorded across all loci. The DNase hypersensitivity profile was different to $CD4^+$ T cells with the absence of DHSs at the *CXCL13* loci. *IFNAR1* had considerably more DHSs then *IL23*. IGV visualisation showed crossover of multiple histone marks at both loop anchor sites in *IFNAR1* (Figure 4.10). However, in the absence of inhibitory histone marks, unlike the suggested poised state of genes shown in PBMCs and $CD4^+$ T cells, this data here suggests *IFNAR1* is an active enhancer site. This data was also supported by the availability of transcription factor data from the DeepBlue Epigenomic server. DHSs sites overlap with CTCF, STAT1 and IRF at site a and b, respectively. This further supports the suggestion of likelihood that there could be enhanced transcription of this gene. This is supported by previous work in mice and humans, particularly in the locus control region (LCR). In transgenic mice, in a 1.9kb region with a DHS, the human beta-globin gene expression was increased 100-fold[325]. Previous work has implicated the role for DHSs in protein interactions at the LCR. Data suggests that DHS properties are more responsible for protein interactions than to other LCR regulatory features[326]. In the region 500kb upstream, all 3 loci appear to have similar enrichment of all histones (Figure 4.11). *IL23* differs with the higher enrichment of H3K4me3 and H3K36me3. Based on the low enrichment of H3K9me3 and H3K27me3, all three loci in this region appear to be in an environment which would support active

transcription. This environment appears to be similar 500kb downstream, based on higher enrichment of the same histones as upstream. The data suggests active transcription would be more likely in the downstream region based on the higher enrichment of H3K36me3.

Exploring data in the differing cell types gave the opportunity to observe if the epigenetic environment may be different between them, a question that could not be answered from data explored in Chapter 3. Overall, the data suggests that the environments at CCS loci are similar in $CD4^+$ T cells, $CD14^+CD16^-$ monocytes and PBMCs. Similar regulatory profiles between $CD14^+$ monocytes and $CD4^+$ T cells alludes to their interaction *in vivo*[327]. However, as this data is for healthy cells, samples from RA are very likely to differ. In RA, elevated *IFNAR1* expression is known, which would parallel our data[328]. However, it could be hypothesised that there may be more enrichment of enhancer marks at the *IFNAR1* locus in an inflammatory cell than found in healthy populations studied here. Moreover, a study has shown that the whole blood *IFNAR1* signature is mostly contributed to by peripheral granulocytes, such as neutrophils. This is an indication of another cell type of interest for future work[257]. While our data suggest that *CXCL13* expression could be repressed, *CXCL13* levels have been suggested to be a potential candidate for measuring RA disease severity[234]. Furthermore, while *IL21R* expression is known to contribute to osteoclast formation in multiple myeloma, its expression in RA is less understood. Our data from healthy subjects would indicate the potential for increased expression of this gene, which exemplifies the need for caution when interrogating the data and the importance of exploring these results in RA patients[329].

The overlap of CTCF and cohesin has been widely discussed in the literature and has been found to be indicative of stable chromatin loop formation[330]. From the datasets available, interactions between CTCF sites and cohesin proteins were discovered. Most interactions were found at the *IL17A* loop site, which was the only loop site to have both RAD21 and SMC3 cohesin complex proteins (Figure 4.12). The *IL21R* loop site had least CTCF cohesin protein overlap, with only SMC3 presence at one anchor point on the loop. Li *et al* carried out a study to demonstrate constitutive CTCF and cohesin interaction in the human genome. Using multiple cell lines they illustrated that RAD21 and SMC3 overlap in 90% of

cases[331]. This may suggest the CTCF-cohesin interaction at the MTX CCS loop sites are not constitutive, which would relate to the findings that the loops are largely dynamic (Section 3.2.4). The CTCF and cohesin proteins are not close to promotors of transcription start sites (TSS) in the *IL17A* loop region, suggesting a structural role. The interactions at *IL21R* and *IFNAR1* loop sites may play more of a transcriptional role based on the localisation to the gene itself. If this co-localization was indicative of enhancer transcription, and results could be translated from cell line to B cells in the context of RA, this may suggest contribution to B cell dysregulation in those who respond to MTX[239 332]. B cells would be of interest to explore further as they have been heavily implicated in RA pathogenesis. Specifically, they can act as antigen-presenting cells leading to the activation of autoreactive T cells, and can also contribute to the production of autoantibodies[333,334]. Contrastingly, some studies report that peripheral B cell level in RA blood is comparable to healthy blood. Moreover, some new roles for B cells in RA have been found which suggest B cells within bone marrow aggregates contribute to the upregulation of bone-resorbing osteoblasts, and thus restore bone homeostasis[335]. Preferably, exploration of these concepts would be done in primarily cells, although data from a B cell representative cell line, GM12878 would be an appropriate surrogate.

Further research has shown that CTCF-cohesin protein interactions are highly cell specific. Cohesin was found to co-localize with master regulators such as ER in breast cancer cell lines and HNF4A in liver cell lines[336], which suggests their role in transcription. As our data was taken from the GM12878 cell line, it would be of interest to understand if any of the cohesin proteins at the loop sites were also in proximity to master regulators for B cells, such as Pax5[337]. There was not sufficient time to identify publicly available datasets for this information, but it should be done going forward to build on the data already found. Moreover, the interpretation of these results is limited by the lack of datasets available to explore the presence of other cohesin proteins such as SMC1 and STAG. Research should be done to ascertain if other online datasets exist obtain this information.

Data shown in Figure 4.13 revealed that significant interactions at promotor sites are present within the regions of the MTX CCS loci. These results indicate that

the formation of signature loops in R and NR to MTX are potentially playing a vital role in transcriptional regulation. *IL23* and *IFNAR1* have most interactions, and the biggest range across the 17 cell types tested. These findings closely relate to the findings from mapping of other epigenetic marks at MTX CCS sites and further support the hypothesis that the loops in *IL21R*, *IL23* and *IFNAR1* are acting to enhance gene expression. Lymphoid subsets have more interactions than the myeloid subset, which would correspond to the cytokines and receptors being investigated. $CD4^+$ T cells had high numbers of significant interactions in all loci. Earlier data shown in this chapter suggests that some loci may be in a poised state based on the co-localisation of certain histone marks. Interestingly, a study has linked the formation of new promotor-enhancer loops, identified by PCHiC, with activation of poised genes[338]. Moreover, a recent study has shown an adapted 3C method has facilitated identification of over 7000 active promotors[339]. Specifically, by removing the noise created by some 3C methodologies, it has allowed the capture of more information, such as hubs. Another newly developed enhancer exploratory network, HACER, allows exploration of cell-specific enhancers at loci of interest[340]. Association analysis with these datasets would be useful to compliment the work described in this chapter. Our work shows where possible poised genes were identified is also where promotor loop interactions were found. No interactions were captured at the *CXCL13* locus in our area of interest, however, other datasets may exist to find out information about the epigenetic marks in this region. In future studies, it will be important to explore this dataset further and capture the relationship between promotor sites and eQTLs. Notably, a recent study has explored this concept in heart disease and was able identify new candidate genes in heart disease through exploration of the promotor interactome and eQTLs in embryonic stem cell-derived cardiomyocytes[341]. Out with the scope of this chapter was exploration of the possible orientation of transcription. This may be an appropriate follow up experiment.

So far, based on availability, only data from healthy cells had been analysed. To translate these findings to further understand disease, data from cells in an inflammatory environment was considered beneficial. Data from inflammatory macrophages was of interest in the exploration of possible loop function. Mapping of epigenetic marks in macrophages, allowed direct comparison

between a non-inflammatory and inflammatory environment. Like PBMCs, the histone enrichment at loop sites suggests the region surrounding the *CXCL13* locus is more inhibitory, and the region around *IL23* and *IFNAR1* may be more likely to enhance transcription (Figure 4.14). However, there was only 1 H3K237ac protein identified at the *CXCL13* loop site, therefore it is difficult to be conclusive. Multiple experiments identified a peak of H3K36me3 at the *IL23* loop site, indicating potential for transcription. Regulation of transcription in macrophages can be largely attributed to enhancers[342]. The H3K4me3^{hi}H3K27ac^{hi} status of site a in *IFNAR1* indicates enhancer presence. An enhancer at this site would result in the downstream increase of inflammatory cytokines[343]. There was no methylation data extracted at the *IL23* site, while *CXCL13* loop site was more enriched for hypermethylation, than *IFNAR1*. In this case, the histone data and methylation data are partially aligned with each other in terms of functional implications. Once more, further data on other inflammatory cell types would provide more insight.

In summation, the high-throughput mapping of histone marks, methylation and DNA hypersensitive sites consistently suggests that NR loops are more likely to have an inhibitory function, while R loops are more likely to enhance transcription. This is consistent with the eQTL presence at only R loops, which show where single nucleotide polymorphisms (SNPs) are likely to have a functional impact. It is important to note that variation in results and the variation between samples in each cell type may not be due to biological reasons, but down to differing experimental protocols. As mentioned previously, ChIP-Seq experiments are easily adaptable and vary in sequencing depth. As such, a peak measured in one region in one experiment and not another, may be due to sequencing depth[344]. Moreover, variability with methylation quantification can arise from the incomplete bisulphite conversion resulting in a hyper methylated region being interpreted as a hypo methylated region[345].

Further to the analysis described, interpretation of the data could be enhanced by quantifying the signal of each peak, mapping the coverage of each epigenetic mark, quantifying the distance to nearest to TSS and finding more TF data. Previous research has identified that signal quantification is more applicable to transcription factor binding as there is a large peak over a small range, as

opposed to a histone modification which may span a large region, as this data has shown (Figure 4.10)[346]. HACER, as mentioned above, would be an appropriate tool to explore this concept further. HACER could facilitate association analysis with functional SNPs and TF binding sites. Moreover, going forward it would be useful to look at B cells and DCs. Work by Karlic *et al* demonstrated that gene expression predictions based on the predictions of certain histone modifications could be translated successfully from one cell type to another[347].

Although data from inflammatory macrophages may be more easily applied to RA, the data was still from healthy donors. Data from RA patients would be extremely valuable to understand the epigenetic landscape in RA and observe if there were similarities or differences with the data gathered from healthy datasets. Thus, data from a previous experiment carried out in the lab on $CD14^{+}$ monocytes was used. It was found that H3K4me3 peaks had similar profiles in both RA and healthy samples (Figure 4.15). Across *IL17A* and *CXCL13* loci, there were few H3K4me3 peaks, in comparison to the *IL21R*, *IL23* and *IFNAR1* loci. Whilst this data did not provide much insight into the influence of RA on this histone profile, the data did align with other findings discussed earlier in this chapter (Figure 4.6, 4.9, 4.12) and suggest at some loci where loops are associated with R (*IL21R* and *IFNAR1*), there is an environment that could enhance gene expression activity. A recent study demonstrated that increased *IFNAR1* expression could be associated with patients less likely to respond to TNFα inhibitor treatment[348]. Most of these patients were also being treated with MTX, so it is interesting to consider the influence of increased *IFNAR1* expression and contribution to treatment response from MTX. As *IL21R* is not expressed on classical monocytes, the result of increased expression in RA in this cell type is less understood. However, increased expression in monocyte derived macrophages and Th1 cells, is known to contribute to osteoclastogenesis and cytokine production, respectively, which ultimately contributes to RA disease progression if not controlled[349,350]. The caveat to this section of the study is that only the H3K4me3 histone modification was explored, and therefore, it is very likely there are other epigenetic marks at these sites in RA patients would provide more insight into the pathogenic consequences of loop formation in R and NR.

One of the significant limitations of this chapter is the varied availability of data for each epigenetic mark and cell type. As mentioned above, data from DCs, B cells, and regulatory T cells, from both health and disease would of benefit to provide a more informed picture on RA pathogenesis implications. Not all cell types had data available for every epigenetic mark of interest. Moreover, most data came from healthy samples, therefore the implications of disease have yet to be fully understood. There was also differing number of samples and experiments between each locus and epigenetic mark, which limited statistical analysis capabilities. Other datasets that would strengthen interpretation of functional implication of loop formation may be those that identify the function of SNPs which co-localise with loop anchor points and epigenetic marks. Further work *in-silico* to gain more of an understanding may lead to useful *in-vitro* studies.

4.4 Conclusion

The work in this chapter builds on the findings of the MTX CCS and has shed light on the possible functional implications of loop formation in RA patients pre-MTX treatment. As with the previous chapter, this work demonstrates the complexity of investigating the 3D epigenetic environment in disease. The breadth of publicly available data analysed in this work suggests that R and NR loops are present in contrasting epigenetic environments, that may contribute the ability to respond to MTX. Data from RA patients implied that the findings from healthy data collected could be applicable in disease, however this is likely to change with analysis of more RA datasets.

In future studies, the work in this chapter should be extended to further publicly available datasets, or the generation of new bespoke data, to assess the epigenetic environment in other disease relevant cell types such as B cells and DCs, as well as RA T cells and monocytes. Ultimately this work should be translated *in-vitro* to validate the findings found *in-silico*. Overall, this work has shown the potential to gather a more informed picture of CCS loops and their functional potential, which may ultimately provide insight into disease pathogenesis and MTX response.

Chapter 5 Identification of Early RA Epigenetic Endotypes

5.1 Introduction

The work in previous chapters explored the stratification potential of the methotrexate (MTX) chromatin conformation signature (CCS) in more detail and attempted to understand the functional consequences of loop formation and its relation to disease pathogenesis. Whilst exploring stratification around MTX response is vital, further baseline stratification signatures would be beneficial and may provide useful insight into the 3D architecture underpinning different responses to rheumatoid arthritis (RA) therapy.

There has been a regulatory role suggested for genome-wide association study (GWAS) loci, which has been strengthened by the observation that single nucleotide polymorphisms (SNPs) found through GWAS are abundant at DNA variants associated with gene expression alterations[209]. These are known as expression quantitative trait loci (eQTL) and have been found in the loci of people with autoimmune diseases, including RA. These eQTLs could provide the link between suggested causal SNPs and the consequent abnormal gene expression which can lead to disease.

By most, RA is now considered a heterogeneous syndrome, based on molecular and clinical endotypes[351]. With the development of new technologies, identifying endotypes in RA has attracted growing research interest. Various methods have been used to define RA endotypes, such as identification of serum biomarkers[351], as well as RNA-seq of blood and synovial tissue from RA patients[352]. Given that a prior proof-of-principle study in leukaemia patients provided evidence that chromosome conformation could classify leukaemia subtypes[294,353], we hypothesised that using 3D chromosomal conformation could be a way to define endotypes in RA.

For the work in this chapter, custom microarrays were designed based on findings by Walsh *et al* [209], with the aim of capturing the differences in the 3D epigenetic environment underpinning different response states in RA. This chapter explores the process of identifying appropriate longitudinal early RA

clinical samples for use on the custom microarrays and determining the most informative analysis methods.

To achieve this the aims were:

1) Identify patients from Scottish Early Rheumatoid Arthritis (SERA) cohort with varying 12-month response trajectories

2) Use samples on custom microarrays to identify stratifying EpiSwitch™ loops between groups

3) Statistically refine microarray data to find informative stratifying loops and use these to shed light on underlying pathogenesis between endotypes

4) Identify potential candidates for a new CCS that could predict response trajectory/endotype at baseline

5.2 Results

5.2.1 Distinct trajectories of early RA exist in SERA cohort

In addition to identifying prediction signatures for response to specific therapies, being able to identify if someone will follow a certain response trajectory is valuable. The SERA cohort was interrogated to identify the 12-month response trajectories of early RA patients (Figure 5.1). Three main trajectories were observed over this time-period. The first of these are the 'responders' (R) (Figure 5.1A). These are patients who achieve clinical remission or low disease activity (LDA), CDAI <2.8, by 6 months and maintain this state by 12 months. The second common group are the 'non-responders' (NR). These are patients who are do not reach LDA (CDAI 2.8>10) or remission, regardless of therapy (Figure 5.1B). The third group identified in this cohort are the 'initial responders' (IR) (Figure 8.1C). These are patients who achieve remission by 6 months, however by 12 months, these patients have increased disease activity that varies from low to high (HDA). By examining the epigenome from patients from these three groups, we hypothesised that there may be a differing 3D chromatin profile, which could allude to differing underlying pathogenesis. To do this, the 'extremes' from the 3 trajectories were chosen (Figure 5.1D-F), and the demographics of these patients were assessed to determine if there was indication of which trajectory a patient would follow (Table 5.1). Of the chosen 'extreme' patients, the R group were made of patients who all had HDA at baseline, reached remission and remained there for the period of observation. Those in the NR group all had HDA over the 12-month period. In the IR group, 4 patients had HDA at baseline, of which 2 achieved LDA and 2 achieved remission by 6 months. The 2 patients in remission had moderate disease activity by 12 months and the 2 with LDA at 6 months, had HDA at 12 months. The other 2 patients in this group began with moderate CDAI at baseline, reached low CDAI at 6 months and returned to HDA at 12 months.

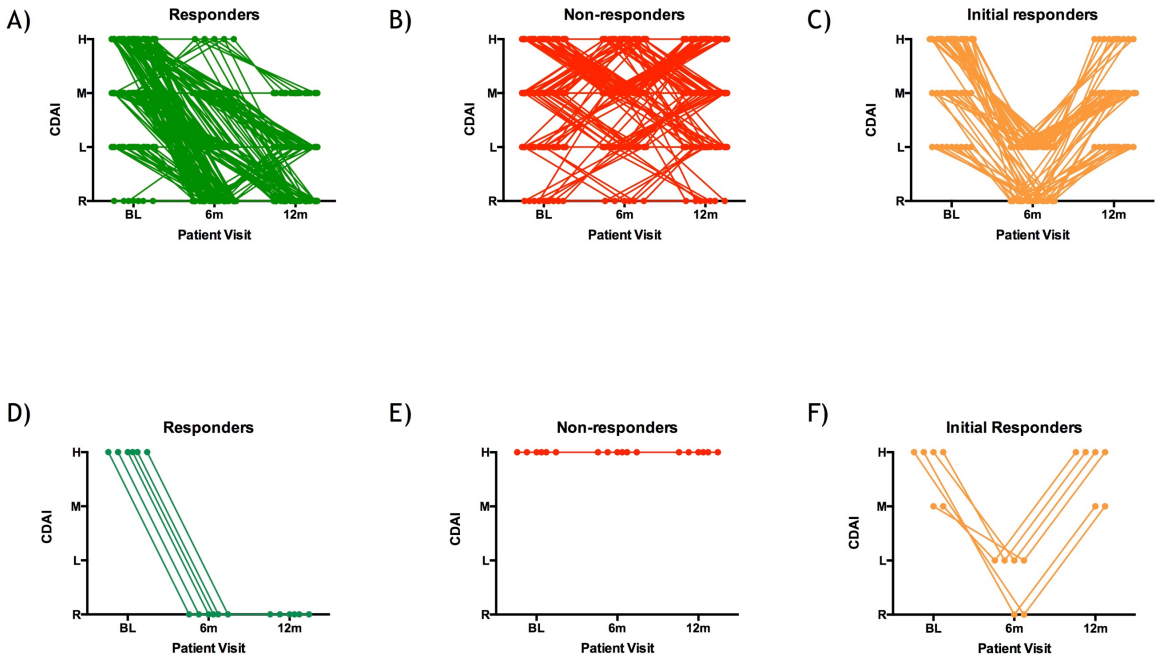


Figure 5.1 Identification of Different Response Trajectories in SERA cohort
Trajectory of patient response over 12 months. CDAI defined as (TJC28/10)+(SJC28/10)+ patient global assessment + physician global assessment. A) Responders. B) Non-responders. C) Initial responders. D-F) Trajectory of patients selected for use on arrays (N=6 for each group, each time point =54 samples in total). DAS28 scores were also used to confirm disease activity; data not shown
BL, baseline; CDAI, clinical disease activity index; DAS, disease activity score; N, number of samples; TJC, tender joint count; SJC, swollen joint count

Most baseline characteristics of chosen SERA patients were similar between the 3 groups, with the biggest difference between groups observed with clinical disease activity scores (Table 5.1). NR and IR had very comparable average ages (54.6 and 57.6, respectively), with R having the highest average age (61.3). Sex, race and BMI were very comparable between groups. NR and IR had the same percentage of rheumatoid factor (RF) positive patients; and R and IR had the same percentage of anti-citrullinated protein antibodies (ACPA) positive patients. Overall, there was no baseline demographic, serum protein or disease activity score that could successfully predict what trajectory someone would follow from baseline. The baseline characteristics between groups were tested for statistical significant differences. DAS28 CRP was the only characteristic to show statistically significant differences between the 3 responder groups. As there was no statistically significant differences between the groups for CDAI score, this meant that the groups were comparable in terms of baseline disease activity. This warranted the investigation via custom microarrays to identity the 3D epigenome in all 18 patients.

Table 5.1 Baseline Characteristics of RA Patients used for Custom Microarray
ACPA, anti-citrullinated protein antibodies; CDAI, clinical disease activity index; CRP, C-reactive protein; DAS, disease acidity index; RF, rheumatoid factor

	Responders	Non-responders	Initial responders
Number	6	6	6
Age Years Mean (SD)	61.3 (11.7)	54.6 (17.3)	57.6 (15.2)
Sex (% female)	66.6	50	66.6
Alcohol excess (%)			
Males	1 (50)	0	1 (20)
Females	1 (25)	0	0
BMI (SD)	31.5 (1.5)	29.98 (3.1)	31.69 (7.3)
Current smoker (%)	0	0	16
Race (% white)	100	100	100
RF+ (%)	50	66.6	66.6
ACPA+ (%)	66.6	50	66.6
CDAI Mean (SD)	36.8 (15.8)	44.7 (8)	27.9 (10.9)
DAS28 CRP Mean (SD)	4.8 (0.7)	6.15 (0.6)	4.4 (1.2)

There was an effort to select patients for the microarray that had the same treatment regimens. However, the priority was the availability of patients with 'extreme' trajectories that had samples from all three time points available from the SERA biobank. As such, chosen patients had a variety treatments over the 12-month trajectory (Figure 5.2). All patients were treated with at least 1 conventional synthetic disease modifying anti rheumatic drug (csDMARD) over the 12 months and only 1 patient did not start treatment on MTX. 4 of 6 R were treated with monotherapy MTX for the year, with 1 other on combination MTX and HCQ therapy and the other combination MTX and SSZ. Four NR were also only treated with monotherapy MTX over 12 months. One patient was treated with triple csDMARD therapy. The other NR was treated with MTX for 3 months before switching to SSZ monotherapy, then SSZ and HCQ in combination. Most IR were treated with a combination of csDMARD therapy over the 12-month period.

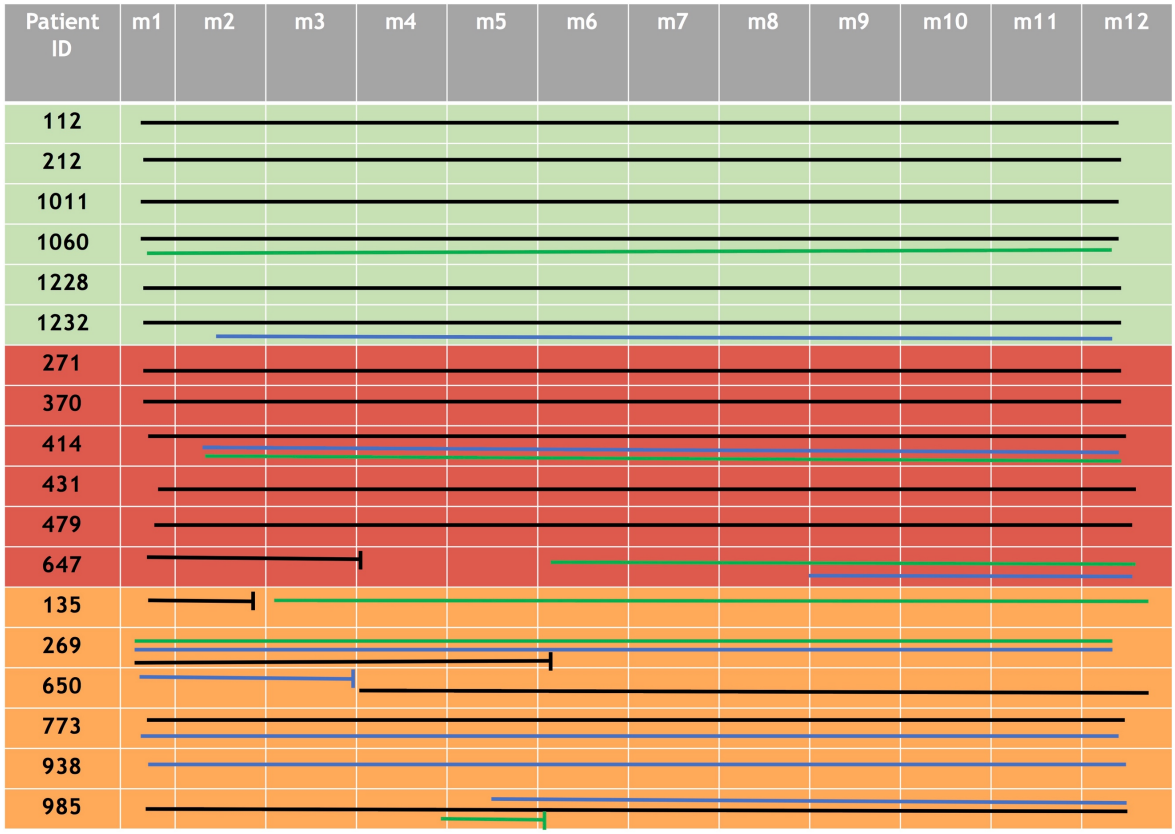


Figure 5.2 Treatment Timeline of Chosen RA Patients for Microarray Patient Treatment Timeline
Treatment trajectories shown for patients selected for endotype microarray analysis. Patient treatment information from SERA cohort. Trajectory represents csDMARD therapy over 12 months from baseline. R shown in green, NR shown in red, IR shown in orange. Black arrows represent MTX, blue for HCQ and green for SSZ.
csDMARDs, conventional synthetic disease modifying anti-rheumatic drug; HCQ, hydroxychloroquine; NR, non- responder; m, month; MTX methotrexate; R, responder SERA, Scottish Early rheumatoid arthritis cohort; SSZ, sulphasalazine

5.2.2 Microarray quality control

The custom microarrays were designed by Oxford BioDynamics Plc (OBD), based on results from the study by Walsh *et al*. In brief, EpiSwitch™ loops in proximity to eQTLs identified by Walsh *et al*[209], which were biologically relevant for RA, were chosen for the array (Appendix). After clinical samples were chosen, they were subsequently sent to the facilities at OBD to be run on the microarrays. The first step in the microarray analysis process was to measure the quality of the data. Firstly, this involved observing the red / green dye distribution of the data. Preliminary analysis before all batches of the array had been run, revealed that 1 array was an outlier, as the dye distribution was not in-line with the other arrays (data not shown). This resulted in that array being included in the 4th batch where the error was rectified, and all densities were uniform (Figure 5.3A). Loess normalisation was successful, as shown in Figure 5.3B. Another

quality control (QC) measure was to plot slide number on a PCA plot. To reduce batch effects, samples of different time points and different response trajectories were spread across slides, therefore slides would not be expected to cluster on the PCA. Samples from slide 14 are highlighted as a representative of this distribution which shows this QC measure was successful (Figure 5.3C).

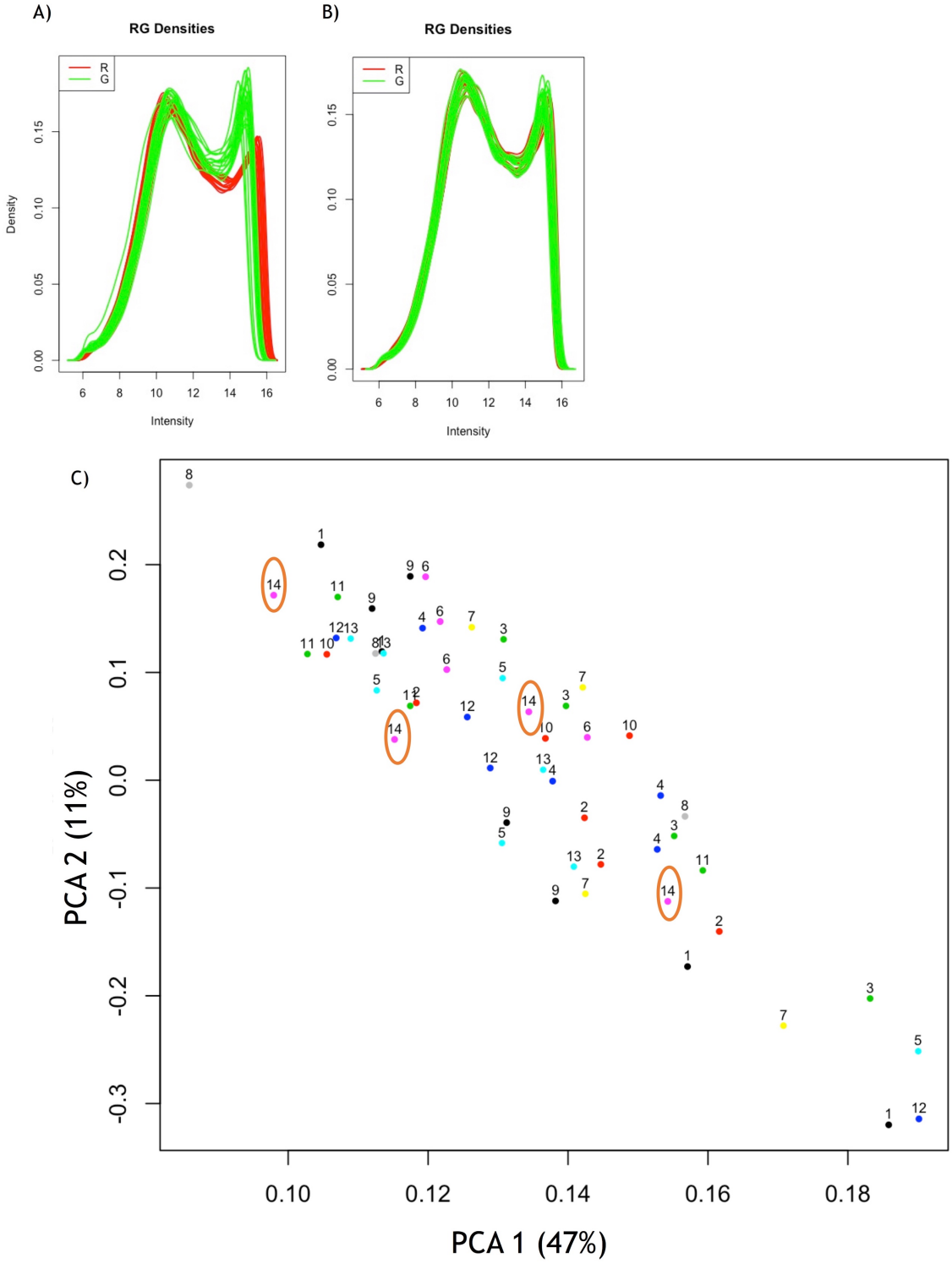


Figure 5.3 Quality Control Assessment of Microarray Data
Series of plots demonstrating quality of raw microarray data and the influence of

normalisation. 4 dual-colour arrays in total, 4 slides per array, 4 samples per slide. Each slide had samples of different responder types and time points. A) Red-green density histogram before normalisation. B) Red-green density histogram after within array locally weighted polynomial regression (Loess) normalisation. C) PCA plot with numbers of slides labelled, and samples from slide 14 circled in orange. Plots created using Limma package on R studio.

Based on the success of the quality control, it was deemed appropriate to continue with further analysis. In the first instance this involved the use of Limma software (Section 2.5.3). Clustering through PCA plotting was performed as a global level analysis of the data (Figure 5.4). At baseline, R appeared to cluster together in PC1, with some more overlap between NR and IR (Figure 5.4A). At 6m, NR and IR more closely clustered together and R did not cluster as 1 group, but 2 groups (Figure 5.4B). At the 12-month time point all groups were clustered together, with only 3 IR samples shown in proximity to each other (Figure 5.4C).

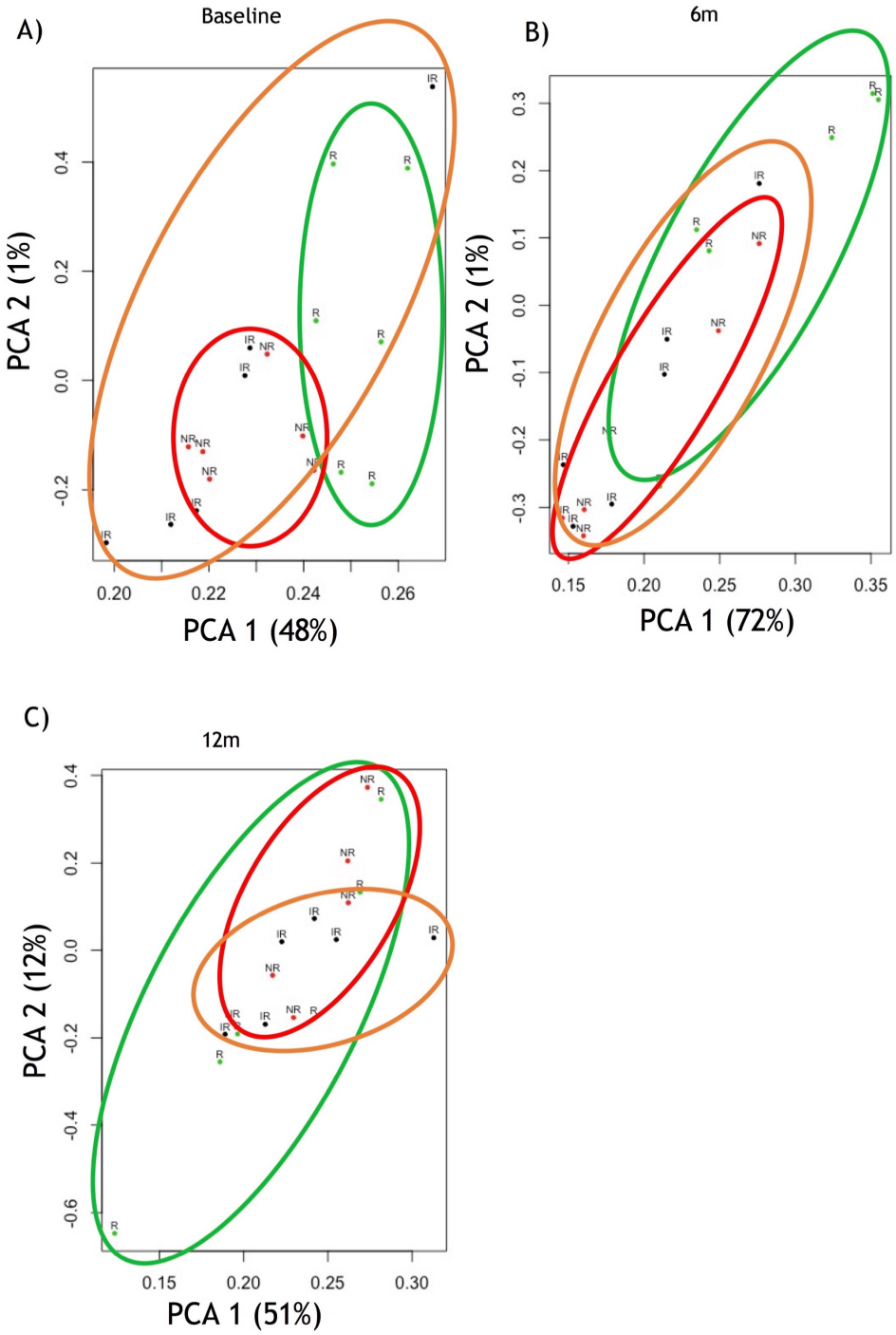


Figure 5.4 Global Analysis of Microarray Data
Pre-linear model analysis of microarray data. 4 dual-colour arrays in total, 4 slides per array, 4 samples per slide. N=6 for each responder group at each time point. A) PCA labelled by responder type at baseline. B) PCA labelled by responder type at 6m. C) PCA labelled by responder type at 12m. Plots created using Limma package on R studio.

5.2.3 Limma linear modelling

After global level analysis, a linear model within Limma was used to extract informative contrasts between endotype groups, as well as between disease and pooled healthy controls (HC) (Table 5.2). An informative contrast was one that adj.P.Val <=0.05 and abundance score (AS) -1.1<= or >=1.1. More details on this analysis are found in Section 2.5.3. Disease-HC contrasts produced more informative loops than endotype comparisons at all time points. All disease-HC contrasts had over 10,000 informative loops. The maximum number of informative loops was found from the R-HC contrast (23131); R in the R-IR comparison at 12m had the least informative loops (1129).

Table 5.2 Informative loops from Limma contrasts
Data generated from contrasts made in the Limma linear model. Informative loops defined as a statistically significant difference from 0 on a log 2 scale with adj.P.Val ≤0.05 and abundance score (AS) -1.1≤ or ≥1. Positive (+) AS associated with sample on left of the contrast model and negative (-) AS associated with sample on right of the contrast model.
6m, 6 months; 12m, 12 months; HC, healthy control; IR, initial responder; NR, non-responder; R, responder

Comparison	Time point	+	-
R-HC	Baseline	14095	10831
NR-HC	Baseline	12611	11732
IR-HC	Baseline	10608	10057
R-NR	Baseline	7383	4811
IR-NR	Baseline	2744	1602
R-IR	Baseline	5341	5569
R-HC	6m	23131	21894
NR-HC	6m	14481	12253
IR-HC	6m	12034	11233
R-NR	6m	11238	13708
IR-NR	6m	1788	1392
R-IR	6m	9182	9255
R-HC	12m	12022	12506
NR-HC	12m	15146	13532
IR-HC	12m	13880	12184
R-NR	12m	2001	2796
IR-NR	12m	2614	3200
R-IR	12m	1129	2081

Reflecting on the data from the Limma contrasts, I debated whether the model was extracting many more ‘meaningful’ loops than expected. This led me to consider that the model may not be stringent enough to find true stratifying, and disease-informing loops. With some comparisons extracting over 20,000 ‘informative’ loops, approximately 10% of the total loops captured on the array, it was decided that a more stringent method of identifying stratifying loops was needed. The hope with a new model is that we would find true biologically meaningful results, and it would also reduce the number of loops to take forward for further analysis.

5.2.4 RankProd analysis of microarray data - Responders

To overcome the possible issues of statistical filtering with the linear parametric testing using Limma, an alternative analysis method was chosen. Namely, RankProd 2.0, a Bioconductor package used to find differentially expressed molecular profiles based on two non-parametric statistics (rank product and rank sum). This method has been widely used to detect variables consistently upregulated (or downregulated) in replicate experiments and developed with gene expression microarrays in mind. As such, RankProd 2.0 was chosen as the analysis method going forward.

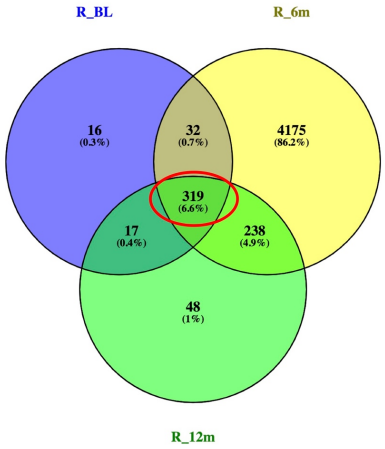
The RankProd approach produced many significant loops, but considerably less than the Limma method, with a maximum number of statistically significant loops of 4765 in the R-HC contrast at 6m (Table 5.3). Disease-HC contrasts were most significant at 6m for R and NR groups, and at 12m for IR.

Table 5.3 Informative Loops from RankProd Contrasts
Analysis conducted under supervision of Dr Ewan Hunter and Christina Koutsothanasi (OBD). Significant loops defined as loops adj P value ≤0.05 and AS -1.1≤or ≥1.1 For each patient group at each time point, N=6, pooled HC, N=20
6m, 6 months; 12m, 12 months; HC, healthy control; IR, initial responder; NR, non-responder; R, responder

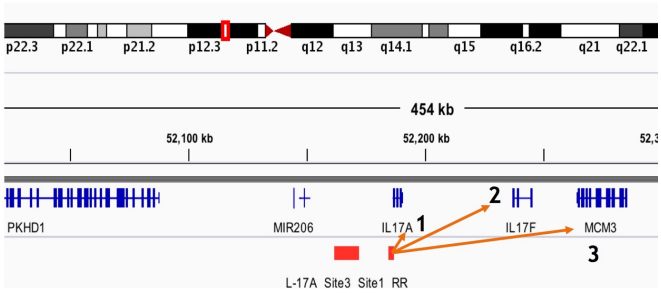
Comparison	Time point	Significant disease loops
R-HC	Baseline	384
NR-HC	Baseline	739
IR-HC	Baseline	300
R-HC	6m	4765
NR-HC	6m	1791
IR-HC	6m	1513
R-HC	12m	621
NR-HC	12m	967
IR-HC	12m	1566

The significant loops that had the potential to stratify disease and pooled HC samples were taken for further analysis. Initially, a Venn diagram was generated from the significant loops at baseline, 6m and 12m in R (Figure 5.5A). This allowed the visualisation of stratifying loops common between pairs of time points, and importantly loops that were common to all time points. 319 significant loops were common to all time points, and could be considered the 'stable' loops. 6m had the highest number of time-point unique loops (4175), followed by 12m (48) and baseline (16). Using analysis software on the EpiSwitch™ data portal, a new interactive interface (https://episwitch3dgenomicsportal.com) to interrogate EpiSwitch™ analysis data, the closest 3 genes to the 319 'stable' loops were identified (Figure 5.5B)(Section 2.5.10). This list of genes was then entered into the Gene Analytics platform to understand the most significant pathways enriched based on these genes (Figure 5.5C). All pathways had medium score matches for the genes in each pathway which indicated a corrected P-value of 0.05 to 1. The 'phagosome' pathway was the most significant of this group, with 16 genes matching this pathway. Gene Ontology (GO) terms were also explored (Figure 5.5D). In contrast to the pathways, all GO terms had a high score match for the genes associated with each term, suggesting better ontologies with defined genes, in addition to a corrected P value of <0.05. The 'ER to golgi transport vesical membrane' had the highest match score in this group. This may suggest that the effective transport of intracellular proteins is important in responding to RA therapy.

A)



B)



C)

Score	SuperPath Name	Matched genes
13.16	Phagosome	16
12.02	Translocation of ZAP-70 to Immunological Synapse	8
11.22	TCR Signaling (REACTOME)	13
10.48	Pentose Phosphate Pathway	6
9.55	G-protein Signaling N-RAS Regulation Pathway	8
9.35	Toxoplasmosis	13
9.08	Tuberculosis	15
8.70	ICos-ICosL Pathway in T-Helper Cell	12
8.34	Nitrogen Metabolism	4
7.96	Complement and Coagulation Cascades	9
7.57	Interferon Gamma Signaling	15
6.56	MHC Class II Antigen Presentation	9
6.50	CD28 Co-stimulation	8
6.30	CTLA4 Signaling	7
6.24	Lysosomal Oligosaccharide Catabolism	2
5.89	Rheumatoid Arthritis	8
5.74	Pentose Phosphate Pathway (Erythrocyte), Pharmacodynamics	2
5.65	Staphylococcus Aureus Infection	8
5.51	Fanconi Anemia Pathway	6
5.42	Hematopoietic Cell Lineage	8

D)

Score	Name	Matched genes
15.40	T Cell Receptor Signaling Pathway	19
13.78	Antigen Processing and Presentation of Peptide or Polysaccharide Antigen Via MHC Class II	5
13.59	Microtubule Cytoskeleton Organization	16
16.28	ER to Golgi Transport Vesicle Membrane	11
14.71	Trans-Golgi Network Membrane	13
14.40	MHC Class II Protein Complex	6
13.88	Transport Vesicle Membrane	9
15.01	MHC Class II Receptor Activity	5

Figure 5.5 Pathway Enrichment of Significant Stable Loops in R
List of loops generated from RankProd analysis. A) Venny 2.1 used to generate Venn diagram of significant loops in the R group at all time points, significant ≤0.05 and AS -1.1≤ or ≥1.1. B) Schematic representing how EpiSwitch™ data portal captures 3 closest genes to an anchor site, *IL17A* locus used as a representative image. Gene Analytics then used to generate list of significant pathways based on matched genes. Scores are given to each pathway to reflect their matching quality to the set of genes entered to the analysis platform. An algorithm is used to determine the threshold for high, medium

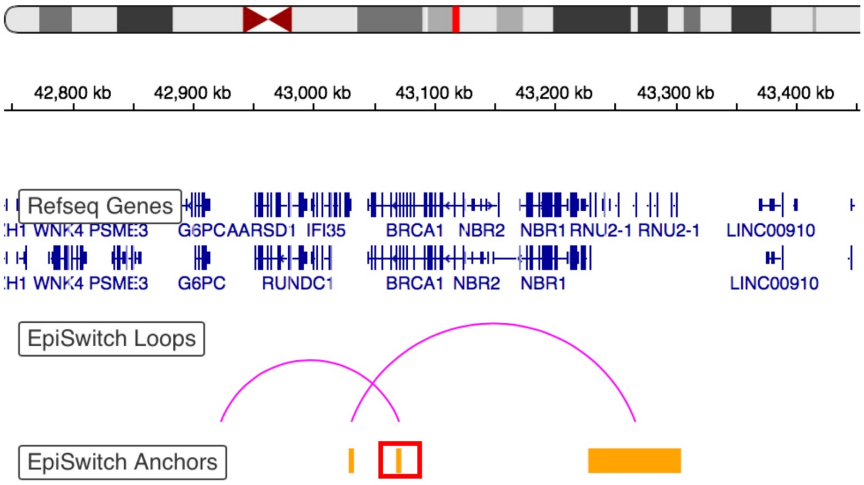
and low scores in each dataset. Cells in green = high score match, corrected P-value of <0.05, cells in orange = med score match, corrected P-value of 0.05-≥1. Tables generated for C) Top pathways, D) Top GO terms.
6m, 6 months; 12m, 12 months; BL, baseline, kb, kilobase;
RR, reverse-reverse orientation

The gene list used for Gene Analytics analysis was then used to make a protein network to understand the relationship between the genes found in the region of stable loops in the R group (Figure 5.6). The online STRING platform (https://string-db.org/) was used to generate this network (Appendix). which could subsequently be transferred to a programme called Cytoscape for further analysis (Section 2.5.7, 2.5.8). Network analysis tools (topology statistics) were then employed on Cytoscape to find the most connected genes, represented by network 'nodes' and 'edges'. This was carried out with a view of considering the more connected nodes, the most contributing loci in the pathways associated with R. The top 9 nodes were noted (Figure 5.6A). The most connected node for the R network, with 23 connections to other genes from the whole Gene Analytics list, was *BRCA1*. Finally, the EpiSwitch™ data portal was used to visualise the genomic environment around this locus, an approach shown in Chapter 4 (Figure 5.6B). Two loops could be visualised in this region; the anchor point of the loop with statistical significance in R is highlighted. For closer characterisation, specifically to visualise the other epigenetic marks that surround this locus, a circos plot was generated (Figure 5.6C). The ability to generate circos plot within the EpiSwitch™ data portal was extremely useful, as this platform contained data from other experiments which investigated histone modifications in immune cells. The anchor point that resides within the *BRCA1* gene overlapped with H3K27ac, (found in CD8, CD4 cells) as well as multiple clinically important SNPs. The other anchor point that is part of this loop lies within *G6PC* locus. At this site, there is also the presence of H3K27ac from CD8 and CD19 cells. These epigenetic marks illustrate an environment that may enhance transcription of the genes in their proximity.

A)

Gene	Number of connections
BRCA1	23
PPP2CA	19
WDR12	16
NFKB1	16
ANAPC4	15
PSMC6	13
NSA2	13
FBXO11	13
USO1	13

B)



C)

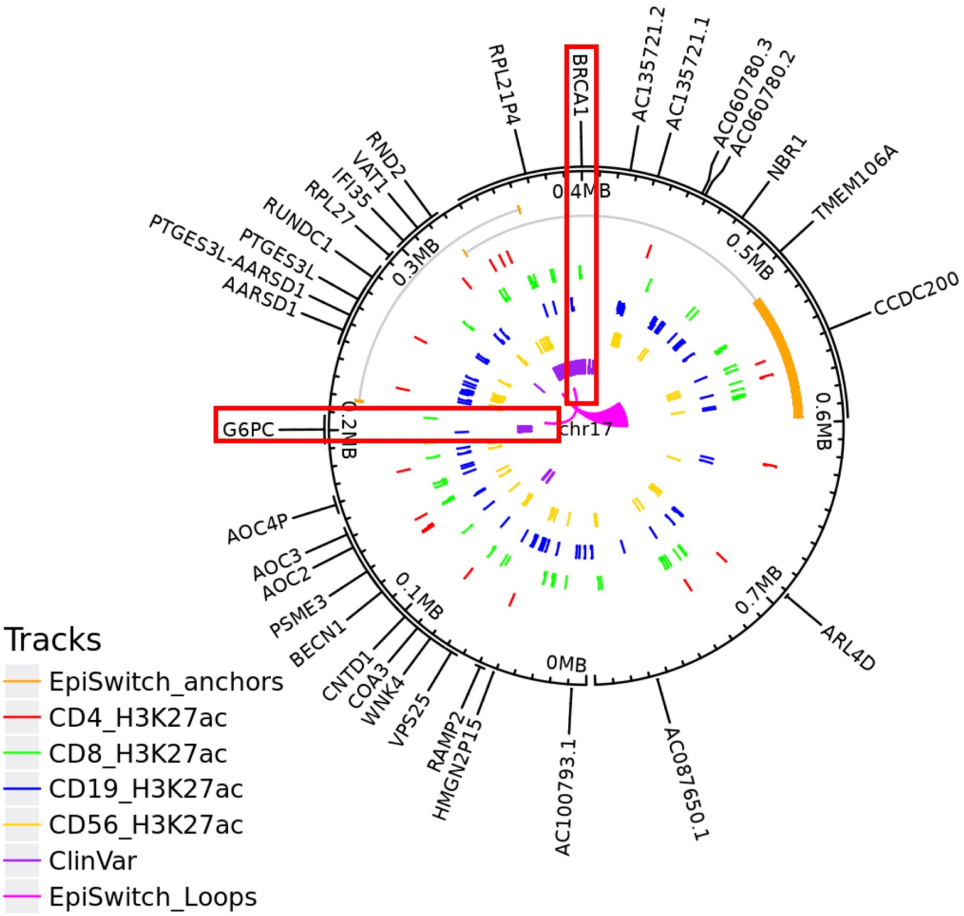
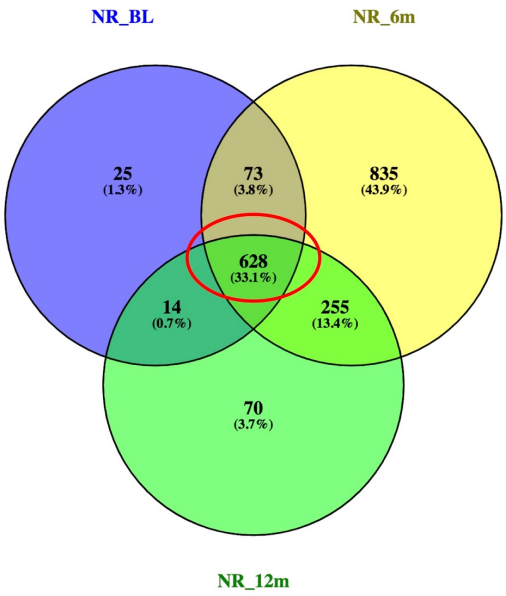


Figure 5.6 Environment Surrounding Central Enriched R Loop
Network analysis used to understand central players in significant pathways. A) Top most connected nodes determined by most directed edges of a node in Cytoscape. Gene with most connections highlighted in red. B) Genome browser view from EpiSwitch™ data portal showing *BRCA1* gene genomic environment. Genes in dark blue, loops in pink and EpiSwitch™ anchor points in orange. Red boxes illustrate anchor points statistically significant loops in this group. 1 anchor point present behind 'EpiSwitch™ Anchors' label. C) Circos plot with the addition of ClinVar representing disease associated SNPs, and H3K27ac marks from publicly available datasets. Red box indicates gene of interest and gene which other loop anchor point lies within and where anchor points of interest interact with other epigenetic features.

5.2.5 Rank Prod analysis of microarray data – Non-Responders

The same approach used for significant R loops was used for NR loops (Figure 5.7). 625 loops significantly differentiated disease and HC at all 3 time points (Figure 5.7A). 825, 70 and 25 loops significantly differentiated between disease and HC at 6m, 12m and baseline, respectively. The 625 'stable' loop list was entered into GeneAnalytics software. All pathways had a med match gene score (Figure 5.7B). The top scoring pathway was the 'Phagosome' pathway. There were 2 high score matching GO terms, namely 'Interferon-gamma-mediated signalling pathway' and 'ER to golgi transport vesical membrane' with 13 and 14 matched genes, respectively (Figure 5.7C). This data suggests regulation of interferon gamma signalling may be important to NR, and the enrichment of genes in the 'ER to golgi transport vesical membrane' pathways may suggest that it is not a R specific pathway, but more important in RA as a whole.

A)



B)

Score	SuperPath Name	Matched genes
11.71	Phagosome	19
8.70	Translocation of ZAP-70 to Immunological Synapse	8
8.47	Rheumatoid Arthritis	12
8.24	Interferon Gamma Signaling	20
7.89	Pentose Phosphate Pathway	6
7.52	Toxoplasmosis	15
7.20	ICos-ICosL Pathway in T-Helper Cell	14
6.78	TCR Signaling (REACTOME)	13
6.66	Tuberculosis	17
6.54	Nitrogen Metabolism	4
6.48	G-protein Signaling N-RAS Regulation Pathway	8
6.37	Immune Response Role of DAP12 Receptors in NK Cells	18
6.16	Metabolic States and Circadian Oscillators	3
6.04	CLEC7A (Dectin-1) Signaling	14
6.02	MHC Class II Antigen Presentation	11
5.77	3-phosphoinositide Degradation	4
5.75	Glucagon Signaling Pathway	11
5.69	Deadenylation-dependent MRNA Decay	10
5.67	Antigen Processing-Cross Presentation	12
5.55	Regulation of Cholesterol Biosynthesis By SREBP (SREBF)	7

C)

Score	Name	Matched genes
13.48	Interferon-gamma-mediated Signaling Pathway	13
18.13	ER to Golgi Transport Vesicle Membrane	14

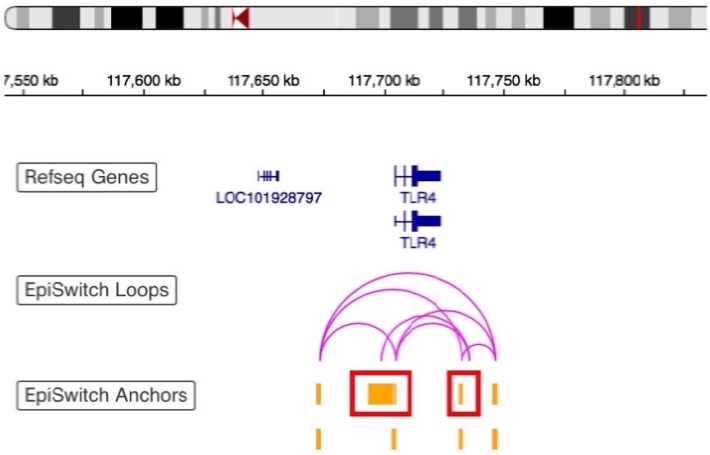
Figure 5.7 Pathway Enrichment of Significant Stable Loops in NR
List of loops generated from RankProd analysis. A) Venny 2.1 used to generate Venn diagram of significant loops at all time points, significant adj. P value ≤0.05 and AS -1.1≤ or ≥1.1. Gene Analytics then used generated list of significant pathways based on matched genes. Scores are given to each pathway to reflect their matching quality to the set of genes entered to the analysis platform. An algorithm is used to determine the threshold for high, medium and low scores in each dataset. Cells in green = high score match, corrected P-value <0.05, cells in orange = med score match. P-value 0.05-≥1. Tables generated for B) Top pathways and C) Top GO terms.
6m, 6 months; 12m, 12 months; BL, baseline, kb, kilobase;
RR, reverse-reverse orientation

Once again, the gene list generated from stable loops was used for network analysis to further characterise the significant pathways differentiating NR and pooled HC (Figure 5.8). The top 10 nodes had at least 25 connections in the network, with the top a total of 34 (Figure 5.8A). The most connected node represented the *TLR4* gene. This was an interesting find based on it's implications in RA pathogenesis. *TLR4* is expressed on a number of immune cells involved in RA pathogenesis, including peripheral monocytes and synovial macrophages. Activation of *TLR4* can lead to down-stream production of interferons, cytokines and chemokines. As such, the EpiSwitch™ data portal was used to visualise the genomic area around this gene. The genome browser view clearly demonstrated this gene was enriched with many EpiSwitch™ anchor points, hinting that it is a highly-regulated region (Figure 5.8B). Of note, not all loops in this region are associated with the NR endotype. The largest central anchor point, connecting several of these loops, was the EpiSwitch™ loop with significance in this NR group. To look at this area in more detail, a circos plot was used (Figure 5.8C). The second anchor point of this loop does not lie within another locus, but does overlap with H3K27ac, recorded in CD19 cells. Visualising in this way showed that the anchor point within the *TLR4* locus can overlap with H3K27ac marks which have been found in CD4, CD8, CD19 and CD56 cells. Interestingly, there are no clinically relevant SNPs residing in this area. The histone marks present at this locus suggest the potential for enhanced transcription, which would be in line with what we know about this locus in RA pathogenesis.

A)

Gene	Number of connections
TLR4	34
PPP2CA	33
ALDH18A1	33
ITGB2	32
HLA-DRB1	29
CCR5	26
NFKB1	26
HLA-DQA1	25
WDR12	25
CLTC	25

B)



C)

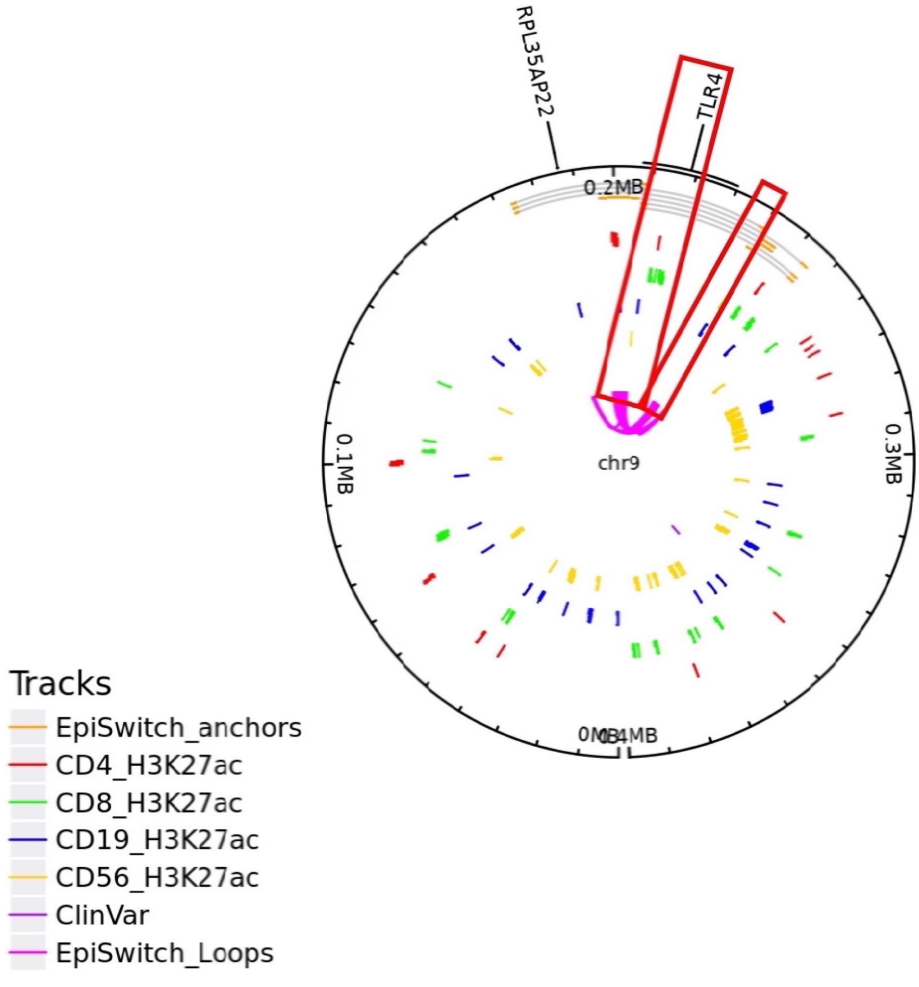
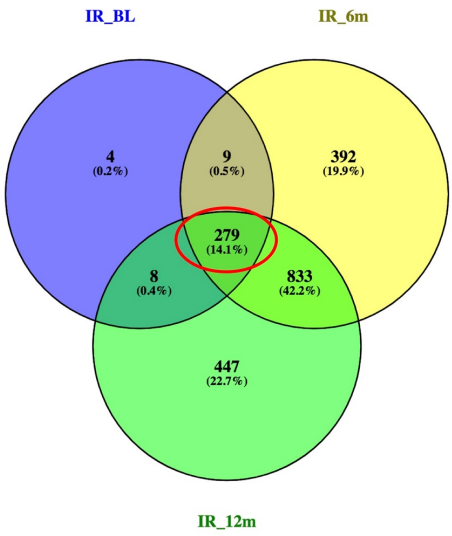


Figure 5.8 Environment Surrounding Central Enriched NR loop
Network analysis used to understand central players in significant pathways. A) Top most connected nodes determined by most directed edges of a node in Cytoscape. Gene with most connections highlighted in red. B) Genome browser view from EpiSwitch™ Data Portal showing TLR4 gene genomic environment. Genes in dark blue, loops in pink and EpiSwitch™ anchor points in orange. Red boxes illustrate anchor points statistically significant loops in this group. C) Circos plot generated from genome browser with the addition of ClinVar representing disease associated SNPs and H3K27ac marks from publicly available datasets. Red box indicates gene of interest and region which other loop anchor point lies within and where anchor points of interest interact with other epigenetic features.

5.2.6 RankProd analysis of microarray data – Initial Responders

As with the other endotype groups, multiple analysis tools were employed to understand the highly-regulated areas the genome unique to the IR disease group (Figure 5.9). 279 loops could stratify disease and pooled HC at all time points (Figure 5.9A). 4, 392 and 447 were unique to baseline, 6m and 12m, respectively. As with the other endotype groups, the closest 3 genes to each of the stable loops were carried forward for gene enrichment analysis. The pathways enriched had a mix of high scoring and med scoring enrichment scores (Figure 5.9B). The high scoring pathways were ‘translocation of ZAP-70 to immunological synapse’, and ‘TCR signalling’. There were 8 GO terms scoring a high match score (Figure 5.9C). The pathway with the most matched genes was the ‘T cell receptor signalling pathway’. Deficient TCR signalling has been shown to contribute to RA pathogenesis and it is interesting that this pathway was enriched in the IR group. Exploring the genomic region of genes in this pathway would be of use to understand this mechanism more.

A)



B)

Score	SuperPath Name	Matched genes
15.10	Translocation of ZAP-70 to Immunological Synapse	8
13.51	TCR Signaling (REACTOME)	12
12.47	G-protein Signaling N-RAS Regulation Pathway	8
11.62	Toxoplasmosis	12
10.67	Phagosome	12
10.31	Interferon Gamma Signaling	14
10.23	Tuberculosis	13
9.97	Nitrogen Metabolism	4
9.43	MHC Class II Antigen Presentation	9
9.13	CD28 Co-stimulation	8
8.66	CTLA4 Signaling	7
8.45	Rheumatoid Arthritis	8
8.17	Staphylococcus Aureus Infection	8
7.19	ICos-ICosL Pathway in T-Helper Cell	9
7.06	Pentose Phosphate Pathway	4
6.13	Cell Adhesion Molecules (CAMs)	9
5.72	Peginterferon Alpha-2a/Peginterferon Alpha-2b Pathway (Hepatocyte), Pharmacodynamics	3
5.72	Regulation of Apoptosis By Parathyroid Hormone-related Protein	3
5.40	HIF Repressors	3
5.39	Th17 Cell Differentiation	9

C)

Score	Name	Matched genes
15.90	Antigen Processing and Presentation of Peptide or Polysaccharide Antigen Via MHC Class II	5
13.79	T Cell Receptor Signaling Pathway	15
20.52	ER to Golgi Transport Vesicle Membrane	11
17.37	Transport Vesicle Membrane	9
16.96	Trans-Golgi Network Membrane	12
16.91	MHC Class II Protein Complex	6
15.23	Clathrin-coated Endocytic Vesicle Membrane	7
17.16	MHC Class II Receptor Activity	5
13.41	Non-membrane Spanning Protein Tyrosine Phosphatase Activity	4

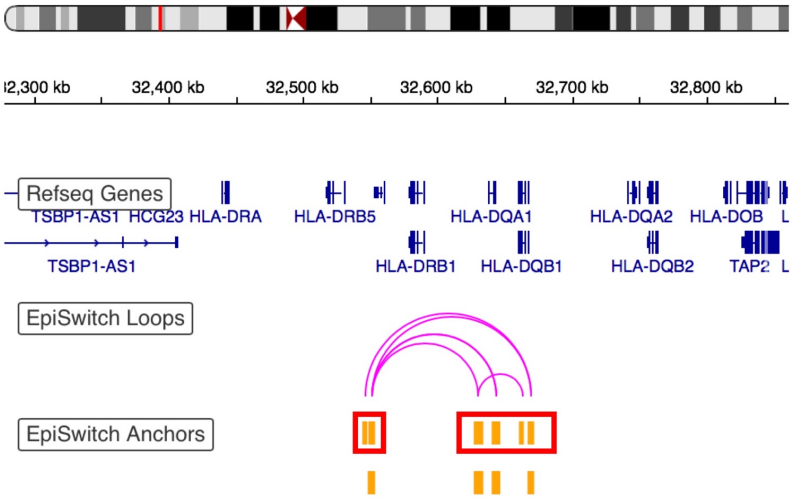
Figure 5.9 Pathway Enrichment of Significant Stable loops in IR
List of loops generated from RankProd analysis. A) Venny 2.1 used to generate Venn diagram of significant loops at all time points, significant adj P value ≤0.05 and AS -1.1≤ or ≥1.1. Gene Analytics then used generated list of significant pathways based on matched genes. Scores are given to each pathway to reflect their matching quality to the set of genes entered to the analysis platform. An algorithm is used to determine the threshold for high, medium and low scores in each dataset. Cells in green = high score match, P-value <0.05 cells in orange = med score match, P-value 0.05-≥1. Tables generated for B) Top pathways and C) Top GO terms.
6m, 6 months; 12m, 12 months; BL, baseline, kb, kilobase;
RR, reverse-reverse orientation;

As with previous analysis approaches, network analysis tools were used to further characterise genes from significant pathways (Figure 5.10). Once again, in Cytoscape, network analysis was used to identify the most connected nodes. The most connected nodes ranged from 14 to 18 connections, with HLA-DRB1 having the most connections (18) (Figure 5.10A). Again, this is a gene with known implications in RA pathology. Specifically, this allele is associated with susceptibility to the development of RA. This gene was explored further using the EpiSwitch^{TM} data portal. Exploration of the surrounding genomic area revealed that 5 loops resided in this region (Figure 5.10B). All loops were statistically significant in the IR group. A circos plot was used again to visualise the epigenomic environment in more depth (Figure 5.10C). The area was clearly enriched with H3K27ac marks which can be found in CD4, CD8, CD19 and CD56 cells. This data illustrates the capacity for enhanced gene transcription at this site. Of note, 4 loops stem from 1 anchor point in this region, which appears to lie within the HLA-DRB5 locus, hinting at the significant regulatory role of this gene. Interestingly, this allele has been shown to be play a protective role in RA. Anchor sites residing within loci with differing functional consequences in RA is interesting, particularly as this is the IR group.

A)

Gene	Number of connections
HLA-DRB1	18
HLA-DQA2	17
HLA-DQA1	16
CLTC	16
HLA-DQB1	15
WDR12	15
HLA-DQB2	15
PPP2CA	15
CSK	14
HLA-DRB5	14

B)



C)

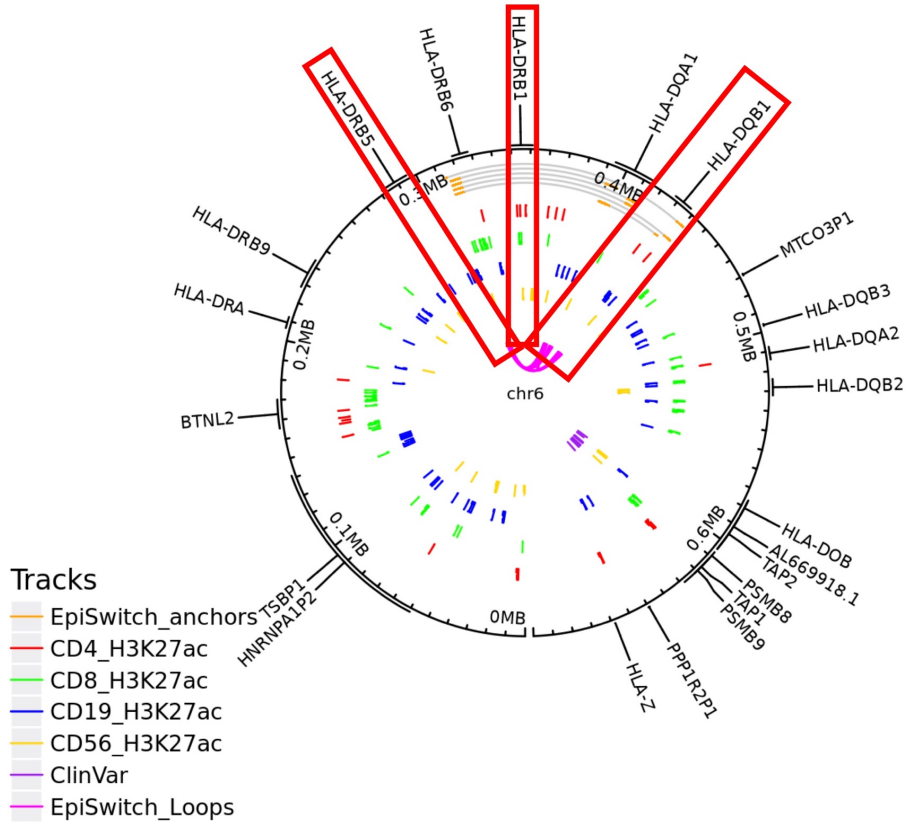
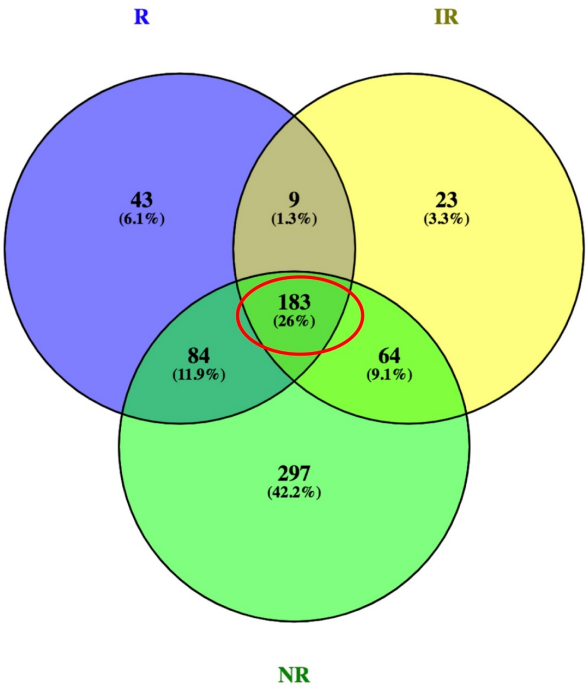


Figure 5.10 Environment Surrounding Central Enriched IR Loop
Network analysis used to understand central players in significant pathways. A) Top most connected nodes determined by most directed edges of a node in Cytoscape. Gene with most connections highlighted in red. B) Genome browser view from EpiSwitch™ Data Portal showing HLA-DRB1 gene genomic environment. Genes in dark blue, loops in pink and EpiSwitch™ anchor points in orange. Red boxes illustrate anchor points statistically significant loops in this group. C) Circos plot generated from genome browser with the addition of ClinVar representing disease associated SNPs and H3K27ac marks from publicly available datasets. Red box indicates gene of interest and gene which other loop anchor point lies within and where anchor points of interest interact with other epigenetic features.

5.2.7 RankProd analysis of microarray data – RA

While it was of interest to understand the pathways enriched in each endotype group, it was also of interest to determine the loops common to all endotypes that could stratify disease and HC at all time points. As such, all the stable loops from each endotype group were used to generate a fourth Venn diagram (Figure 5.11A). One hundred and eighty-three loops were common to all groups, and collectively made an RA-specific loop signature group. Forty-three, 23 and 297 loops were unique to R, IR and NR, respectively. The closest 3 genes to the 183 RA loops were analysed further using GeneAnalytics as before. This produced a set of pathways all with a med gene match score (Figure 5.11B). The most significant enriched pathway was the pathway of the 'regulation of apoptosis by parathyroid hormone related protein'. Once again, GO terms were also interrogated. Seven GO terms in total had high gene match scores (Figure 5.11C). The 'cytosol', 'nucleus' and 'cytoplasm' were GO terms with many matched genes. This alludes to the highly-regulated environment of an RA peripheral blood cell.

A)



B)

Score	SuperPath Name	Matched genes
9.92	Regulation of Apoptosis By Parathyroid Hormone-related Protein	4
9.44	NAD Metabolism, Sirtuins and Aging	3
9.09	Uricosurics Pathway, Pharmacodynamics	3
8.45	Aryl Hydrocarbon Receptor	5
8.17	Mitotic Metaphase and Anaphase	10
7.99	Integrated Breast Cancer Pathway	9
7.85	Lysosome	8
7.69	Nitrogen Metabolism	3
7.28	Thyroxine (Thyroid Hormone) Production	2
6.88	MET Promotes Cell Motility	7
6.67	Peginterferon Alpha-2a/Peginterferon Alpha-2b Pathway (Hepatocyte), Pharmacodynamics	3
6.59	Tuberculosis	9
6.41	Monoamine GPCRs	4
6.41	Negative Regulation of MAPK Pathway	4
6.11	Osteoclast Differentiation	7
5.88	Assembly of RNA Polymerase-I Initiation Complex	2
5.71	Hepatitis C and Hepatocellular Carcinoma	4
5.69	Herpes Simplex Virus 1 Infection	17
5.50	Pentose Phosphate Pathway	3
5.45	Innate Immune System	55

C)

Score	Name	Matched genes
19.26	Cytosol	146
17.65	Nucleus	176
16.09	Cytoplasm	179
16.19	Sialic Acid Transmembrane Transporter Activity	4
14.79	Non-membrane Spanning Protein Tyrosine Phosphatase Activity	4
14.40	Nucleotide Binding	61
13.41	Phosphatase Activity	11

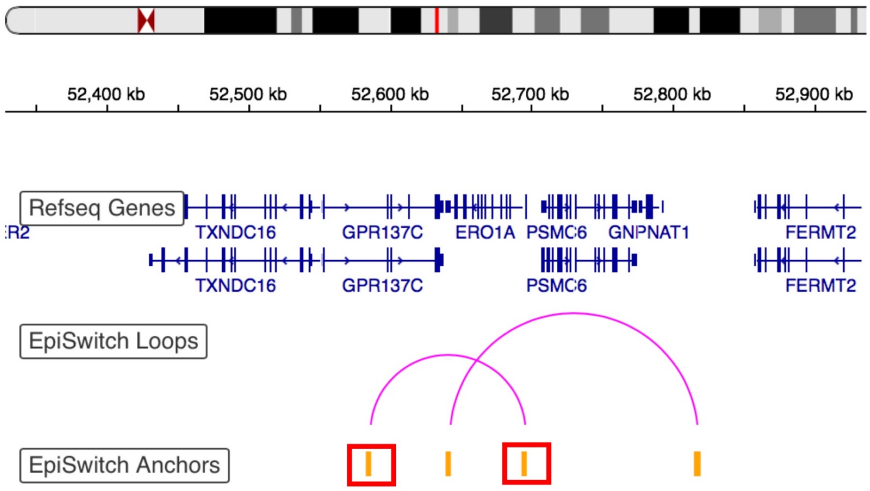
Figure 5.11 Pathway Enrichment of Significant Stable Loops in Disease
A) Venny 2.1 used to generate venn diagram of significant loops at all time points, significant ≤0.05 and AS -1.1≤ or ≥1.1. Gene Analytics then used generated list of significant pathways based on matched genes. Scores are given to each pathway to reflect their matching quality to the set of genes entered to the analysis platform. An algorithm is used to determine the threshold for high, medium and low scores in each dataset. Cells in green = high score match, corrected P-value <0.05 cells in orange = med score match, P-value 0.05 ≤ 1.Tables generated for B) Top pathways and C) Top GO terms.
6m, 6 months; 12m, 12 months; BL, baseline, kb, kilobase;
RR, reverse-reverse orientation;

As before, network analysis was used to further interpret the data (Figure 5.12). Using network tools, the most connected nodes were identified, with the top nodes having 13 or more connections. The most connected node was shown to be PSMC6 with 18 connections (Figure 5.12A). Using the EpiSwitch™ data portal, the genomic environment around the PSMC6 gene was revealed (Figure 5.12B). Two loops were found to be in this region. For increased characterisation, a circos plot was used (Figure 5.12C). H3K27ac marks found in CD4, CD8, CD19 and CD56 cells were found at all anchor points, once again suggesting that this region of the genome may be subject to enhanced transcription in T cells, B cells and NK cells. This enhanced transcription in RA could lead to increased immune pathology and exacerbation in disease. The circos plot also illustrates that the second anchor point of the significant loop lies in proximity to the GRP137C gene, known for cell proliferation, but limited evidence of its implications in RA.

A)

Gene	Number of connections
PSMC6	18
ANAPC1	18
WDR12	18
CCT6A	16
MAD2L1	15
PSMA6	14
NSA2	14
MRPS14	14
POLR1B	13
RARS2	13
SIL1	13

B)



C)

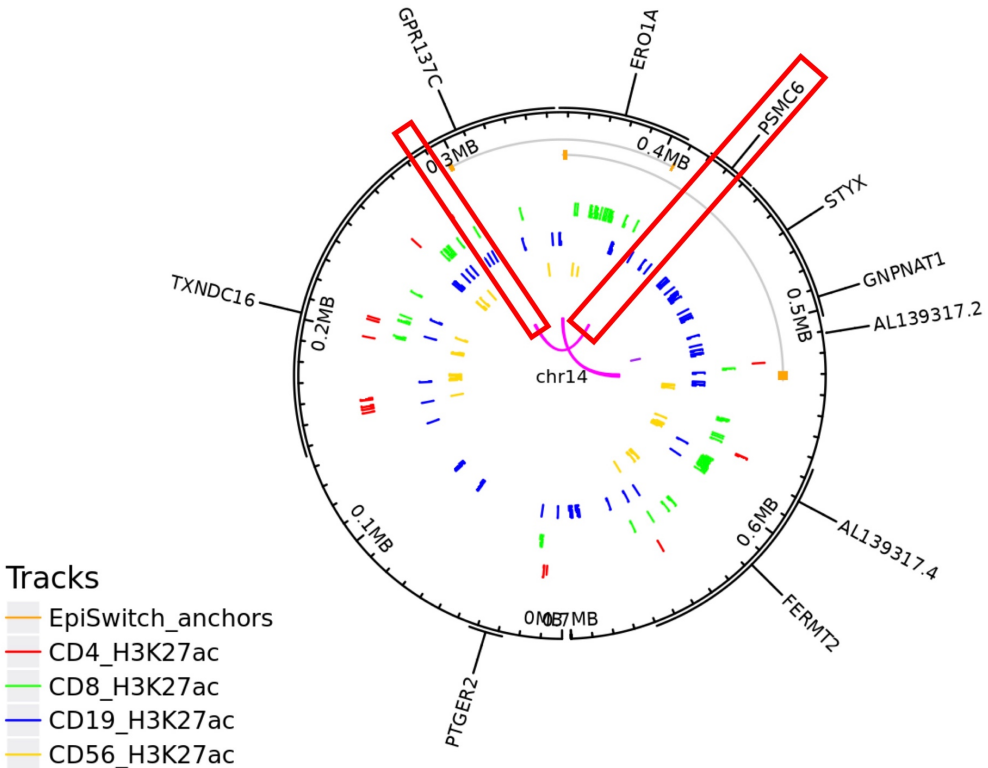


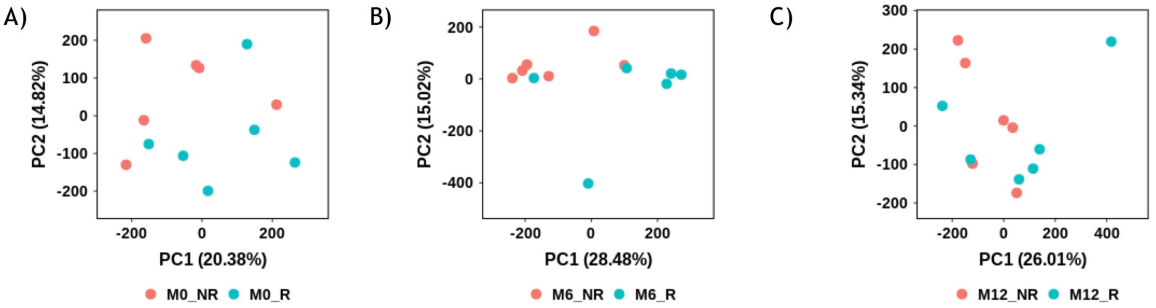
Figure 5.12 Environment Central Enriched Disease Loop
Network analysis used to understand central players in significant pathways. A) Top most connected nodes determined by most directed edges of a node. Gene with most connections highlighted in red. B) Genome browser view from EpiSwitch™ Data Portal showing PSMC6 gene genomic environment. Genes in dark blue, loops in pink and EpiSwitch™ anchor points in orange. Red boxes illustrate anchor points of statistically significant loops in this group. C) Circos plot generated from genome browser with the addition of ClinVar representing disease associated SNPs and H3K27ac marks. Red box indicates gene of interest and where anchor point of interest interacts with other epigenetic features.

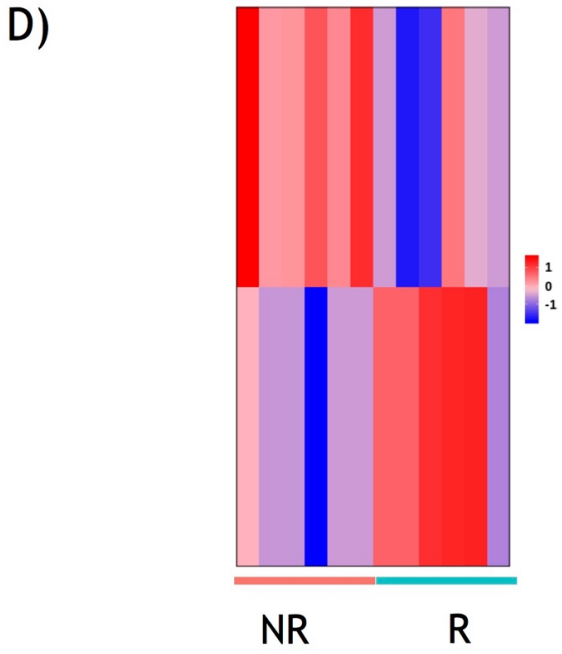
Whilst the network analysis data extracted genes that did appear to heavily be regulated, many of the pathways from GeneAnalytics analysis had limited significance with corrected P-values between 0.05 and 1, suggesting there may be more significant pathways to find. With this in mind, it was of interest to employ a second analysis method, as a way to compare evidence and observe if similar pathways were found in both methods.

5.2.8 SearchLight as second approach to analysis of microarray data

To potentially identify other important loops that could be used to stratify RA patients at baseline, and to possibly strengthen our confidence in the findings from the RankProd method, it was decided that another method could be used to validate and further explore findings. A computational analysis method, Searchlight, primarily used for RNA-seq data, was adapted to analyse our microarray data. While analysis with RankProd focused on contrasts between disease and HC, Searchlight was used to focus on contrasts between pairs of endotype groups. Using linear modelling, differences between endotype groups at all 3 time points were extracted using the contrasts from the model within Limma. A 1 to 1 comparison between pairs of endotypes was done to attempt to understand the degree of difference between each group. This pairwise comparison was done first as a thorough approach to find regulatory differences and hopefully capture the nuances between groups, that may not be captured if a three-way analysis was done in the first instance. This would be carried out later in the analysis pipeline. Firstly, the stratification of NR and R was assessed (Figure 5.13). PCA plots were generated to assess differences between groups at each time point (Figure 5.13A-C). The largest separation between NR and R was seen at baseline (Figure 5.13A). Observing the plots, the least separation seems to appear at 12 months (Figure 5.13C). The reduction in difference between

baseline and 12m may be due to the effect of treatment on individuals. The next step was to evaluate whether there were any loops that were significantly different at the various time points, and observe if any signatures emerged from the data. Notably, there were no significant differentiating loops at 12m, however, at both baseline and 6 months, differences were observed (Figure 5.13D-E). At both the baseline and 6m time points, there were 2 loops which had differential fold change in abundance between NR and R. One loop had higher fold change in NR and the other in R. Both loops at 6m also had differential fold change. At 6m, all 6 individual patients appeared to have more varied fold change values that the 6 patients at baseline. Collectively these data suggested at the 3D epigenome level, R and NR had few significant differences.





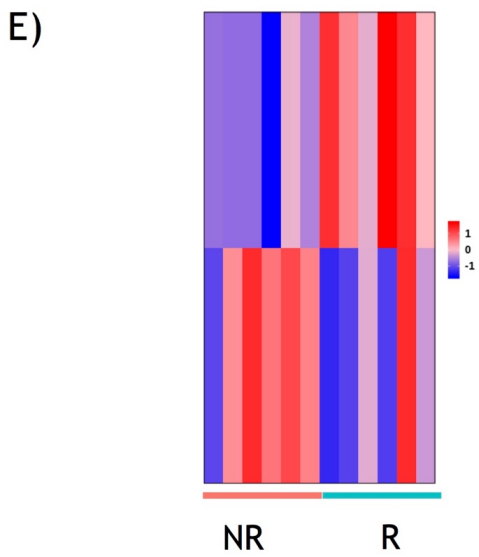
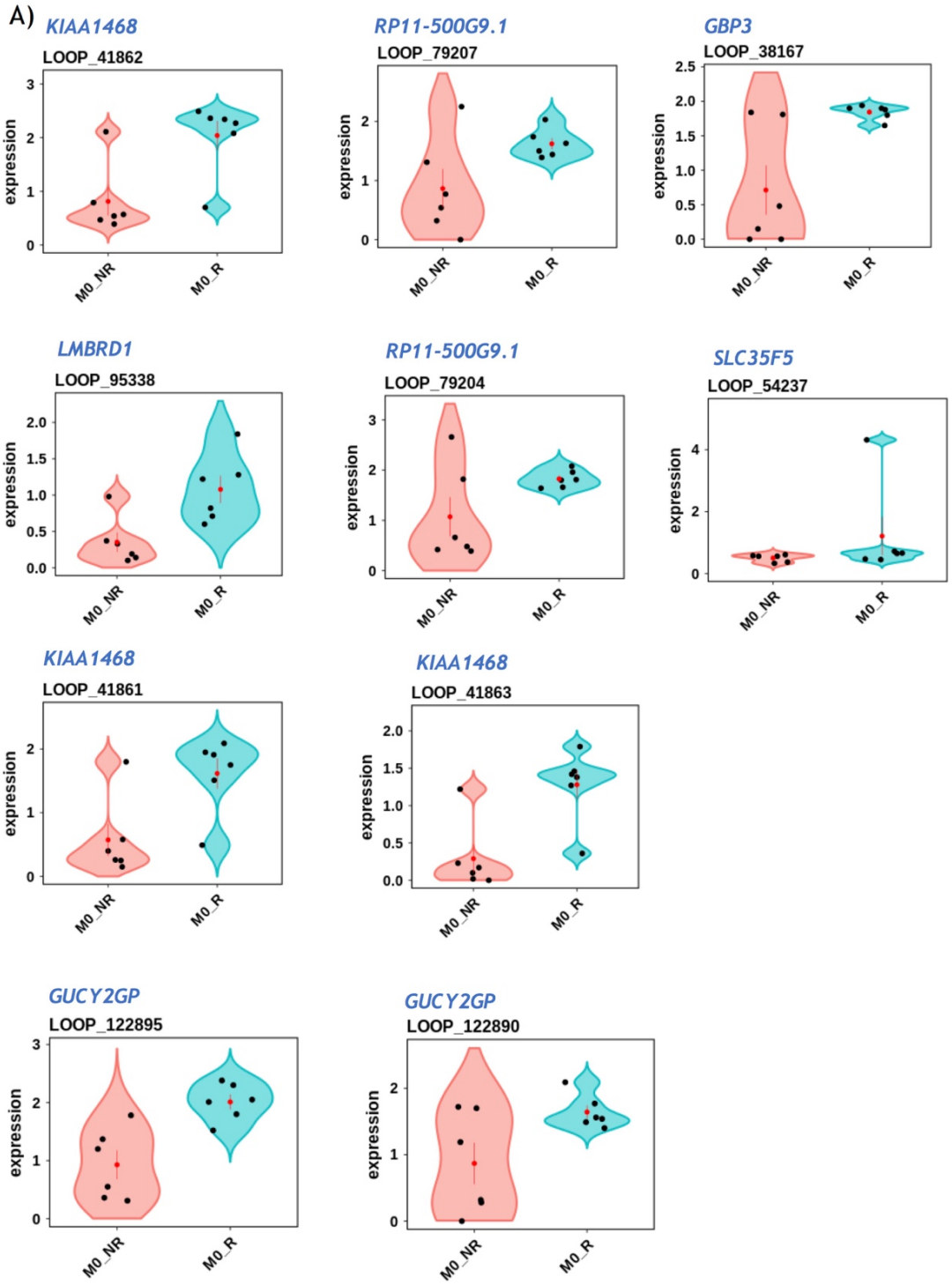


Figure 5.13 Differential Loops Between R and NR at Each Time-point
Data analysis was performed by John Cole, University of Glasgow. Principal component analysis (PCA) of microarray data at (A) M0 (baseline), (B) 6 months, (C) 12 months. Data scaled by Z-score transformation. Each dot represents 1 patient sample. Pink = NR, blue =R. N=12. (D&E) un-clustered heat map representing significantly different loops between NR and R at (D) baseline and (E) 6 months. Log2 fold > 1, adj. P value ≤ 0.05. Expression levels of loops represented as z-scores, -1 -1 = loop abundance fold change

Whilst analysis did not identify stratification signatures at each time point, the most differential individual loops, based on fold change in abundance, were identified to discover if any individual loops had differences at the patient level (Figure 5.14). The 10 most up and down regulated loops were plotted and those loops were then mapped to the closest gene, with a view to understand possible functional consequences of loop formation. To investigate the extent of change in each loop across samples, we evaluated the loops that were most up or down regulated in R compared to NR. The loops that appeared to be most upregulated in R were Loop_41682 (KIAA1468), Loop_79207 (RP11-500G9.1) and Loop_38167 (GBP3)(Figure 5.14A). Two other loops in this set also mapped to KIAA1468 and GUCYGP2, implying that expression of these genes has implications for the R group. The loops that appeared most downregulated in R compared to NR were LOOP_105854(UBE2H), LOOP_36217 (LINC00854) and LOOP_88219 (BTLN8)(Figure 5.14B). It was demonstrated that 2 other loops in this group mapped to UBE2H, 2 others mapped to BTLN8 and another 2 to TSNAX-DISC1. As 3 loops in this group mapped to UBE2H, it is plausible to consider that this gene may be an important gene in NR.


A)
KIAA1468
LOOP_41862
RP11-500G9.1
LOOP_79207
GBP3
LOOP_38167
LMBRD1
LOOP_95338
RP11-500G9.1
LOOP_79204
SLC35F5
LOOP_54237
KIAA1468
LOOP_41861
KIAA1468
LOOP_41863
GUCY2GP
LOOP_122895
GUCY2GP
LOOP_122890
expression
M0_NR
M0_R

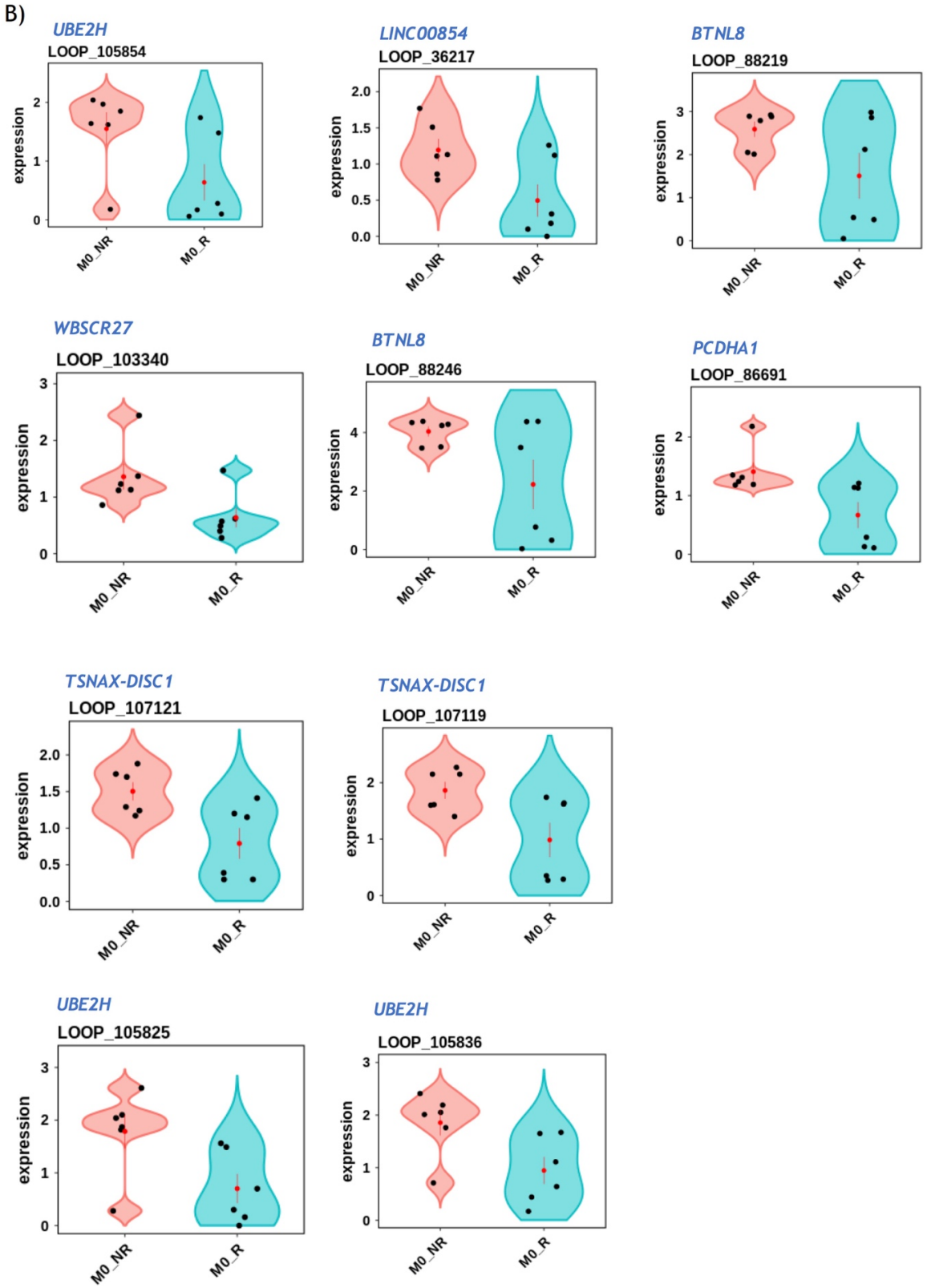


Figure 5.14 Significantly Up and Down Regulated Loops in R (R vs NR comparison) Data analysis was performed by John Cole, University of Glasgow Violin plots of significantly up and down regulated loops, (p.adj ≤ 0.05, absolute log2 fold >1). N=12. A) Significantly upregulated genes in R at baseline time-point. B) Significantly downregulated loops in R at baseline. Loop number highlighted above plot, with closest gene to loop site above. Closest gene determined by Bedtools *closest* function. Pink = NR, blue = R

Stratification of IR and R was next to be analysed (Figure 5.15). There was a degree of separation between IR and R at baseline when using PCA analysis (Figure 5.15A). This separation was partly lost at 6m (Figure 5.15B) and at 12m, IR and R samples were plotted amongst each other (Figure 5.15C). As with the NR and R analysis, heat maps were used to visualise the data and determine if any signatures emerged from the data that could differentiate between IR and R. There were more significantly different loops between these 2 groups than NR and R at baseline and 6m time points (Figure 5.15D,E). Again, there were not enough significantly different loops to plot a heat map at 12m. The fold change of the signature loops at baseline was more than at 6m. At both time points, there were 2 main signatures which differentiated the endotypes; signature 1 had upregulation of loops in IR and downregulation of loops in R, signature 2 had the opposite.

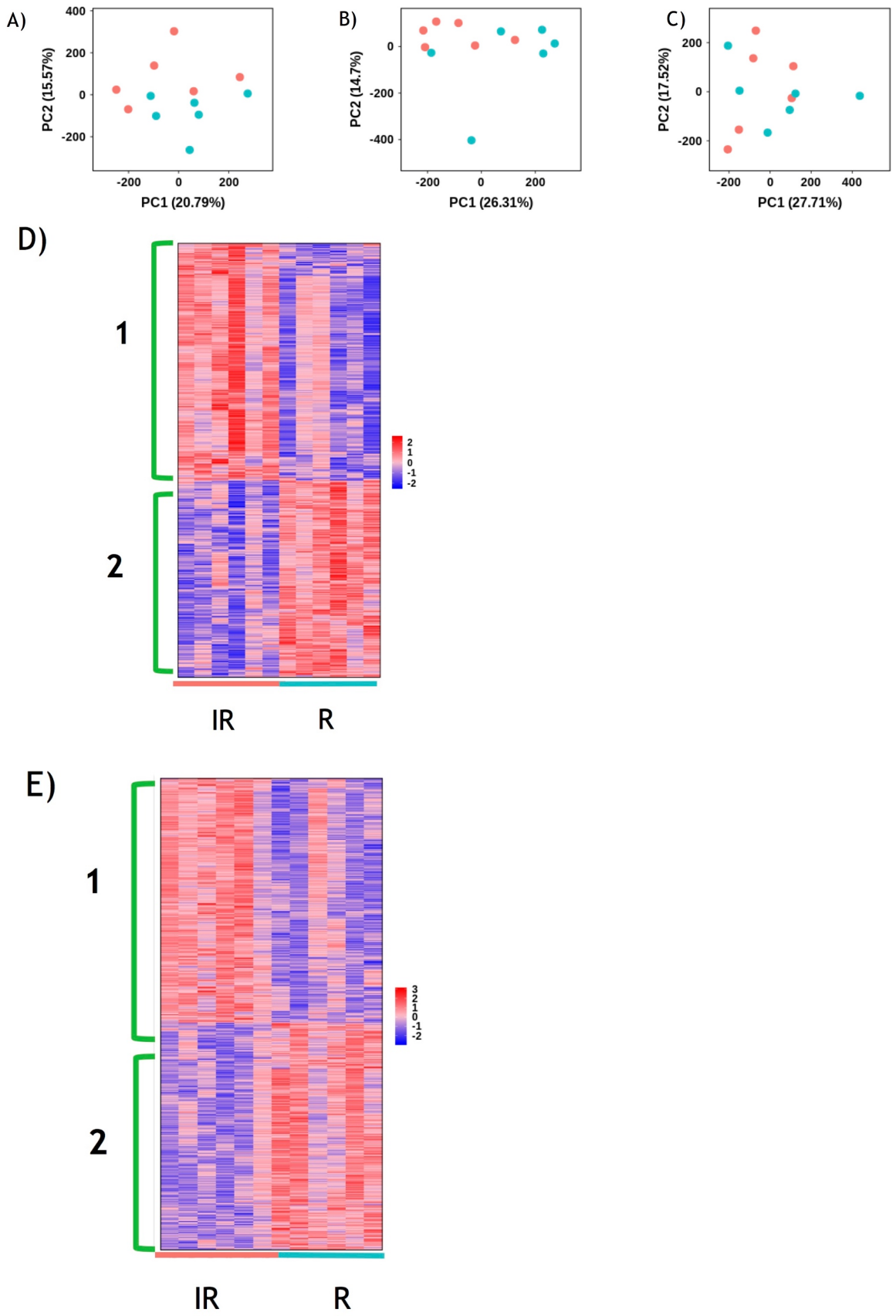
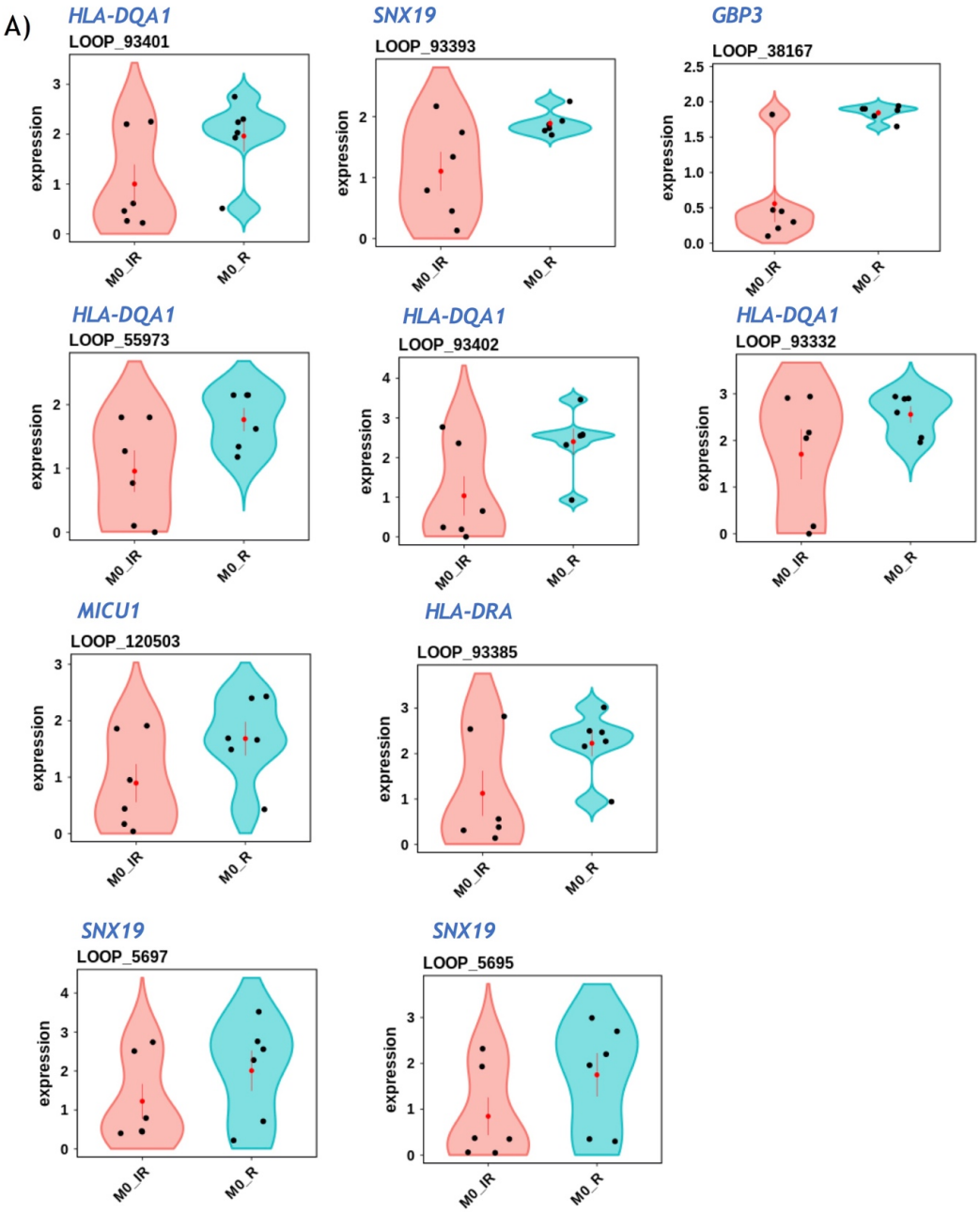


Figure 5.15 Differential Loops Between IR and R at Each Time-point
Data analysis was performed by John Cole, University of Glasgow. Principal component analysis (PCA) of microarray data at A) M0 (baseline), B) 6 months, C) 12 months, Data scaled by Z-score transformation. Each dot represents 1 patient sample. Pink = IR, blue =R. N=12. D) un-clustered heat map representing significantly different loops between NR and R at baseline, E) 6 months. Log2 fold > 1 adj. P value ≤ 0.05. Expression levels of

loops represented as z-scores. -2 -3 = loop abundance fold change. Green brackets and numbers indicate different signatures.

The most significantly up and down regulated loops between IR and R were plotted and mapped to their closest gene (Figure 5.16). Across samples, the loops that appeared to be most consistently upregulated in R were Loop_93401 (HLA-DQA1), Loop_93393 (SNX19) and Loop_38167 (GBP3) (Figure 5.16A). Two other genes in the group mapped to SNX19 and HLA-DQ1. Others in the group mapped to MICU1 and HLA-DRA. The loops that appeared most downregulated in R compared to NR were LOOP_107119 (TNSAX-DISC1), LOOP_107121(TSNAX_DISC1) and LOOP_88225 (BTLN3)(Figure 5.16B). Other loops in this downregulated group mapped to VAMP4, RP11-345I18.6, FRAS1 and BTLN8. Two loops mapping to TSNAX-DISC1 may be an indicator that repression of expression is important in R.

A)


HLA-DQA1
LOOP_93401
SNX19
LOOP_93393
GBP3
LOOP_38167
HLA-DQA1
LOOP_55973
HLA-DQA1
LOOP_93402
HLA-DQA1
LOOP_93332
MICU1
LOOP_120503
HLA-DRA
LOOP_93385
SNX19
LOOP_5697
SNX19
LOOP_5695
expression
M0_IR
M0_R

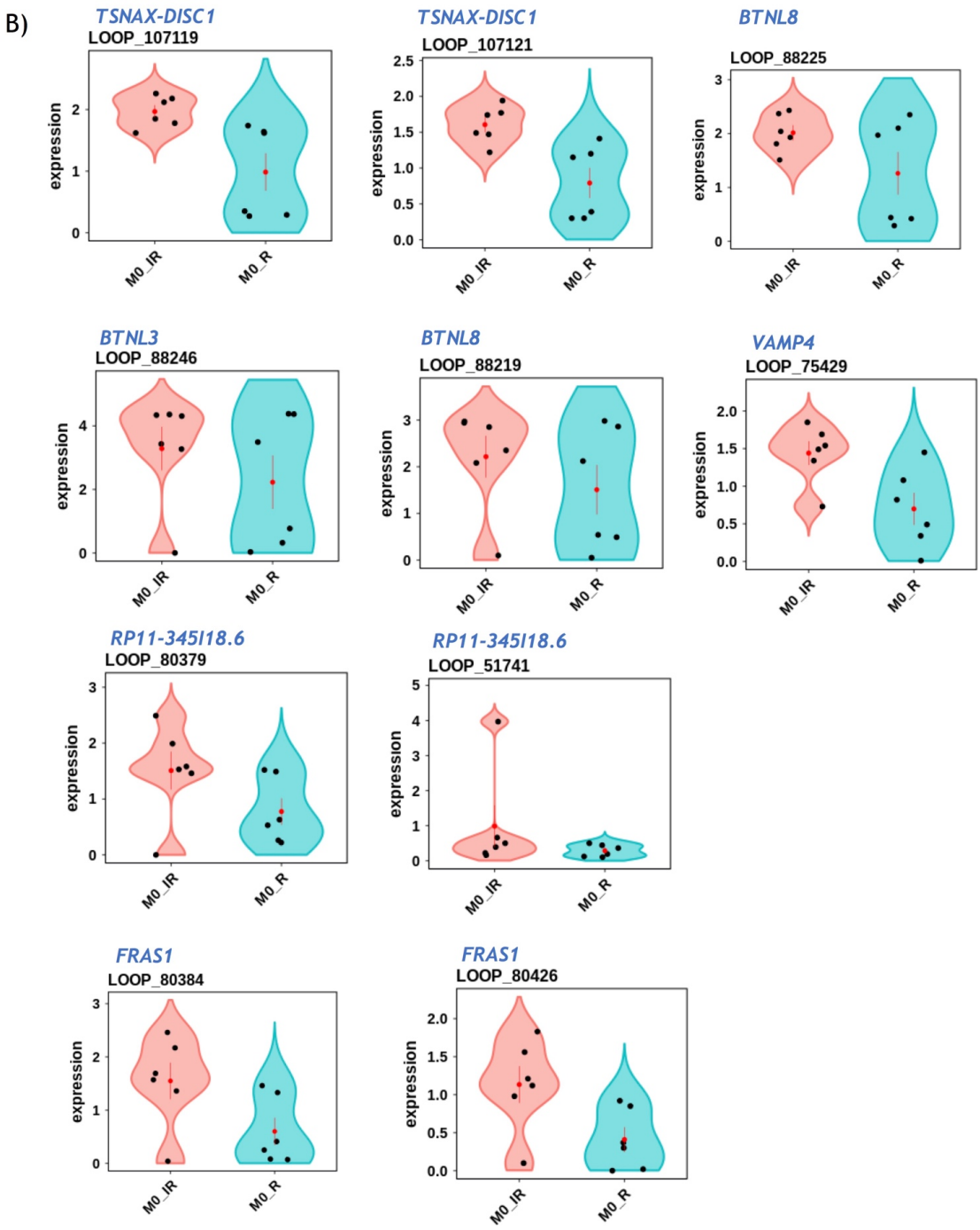


Figure 5.16 Significantly Up and Down Regulated Loops in R (IR vs R comparison) Data analysis was performed by John Cole, University of Glasgow Violin plots of significantly up and down regulated loops, (adj. P value ≤ 0.05, absolute log2 fold >1). N=12. A) Significantly upregulated genes in R at baseline time-point. B) Significantly downregulated loops in R at baseline. Loop number highlighted above plot, with closest gene to loop site above. Closest gene determined by Bedtools *closest* function. Pink = IR, blue = R

The last comparison to be made was between NR and IR (Figure 5.17). PCA analysis was once again used to visualise clustering of the 2 groups at baseline, 6m and 12m. There was a lot of overlap between endotypes at baseline (Figure

5.17A). The 2 groups became more separated at 6m, however the endotypes were not clustered separately (Figure 5.17B). Similar clustering was observed at 12m (Figure 5.17C). Once again, heat maps were used to picture the data and identify stratifying signatures between endotypes. At baseline, 2 signatures emerged from the heat map (Figure 5.17D). The smaller of the 2 signatures (1) showed upregulation of loops in NR. The other signature shows upregulated loops in IR. At the 6m time point there were less distinct signatures shown in the heat map (Figure 5.17E). From the heat map only 1 signature emerged. This signature showed a small group of highly upregulated loops in NR. The rest of the heat map did not highlight a high level of differentiation between the endotypes. As other comparisons have shown, there were not enough significant loops at 12m to plot a heat map.

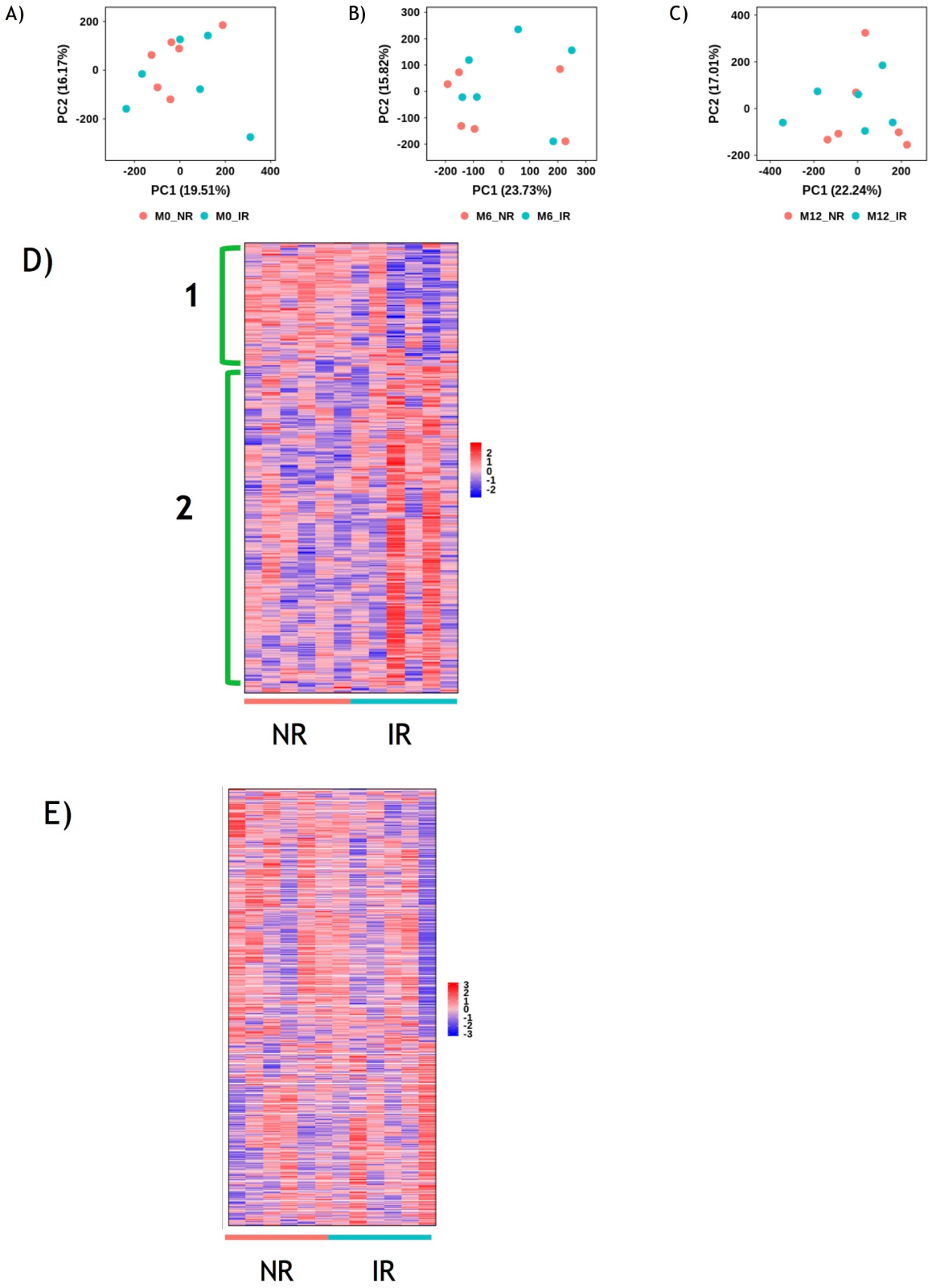
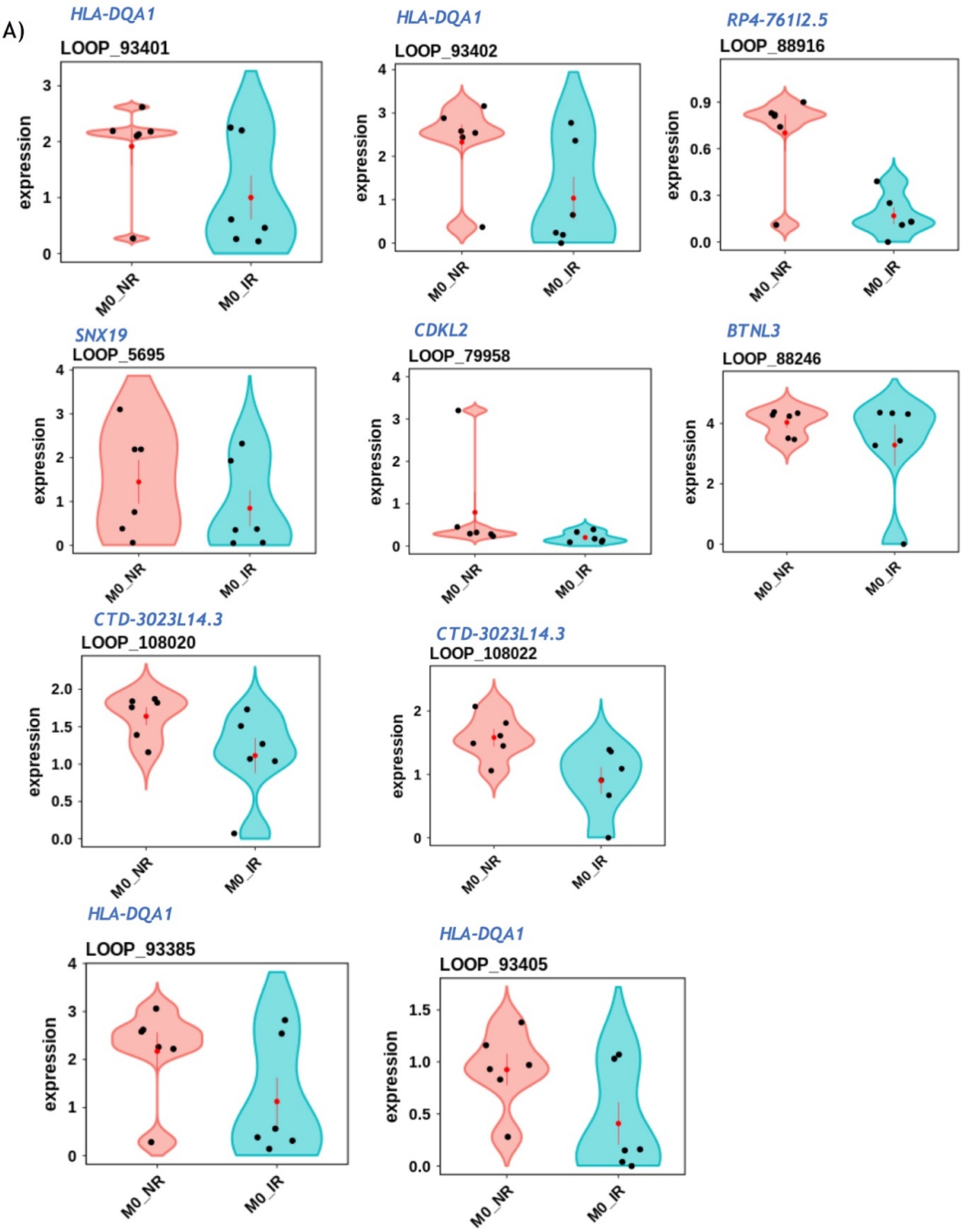


Figure 5.17 Differential Loops Between NR and IR at Each Time-point
Data analysis was performed by John Cole, University of Glasgow. Principal component analysis (PCA) of microarray data at A) M0 (baseline), B) 6 months, C) 12 months, Data scaled by Z-score transformation. Each dot represents 1 patient sample. Pink = NR, blue = IR. N=12. D) un-clustered heat map representing significantly different loops between NR and R at baseline, E) 6 months. Log2 fold > 1 adj P value ≤ 0.05. Expression levels of loops represented as z-scores. -2 – 2 = loop abundance fold change

As before, the most significantly up and down regulated loops between the endotypes were plotted and closest gene to each, identified (Figure 5.18). The loops that appeared to be most consistently upregulated in NR were Loop_93401 (HLA-DQA1), Loop_93402 (HLA-DQA1) and Loop_88916 (RP4-761I2.5)(Figure 5.18A). Two other loops in the group mapped to HLA-DQ1. This data suggests importance of the expression of this gene in the NR. Others in the group mapped to MICU1 and HLA-DRA. The loops that appeared most downregulated in NR compared to R were LOOP_113951 (PRUNE2), LOOP_122895 (GUCY2GP) and LOOP_99942 (AC018641.7) (Figure 5.18B). Other loops in this downregulated group mapped to PDE8B, WDR41, and RP11-551L14.1.

A)


HLA-DQA1
LOOP_93401
HLA-DQA1
LOOP_93402
RP4-761I2.5
LOOP_88916
SNX19
LOOP_5695
CDKL2
LOOP_79958
BTNL3
LOOP_88246
CTD-3023L14.3
LOOP_108020
CTD-3023L14.3
LOOP_108022
HLA-DQA1
LOOP_93385
HLA-DQA1
LOOP_93405
expression
M0_NR
M0_IR

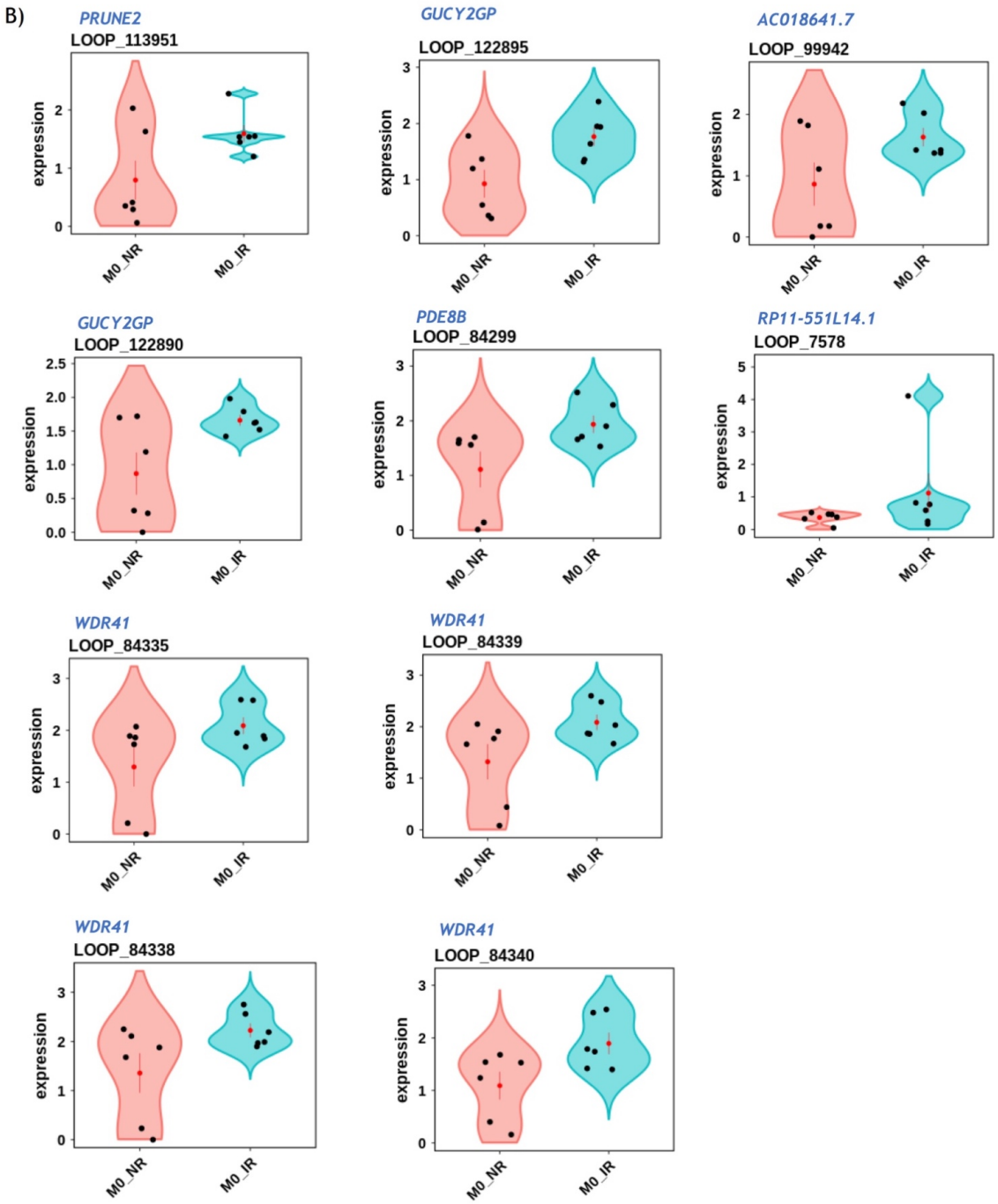


Figure 5.18 Significantly Up and Down Regulated Loops in NR (NR vs IR comparison) Data analysis was performed by John Cole, University of Glasgow. Violin plots of significantly up and down regulated loops, (p.adj ≤ 0.05, absolute log2 fold >1). N=12 A) upregulated genes in IR at baseline time point. B) Significantly downregulated loops in IR at baseline. Loop number highlighted above plot, with closest gene to loop site above. Closest gene determined by Bedtools *closest* function. Pink = NR, blue = IR

5.2.9 SearchLight as second approach to analysis of microarray data – RA

As well as identifying stratifying loops and signatures between pairs of endotypes across the time points, analysis was conducted to compare all 3 endotypes

together at each time point (Figure 5.19). Global PCA analysis of all 3 endotypes at baseline demonstrated 2 large clusters, consisting of patients of all endotypes (Figure 5.19A). This hints as shared RA pathology between endotypes, but highlights the heterogeneity of the RA population. There was no distinct cluster based on endotype. To try and identify stratifying signatures between the groups, we analysed the data with the use of a heat map (Figure 5.19B). Four clear signatures emerged from the heat map. The first clear signature (1) shows upregulation of loops in IR and NR, with downregulation of loops in R. The second signature (2) demonstrated highly significant upregulation of loops in IR, with downregulation in the other 2 endotype groups. The third signature (3) shown upregulation of loops in R with downregulation in NR and IR. The fourth signature (4) shows upregulation in R with downregulation in the other 2 endotype groups, with 2 patients in the IR groups showing upregulation of loops at a similar significance to the R group.

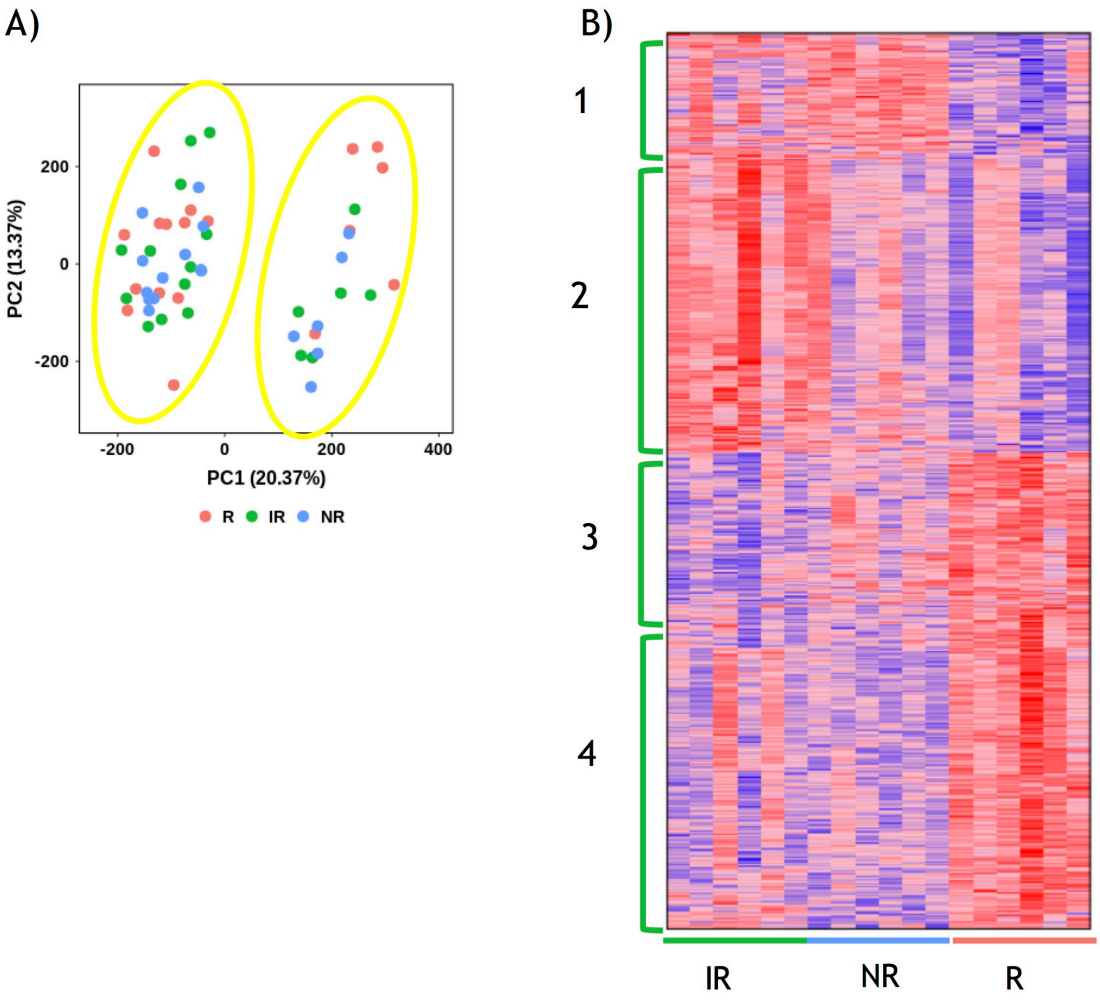


Figure 5.19 Differential Loops Between R, NR and IR at Baseline
Data analysis was performed by John Cole, University of Glasgow. Principal component analysis (PCA) of microarray data at A) baseline, Data scaled by Z-score transformation.

Each dot represents 1 patient sample. Green = IR, blue = NR, pink =R. N=18. B) un-clustered heat map representing significantly different loops between IR, NR and R at baseline, Log2 fold > 1 adj.P value ≤ 0.05. Abundance of loops represented at z-scores. Green brackets and numbers indicate different signatures.
Nos 1-4 represent individual signatures within the heat map

Alike to the analysis conducted between pairs of endotype groups, the most significantly different loops between the three groups were investigated, and the top 8 plotted (Figure 5.20). LOOP_14360 and LOOP_14364, both mapping to DOCK9, appeared to be upregulated in IR and NR samples, compared to R. This difference was consistent across time points. This could mean down regulation of this gene is important in responding to RA therapy. In contrast, LOOP_24104 (RP11-282M16.1) and LOOP_64070 (DSCR3) showed highest loop expression in R, with similarly lower levels in IR and NR. Overall, the difference between groups remained similar across all 3 time points. However, it appeared at baseline, some loops showed the biggest difference, as observed in LOOP_24101 (RP11-282M16.1), LOOP_38537 (PRR11) and LOOP_950095 (SUPT3H). Interestingly, no loops that were significantly different in the pairwise analysis, were the most significant in the comparison between all 3 groups. This may suggest findings are an artefact, or could mean that we captured the slight differences between groups, that can't be captured by analysing the groups all together.

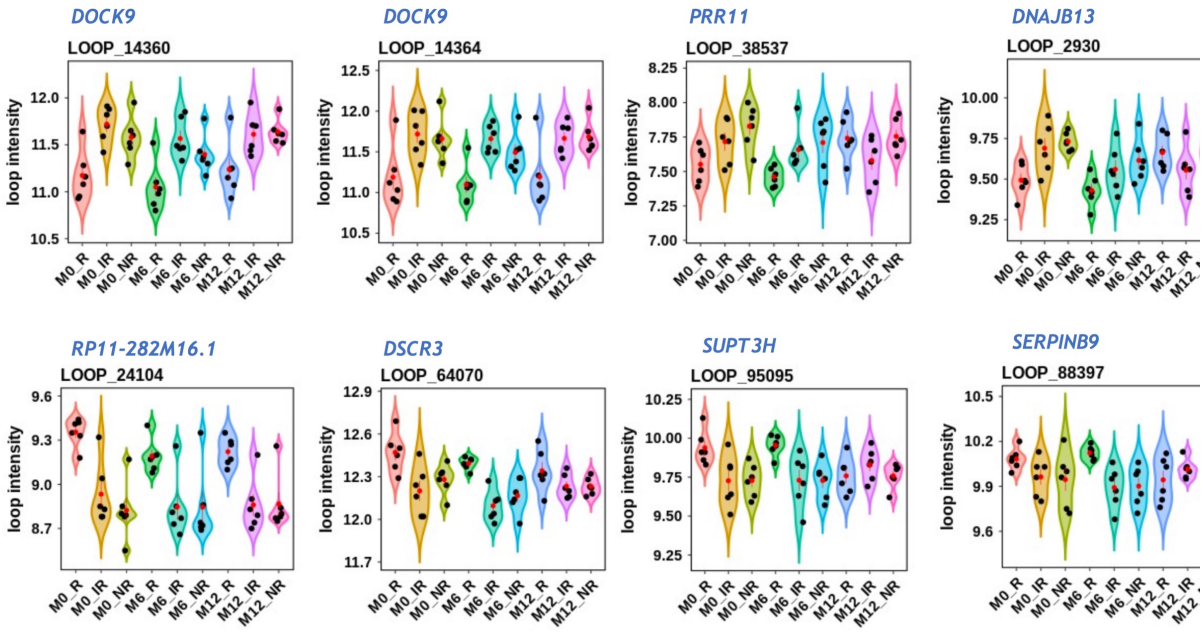


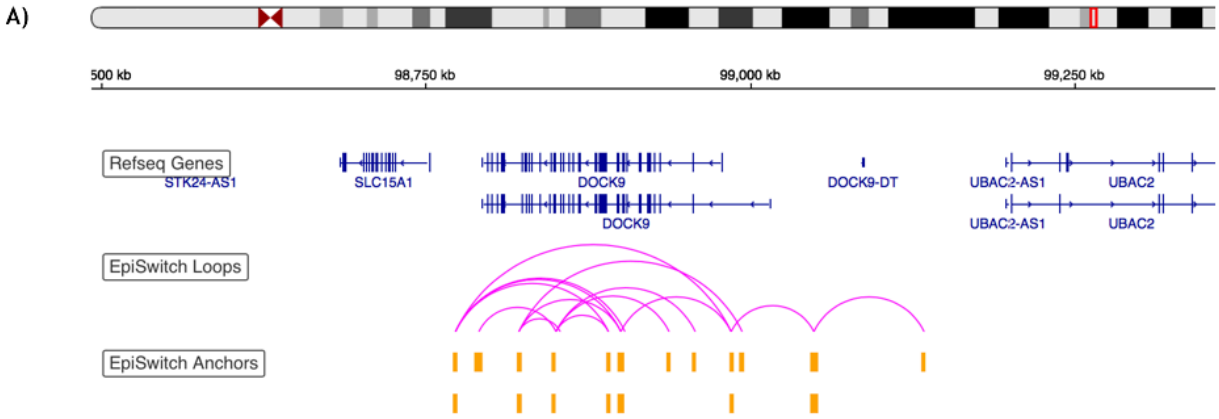
Figure 5.20 Significantly Different Loops between R, NR, IR at All Time points
Data analysis was performed by John Cole, University of Glasgow. A) Violin plots of significantly different loops in between R, NR and IR across all time points. Loop number

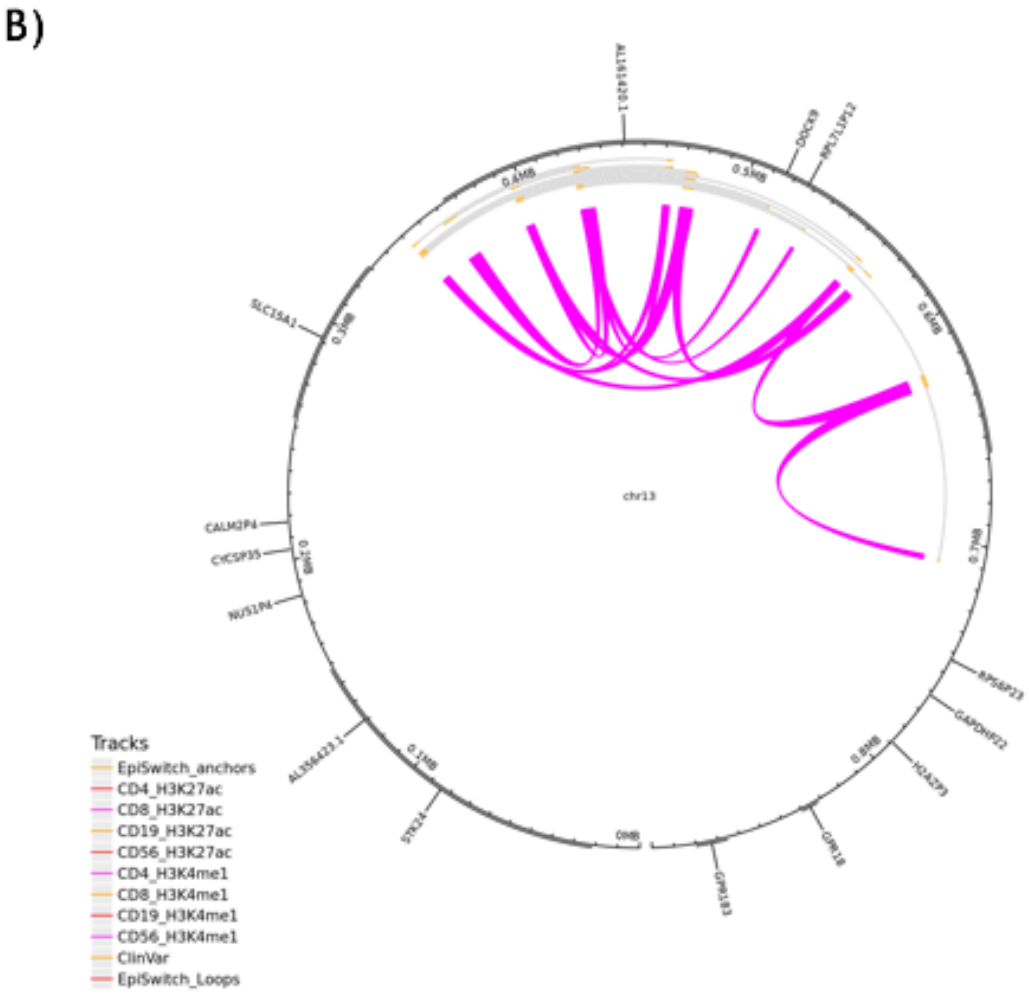
highlighted above plot, with closest gene to loop site above. Closest gene determined by Bedtools *closest* function.

5.2.10 Comparison of analysis approaches

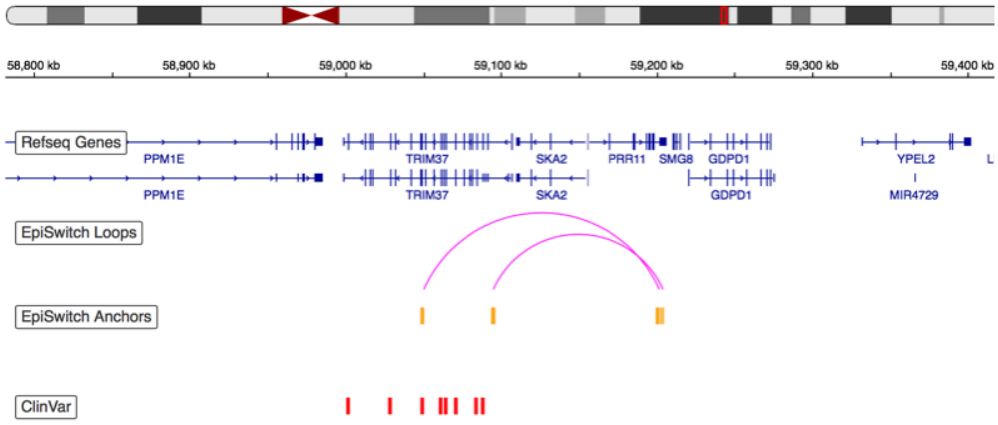
Direct comparison of the analysis pipelines was complex due to the RankProd approach focusing on enriched biological pathways, while analysis with SearchLight focused on individual genes. Moreover, RankProd analysis focused on differences between disease and HC, whilst SearchLight focused on differences between endotypes. Yet notably, this was how I approached the analysis, and both platforms could be used to approach analysis in different ways. However, the EpiSwitch™ data portal visualisation software was used to gather more detail about the most significantly changed loops between the endotypes identified through the Searchlight pipeline, and identify if genes from Searchlight were in regions of other EpiSwitch™ loops (Figure 5.21). Data was available in the portal for DOCK9, PRR11 and DNAJB13. In the region of DOCK9, multiple loop anchor points could be found, with over 10 loops visualised, interacting between them (Figure 5.21A). Of these 3 genes with data in the portal, only DOCK9 was a gene that was found to be significant with the RankProd analysis pipeline. This was significant in NR and IR. DOCK9 can be seen with an anchor site within the RPL17L1P12 locus (Figure 5.21B). Interestingly, there was no interaction with histones or SNPs in this region. Only 2 loops could be visualised in the region of PRR11 (Figure 5.21C). Both loops have anchor points within the RP11 gene region, and loop to anchor points in TRIM37. The loops also overlap with SNPs. Again, the circos plot visualisation provides further detail and highlights that the loops span 3 and 4 genes (Figure 5.21D). Three loops could be shown in the region of DNAJB13 (Figure 5.21E). Two anchor sites within the gene region interact with one other anchor site in the region of CDA4. The other loop in this region has anchor points either side of the DNAJB13 gene. These loops also overlap with SNP sites. The circos plot replicates this data and highlighted one loop spans 0.2mb (Figure 5.21F). DOCK9 is clearly an enriched region, and since it was a significantly differential loop in both analysis pipelines, it is highly likely that this gene is important in RA pathogenesis. Furthermore, using the EpiSwitch™ visualisation software, it allowed us to capture this enriched region, which may not have been as clear otherwise. However, it should be noted that

only comparing 10 genes across platforms is a small number and may not reflect the comparability of the techniques as a whole.

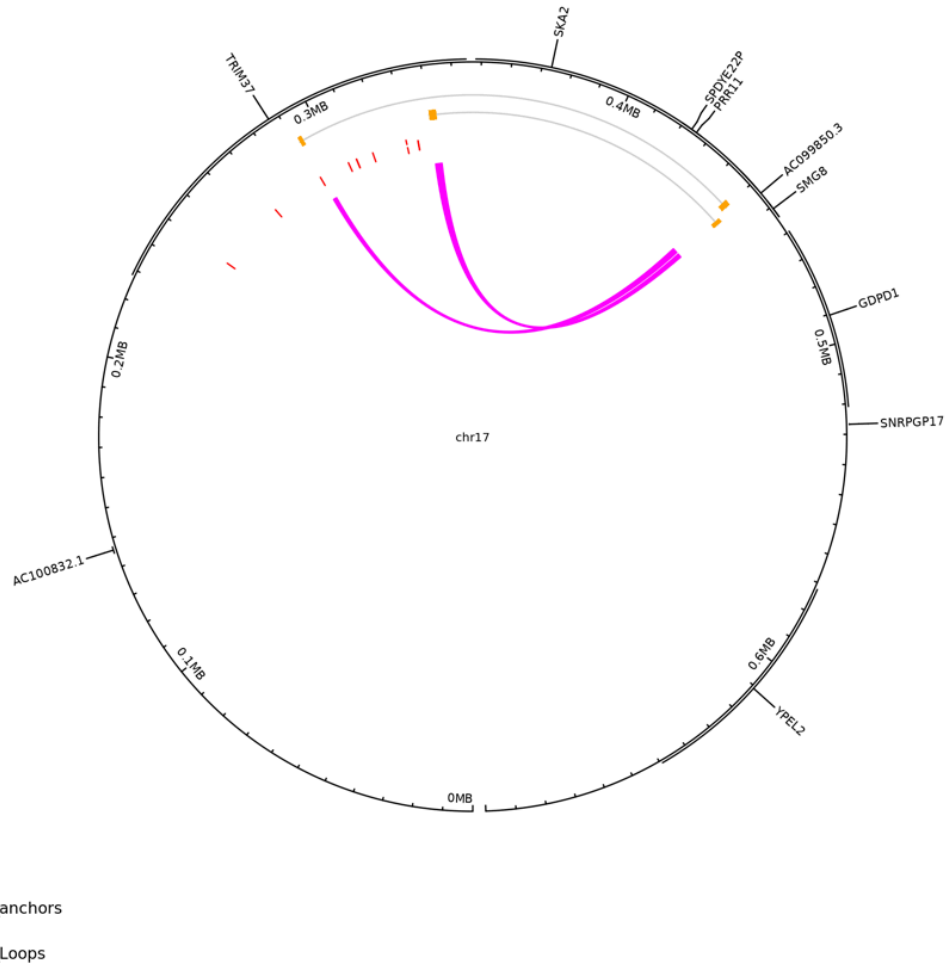




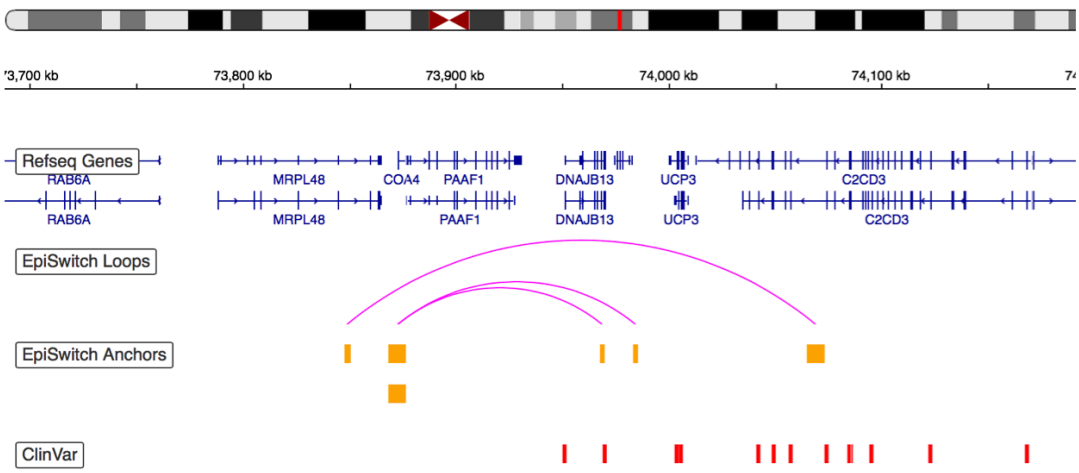
C)


58,800 kb
58,900 kb
59,000 kb
59,100 kb
59,200 kb
59,300 kb
59,400 kb
Refseq Genes
PPM1E
TRIM37
SKA2
PRR11
SMG8
GDPD1
YPEL2
MIR4729
EpiSwitch Loops
EpiSwitch Anchors
ClinVar

D)


SKA2
TRIM37
0.3MB
0.4MB
SPDYE22P
PRR11
AC099850.3
SMG8
GDPD1
0.5MB
SNRPGP17
0.2MB
chr17
AC100832.1
0.1MB
0.6MB
YPEL2
0MB
Tracks
EpiSwitch_anchors
ClinVar
EpiSwitch_Loops

E)



F)

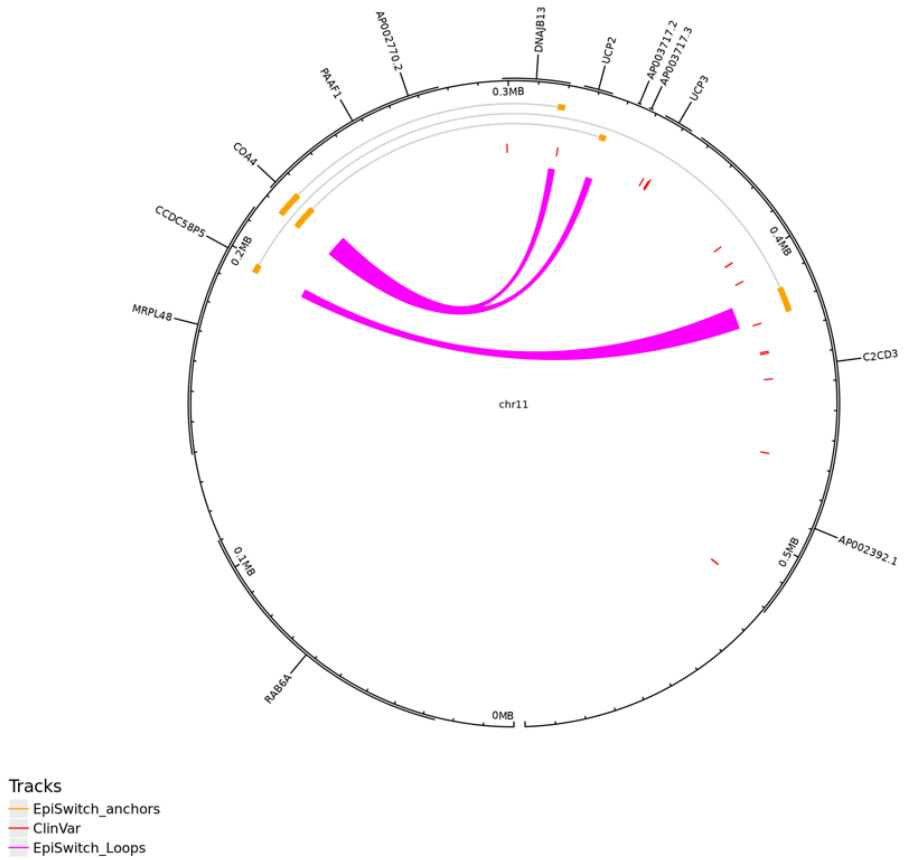
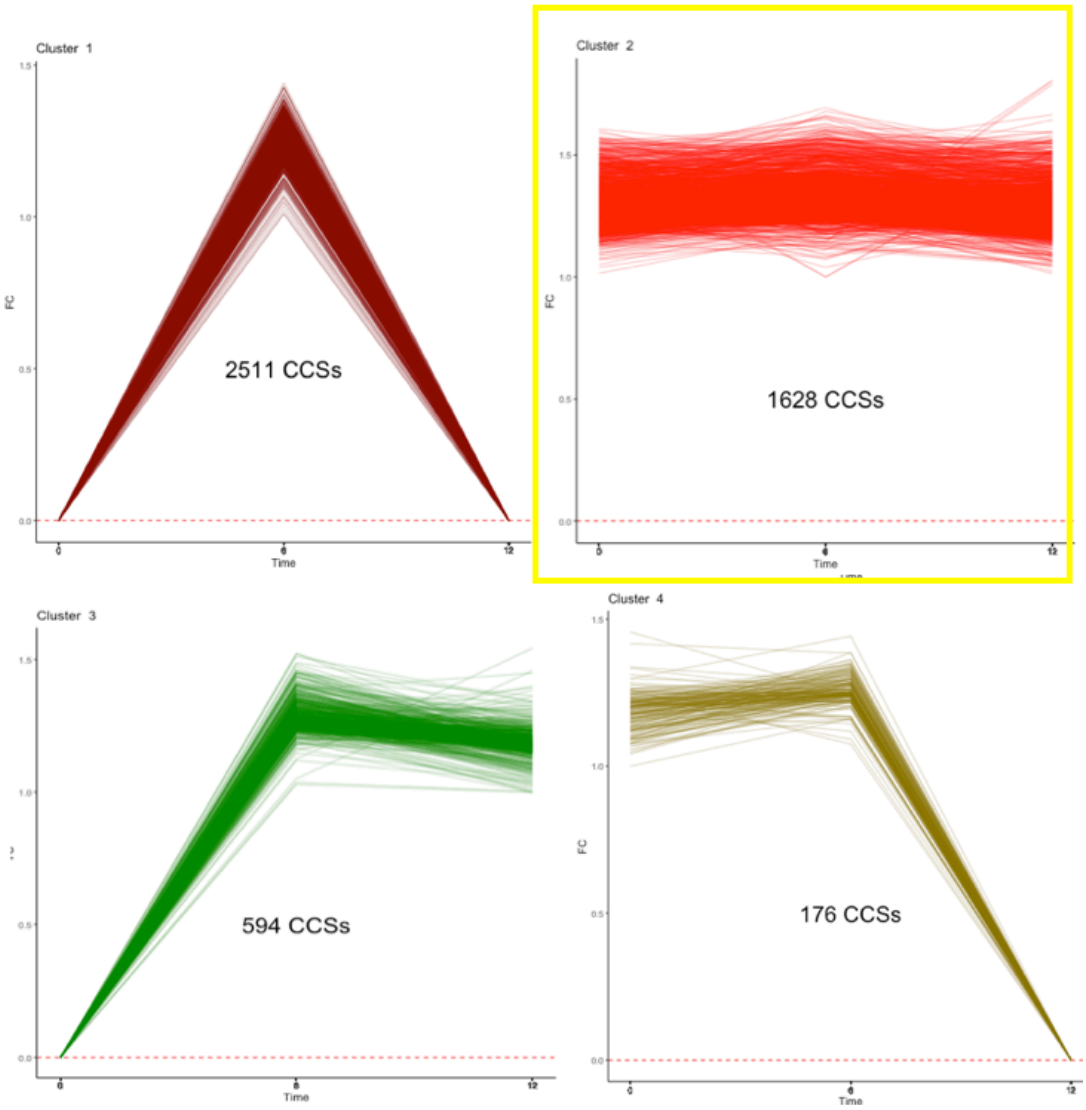


Figure 5.21 EpiSwitch™ Data Portal Visualisation of Searchlight Genes
Most significantly changed genes from Searchlight analysis were searched for in EpiSwitch™ data portal. Data was available for 3 genes. Visualised in IGV simulation and replicated in circos plot for A+B) DOCK9. C+D) PRR11. E+F) DNAJB13

5.2.11 Identification of candidates for endotype CCS

As one of the aims from this part if the study was to identify loops that have the potential to become part of an effective predictive CCS, it was important to establish the dynamics of loop significance across the 3 time points. To identify the loops with the strongest stratification potential for prediction of endotypes, the quality threshold (QT) clustering algorithm was used on the RankProd data. This algorithm does not specify the number of clusters *a priori*, and clusters must pass a user-defined quality threshold. To be included in the cluster, a loop must have had to be significant in at least 1 time point. Data generated from RankProd, which has been explored above, was used in the clustering algorithm. Three groups of clusters were produced based on comparison between each endotype and pooled HC. Comparison between R and HC produced 9 clusters (Figure 5.22). From the 9 clusters, 7 loop dynamic patterns can be observed. Two clusters (2 and 5) have statistically significant loops at all 3 time points and 2 (6 and 9) clusters have significance at baseline and 12m but not at 6m. The cluster with the most loops (2511), had loops with statistical significance only at 6m. In the interest of stratification significance over 12 months, cluster 2 would be the choice cluster to be taken for further analysis. Loops from this cluster would be the most likely to generate a significant stratifying signature.


Cluster 1
2511 CCSs
Cluster 2
1628 CCSs
Cluster 3
594 CCSs
Cluster 4
176 CCSs
FC
Time

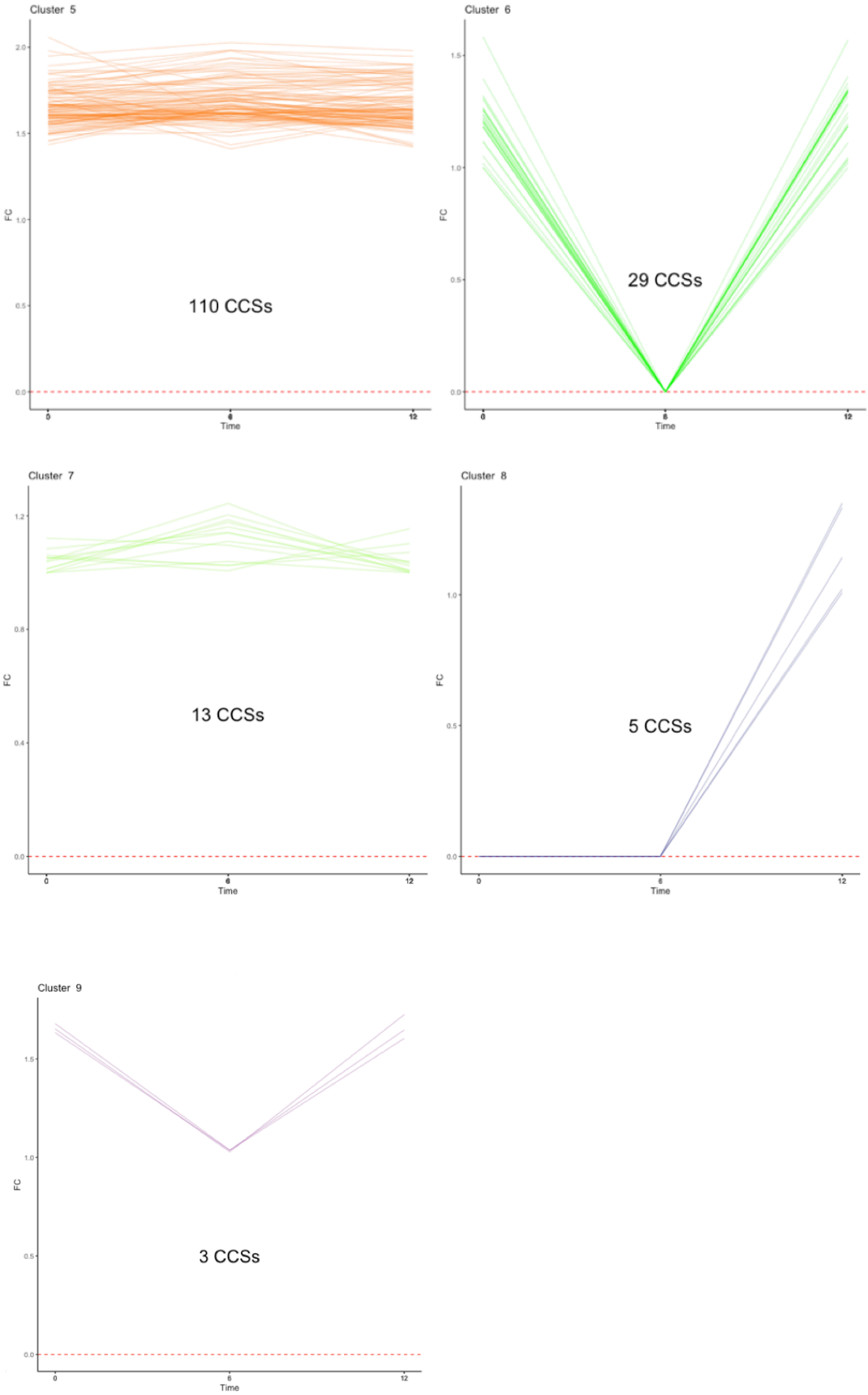
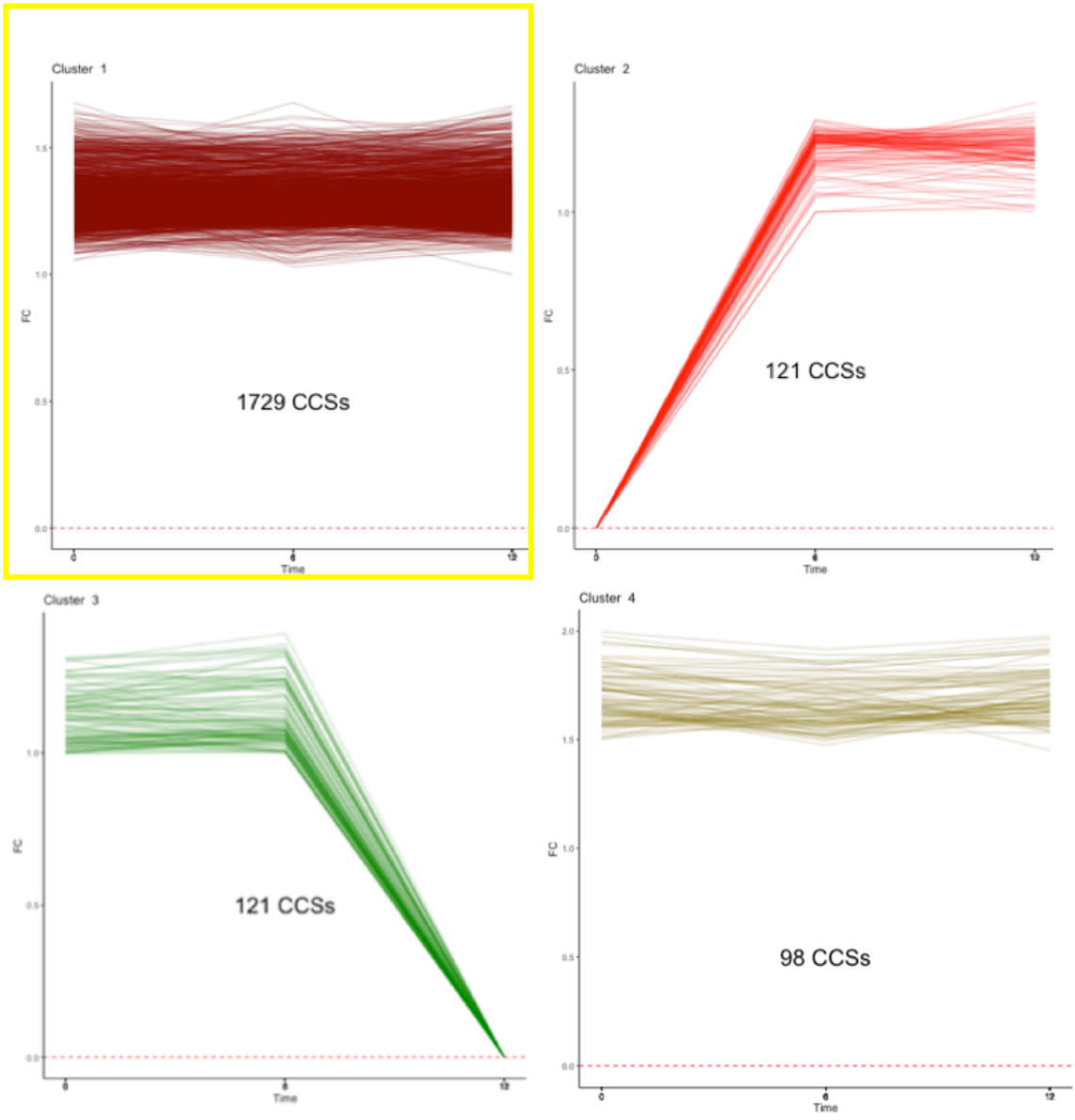


Figure 5.22 QT clustering of Loops in R
Data analysis conducted by Ewan Hunter, OBD. Raw microarray data from R and pooled HC comparisons was used for the QT clustering algorithm. To pass the quality threshold loops had to have significance, ≤0.2 FDR. Data with most potential for an endotype stratification signature highlighted in yellow square.

Comparison between NR and pooled HC revealed 9 clusters (Figure 5.23). Clusters 2-9 had 121 loops or more identified. In contrast, cluster 1 had 1729 loops identified. This cluster, along with cluster 4 and 6 are made of loops with statistical significance at all time points. The 6 remaining clusters have unique dynamics ranging from significance only at baseline to significance only at 12m These clusters are made of 4 and 3 loops respectively. Cluster 1 would be the cluster of choice to be taken forward for further analysis based on strong significance at all time points in many loops.


Cluster 1
1729 CCSs
Cluster 2
121 CCSs
Cluster 3
121 CCSs
Cluster 4
98 CCSs
FC
Time

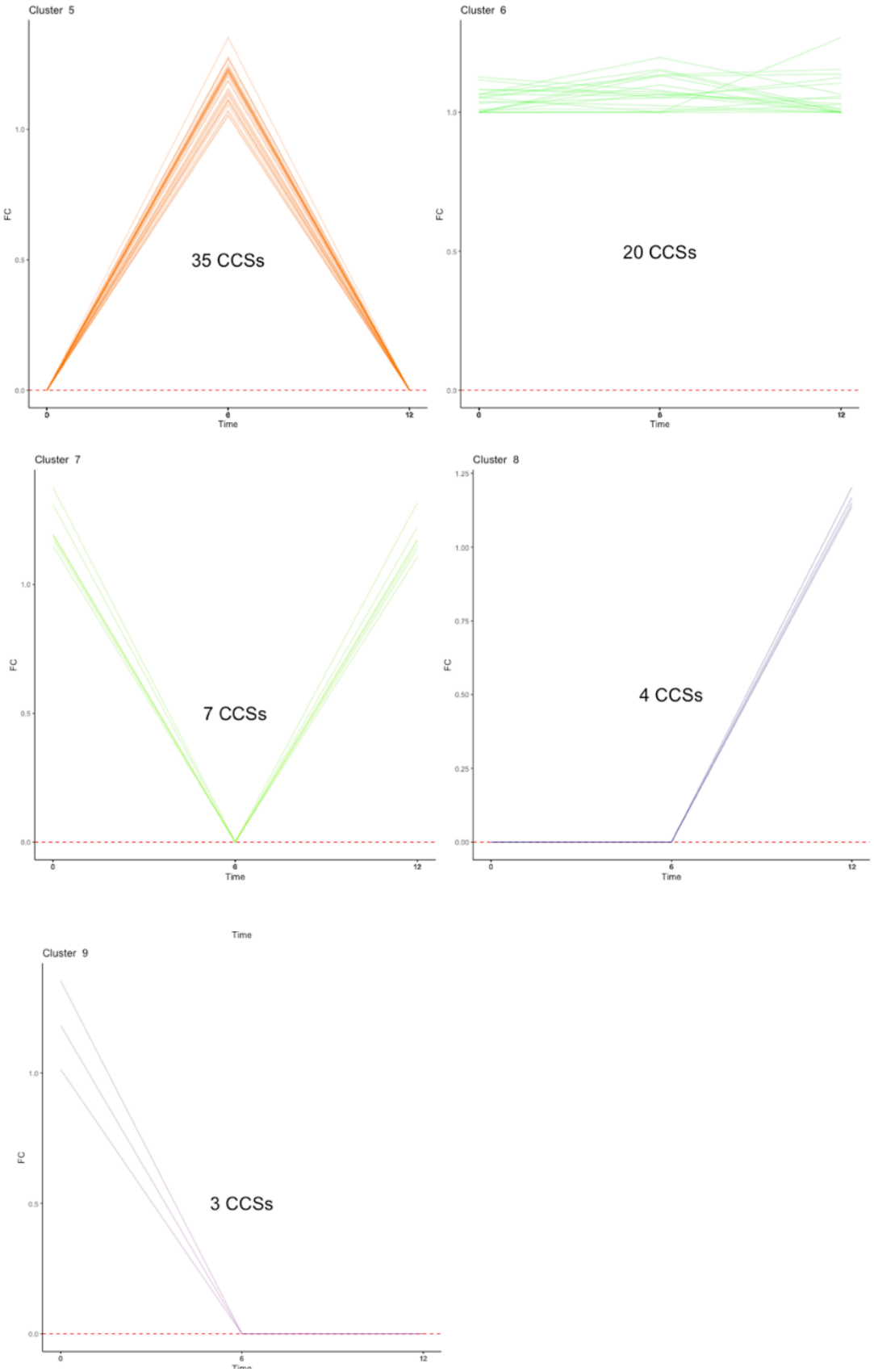
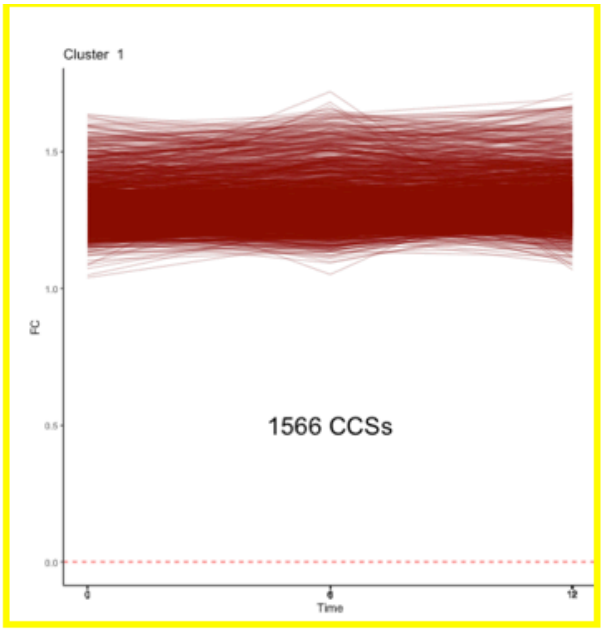
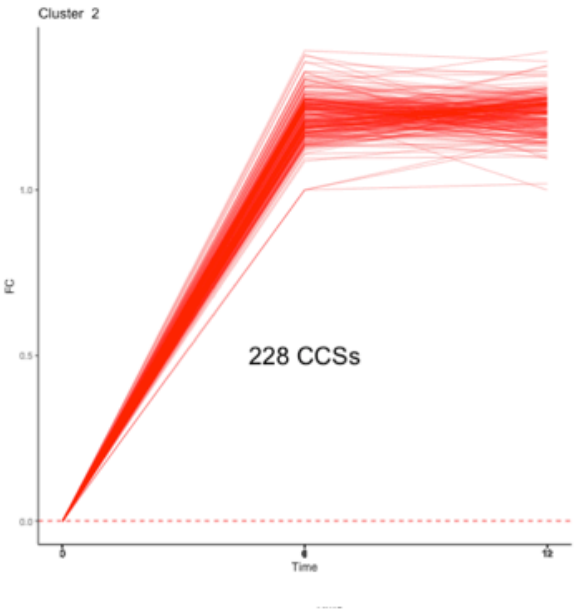
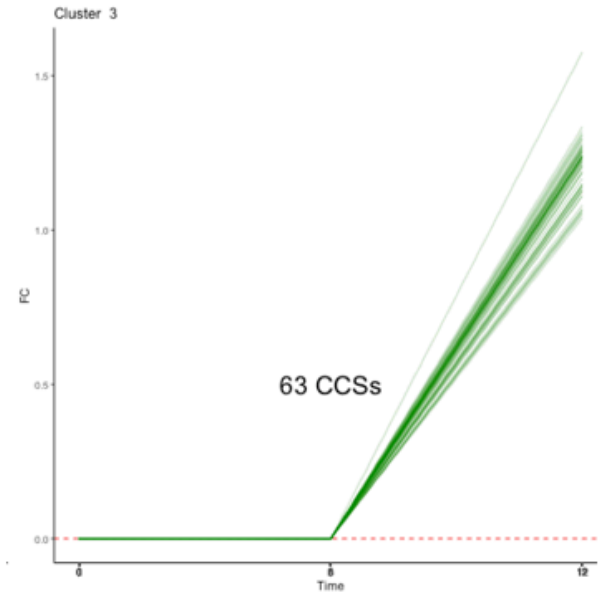


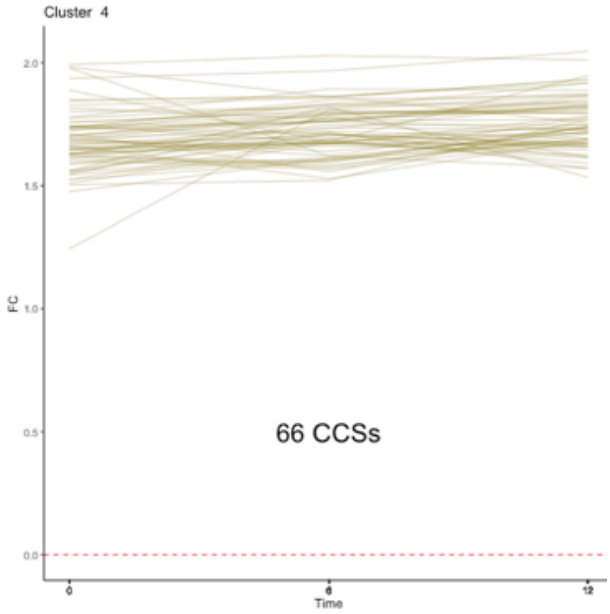
Figure 5.23 QT clustering of Loops in NR
Data analysis conducted by Ewan Hunter, OBD. Raw microarray data from NR and pooled HC comparisons was used for the QT clustering algorithm. To pass the quality threshold loops had to have significance, ≤ 0.2 FDR. Data with most potential for an endotype stratification signature highlighted in yellow square.

Clustering of IR and pooled HC comparison data revealed 8 clusters (Figure 5.24). Again, cluster 1 had considerably more loops than the other clusters with 1566 loops compared to the second highest cluster made up of 228 loops. Cluster 1, 4 and 8 had loops with statistical significance at all time points. The remaining 5 had dynamics unique to each cluster. Of these 5 clusters, the cluster with significance at baseline and 12m had the least loops (18). Based on stable significance throughout 12m, data from cluster 1 would be the most promising cluster to take forward for further analysis to determine an endotype signature.


Cluster 1
1566 CCSs
FC
Time


Cluster 2
228 CCSs
FC
Time


Cluster 3
63 CCSs
FC
Time


Cluster 4
66 CCSs
FC
Time

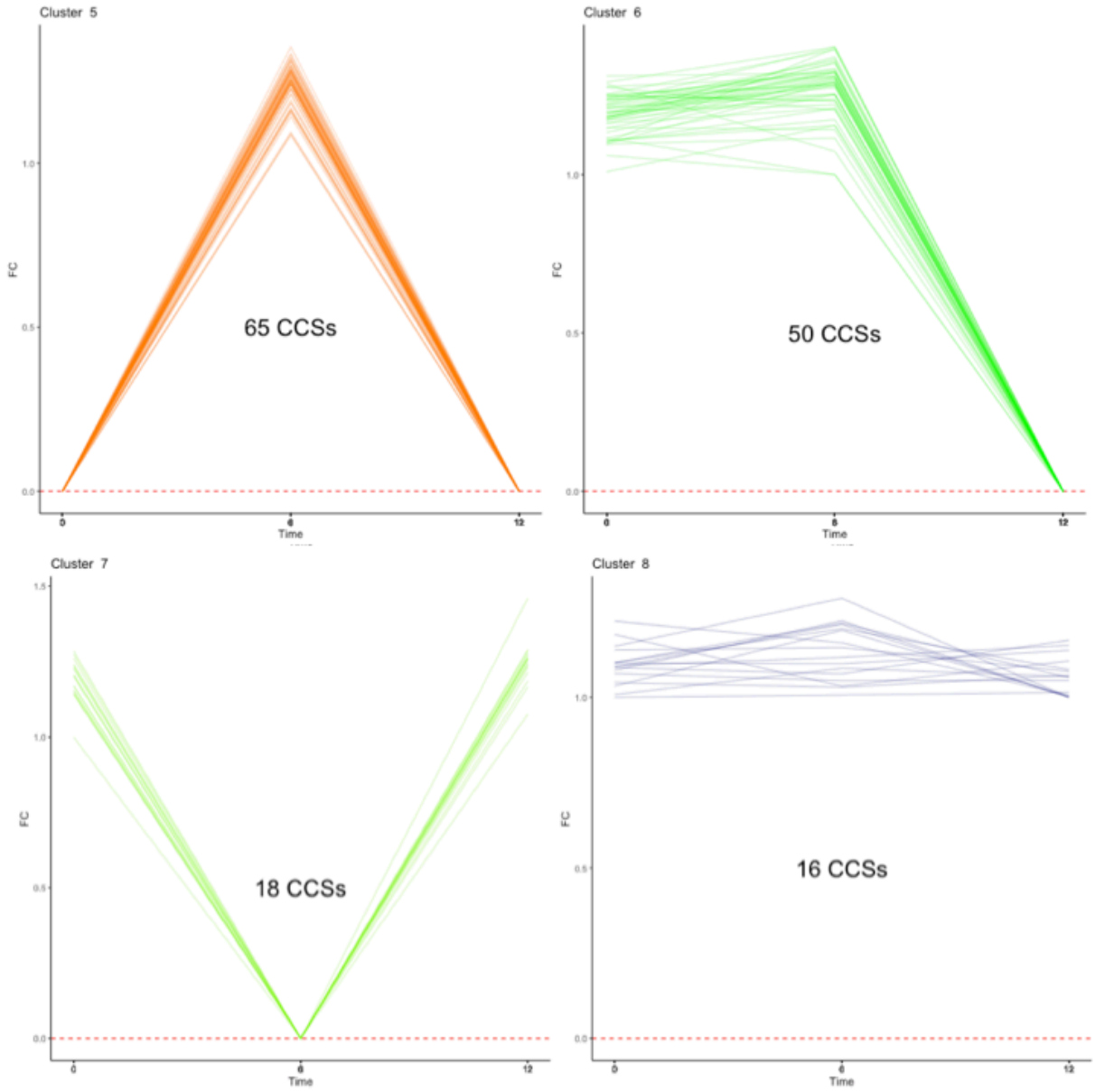


Figure 5.24 QT clustering of Loops in IR
Data analysis conducted by Ewan Hunter, OBD. Raw microarray data from IR and pooled HC comparisons was used for the QT clustering algorithm. To pass the quality threshold loops had to have significance, ≤ 0.2 FDR. Data with most potential for an endotype stratification signature highlighted in yellow square.

5.3 Discussion

The work in this chapter has alluded to the presence of several molecular endotypes, specifically 3D chromatin endotypes, that exist in early RA. Moreover, the data suggests that there may be loops present in each endotype at baseline that may be able predict which trajectory a patient will follow. Analysis also revealed that some loops were present in all endotype groups at all time points, revealing a RA signature that suggests a baseline level of dysregulation dictated by the 3D structural epigenome. Moreover, as many the loops in this group were determined 'stable', it implies a core RA state, from which important disease relevant pathways can be found. The loops in disease were of most interest based on the aim of using stratifying loops to understand underlying pathogenesis.

Analysis of the SERA dataset revealed the presence of 3 main response trajectories (Figure 5.1). This analysis builds on the interrogation of the SERA cohort for the identification of the MTX CCS. Where R, and NR were assessed on their clinical scores at 6m, extending analysis to include the 12m time point revealed the IR response group, that at 6 months have the same trajectory as R. Using these longitudinal samples is a great advantage to this work. These trajectories are similar to findings from other RA cohorts. Other studies often identify 3 groups of responder; named fast/rapid, slow/gradual or non/inadequate-responders[354,355]. Some studies have even characterised 5 groups established on baseline DAS[356]. However, the R group from our cohort were not split into fast or poor responders. Most studies identify the majority (82.6%) of patients to be in one of the rapid or gradual responder categories, with only a small proportion of patients classed as non-responders (3.3%). In the SERA cohort, a similar proportion of patients are R and NR, with a small proportion of patients classed as IR. The difference with the SERA cohort, and the patients investigated, is the absence of biologic treatment during the 12 months explored. The study by Siemons *et al*[355] followed a treat-to-target strategy that included the introduction of anti-TNFα biologic, adalimumab, at Week 24, and etanercept at Week 48, which would most likely contribute to the increased good responder rates. However, one observational cohort study, with patients commencing MTX treatment for the first time found non-response rates at 6 months to be 43%[99]. Due to the 'extreme' responders being chosen for the array,

all 6 in the R group could be considered fast responders. The most extreme trajectories were chosen as best as possible, however, sample availability from the SERA biobank did impact the choice of patients. It was desirable to have all patients in each group with as similar a trajectory as possible, to reduce confounding factors in the analysis, and this was achieved with R and NR groups. Unfortunately, the IR group had different trajectories, with some variation in disease activity at all time points. Specifically, there was a combination of high and moderate, remission and low, and high and moderate disease activity at baseline, 6m and 12m, respectively. Subgroup analysis of the IR trajectories was out with the scope of this work, however, it would be of interest to stratify this group alone to reveal if differences in disease activity were reflected by underlying 3D chromatin structures.

Analysis of baseline demographic and clinical characteristics showed all groups were comparable at baseline (Table 5.1). Autoantibody presence was also comparable between groups. One of the clearer indications of which trajectory a patient would follow was clinical scores at baseline, with NR having the highest CDAI and DAS28 CRP at this time point. These findings have been widely described in other RA studies[357 358 359] with some studies isolating individual joint scores as predictors of poorer response[360,361]. Other studies exploring baseline predictors to MTX have indicated female gender and current smoker status as associated with a decreased likelihood of achieving a EULAR response after 4 months[362]. Moreover, other studies have shown baseline depression and anxiety scores to have a negative impact on response to DMARD therapy[363]. This wasn't a factor that was explored in this work, but would be useful to consider. The only current smokers in our array cohort were in the IR group, who all achieved good response at 6m. However, with only 6 patients, the effect may not be obvious and not all 6 were treated with monotherapy MTX. Interestingly, IR were the group with the lowest CDAI and DAS28 CRP at baseline. DAS28 ESR was not documented based on values needed to accurately calculate the score missing from several patients.

It was also of importance to record the treatment trajectory of each patient as a consideration of any inter-endotype differences (Figure 5.2). Understandably, most R were treated on monotherapy MTX throughout the 12 months. The

differing treatment trajectories could be considered a limitation to this work, as it introduces variability in the data. However, it is unrealistic for a NR or IR to stay on a drug or combination of drugs that isn't having a clinically meaningful improvement. Of the patients changing drug in the 12m period, there was considerable variation in the treatment regimen between patients, which is reflective of real-world disease management[364]. Surprisingly, 3 NR remained on monotherapy MTX despite showing no clinical response. On further investigation, all patients that remained on a therapy without showing a clinical response were found to be enrolled at different study sites, therefore that is unlikely to explain the reason for remaining on the drug given at baseline. There are a number of other reasons a patient may have had to remain on a given drug, including avoidance of exacerbating other conditions, such as cardiovascular disease or liver disease[365]. Moreover, studies have varying conclusions on the best way to approach treatment with the 3 csDMARDs, such as the parallel use[117] and the step-up approach[366]. Furthermore, the patient perspective should be considered. One qualitative study analysed the reasons for patient refusal of DMARDs in RA[367]. Some reservations were due to dangers of medications, disappointment with other treatments and denial of disease. Fraenkel *et al* have demonstrated that these feelings can be especially prominent when patients have highly active disease[368]. Even for the small number of patients in our study, trying to fully understand the reasons behind treatment assignment could be a complex analysis and was out with the scope of the aims of this chapter.

The patients were chosen, and the microarrays run at OBD. Subsequently, the first step of analysis was to assess the quality of the data (Figure 5.3). Recording the red-green densities of arrays confirmed no outliers were present and that Loess normalisation was successful. Moreover, these steps helped to ensure information was preserved and no variations in the data were wrongly inflated[369]. Another important QC check was to assess the slide distribution; this was done by the generation a PCA plot with the slides labelled. The issue of batch effects is widely recognised in microarray studies, especially with human samples[370]. Strategically, patients with the same response trajectory and samples from the same time point were not grouped together on the same slide to reduce the chance of a batch effect associated with a particular group of samples. The PCA plot showed that slides did not cluster together, which was

expected. This provided assurance that measures taken to reduce batch effects were successful and provided confidence for further data analysis.

Top level analysis of each group, prior to linear modelling revealed some grouping based on endotype through PCA plots (Figure 5.4). The results suggested that the largest differences between disease groups was at baseline and the epigenome would gradually lose stratification ability at subsequent time points. The PCA plots also suggested most variance in the dataset came from the R group. Of note, PCA visualisation were ultimately an approximation of the data distribution and more in-depth analysis was required.

The Limma programme was carried forward and a linear model implemented to find contrasts between responder groups, as well as between RA and pooled HC (Table 5.2). Importantly, this approach has been used successfully in many studies including the MTX CCS that formed the basis for this work and provides a flexible platform to analyse experiments with multiple parameters[219,371]. This analysis produced thousands of differentiating loops between groups at all time points. The data suggested that there more differences between RA and HC than between different RA responder groups, which is not surprising. However, there were differences between each pair of endotype group at each time point, alluding to the existence of epigenetic endotypes. Further, data suggests the biggest difference in the structural epigenome between disease and pooled healthy exists at 6m followed by 12m. This is in contrast to the PCA results, emphasising the importance of those plots as a guide only.

On reflection of this data and the total number of significant loops, I considered that the Limma model was not stringent enough to reveal truly statistically significant and biologically relevant stratifying loops. Furthermore, in the interests of utilising the data to integrate the underlying biology of all groups, it would be useful to begin work with a smaller list of loops to streamline the analysis process and find meaningful results. The Limma method of analysis is still of importance and produced results which lead to the MTX CCS findings in previous work. However, the microarray for this endotype work had a more complex design, and therefore it was plausible that an alternative, more stringent method was needed. Based on the data available, analysis focused on

the differences between each disease group and the pooled HC. Analysing these comparisons meant that the disease relevant loops, with removal of any loops possibly important in a healthy state would maximise disease relevant information.

The analysis method chosen was RankProd 2.0, a Bioconductor package used to find differentially expressed molecular profiles. RankProd has several advantages which made it an appropriate method for this work, in particular the ability to analyse data with a small sample size and sample heterogeneity, which is of particular relevance in this study[372]. These features have been effectively demonstrated using a wide variety of sample types, from plants and mouse models, to human acute leukaemia samples[373].

A similar table to that produced from the Limma model was produced with the RankProd results (Table 5.3). As previously discussed, the disease loops were of most interest based on the aim of using stratifying loops to understand underlying pathogenesis. As such, the number of loops able to stratify disease and pooled HC in each endotype group at each time point were quantified. Clearly, the RankProd analysis was more stringent and reduced the number of stratifying loops considerably, with the highest number of stratifying loops being 4765 in the NR 6m group and the lowest number of 384 loops in R at baseline. In contrast to the global level analysis that suggested the most difference between endotype groups was at baseline, this RankProd analysis suggested that the biggest difference between groups is present at 6m, which is line with the Limma analysis. While the analysis approaches differ, this provides an internal validation of the results up to this point. This data highlighting the 6m time point alludes to the influence of treatment on the 3D epigenome. Based on the number of significant loops, HC appeared most different from NR at baseline, R at 6m and IR at 12m. While these results showed that disease and pooled HC are epigenetically different throughout the first 12m of treatment, the results also revealed the existence of endotypes, as different number of stratifying loops between disease and HC exist at each time point, highlighting that some loops must have unique statistical significance in each group. As reducing the number of loops to analyse was considered beneficial, it was positive that analysis using

the RankProd method also reduced the number of loops to investigate for biological inference and would streamline the analysis process going forward.

The approach to understanding the biological relevance of the loops found to be significant through the RankProd approach was to look at each endotype group separately in the first instance. It is important to note that each loop may not necessarily be present at each time point in each group, but more likely to be compared to HC. And as explored in Chapter 4, the presence of a loop does not mean that expression of a gene is increased, a loop may be inhibitory. Yet, understanding the pathways for which the genes in proximity to statistically significant loops exist is important. For all endotype groups, a Venn diagram was generated to show the number of loops that could stratify disease and HC at each time point (Figure 5.5, Figure 5.7, Figure 5.9). The Venn diagram allowed visualisation of the loops that stayed stable throughout the 12 months of csDMARD treatment and also revealed the loops that were unique to a particular time point. To focus the analysis, the loops in the centre of the Venn diagram, i.e. the most 'stable' loops were used for gene enrichment analysis. The closest 3 genes to each loop were used as an input for this analysis. This approach revealed that many of the same pathways were enriched in all endotype groups. However, each endotype group had some unique pathways enriched, further strengthening the idea that endotypes exist and pathogenesis may be different in each group. The overlap of significant pathways between endotypes is because we took the loops that were stable over 12 months in each group, but not unique to that endotype. Of note, common to all endotype groups was enrichment of 'phagosome', 'TCR signalling', 'translocation of zap70 to immunological synapse', 'rheumatoid arthritis', 'interferon gamma signalling' and 'MHC class II antigen presentation' pathways. These results are not surprising, as these pathways are known to be involved in RA pathogenesis.

Some unique pathways of interest to each group were 'haematopoietic cell lineage' in R, '*CLEC7A* and glucagon signalling' pathway in NR and 'HIF repressors' and 'cell adhesion molecules' in IR. *CLEC7a*, or Dectin-1 as it's otherwise known, has been shown to have increased expression in RA synovial tissues, potentially contributing to disease severity[374]. The enrichment of this pathway in NR suggests this pathway could be stopping the ability to respond to

treatment. Glucagon signalling in RA has been mostly explored in fibroblast-like synoviocytes, demonstrating that expression of glycogen metabolites can contribute to chronic inflammation[375]. Our understanding of this in PBMCs has yet to be fully explored. Again, this data implies enhanced inflammation could be a contributing factor to the non-response endotype. HIF is a transcription factor associated the hypoxic environment in RA joints. Interestingly, HIF repressors are known to facilitate repression through chromatin remodelling[376]. As this is a pathway we've found in the IR group, it alludes to an attempt of regulation in the underlying cellular biology.

It is intriguing that in all endotype groups, many of the pathways that are enriched are involved in metabolism. In recent years, immunometabolism in RA has gained increased interest[377]. Many metabolic checkpoints are now being considered as therapeutic targets[378].

Further to the pathway enrichment analysis, as with the approach in Chapter 4, we were interested to drill down on some loops and observe the epigenetic/genomic environment around these loops to begin to understand the possible consequence of loop formation (Figure 5.6, Figure 5.8, Figure 5.10). The EpiSwitch™ data portal platform facilitated such an investigation. To narrow down on loops we wanted to focus on, we used a network analysis approach to help determine some of the central genes involved in the pathways. Cytoscape software allowed central nodes to be identified, and this was used as a way of finding genes involved in many of the pathways enriched for each group. In contrast to the GeneAnalytics analysis, which showed many pathways were the same between endotype groups, network analysis produced a more unique dataset. Only *PPP2CA* and *WDR12* were common to all endotype groups. PPPC2A has been mainly implicated in systemic lupus erythematosus[379,380]. On the other hand, WDR12 polymorphisms have been implicated in cardiovascular events in RA patients[381].

The top genes unique to each group were BRCA1, TLR4, HLA-DBR1, associated with R, NR, IR, respectively. BRCA1 is a tumour suppressor and a mutation in BRCA1 is widely recognised as a risk factor for breast cancer, specifically a lifetime risk of 80%. Monitoring of this mutation could be considered precision

medicine in the oncology field[382] [383]. However, it's role in RA has not been extensively explored. One study has shown miR146a binds to the same site as the 3' UTR in the *BRCA1* gene and subsequently down-regulated the gene. Interestingly, mir146a has been shown to be differentially expressed in inflammatory disease, including RA[384]. As such, it is being considered as a potential new therapeutic target. Moreover, similar therapies are being considered to treat cancer and inflammatory disease, namely PARP inhibitors[385]. The ability to inhibit DNA repair has made these drugs considered for both indications. Based on these findings, this gene being significant in the R group in our study may suggest an enhanced inflammatory response mediated by *BRCA1* may contribute to good response to therapy.

The significance of TLR4 in the NR group is interesting. Activation of TLR4 is known to exacerbate RA through activation of serval components in the innate immune system[386]. Furthermore, blockade of TLR4 has been investigated as a drug target in RA[387]. Moreover, the endogenous TLR4 receptor agonist, MRP8/14 was shown to be a promising candidate for the prediction of biologic response in RA, with baseline serum levels correlated with HDA[388]. After defining the central gene, this was taken into EpiSwitch™ data portal to visualise other genetic and epigenetic features surrounding the loop. Some loops were revealed to be part of a 'hub' with multiple loops in the same region. Furthermore, some loop anchor points overlapped with H3K27ac marks indicating potential for enhanced gene transcription activity. Interestingly, disease associated SNPs, did not overlap with anchor points, although they were in close proximity to them. Evidence has suggested that the closer in proximity epigenetic elements are to each other, the more likely they will be to impact each other and influence gene regulation[389]. Further work could be conducted on this data to quantify the number of bases each anchor site is from a SNP.

Both R and NR groups had considerably more loops at the 6m time point, reflecting the results presented in Table 5.3. However, the IR group had most loops at the 12m time point and most loops were shared between the 6m and 12m time points. Moreover, only 4 loops were present at baseline alone, and only 9 common to baseline and 6m. These results are of great interest and allude to a major change in the epigenome after 6m. This is reasonable considering the

change from remission or LDA to MDA or HDA by 12m. This type of change has been mostly explored in the oncology field. Global changes to chromosome conformation have been shown to contribute to development of leukaemia, and thus a target for therapy[390]. However, there is evidence for RA drug therapies manipulating the epigenome, such as the effect of etanercept and adalimumab on multiple histones at the *CCL2* promotor site[391]. Moreover, a recent study has shown inhibition of histone enzymes could reduce cytokine production and osteoclastogenesis *in vitro*[392]. These data also allude to the effect of DMARD therapy on the IR group particularly. Furthermore, this indicates the importance of this group of significant loops in understanding loss of response in RA, as well as flare. A lot remains to be understood about RA flare, and like RA itself, it is a multifactorial and heterogeneous process. One of the most interesting loops from the group in IR was HLA-DBR1. This gene has been largely explored in RA but not in the context of flare or loss of response. The effect of HLA-DRB1 on susceptibility to RA has been widely described. It has been shown to contribute to radiographic progression and treatment response[16]. The *HLA-DRB1*13* allele has been shown to have protective effects in some stages of RA in ACPA positive patients[393].

Through this analysis, it was discovered that 183 loops were common to all 3 endotypes, at all 3 time points, creating a 'stable' RA signature (Figure 5.11). Pathways of interest enriched in this group were osteoclast differentiation, negative regulation of MAPK pathway and innate immune system. These pathways are well known to be involved in RA pathogenesis, with osteoclastogenesis showing evidence of mediation though the MAPK pathway[394]. Enrichment of these pathways indicates that these processes are likely driving pathogenesis. Pathways that have been explored less in RA and have appeared in this group were nitrogen metabolism and thyroxine production. A recent study investigated the role of the thyroid hormone network on RA synovial fibroblasts[395]. They found evidence to suggest TNFα may have a role in the degradation of thyroid hormones in the synovial environment. Studies of nitrogen metabolism date back decades and did not appear to have an effect of disease activity[396].

In the RA associated genes, the most connected node was *PSMC6*. This gene is an ATPase subunit and inhibition of this has been recently explored in colon cancer [397]. This is a gene previously has also been shown to be associated with juvenile idiopathic arthritis[398] and asthma. Upon visualisation, only 2 loops were in the region of PSMC6. Loops can be seen in proximity to GPR137C, ERO1A and GNPNAT1. Like loops in the different endotypes, anchor point overlap with H3K27ac marks observed in with T cells, B cells and NK cells. Similar to endotype loops, there was no direct SNP overlap with RA associated genes. ERO1A has been reported to have biomarker potential in pancreatic cancer. Data mined from microarrays suggested that expression of this gene was negatively correlated with poor prognosis. GPR137C encodes a G-protein coupled receptor, the downregulation of which has been implicated in several cancers[399]. Studies on these 3 genes in RA are limited, thus further work may be justified to understand their context in RA. The lack of studies in the literature which implicate PSMC6 in RA, may indicate a novel driver of disease has been found.

Following on from this analysis, the Searchlight platform was utilised compare findings and explore if other loops were found to be significant (Figure 5.13 – Figure 5.20). While RankProd and Searchlight used different angles to the analysis, with RankProd focusing on healthy and disease comparisons, and Searchlight comparing pairs of endotypes at all endpoints, they had the same aim of understanding the 3D epigenome underlying disease and differences between groups of patients. The RankProd and Searchlight data did show some broad similarities, but not at the individual loop level. Both analysis approaches identified 6m as the time point with the most significant stratifying loops. However, the loops found through each method map to different genes, which means that the results could be an artefact and the true epigenetic differences between endotypes remain unknown. Moreover, for all responder contrasts, Searchlight analysis did not identify significant differences at 12 months. This may suggest that results from the 12m time point using the RankProd pipeline should be interpreted with great caution.

There was some validation that the RankProd and Searchlight data were capturing comparable answers when the genes associated with most significant loops found in Searchlight analysis, were entered into the EpiSwitch™ data

portal and enrichment of loops were found (Figure 5.21). The 3 genes explored could be described as 'hubs' where multiple loops are present. This alludes to a high level of regulation of these genes. Regulatory hubs have been described in the literature, where they have also been referred to as 'cliques' or frequently interacting regions (FIRES)[400 401,402]. However, it is widely acknowledged that defining the role of these hubs in disease mechanisms remains a challenge, and additional studies are needed.

DOCK9 was clearly significantly different between endotype groups, and on inspection in the data portal, it appears this region is a hub of regulation with over 10 loops, and multiple other anchor points present in this region. These data suggest this is a highly regulated and important gene in RA. It has been previously identified as having increased expression in mouse models of RA[403]. Moreover, a SNP within the DOCK9 locus was shown to be significantly associated with RA in a North Indian cohort of RA patients[404]. Interestingly, compared with the other 2 genes visualised in this way, this region has no presence of SNPs. The other 2 genes investigated revealed 3 loops in the region, indicating a similar, yet slightly less of a regulatory region. In contrast to the DOCK9 region, these loops were in the region of SNPs. This visualisation of data provided a validation of the importance of Searchlight findings by identifying regulatory hubs which may impact pathogenesis. Overall, this demonstrates the valuable resource of the 3D epigenome and EpiSwitch™ platform to find novel genes which may drive disease or subtypes of disease.

Understanding the loop dynamics over the 12 months was of great interest to not only find candidates for a stable biomarker for predicting patient endotype, but also to understand loops that lose stratification potential at certain time points, which may reveal the influence of csDMARDs on the 3D epigenome (Figure 2.22 – Figure 2.24). Recent research has acknowledged determining stability of biomarkers in complex disease a challenge and has attempted the development of 2 assays to measure this[405]. We took the decision to use the QT clustering algorithm for our analysis. This was considered advantageous based on the quality control thresholds required to identify statistically significant findings, and the lack of need to identify the number of clusters prior to analysis. All 3 comparisons revealed similar clustering dynamics. All groups had at least 1

cluster with loops that were significant at stratifying disease and HC at all time points. This holds potential to take forward for further analysis for a potential biomarker. This clustering approach has been used previously to identify miRNA combinatorial biomarkers using breast cancer cell lines[406]. The algorithm has also demonstrated it's precision in the application of mass spectrometry[407]. As other data has shown, R and NR have similar proportions of stable and dynamic loops. IR was shown to have alternative dynamics, and this can also be seen with the clustering. Both R and NR had 9 clusters, and IR had only 8. Interestingly, IR had the largest number of loops that gain significance at 6m, shown in cluster 3 in the IR group. This data makes sense based on other findings that most loops are significant at 6m, with many loops significant and common to 6m and 12m time points.

There are several limitations that need to be considered with work in this chapter. While significant work and proprietary information was used to design an array that would capture meaningful data for RA; by designing the array, there was the chance that significantly stratifying loops could be missed. Another limitation was the lack of analysis into the loops significant in the HC. While disease loops were of most interest, understanding the loops that had stratification potential in pooled HC and did not have that ability in disease may indicate relevant RA pathogenesis information. Ultimately, this chapter was a discovery process and exploratory in nature, and many further exploratory and validation steps are required, some of which have already been discussed above.

There is a considerable amount of future investigation that could be carried forward from this work. Primarily, biomarker candidates should be taken forward for analysis. In the first instance this involves identifying a number of loops for PCR analysis from which loops can be statistically refined for a predictive endotype CCS. For further validation steps, multiple other samples would have to be sourced. In this instance, baseline samples alone would be sufficient, which may be available from the SERA cohort. Other cohorts could be sourced for this work, such as the TACERA cohort, used in the attempted validation the MTX CCS explored in Chapter 3. Furthermore, visualisation of epigenetic marks using EpiSwitch™ data portal was limited by the data that was included in the system, meaning there could, and very most likely would be

other epigenetic marks of interest that would shed more light on the potential consequences of loop formation, and association with each endotype. It should be noted, that since the generation of the data for this chapter, the data portal has been updated with more information, therefore offering the opportunity to explore our data further.

Recently published work has built on other studies suggesting that autoantibody positive and autoantibody negative RA are 2 distinct diseases and stratifying by presence of autoantibodies would help considerably for precision treatment[408]. Overall, around 50% of patients in each endotype group were autoantibody positive. It would therefore make sense to pre-stratify patients that way and work through the analysis pipeline to find if antibody presence effects stratification by chromosomal loops. One group has produced a bioinformatics framework to profile biomarkers in ACPA positive and negative patients[409]. This group revealed that differentially methylated regions were found between patients with opposite APCA status, as well as common differentially methylated regions between the patient groups. This mix of similarities and differences in the epigenome is similar to the results produced in this study and enforces the key the epigenome may play in ascertaining the underlying pathogenesis in RA. Their work was optimised for small samples of twins, so it may be applicable to small samples of RA patients, albeit not twins. This methodology also has the advantage of a deconvolution to account for the differences in epigenome between cell types, primarily T cells, NK cells and neutrophils. It would be worth further exploration for this work.

Furthermore, there is extensive other analyses that could be conducted for the data gathered. Each endotype group and time point revealed a long list of pathways and genes that may be relevant in disease. It would be interesting to explore more in depth the genes not already known to be associated with RA pathogenesis. Moreover, while the clusters of loops that retain significance over 12 months is important for biomarker discover, studying the loops that lose significance could reveal more about treatment influence on the epigenome. It would also be extremely valuable to determine transcription factor binding site locations in relation to the anchor points, and ultimately characterise the

transcription factors that bind there, to further understand the gene regulatory process in RA patients.

5.4 Conclusion

In conclusion, this chapter clearly demonstrates the complexity of RA heterogeneity and suggests that 3D epigenetic endotypes exist in the early RA population. Interrogation of the biological relevance of stratifying loops found known contributors to RA pathogenesis were more likely driving disease in some endotypes than others. Moreover, this analysis revealed genes that may be driving different endotypes, and RA as a whole. The data presented here provides a great basis for development of a CCS biomarker that could predict endotype at baseline and provides a chance to understand the complex pathogenesis further.

Chapter 6 General Discussion

6.1 Discussion

Overall, this thesis sought to investigate the stratification potential of chromosome conformation signatures (CCS) in rheumatoid arthritis (RA). In doing so, this work also aimed to use CCS to investigate the underlying pathogenesis of the disease. Therefore, this thesis aimed to strike a balance between clinically important experiments for the implementation of a biomarker and exploratory experiments that may allude to novel regulatory pathways of disease. In terms of the biomarker discovery pipeline, Chapter 3 demonstrated the validation stage of the process, Chapter 4, some exploration and Chapter 5 illustrated the discovery stage[410,411].

The first aim of the work in this thesis was to validate the methotrexate (MTX) CCS biomarker through bioinformatics approaches and in an independent clinical cohort. Results showed that in my hands, the biomarker could be validated computationally and replicated the high sensitivity and specificity scores that would be desirable for a biomarker[412]. Unfortunately, data gathered from testing the MTX CCS in an independent cohort of early RA patients suggests that it was not validated, and further exploratory work should be done. Whilst disappointing, it is not entirely surprising that it was not validated in the first SERA-independent cohort. There is the theory that differences in protocol could have contributed to the differences, however, until further work is done this cannot be concluded. Moreover, throughout this thesis, it has been demonstrated on a number of occasions that more precision medicine studies have been conducted in the oncology field than other fields. Yet, despite the concentrated work in this area, few biomarkers have made it into the clinic[413] [414]. The complex biomarker discovery process has also been reported for acute liver injury[415] and heart failure[416,417]. These studies exemplify the complexity of biomarker discovery and highlight the need for future work to ascertain the clinical potential for the MTX CCS.

Complexity of ascertaining the stratification ability of the MTX CCS was further demonstrated when testing in a cohort treated with several conventional

synthetic disease modifying anti-rheumatic drugs (csDMARDs). The data have shown that the 5-loop MTX CCS is specific for MTX alone. Our findings suggest that to achieve a clinically significant biomarker for all csDMARDs, a new biomarker will have to be found, likely by using a systematic approach similar to the discovery process of the MTX CCS. On reflection, as a highly systematic approach was employed to find the MTX CCS, it is very plausible that new CCS would need be developed for other therapies[219]. This has been demonstrated in prostate cancer, where dynamic chromatin conformation resulted in over expression of *UBE2C*, which could be targeted by several drugs including carvacrol[418] and Ipatasertib[419]. However, in the same disease, it was clear that other dysregulated regions within the genome had to be rectified with alternative therapies, such as cisplatin and niclosamide[420,421]. Ultimately, I believe that with a combination of the data from Carini *et al*[219], and the new data gathered here, a more specific, informative biomarker may be found. This integrated approach to biomarker discovery has been successfully demonstrated by Spiliopoulou *et al*[422].

Whilst understanding how well the MTX CCS could predict response to baseline csDMARDs, it was of interest to understand the relationship between the MTX CCS and underlying cellular biology. Based on evidence from Walsh *et al*[209] that there was a functional difference between responder (R) and (NR) loci, we hypothesized that the regulatory environment surrounding these loci may differ[209]. The exploratory approach in Chapter 4 revealed some evidence that the regulatory environment within and surrounding the loci of R and NR loops differed at the epigenetic level. Specifically, across cell types, loci where loops form in NR suggested an inhibitory environment, based on the presence of histone modifications that are associated with repression of gene expression. Conversely, at loci of loops that form in R, quantification of epigenetic marks suggested an environment that could enhance gene expression (Figure 6.1B). Evidence of the ability of molecular signatures to differentiate R and NR to RA therapy has been demonstrated in a recent study by Tao *et al*. They showed that transcription signatures in peripheral blood monocular cells (PBMCs) differed between R and NR to two biologic therapies, Adalimumab and Etanercept[423]. The caveat to our work was that it largely came from healthy samples, and it has yet

to be validated by other datasets, either existing, or yet to be created through novel wet-lab experiments.

Through observation of disease activity from the SERA cohort, various response trajectories were identified, and we hypothesized that 3D epigenetic endotypes exist in the early RA population (Figure 6.1A). We employed a systematic approach using a novel, custom microarray to ascertain if 3D epigenetic endotypes existed, and to identify candidates for a new CCS. The data shows that we were successful in identifying 3D epigenetic endotypes in the early RA population and we have statistically significant loops that can be taken forward for development of a new CCS. Precision medicine remains highly desirable in RA, with patients showing non-response to not only csDMARDs but also biologic (b)DMARDs. Moreover, a recent study concluded that there is a lack robust evidence on how to pharmacologically manage difficult-to-treat RA patients[424]. It could be speculated that many RA patients, if given the correct csDMARD at the start of treatment, would be subject to less joint damage, and therefore maybe more susceptible to responding to bDMARDs in the future. This exemplifies the importance of identifying the endotype of each patient as early as possible to facilitate appropriate clinical intervention.

An interesting finding from this data was the 'stable' RA loop profile that was found through our discovery microarray (Figure 6.1C). This suggested that a baseline level of dysregulation exists in all RA patients, regardless of 3D endotype. A recent study suggested that a baseline dysregulation of B cells exists, which may contribute to autoimmunity in RA[312]. There is limited data on B cells throughout this thesis, but data gathered from a PCHiC dataset, shown in Chapter 4, suggested there were significant promotor interactions at the *IFNAR1* loop site in B cells, which is suggestive of enhanced gene expression with loop formation in R. There was further evidence to suggest that histone marks associated with enhanced gene expression activity are present around other loci that were statistically significant in the 'RA' group. Our gene enrichment analysis suggested that genes in proximity to statistically significant EpiSwich™ loops were associated with regulation of the cytoplasm, nucleus and cytosol. As the transcription of mRNA into proteins through the nucleus to the cytoplasm is crucial for normal gene regulation and physiological function, it may suggest

that this is a contributor to the dysregulation in RA patients[425]. Interestingly, nucelo-cytoplasmic transport has recently been suggested to have a role in autoimmune neurodegenerative diseases[426]. Moreover, one of the most significant genes found through our analysis was *PSMC6*, a gene which encodes part of the ATPase subunit and is involved in regulation of the proteasome[427]. As such, regulation of this gene has implications on antigen presentation and the immune response. Notably, this is a gene that has had reported involvement in juvenile idiopathic arthritis (JIA), and asthma in Asian populations[428,429]. These findings suggest further investigation into this gene in RA cohorts is warranted.

It should also be considered that insight from other autoimmune diseases could be combined with our findings to provide a more informed picture of pathogenesis. Our data suggest this is plausible as statically significant stratifying loops were found in regions known to be implicated in other autoimmune diseases such as systematic lupus erythematosus (SLE)[379,380]. This is of research interest to many, evidenced by the formation of The Immune-Mediated Inflammatory Disease Biobanks in the UK (IMID-Bio-UK). This aims to bring together biobanks of clinical information and samples from patients with a host of autoimmune diseases including RA and SLE into one cohort. It is thought that shared pathology can be found through bringing this cohort together.

The strengths of the work detailed in Chapter 5 include the use of longitudinal samples[430,431,432]. Firstly, observing SERA patients past their 6-month time point revealed the initial responder (IR) endotype. Based on their disease activity, this group appear like the R group at 6 months. Moreover, having epigenetic data from 3 time points allowed the possibility of understanding the changing pathogenesis over time, and the ability to filter CCS candidates for the most statistically significant through time. Furthermore, this study used more patient samples per condition than in the study by Carini *et al*[219]. Whilst the sensitivity and specificity was not replicated with our analysis of the MTX CCS in an independent cohort, the work by Carini et al[219], the work is still promising. This provides confidence that with further work, an even more sensitive and specific biomarker will be found through our endotype data.

Another strength of this work as a whole is the use of peripheral blood, which is a major advantage for future clinical implementation. Recently, there has been a focus to transition to the synovium in search for predictive RA biomarkers[433]. However, this often involves using a needle biopsy to retrieve the synovial sample. Whilst these biomarkers may show promise, a biomarker found from the blood would rely on a less invasive procedure[434,435].

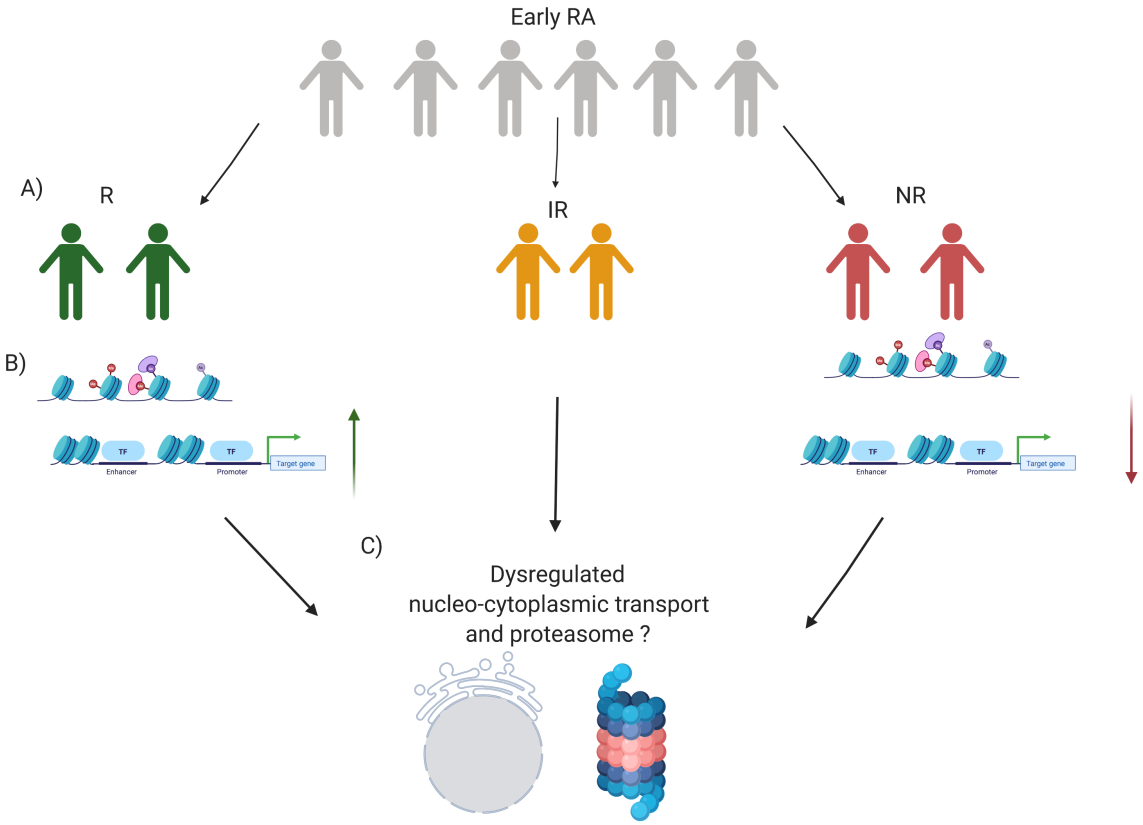


Figure 6.1 Suggested 3D Epigenetic Regulatory Differences and Similarities between RA Endotypes

Schematic representation of early RA 3D endotypes and the possible functional differences and similarities between them, collated from findings in Chapter 4 and Chapter 5. A) 3 response trajectories (R, NR, IR) from SERA cohort, suggestive of different endotypes. B) Epigenetic environment surround MTX CCS loci suggests possibility of increased gene expression in loci of R associated loops, and suppression of gene expression in NR associated loops. C) Custom microarray and analysis with Rankprod revealed 183 shared statistically significant loops between all endotypes. Gene enrichment analysis suggested these loops were in proximity to genes that were part of the nucleus, cytosol and cytoplasm pathways

6.2 Limitations

The work throughout this thesis has demonstrated the challenges of researching the 3D epigenome; this was most obvious in Chapter 3. The attempt to transition to a new platform to measure 3C loops was complex and involved many optimisation processes, which impacted on time. However, after optimisation, I was confident with the robust nature of the protocols.

PMBC heterogeneity is another limitation that has been discussed in all chapters. Due to the nature of sample collection, splitting cells was not possible. This is a well-recognised limitation, particularly when investigating the epigenome[436]. Studies in the RA field are now taking this into consideration in the sample collection process and separating cells into their subtypes before cryopreservation to facilitate more informative 3D epigenome analysis. Moreover, software is being developed to take into account this heterogeneity[437]. In contrast Liu *et al* and Glossop *et al* have argued that a mixed cell population would provide an overall accurate picture of the RA epigenetic profile[438,439].

Sample sizes for different parts of this work have been recognised as a limitation, particularly when assessing the stratification ability of the CCS as a pan-DMARD predictor. Having such a small sample size, with many patients on combination of csDMARDs, the predictive ability of HCQ and SSZ alone could not be determined.

6.3 Future Directions

The data described in this thesis, particularly the findings from Chapter 5, provide an exciting basis for the development of a baseline CCS that can predict endotype at baseline of RA treatment. Using the data discussed in Chapter 3 and Chapter 5, this can inform future collaborative studies with OBD to refine the CCS for RA.

It would also be useful to consider the use of clinical information in the biomarker model. Studies have shown markers of bone metabolism and signalling molecules could aid in the prediction of treatment response in RA[440,441].

Some of the findings from Chapter 4, which revealed different epigenetic landscapes in at the sites of R and NR loops, could be tested *in vitro* using appropriate cell lines and subsequently primary RA cells. Interestingly, some researchers are now modelling whole cells *in-silico*[442]. It is recognised that modelling cells in this way requires accurate knowledge of the biology, have accurate mathematics and an appropriate simulation platform. However, RA *in-vitro* models have been successfully established[443]. These models have ranged from 2D co-cultures and 3D cultures to the less common organ-on-a-chip[444,445]. [446]. ChIP-Seq is another accessible technique that could hopefully be used to replicate our findings. Asadipour *et al* have used ChIP-Seq successfully to demonstrate that that chromatin is accessible in monocytes and lymphocytes. Furthermore, ATAC-seq is a technique that has been developed in recent years and could complement this work. This is a technique that allows identification of open areas of chromatin throughout the genome[447]. This technique has been used to report regulatory landscape in $CD4^+$ T cells[448].

As a whole, the clinical application of precision medicine still has a number of barriers. One of the substantial barriers is the highly complex technologies and methods needed to interrogate the genome and identify the relevant genes contributing to drug response, as evidenced through the work in this thesis. To overcome this, there has been a rise in the computational methods used to interpret this and an increasing number of people with the desired skills[188]. It has been suggested that the burden of cost is not associated with the genomic technology itself, but in the interpretation of the data produced and the linking of this information with other patient characteristics to make this a relevant clinical biomarker. Other challenges include storage of the data that is produced from this research, issues with security and ownership of data and the cost to the healthcare system [187]. Additionally, as previously discussed, the era of precision medicine will see a rise in bio-banks that will be essential for biomarker studies. This will involve the public being on board with the concept and trusting that any samples donated to a bio-bank will be used appropriately.

As alluded to above, this thesis has approached experiments from an exploratory and clinically meaningful angle. Going forward, both angles should be

considered. Experiments such as HiC and other ‘C’ derivatives could provide a wealth of information that could be mined for biologically important findings. However, from a clinically implementable biomarker perspective, HiC would not be the best method economically. Moreover, 3C has been shown to have sufficient specificity and sensitivity, and can be successfully carried out over 48 hours; it is also financially realistic in a clinical setting. Ultimately, EpiSwitch™ CCS continue to demonstrate the applicability across a breadth of therapy areas, with a recent study demonstrating its ability in sports and exercise[449].

6.4 Conclusions

In summation, this work has clearly demonstrated the heterogeneity of RA at the clinical and epigenome level, and highlighted that study of the 3D epigenome may provide a novel opportunity to provide insight into RA. It has also shown the complexity, and at times difficulty in successfully examining the epigenome. Taken together, data has shown that using the EpiSwitch™ CCS platform provides an integrated view of gene regulation, providing a more informed picture than studying epigenetic modifications separately. This thesis has shown that RA endotypes exist at the 3D epigenome level and implies that there is dysregulation that underpins RA as a whole. Further work is warranted to take this data further and identify if a CCS can be developed that has the capacity to stratify treatment naive patients at baseline. It is hoped that this work will contribute to the more tailored treatment of RA patients in the future.

Appendix

Sequences for each loci for qPCR primer design

IL17A

ATTCATAATATGCATTTATAATTTGCTTAAAATATTGTGAAGACCTTTGTCTATGGAAAG
TACCTAAAAGGAAAGTGGATTAGTCAATTTTCACACTGCTATAAAGAACTACCTGAGACT
GAGTAATTTATAAAGAAAAGAGGTTTGATTGACTCATAGTTCCTCAAGGTTAGGGAGACC
TCAGGAAACTCAGAGTCACAGCAGAAGGGTAAGGGGGAAGTAAGGCATGTCTTACGTGGC
AGCAGGAGAGAAAGAGAGAATGAAGGGGGAAGTGCTACACCTTTAAACCATCGGATCTCT
TGAGAACTCACTATCACAAGAACAGCATGGGGGAAACTGCCCTCATAATCCAATCACCTC
CCACCAGATCCCTCCCTCAACATGCAGGGATTACAATTCGAAGATGGTCTGAAGGAAGCA
ATTGGGAAAAGCAAGCATAGCTCATCCAAGTGGGTGAGAGTGTGAGTTAGAGGAAGCTTG
GAAATTGGTGATGTGAGAGATGCTGCAGCTTCTGGGATTGCTGCCTGGTCGTGTGTAGAG
GAGGGGCAGTAGGGCTCATTCTGAATCTTGTCTTGAAAAGCACATAGATAGTGATGCCAA
AACCAGGACTACGGAAATCACTTGAAGCTGTATCCTACCTCCTCCTCCATCTGTATCTGC
TTCACCTATCAAGGATATCTACTATTGCCACTAAAATTCAGGTGCTTATGGCCTCCCTCA
TTCATCAGCCAGGGTTTATCTGGCCAGGAAAGAGAAGCCCCTTCAGGCATTTGCAACAGA
GGGAGTTTAAGTCAGAGA

CXCL13

CTTCCAGAGACAACAGGGCTTGTTTCAGATTTCATTGAAAGATCAAGATTTCCCTTCTAC
TAGCTTAAGCCTGCTCCCGAGGCACAGACAGCTTCTCCATGCCCCTGTCCTGAAGCTCTG
CCCCTCTTCTCCCACTCACCCGGCTGCATCTGGTTCCTCTAAGGCTGGAGCCCAAGGTGA
AGGTGAGGGCTGGAGTTCTAAGAGGCAGGAAAGGGCTTGGCTGGCTTTCAGGTAATCTCA
TGTTGGCTCCCTTGGGCCAGGGCAAGTGTTTGACTGACAGTCTGTCACGTGGGTTATTTG
AAGGCTCCCCGTAGGAATGCAAAGCAAACTGCTCTGACATGAGTGATGCGTTTTCTTTCC
AGTGGCTGCTTATATCTCCTACCTCCAAGCCTGGCAGTCGATTCCAAAGTGAAGCAAAAA
AAAAACTTCCCTATAATTTTCTCATTTTATGTATCTGATGATAAAAATCAGGTTTGATAT
AAACCAAGTTATGTATTTTTATGTTTCTAAACAAGGGTATCTATAAAATCTATTTTAAAA
ATAAATTGTTTCCATTTGAGTTTTAAATCTATTTGTTTAAATATTTTTTCTTCTTGTCAT
CAGAGTTTTTGAAATTTATACCTCTACCATAAGATATCATGTAGTGTTGGACTTGAAACT
TCTAGAATCATCTGGCTATCAGAGAGCTGCCCAACCTCCAGCCCCATCTTAGGGCAGACA
TGAATAAGGCTTCTTTGGTCTAAATAAGCTAAATCCAGAGCTCTGGGCATTAAAGAAACT
GAAGTCAGATTTTAGAAT

IL21R

GCAAATTCCTCTGGCTCCTACCACATGCCCCCAGATGCCCCCAGTTGCCCAGGCACCCAG
CGTTGTAAGGGGATGTGGGATGACTCCATCTTGCTTATCCCTGGCCAGCTCATTCTCCCC
AGTCAGGGCCCCCCTCAAATTTTCCCTCCTTGGAGAGCCTTATCCTGGTCCCCCTTGCAA
AAAGCACCCCAGAAGATCTCTGGCCTGCACCCTGGTCTAGTGCATTCAGAGAGTGGGTCA
CTATGGAAAACCACCTGTTTGTGATTGATATGTATTTTCAGTCTCCCCAGTAGGATGCCG
GTTTTAGAAATGTGGAGTCCTGGTTGGGTGCAGTGGCTCACACCTGTAATCCCAGCACTT
TGGGAGGCTGAGGCAGGCAGATCATGAGGTCAGGAGTTCGAGCCCTGGACCCCAGGCCAG
CTAATGAGGCTGACTGGATGTCCAGAGCTGGCCAGGCCACTGGGCCCTGAGCCAGAGACA
AGGTCACCTGGGCTGTGATGTGAAGACACCTGCAGCCTTTGGTCTCCTGGATGGGCCTTT
GAGCCTGATGTTTACAGTGTCTGTGTGTGTGTGTGCATATGTGTGTGTGTGCATATGCAT
GTGTGTGTGTGTGTGTGTCTTAGGTGCGCAGTGGCATGTCCACGTGTGTGTGTGATTGCA
CGTGCCTGTGGGCCTGGGATAATGCCCATGGTACTCCATGCATTCACCTGCCCTGTGCAT
GTCTGGACTCACGGAGCTCACCCATGTGCACAAGTGTGCACAGTAAACGTGTTTGTGGTC
AACAGATGACAACAGCCG

IL23

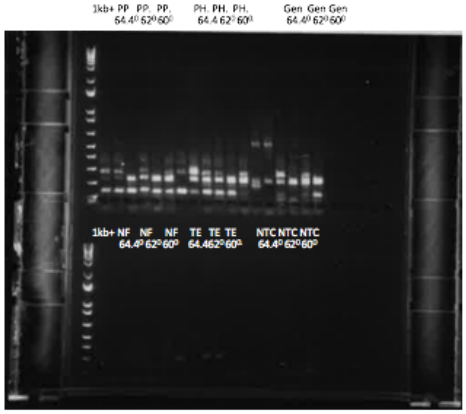
GAAGCGCAGTAGAAAGGTGCCAGACATGGTCTTCTTCAGCAGCCGGCGCTCCTGGCTCCG
ACTCACAAAGCCCATGATGCGTCTGGAGCACAGAGAGCAGCTGTGAGACACCGCCCAACA
CCCTGCCCCACCAGGCCCCTGCCTCCCTGCTCCCCTTGTATGGAGAAACAGCCCAGGTTT
GGAATCCAGGCTTTGGACCAAGCCCTGCCACTTAGCTTTTTTTTATCTTTTACAAGTCAC
TTCACTTTGCTGACAATAGCCACCTTCCGAGGGTTTGAAGTACGAAGATTCAGTGAGATC
ACACAGATGAACAGCACTTTTTTTTTTTTCTGGGAGACAGGGTGTCATTCTGTCACCCAG
GCTGGAGAATAGTGGCATGATCACAGCTCACTGCCACCTCGAAACCAAACCCTGTGACTT
CAACACCCAAACATTTACGTTCTTACGTTTTACTGTACAGCCTTCCTCCTGGATAAATCA
GATTAAAATTTTGAAGACATGTATATAGCTTGTCAGGGTAGTACAGTTATTAATTCTGTG
GTTACCACATTCTTTTATATCTCAAGACTCCCAGCCCCAGGATCTAAGTCATAGCTCTTG
ATTATGGCCCACCCCCAGTAGGGAGCTGAACTTACTACTTCTGATATGAAAGAAGCCAGA
GTAGTTGTTTCTTCCAAGTCACTCACATCTGAGATGGCCCTCAAACCCTCCTTGGTGGTC
TCCGGCTGAGAGATGTTTGCGTCTTTCTGATCTTGGTAATACTGGATCAACTGCTGAACC
CCAGACCTTAATGCAGGTT

IFNAR1

CTTAGAAATCGCAGGCCTCCTGCATTCTCTTCTTCCTCACTCTTTCCGGCCTATGGCCGG
TCTGGAGCACTTCACCATCGCATTCCAGGACGACTCCCCCTCACAGGAGGCCTTGACCTT
CACAGGATCGCGGCACACACCTGCGCAGACCCAGCGACTTTGGAGTTCCCGCGTCTCTTC
CCCATGTCTGCAGTCGGAGGGCGCGGGTGGGCTCCTGGTCGCGGCCCGGCGCTCTTCGTC
CTCGCCCTTCTCCGCCCGGGGCGCTTGCTGACCCTAGAGTCTTGGACAGTGGCGTTGGGA
GTCCCCAGCGGCCTGGGAACCAGATGAGGCTGGAAGGAGGAAGTGGCTGAGCGACCAGAC
CGTTGCTGTGTGCAGAGCGAGAGCGGGGCAGAGGCGGTCGAAACTGGGAGAATTCATCTG
AAATGATTATTTATGAAGTTCAAGTATATATTCACCCTGAGCAAGAAGAGTTTATTTGAT
TTCATTAAAATCATAACTGCATGGAAAACAAGTTGGTAAAATAAAGCATTTGTGTTGCAC
GTGTTTTCCTCTAAAATTAAACTGGGATTATACCCAAAAGCTGCCTTCTGTAACCAGCTC
TGTACCAGTTTGGCTGGTATCTACTTCCTGAAACTCACATCTTTATGACAATAATCATCA
AGAATAAATACATGAAGCAAAGCTGGGTCTTCCTAGTTCTGTTGCTGCTTGTCACTTCAG
GGAAAGGAAGTTGAAAGAAAGAAGACCGAGTATATTTGTTCATCCTACAAACTGTGTGAG
TGTATTGTATGCCCTATA

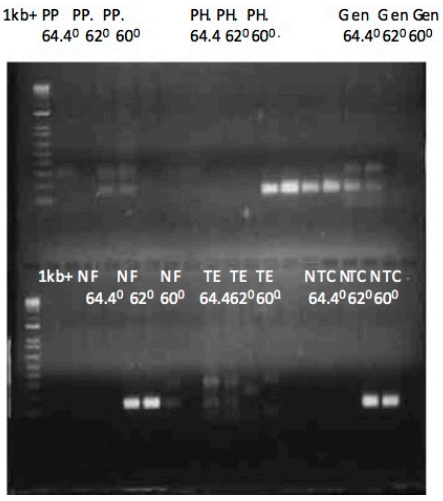
Representative gels from qPCR

Version A (from *IFNAR1* 60°C)



Version B (from *IL23* 62°C)

Version C (From *IFNAR1* 60°C)



Representative Scripts

Limma

```
library(limma)

targets_RAall <- readTargets("targets_RA.txt")
#reading the targets to run analysis as single channel
targets_RAall2 <- readTargets("targets_RA_2.txt")

# find the files to read in and load them with the same order as the targets file.
files <- unlist(lapply(targets_RAall$FileName, function(x) list.files("./", pattern =
x, full.names = T)))
array_data_read <- utils::capture.output(array_data <-
limma::read.maimages(files, columns = list(G = "gMedianSignal", Gb =
"gBGMedianSignal",
                                                                    R =
"rMedianSignal",Rb = "rBGMedianSignal"),
                                                          annotation =
c("ControlType","ProbeName","SystematicName", "PValueLogRatio"),
                                                          source = "agilent"))

######################## 1. Remove agilent control probes
#####################
# Remove agilent control probes and output the number of probes removed and
remain in the log file. ####
rmcnrl <-which(array_data$genes$ControlType == 0)
agilentcnrl <-which(array_data$genes$ControlType != 0)
gb <-array_data[rmcnrl,]


######################## 2. Remove Saturated probes ####################
# Remove probes that have a saturated signal above 65525 ####
satSignal <- 65525

#Find the index of the Red table that has a signal over the satSignal for every
file.
indexR <-unlist(apply(gb$R, 2, function(x) which(x>=satSignal)))
#Find the index of the Green table that has a signal over the satSignal for every
file.
indexG <- unlist(apply(gb$G, 2, function(x) which(x>=satSignal)))
# Create the union of the two indexes.
indexRG <- union(indexR, indexG)
# Keep the elements of the RGList (R or G) that do not have signal values over
the satSignal
if (length(indexRG)>0){
  gbNoSatNew<-gb[-indexRG,]
}else{
  gbNoSatNew<-gb
}
```

```
gb <- gbNoSatNew

######################## 3. Background_correction ######################
# Call the backgroundCorrect function of the limma package to background
correct the expression intensities ####
rg <- limma::backgroundCorrect(gb, method="normexp", offset=50)


######################## 4. Normalise_Within_Arrays
#############################
# Call normalizeWithinArrays function of the limma package to normalize the
expression log-ratios with ####
# the loess method, so that the log-ratios average to zero within each array.
####
RGq<- limma::normalizeWithinArrays(rg, method="loess")


######################## 6.Quality_checks
#######################################
#Unlog the M values
processed_data <- RGq
unlog_M <- 2^(processed_data$M)
processed_data$M <- unlog_M

# create the pca plot
pca <- stats::princomp(stats::na.omit(processed_data$M))

# export and save the plot in the svg graphics device
svg("./Slide_PCA.svg")
plot(pca$loadings[,1],pca$loadings[,2],pch=19,cex=0.5,col=as.factor(targets$Slid
e))
text(pca$loadings[,1], pca$loadings[,2], labels=as.factor(targets$Slide), pos=3,
offset=0.22, cex=0.6)
title("PCA plot of M values for Slide!")
dev.off()

#just produce the plot

lab <- as.factor(targets_RAall$Group)
pcaVar <- round((pca$sdev^2 / sum(pca$sdev^2)),4)*100
plot(pca$loadings[, 1], pca$loadings[, 2], pch = 19, cex = 0.5, col = lab,
    xlab=paste0("PC1 (",pcaVar[1],"%)"), ylab = paste0("PC2 (",pcaVar[2],"%)"))
graphics::text(pca$loadings[, 1], pca$loadings[, 2], labels = lab, pos = 3, offset =
0.22, cex = 0.7)
title(paste0("PCA plot of M values for ", title))


######################## 7. Merge Probes section ####################

source("/Users/caitlinduncan/Desktop/PhD docs/DATA/duplicateCalculation.R")
list_of_matrices <- duplicateCalculation(MAdata = processed_data)

# Export the logMedianMatrix matrix ####
```

```
fwrite(list_of_matrices$logMedianMatrix,
      file = "logMedianMatrix_M_values.txt",
      sep = "\t", col.names = T,
      row.names = T, quote = F,
      eol = "\n", na = "NA", dec = ".")

##########################################
If single channel option is selected, individual channel processing will follow

# Split data into individual G & R channels ####
RGt <-limma::RG.MA(RGq)
split_data_RAall <-(matrix(c(RGt$G,RGt$R),ncol=size <- (length(files)*2)))
rownames(split_data_RAall)=rownames=RGt$genes$ProbeName

split_data_RAall <-as.data.frame(split_data_RAall)
colnames(split_data_RAall)=targets_RAall2$Group
split_data_RAall[,"SystematicName"] <-RGt$genes$SystematicName

list_of_Splitmatrices <- duplicateCalculation(split_data_RAall)

# Export the logMedianMatrix matrix ####
fwrite(list_of_Splitmatrices$logMedianMatrix,
      file = "logMedianMatrix_Channel_values.txt",
      sep = "\t", col.names = T,
      row.names = T, quote = F,
      eol = "\n", na = "NA", dec = ".")

######################## 8. Perform the limma Contrasts

# Use the log median data for the analysis
logMedianMatrix_s <- list_of_Splitmatrices$logMedianMatrix

#create the design of the analysis
design<-model.matrix(~0+factor(targetsR2$Cy3))
colnames(design) <- c("RA", "HC")

#create the contrasts.matrix
contrast.matrix <- makeContrasts(RA-HC,levels=design)

#Fit linear model for each gene or attribute, based on the array files given
linear_model<- lmFit(logMedianMatrix_s, design = design)

#Based on the above linear model fit to microarray data, compute estimated
coefficients and standard errors for a given set of contrasts.
contrast_model <- contrasts.fit(linear_model,contrast.matrix)

#Given the microarray linear model fit, compute moderated t-statistics,
#moderated F-statistic, and log-odds of differential expression by empirical
Bayes moderation
ebay_model <- eBayes(contrast_model)
head(coef(ebay_model))
```

```
#################
# OR to add extra filtration you can also do:
#################
#1. Filter with p.value<=0.05 (or adj.P.Val)
contrasts <- contrasts[which(contrasts$P.Value<=0.05),]


#2. Set a cutOff of 1.1 and -1.1 by setting the LS column to 1 or -1 respectively and remove the probes with in between values!
contrasts$LS <- ifelse(contrasts$FC_1>=1.1,1,ifelse(contrasts$FC_1<=-1.1,-1,0))
Informative <- which(contrasts$LS!=0)
contrasts <- contrasts[Informative,]

#4. Calculate the absolute Fold change and sort the table by that column (abs).
contrasts <- contrasts[order(contrasts$FC_1, decreasing = T),]

write.csv (contrasts, file = ".csv")
```

Bedtools

```
bedtools closest [OPTIONS] -a <FILE> \
                    -b <FILE1, FILE2, ..., FILEN>
```

```
bedtools intersect [OPTIONS] -a <FILE> \
                     -b <FILE1, FILE2, ..., FILEN>
```

DeepBlue Epigenome

```
#Install and download necessary packages
install.packages (“BiocManager”)
BiocManager:: install(“DeepBlueR”)
library(“DeepBlueR”)
#request desired experimental data
experiment = deepblue_list_experiments(type="peaks", epigenetic_mark="H3K27ac", biosource=c("peripheral blood mononuclear cell", project="ENCODE"))
#retrieve names of experiments
experiment
#get data on specified experiments
query_id=deepblue_select_experiments(experiment_name=c("", ""))
```

```
request_id = deepblue_count_regions(query_id=query_id)
requested_data = deepblue_download_request_data(request_id=request_id)

#get data on specific experiments, and at the genome region of interest
query_id = deepblue_select_experiments (experiment_name=c(“”
 chromosome=“chr6”, start=52161697, end=52172165)

#get the regions the epigenomic mark is in, in a readable format
request_id = deepblue_get_regions(query_id=query_id,
output_format="CHROMOSOME,START,END,SIGNAL_VALUE,PEAK,@NAME,@BIOSO
URCE")

regions = deepblue_download_request_data(request_id=request_id)

#list regions
regions

#export data as bed file, to folder of choice
deepblue_export_bed(regions, target.directory =
"/Users/caitlinduncan/Desktop/DeepBlue_Bed_files", file.name =
"311019_H3K27ac_peripheralbloodmononuclearcell_IL17Aa_hg38”)
#export data as tab file, to folder of choice
deepblue_export_tab(regions, target.directory =
"/Users/caitlinduncan/Desktop/DeepBlue_tab_files", file.name =
"311019_H3K27ac_peripheralbloodmononuclearcell_IL17Aa_hg38”)
```

Coordinates for DeepBlue:

hg38
“chr6”, start=52161697, end=52172165
“chr6”, start=52184632, end= 52187067
“chr4”, start= 77510412, end= 77512093
“chr4”, start= 77602625, end= 77605433
“chr16”, start= 27356311, end= 27357534
“chr16”, start= 27449257, end= 27451508
“chr12”, start= 56345719, end= 56347275
“chr12”, start= 56361069, end= 56361825
“chr21”, start= 33324378, end= 33325411
“chr21”, start= 33373955, end= 33376515

hg19
“chr6”, start=52026495, end= 52036963
“chr6”, start= 52049430, end= 52051865
“chr4”, start= 78431566, end= 78433247
“chr4”, start= 78523779, end= 78526587
“chr16”, start= 27367632, end= 27368855
“chr16”, start= 27460578, end= 27462829
“chr12”, start= 56739503, end= 56741059
“chr12”, start= 56754853, end= 56755609
“chr21”, start=34696683, end= 34697716
“chr21”, start=34746261, end= 34748821

hg19 – 500kb upstream
"chr6", start= 51526495, end= 52026495
"chr4", start= 77931566, end=78431566
"chr16", start= 26867632, end= 27367632
"chr12", start=56239503, end= 56739503
"chr21", start= 34196683, end= 34696683

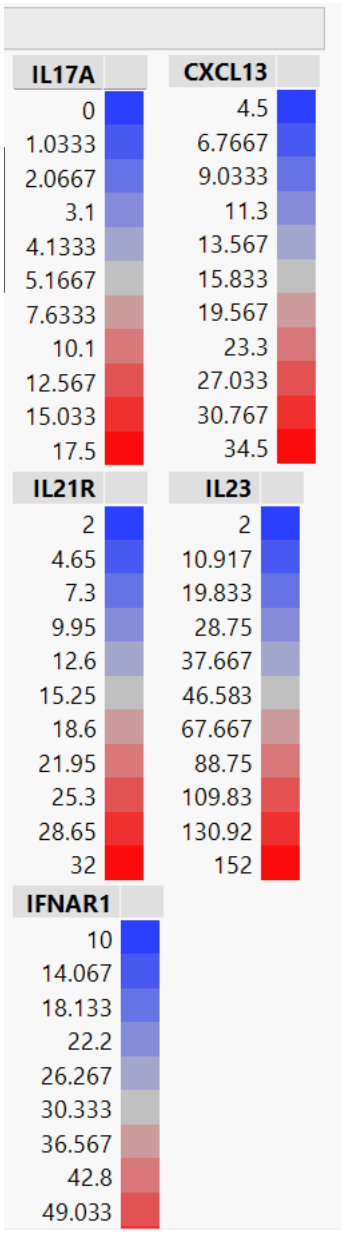
hg19 – 500kb downstream
"chr6", start= 52051865, end= 52551865
"chr4", start= 78526587, end= 79026587
"chr16", start= 27462829, end= 27962829
"chr12", start= 56755609, end= 57255609
"chr21", start=34748821, end= 35248821

hg38 – 500kb upstream
"chr6", start= 51661697, end= 52161697
"chr4", start= 77010412, end= 77510412
"chr16", start= 26856311, end= 27356311
"chr12", start= 55845719, end= 56345719
"chr21", start= 32824378, end= 33324378

hg38 – 500kb downstream
"chr6", start= 52187067, end= 52687067
"chr4", start= 77605433, end= 78105433
"chr16", start= 27451508, end= 27951508
"chr12", start= 56361825, end= 56861825
"chr21", start= 33376515, end= 33876515

Enrichment values for histones

PBMCs



$CD4^+$ T cells

IL17A	CXCL13
1	4.3
1.82	5.2267
2.64	6.1533
3.46	7.08
4.28	8.0067
5.1	8.9333
7.82	10.507
10.54	12.08
13.26	13.653
15.98	15.227
18.7	16.8

IL21R	IFNAR1
7.1	6.8
9.26	8.6067
11.42	10.413
13.58	12.22
15.74	14.027
17.9	15.833
20.54	18.267
23.18	20.7
25.82	23.133
28.46	25.567
31.1	28

IL17A	CXCL13
1.8	6.1
4.3033	7.96
6.8067	9.82
9.31	11.68
11.813	13.54
14.317	15.4
17.053	18.12
19.79	20.84
22.527	23.56
25.263	26.28
28	29

IL21R	IFNAR1
2.4	1.3
4.0167	3.39
5.6333	5.48
7.25	7.57
8.8667	9.66
10.483	11.75
14.987	14.56
19.49	17.37
23.993	20.18
28.497	22.99
33	25.8

$CD14^+CD16^-$ Monocytes

CXCL13	IL23	IFNAR1
4.2	1.5	6.6
5.12	6.67	10.167
6.04	11.84	13.733
6.96	17.01	17.3
7.88	22.18	20.867
8.8	27.35	24.433
10.14	31.66	29.367
11.48	35.97	34.3
12.82	40.28	39.233
14.16	44.59	44.167
15.5	48.9	49.1

IL23	CXCL13	IFNAR1
4.1	3.3	1.3
7.3767	5.33	5.42
10.653	7.36	9.54
13.93	9.39	13.66
17.207	11.42	17.78
20.483	13.45	21.9
23.527	17.36	26.66
26.57	21.27	31.42
29.613	25.18	36.18
32.657	29.09	40.94
35.7	33	45.7

ChiCAGO score (used to determine significance in PCHiC dataset)

Let x_i denote the measured value of a quantitative property (such as CHiCAGO interaction score or gene expression level) for cell type $i \in \mathbf{I}$. Then, the specificity score s_c for a given cell type $c \in \mathbf{I}$ is a weighted mean of the differences $x_c - x_i$ for $i \neq c$,

where the weights $d_{c,i}$ are distances between cell type c and cell types i, calculated using the complete dataset (e.g., CHiCAGO interaction scores for all interactions or expression values for all genes; distances calculated using Euclidean distance metric).

$$s_c = \frac{1}{\sum_{i \neq c} d_{c,i}} \sum_{i \neq c} d_{c,i} (x_c - x_i)$$

Raw GeneAnalytics data - R

Results

Score	SuperPath Name	# SuperPath Total Genes	# SuperPath Matche	Matched Genes (Syr	Evidence URL
13.16	Phagosome	152	16	FCGR2C, CTSS, STX7, TU	http://pathcards.genecards.org/card/phagosome
12.02	Translocation of ZAP-70 to Immunological Synapse	46	8	CSK, PTPN22, HLA-DQA:	http://pathcards.genecards.org/card/translocation_of_zap-70_to_immunological_synapse
11.22	TCR Signaling (REACTOME)	122	13	CSK, PSMB4, PTPN22, H	http://pathcards.genecards.org/card/tcr_signaling_(reactome)
10.48	Pentose Phosphate Pathway	30	6	PFKP, FBP2, GPI, FBP1, I	http://pathcards.genecards.org/card/pentose_phosphate_pathway
9.55	G-protein Signaling N-RAS Regulation Pathway	60	8	CTSS, ITPR1, HLA-DQA2,	http://pathcards.genecards.org/card/g-protein_signaling_n-ras_regulation_pathway
9.35	Toxoplasmosis	141	13	FCGR2C, LAMC1, LAMC:	http://pathcards.genecards.org/card/toxoplasmosis
9.08	Tuberculosis	180	15	CALML4, FCGR2C, CTSS,	http://pathcards.genecards.org/card/tuberculosis
8.70	ICos-ICosL Pathway in T-Helper Cell	131	12	CSK, TXK, ITPR1, HLA-DC	http://pathcards.genecards.org/card/icos-icosl_pathway_in_t-helper_cell
8.34	Nitrogen Metabolism	17	4	CA1, CA13, CA2, CA3	http://pathcards.genecards.org/card/nitrogen_metabolism
7.96	Complement and Coagulation Cascades	88	9	KLKB1, F2, PROCR, CR2,	http://pathcards.genecards.org/card/complement_and_coagulation_cascades
7.57	Interferon Gamma Signaling	202	15	GBP2, KPNA3, GBP1, GB	http://pathcards.genecards.org/card/interferon_gamma_signaling
6.56	MHC Class II Antigen Presentation	103	9	CTSS, HLA-DQA2, HLA-D	http://pathcards.genecards.org/card/mhc_class_ii_antigen_presentation
6.50	CD28 Co-stimulation	86	8	CSK, HLA-DQA2, HLA-DC	http://pathcards.genecards.org/card/cd28_co-stimulation
6.30	CTLA4 Signaling	71	7	PTPN22, HLA-DQA2, HL	http://pathcards.genecards.org/card/ctla4_signaling
6.24	Lysosomal Oligosaccharide Catabolism	5	2	MANBA, MAN2B2	http://pathcards.genecards.org/card/lysosomal_oligosaccharide_catabolism
5.89	Rheumatoid Arthritis	93	8	ITGB2, HLA-DQA2, HLA-	http://pathcards.genecards.org/card/rheumatoid_arthritis
5.74	Pentose Phosphate Pathway (Erythrocyte), Pharmacodynamics	6	2	GPI, PGD	http://pathcards.genecards.org/card/pentose_phosphate_pathway_(erythrocyte)_pharmacodynamics
5.65	Staphylococcus Aureus Infection	96	8	FCGR2C, ITGB2, HLA-DC	http://pathcards.genecards.org/card/staphylococcus_aureus_infection
5.51	Fanconi Anemia Pathway	62	6	RAD51C, FAN1, MLH1, R	http://pathcards.genecards.org/card/fanconi_anemia_pathway
5.42	Hematopoietic Cell Lineage	99	8	CR2, HLA-DQA2, CR1, CI	http://pathcards.genecards.org/card/hematopoietic_cell_lineage

Results

Score	Name	# Genes	# Matched Genes	Matched Genes (Syn	Evidence URL	Ontology
15.40	T Cell Receptor Signaling Pathway	179	19	BTNL3, CSK, TXK, PSMB4	http://amigo.geneontology.org/amigo/term/GO:0050852	GoBiolProc
13.78	Antigen Processing and Presentation of Peptide or Polysaccharide Antigen Via MHC Class II	12	5	HLA-DQA2, HLA-DQB2, H	http://amigo.geneontology.org/amigo/term/GO:0002504	GoBiolProc
13.59	Microtubule Cytoskeleton Organization	148	16	PPP2R3C, DCLK2, CALML	http://amigo.geneontology.org/amigo/term/GO:0000226	GoBiolProc
10.16	Fructose 6-phosphate Metabolic Process	12	4	PFKP, FBP2, FBP1, GFPT1	http://amigo.geneontology.org/amigo/term/GO:0006002	GoBiolProc
10.00	Antigen Processing and Presentation	44	7	CTSS, HLA-DQA2, HLA-DC	http://amigo.geneontology.org/amigo/term/GO:0019882	GoBiolProc
9.88	Interferon-gamma-mediated Signaling Pathway	72	9	GBP2, GBP1, JAK1, HLA-I	http://amigo.geneontology.org/amigo/term/GO:0060333	GoBiolProc
8.96	One-carbon Metabolic Process	37	6	CA1, DHFR, CA13, CA2, C	http://amigo.geneontology.org/amigo/term/GO:0006730	GoBiolProc
8.78	Negative Regulation of Transposon Integration	2	2	ZNF91, ZNF93	http://amigo.geneontology.org/amigo/term/GO:0070895	GoBiolProc
8.78	Histidyl-tRNA Aminoacylation	2	2	HARS1, HARS2	http://amigo.geneontology.org/amigo/term/GO:0006427	GoBiolProc
8.78	Sucrose Biosynthetic Process	2	2	FBP2, FBP1	http://amigo.geneontology.org/amigo/term/GO:0005986	GoBiolProc
8.58	Endoplasmic Reticulum Organization	39	6	CAV2, PEX5, LNPK, ATL2,	http://amigo.geneontology.org/amigo/term/GO:0007029	GoBiolProc
8.46	Negative Regulation of Coagulation	8	3	PROCR, ANXA5, ANXA4	http://amigo.geneontology.org/amigo/term/GO:0050819	GoBiolProc
7.99	Fructose 1,6-bisphosphate Metabolic Process	9	3	PFKP, FBP2, FBP1	http://amigo.geneontology.org/amigo/term/GO:0030388	GoBiolProc
7.76	ER to Golgi Vesicle-mediated Transport	199	15	ERGIC2, F2, COPG2, SPTI	http://amigo.geneontology.org/amigo/term/GO:0006888	GoBiolProc
7.65	Synaptic Vesicle to Endosome Fusion	3	2	EEA1, VAMP4	http://amigo.geneontology.org/amigo/term/GO:0016189	GoBiolProc
7.65	Carnitine Transmembrane Transport	3	2	SLC22A4, SLC22A5	http://amigo.geneontology.org/amigo/term/GO:1902603	GoBiolProc
7.65	Positive Regulation of Mitotic Cell Cycle Phase Transition	3	2	KLHL18, TMOD3	http://amigo.geneontology.org/amigo/term/GO:1901992	GoBiolProc
7.65	Positive Regulation of Hormone Biosynthetic Process	3	2	ARNT, HIF1A	http://amigo.geneontology.org/amigo/term/GO:0046886	GoBiolProc
7.57	Posttranscriptional Gene Silencing By RNA	10	3	TNRC6C, TNRC6B, AGO2	http://amigo.geneontology.org/amigo/term/GO:0035194	GoBiolProc
7.44	Nucleotide-excision Repair	46	6	GTF2H5, FAN1, STN1, AT	http://amigo.geneontology.org/amigo/term/GO:0006289	GoBiolProc
16.28	ER to Golgi Transport Vesicle Membrane	61	11	SREBF2, HLA-DQA2, SEC	http://amigo.geneontology.org/amigo/term/GO:0012507	GoCellComp
14.71	Trans-Golgi Network Membrane	95	13	MMP24, AP4B1, SCAMP	http://amigo.geneontology.org/amigo/term/GO:0032588	GoCellComp
14.40	MHC Class II Protein Complex	18	6	HLA-DQA2, HLA-DQB1, H	http://amigo.geneontology.org/amigo/term/GO:0042613	GoCellComp
13.88	Transport Vesicle Membrane	49	9	ITPR1, HLA-DQA2, HLA-D	http://amigo.geneontology.org/amigo/term/GO:0030658	GoCellComp
12.44	Clathrin-coated Endocytic Vesicle Membrane	33	7	HLA-DQA2, HLA-DQB1, H	http://amigo.geneontology.org/amigo/term/GO:0030669	GoCellComp
10.73	Integral Component of Lumenal Side of Endoplasmic Reticulum Membrane	29	6	HLA-DQA2, HLA-DQB1, H	http://amigo.geneontology.org/amigo/term/GO:0071556	GoCellComp
10.38	Endoplasmic Reticulum Tubular Network Membrane	5	3	LNPK, RTN4, ATL2	http://amigo.geneontology.org/amigo/term/GO:0098826	GoCellComp
9.74	Specific Granule	13	4	ANXA11, ANXA3, STXBP:	http://amigo.geneontology.org/amigo/term/GO:0042581	GoCellComp
9.05	Golgi Membrane	661	38	COPG2, CAV2, GBP2, TRI	http://amigo.geneontology.org/amigo/term/GO:0000139	GoCellComp
8.32	Transport Vesicle	101	10	COPG2, CAV2, TRIP11, SI	http://amigo.geneontology.org/amigo/term/GO:0030133	GoCellComp
7.65	CHRAC	3	2	CHRAC1, BAZ1A	http://amigo.geneontology.org/amigo/term/GO:0008623	GoCellComp
7.41	Lysosome	414	25	CXCR4, CTSS, ADA, SLC1	http://amigo.geneontology.org/amigo/term/GO:0005764	GoCellComp
7.19	Azurophil Granule	11	3	ANXA11, STX7, SNAP23	http://amigo.geneontology.org/amigo/term/GO:0042582	GoCellComp
6.85	Tertiary Granule	4	2	STXBP3, STX7	http://amigo.geneontology.org/amigo/term/GO:0070820	GoCellComp
6.62	Extrinsic Component of Plasma Membrane	24	4	USP8, EEA1, USP50, AAK	http://amigo.geneontology.org/amigo/term/GO:0019897	GoCellComp
6.24	Endolysosome Lumen	5	2	CTSS, CTSK	http://amigo.geneontology.org/amigo/term/GO:0036021	GoCellComp
5.75	RISC Complex	16	3	SND1, AGO2, EIF4E	http://amigo.geneontology.org/amigo/term/GO:0016442	GoCellComp
5.74	Sec61 Translocon Complex	6	2	SEC61B, SEC61A2	http://amigo.geneontology.org/amigo/term/GO:0005784	GoCellComp
5.57	Myofibril	45	5	LRRC10, PSMA6, TMOD3	http://amigo.geneontology.org/amigo/term/GO:0030016	GoCellComp
5.57	Vacuolar Membrane	30	4	VMP1, SLC36A3, SLC36A	http://amigo.geneontology.org/amigo/term/GO:0005774	GoCellComp
15.01	MHC Class II Receptor Activity	10	5	HLA-DQA2, HLA-DQB1, H	http://amigo.geneontology.org/amigo/term/GO:0032395	GoMolecFunc
11.70	Non-membrane Spanning Protein Tyrosine Phosphatase Activity	9	4	ACP1, PTPN22, PTPN12,	http://amigo.geneontology.org/amigo/term/GO:0004726	GoMolecFunc
9.63	Arylesterase Activity	6	3	CA1, APMAP, CA2	http://amigo.geneontology.org/amigo/term/GO:0004064	GoMolecFunc
9.00	Sialic Acid Transmembrane Transporter Activity	7	3	SLC17A3, SLC35A1, SLC1	http://amigo.geneontology.org/amigo/term/GO:0015136	GoMolecFunc
8.78	Fructose 1,6-bisphosphate 1-phosphatase Activity	2	2	FBP2, FBP1	http://amigo.geneontology.org/amigo/term/GO:0042132	GoMolecFunc
8.78	Histidine-tRNA Ligase Activity	2	2	HARS1, HARS2	http://amigo.geneontology.org/amigo/term/GO:0004821	GoMolecFunc
8.65	Carbonate Dehydratase Activity	16	4	CA1, CA13, CA2, CA3	http://amigo.geneontology.org/amigo/term/GO:0004089	GoMolecFunc
7.94	Receptor Tyrosine Kinase Binding	57	7	CPNE3, YWHAG, NRG1, I	http://amigo.geneontology.org/amigo/term/GO:0030971	GoMolecFunc
7.65	Carnitine Transmembrane Transporter Activity	3	2	SLC22A4, SLC22A5	http://amigo.geneontology.org/amigo/term/GO:0015226	GoMolecFunc
7.65	Amino Acid:proton Symporter Activity	3	2	SLC36A3, SLC36A2	http://amigo.geneontology.org/amigo/term/GO:0005280	GoMolecFunc
7.57	Phospholipase Inhibitor Activity	10	3	ANXA5, ANXA3, ANXA4	http://amigo.geneontology.org/amigo/term/GO:0004859	GoMolecFunc
7.52	Hydro-lyase Activity	20	4	CA1, CA13, CA2, CA3	http://amigo.geneontology.org/amigo/term/GO:0016836	GoMolecFunc
7.28	L-ascorbic Acid Binding	21	4	P4HA1, ALKBH3, PAM, P	http://amigo.geneontology.org/amigo/term/GO:0031418	GoMolecFunc
6.85	Quaternary Ammonium Group Transmembrane Transporter Activity	4	2	SLC22A4, SLC22A5	http://amigo.geneontology.org/amigo/term/GO:0015651	GoMolecFunc
6.85	Guanine/thymine Mispair Binding	4	2	MLH1, MSH3	http://amigo.geneontology.org/amigo/term/GO:0032137	GoMolecFunc
6.54	Endodeoxyribonuclease Activity	13	3	ERCC5, BIVM, BIVM-ERC	http://amigo.geneontology.org/amigo/term/GO:0004520	GoMolecFunc
6.54	Mechanosensitive Ion Channel Activity	13	3	TMC8, TMC1, TMC6	http://amigo.geneontology.org/amigo/term/GO:0008381	GoMolecFunc
6.26	AMP Binding	14	3	PFKP, FBP1, ACSS1	http://amigo.geneontology.org/amigo/term/GO:0016208	GoMolecFunc
6.24	Procollagen-proline 4-dioxygenase Activity	5	2	P4HA1, P4HA2	http://amigo.geneontology.org/amigo/term/GO:0004656	GoMolecFunc
6.24	Complement Component C3b Binding	5	2	ITGB2, CR1	http://amigo.geneontology.org/amigo/term/GO:0001851	GoMolecFunc

Raw GeneAnalytics data – NR

Results

Score	Name	# Genes	# Matched Genes	Matched Genes (S	Evidence URL	Ontology
13.48	Interferon-gamma-mediated Signaling Pathway	72	13	GBP2, CAMK2G, PRKCI	http://amigo.geneontology.org/amigo/term/GO:0060333	GoBiolProc
12.93	Somatic Recombination of Immunoglobulin Gene Segments	5	4	MLH1, MSH6, MSH3, U	http://amigo.geneontology.org/amigo/term/GO:0016447	GoBiolProc
10.17	Endoplasmic Reticulum Organization	39	8	CAV2, PEX5, ATL1, LNP	http://amigo.geneontology.org/amigo/term/GO:0007029	GoBiolProc
9.81	Immunoglobulin Production Involved in Immunoglobulin Mediated Immune Response	4	3	FGL2, HLA-DQB1, HLA	http://amigo.geneontology.org/amigo/term/GO:0002381	GoBiolProc
9.81	Endoplasmic Reticulum Tubular Network Membrane Organization	4	3	ATL1, RTN4, ATL2	http://amigo.geneontology.org/amigo/term/GO:1990809	GoBiolProc
9.21	Negative Regulation of Natural Killer Cell Mediated Cytotoxicity	10	4	MICA, LGALS9, HLA-G,	http://amigo.geneontology.org/amigo/term/GO:0045953	GoBiolProc
9.09	Antigen Processing and Presentation	44	8	CTSS, HLA-B, HLA-DRA,	http://amigo.geneontology.org/amigo/term/GO:0019882	GoBiolProc
8.90	Sialic Acid Transport	5	3	SLC17A4, SLC17A3, SL	http://amigo.geneontology.org/amigo/term/GO:0015739	GoBiolProc
8.27	Antigen Processing and Presentation of Peptide or Polysaccharide Antigen Via MHC Class II	12	4	HLA-DRA, HLA-DRB1, H	http://amigo.geneontology.org/amigo/term/GO:0002504	GoBiolProc
8.27	Fructose 6-phosphate Metabolic Process	12	4	PFKP, FBP2, FBP1, GFP	http://amigo.geneontology.org/amigo/term/GO:0006002	GoBiolProc
7.77	Dimethylallyl Diphosphate Biosynthetic Process	2	2	IDI2, IDI1	http://amigo.geneontology.org/amigo/term/GO:0050992	GoBiolProc
7.77	Regulation of Interleukin-4 Production	2	2	LGALS9, HLA-DRB1	http://amigo.geneontology.org/amigo/term/GO:0032673	GoBiolProc
7.77	Sucrose Biosynthetic Process	2	2	FBP2, FBP1	http://amigo.geneontology.org/amigo/term/GO:0005986	GoBiolProc
7.77	Protein Glycosylation in Endoplasmic Reticulum	2	2	ALG2, AQP11	http://amigo.geneontology.org/amigo/term/GO:0033577	GoBiolProc
7.77	Isopentenyl Diphosphate Biosynthetic Process	2	2	IDI2, IDI1	http://amigo.geneontology.org/amigo/term/GO:0009240	GoBiolProc
7.77	Regulation of Kidney Size	2	2	LRRK2, ASXL1	http://amigo.geneontology.org/amigo/term/GO:0035564	GoBiolProc
7.77	MAPK Export From Nucleus	2	2	STYX, DUSP16	http://amigo.geneontology.org/amigo/term/GO:0045204	GoBiolProc
7.54	Immune Response	506	39	EXOSC9, CCR5, MICA,	http://amigo.geneontology.org/amigo/term/GO:0006955	GoBiolProc
7.50	Kinetochore Assembly	14	4	MIS12, PDG2, CENPK,	http://amigo.geneontology.org/amigo/term/GO:0051382	GoBiolProc
7.33	Nucleotide-excision Repair, DNA Incision, 3'-to Lesion	23	5	GTF2H1, GTF2H5, ERC	http://amigo.geneontology.org/amigo/term/GO:0006295	GoBiolProc
18.13	ER to Golgi Transport Vesicle Membrane	61	14	PDCD6, SREBF1, SREBF	http://amigo.geneontology.org/amigo/term/GO:0012507	GoCellComp
12.97	Integral Component of Lumenal Side of Endoplasmic Reticulum Membrane	29	8	HLA-B, HLA-DRA, HLA-	http://amigo.geneontology.org/amigo/term/GO:0071556	GoCellComp
12.93	Endoplasmic Reticulum Tubular Network Membrane	5	4	ATL1, LNPK, RTN4, ATL	http://amigo.geneontology.org/amigo/term/GO:0098826	GoCellComp
10.86	Specific Granule	13	5	ANXA11, ANXA3, STX3	http://amigo.geneontology.org/amigo/term/GO:0042581	GoCellComp
9.76	Trans-Golgi Network Membrane	95	13	MMP24, AP4B1, SCAM	http://amigo.geneontology.org/amigo/term/GO:0032588	GoCellComp
8.81	MHC Class II Protein Complex	18	5	HLA-DRA, HLA-DQB1, I	http://amigo.geneontology.org/amigo/term/GO:0042613	GoCellComp
8.72	Azurophil Granule	11	4	ANXA11, STX3, STX7, S	http://amigo.geneontology.org/amigo/term/GO:0042582	GoCellComp
8.15	Transport Vesicle Membrane	49	8	ITPR1, HLA-DRA, HLA-I	http://amigo.geneontology.org/amigo/term/GO:0030658	GoCellComp
7.94	Endoplasmic Reticulum Membrane	1016	70	ZDHHC6, PIGP, ERGIC2	http://amigo.geneontology.org/amigo/term/GO:0005789	GoCellComp
7.92	Golgi Membrane	661	49	PDCD6, COPG2, CAV2,	http://amigo.geneontology.org/amigo/term/GO:0000139	GoCellComp
7.88	Cytosol	5213	296	KIF3B, FANK1, ERRFI1,	http://amigo.geneontology.org/amigo/term/GO:0005829	GoCellComp
7.77	Interleukin-18 Receptor Complex	2	2	IL18R1, IL18RAP	http://amigo.geneontology.org/amigo/term/GO:0045092	GoCellComp
7.77	Integrin Alpha10-beta1 Complex	2	2	ITGB1, ITGA10	http://amigo.geneontology.org/amigo/term/GO:0034680	GoCellComp
7.24	Clathrin-coated Endocytic Vesicle Membrane	33	6	HLA-DRA, HLA-DQB1, I	http://amigo.geneontology.org/amigo/term/GO:0030669	GoCellComp
6.97	Golgi Apparatus	1366	88	ZDHHC6, PPP2R3C, ERI	http://amigo.geneontology.org/amigo/term/GO:0005794	GoCellComp
6.96	Perinuclear Region of Cytoplasm	720	51	RAD51C, RAPGEF1, CA	http://amigo.geneontology.org/amigo/term/GO:0048471	GoCellComp
6.65	Plus-end Kinesin Complex	3	2	KIF3B, BORCS5	http://amigo.geneontology.org/amigo/term/GO:0005873	GoCellComp
6.56	RNA Polymerase II, Holoenzyme	9	3	POLR2M, GCOM1, RPF	http://amigo.geneontology.org/amigo/term/GO:0016591	GoCellComp
6.29	External Side of Plasma Membrane	404	31	CCR5, F2, MICA, CXCR4	http://amigo.geneontology.org/amigo/term/GO:0009897	GoCellComp
6.27	Endoplasmic Reticulum Tubular Network	18	4	REEP1, ATL1, LNPK, RT	http://amigo.geneontology.org/amigo/term/GO:0071782	GoCellComp
13.28	Natural Killer Cell Lectin-like Receptor Binding	9	5	MICA, ULBP2, ULBP3,	http://amigo.geneontology.org/amigo/term/GO:0046703	GoMolecFunc
11.73	Peptide Antigen Binding	33	8	HLA-B, HLA-DRA, HLA-	http://amigo.geneontology.org/amigo/term/GO:0042605	GoMolecFunc
11.10	Sialic Acid Transmembrane Transporter Activity	7	4	SLC17A4, SLC17A3, SL	http://amigo.geneontology.org/amigo/term/GO:0015136	GoMolecFunc
9.81	Guanine/thymine Mispair Binding	4	3	MLH1, MSH6, MSH3	http://amigo.geneontology.org/amigo/term/GO:0032137	GoMolecFunc
9.77	Non-membrane Spanning Protein Tyrosine Phosphatase Activity	9	4	ACP1, PTPN22, PTPN1	http://amigo.geneontology.org/amigo/term/GO:0004726	GoMolecFunc
9.54	NAD(P)+ Nucleosidase Activity	16	5	IL18R1, IL18RAP, IL1RL	http://amigo.geneontology.org/amigo/term/GO:0050135	GoMolecFunc
9.54	NAD+ Nucleotidase, Cyclic ADP-ribose Generating	16	5	IL18R1, IL18RAP, IL1RL	http://amigo.geneontology.org/amigo/term/GO:0061809	GoMolecFunc
9.21	MHC Class II Receptor Activity	10	4	HLA-DRA, HLA-DQB1, I	http://amigo.geneontology.org/amigo/term/GO:0032395	GoMolecFunc
8.16	Hydro-lyase Activity	20	5	CA1, CA13, L3HYPDH,	http://amigo.geneontology.org/amigo/term/GO:0016836	GoMolecFunc
7.77	Isopentenyl-diphosphate Delta-isomerase Activity	2	2	IDI2, IDI1	http://amigo.geneontology.org/amigo/term/GO:0004452	GoMolecFunc
7.77	Cytosine C-5 DNA Demethylase Activity	2	2	ALKBH3, ALKBH2	http://amigo.geneontology.org/amigo/term/GO:0051747	GoMolecFunc
7.77	Interleukin-18 Receptor Activity	2	2	IL18R1, IL18RAP	http://amigo.geneontology.org/amigo/term/GO:0042008	GoMolecFunc
7.77	Fructose 1,6-bisphosphate 1-phosphatase Activity	2	2	FBP2, FBP1	http://amigo.geneontology.org/amigo/term/GO:0042132	GoMolecFunc
7.59	Chemokine Receptor Activity	22	5	CCR5, CXCR4, CCR3, C	http://amigo.geneontology.org/amigo/term/GO:0004950	GoMolecFunc
7.33	C-C Chemokine Receptor Activity	23	5	CCR5, CXCR4, CCR3, C	http://amigo.geneontology.org/amigo/term/GO:0016493	GoMolecFunc
7.15	Monosaccharide Binding	15	4	ALPK1, PFKP, GPI, FBP	http://amigo.geneontology.org/amigo/term/GO:0048029	GoMolecFunc
7.08	C-C Chemokine Binding	24	5	CCR5, CXCR4, CCR3, C	http://amigo.geneontology.org/amigo/term/GO:0019957	GoMolecFunc
6.84	Carbonate Dehydratase Activity	16	4	CA1, CA13, CA2, CA3	http://amigo.geneontology.org/amigo/term/GO:0004089	GoMolecFunc
6.65	Intramolecular Transferase Activity	3	2	GPI, LSS	http://amigo.geneontology.org/amigo/term/GO:0016866	GoMolecFunc
6.65	Amino Acid:proton Symporter Activity	3	2	SLC36A3, SLC36A2	http://amigo.geneontology.org/amigo/term/GO:0005280	GoMolecFunc

Results

Score	SuperPath Name	# SuperPath Total Genes	# SuperPath Matched Genes	Matched Genes (Sym	Evidence URL
11.71	Phagosome	152	19	FCGR2C, CTSS, STX7, ITGB	http://pathcards.genecards.org/card/phagosome
8.70	Translocation of ZAP-70 to Immunological Synapse	46	8	CSK, PTPN22, HLA-DRA, H	http://pathcards.genecards.org/card/translocation_of_zap-70_to_immunological_synapse
8.47	Rheumatoid Arthritis	93	12	ITGB2, HLA-DRA, ATP6V1	http://pathcards.genecards.org/card/rheumatoid_arthritis
8.24	Interferon Gamma Signaling	202	20	GBP2, CAMK2G, PRKCD, I	http://pathcards.genecards.org/card/interferon_gamma_signaling
7.89	Pentose Phosphate Pathway	30	6	PFKP, FBP2, GPI, FBP1, PG	http://pathcards.genecards.org/card/pentose_phosphate_pathway
7.52	Toxoplasmosis	141	15	CCR5, FCGR2C, LAMC1, Li	http://pathcards.genecards.org/card/toxoplasmosis
7.20	ICos-ICosL Pathway in T-Helper Cell	131	14	PPP3CA, CSK, CAMK2G, PI	http://pathcards.genecards.org/card/icos-icosl_pathway_in_t-helper_cell
6.78	TCR Signaling (REACTOME)	122	13	CSK, PSMB4, PTPN22, HL	http://pathcards.genecards.org/card/tcr_signaling_(reactome)
6.66	Tuberculosis	180	17	PPP3CA, CALML4, FCGR2C	http://pathcards.genecards.org/card/tuberculosis
6.54	Nitrogen Metabolism	17	4	CA1, CA13, CA2, CA3	http://pathcards.genecards.org/card/nitrogen_metabolism
6.48	G-protein Signaling N-RAS Regulation Pathway	60	8	CTSS, SOS2, ITPR1, HLA-D	http://pathcards.genecards.org/card/g-protein_signaling_n-ras_regulation_pathway
6.37	Immune Response Role of DAP12 Receptors in NK Cells	199	18	PPP3CA, MICA, SOS2, NFA	http://pathcards.genecards.org/card/immune_response_role_of_dap12_receptors_in_nk_cells
6.16	Metabolic States and Circadian Oscillators	10	3	CLOCK, LDHC, LDHA	http://pathcards.genecards.org/card/metabolic_states_and_circadian_oscillators
6.04	CLEC7A (Dectin-1) Signaling	145	14	PPP3CA, MUC21, PRKCD,	http://pathcards.genecards.org/card/clec7a_(dectin-1)_signaling
6.02	MHC Class II Antigen Presentation	103	11	KIF3B, CTSS, KLC1, HLA-DI	http://pathcards.genecards.org/card/mhc_class_ii_antigen_presentation
5.77	3-phosphoinositide Degradation	20	4	PMPCA, INPP5J, INPP5B,	http://pathcards.genecards.org/card/3-phosphoinositide_degradation
5.75	Glucagon Signaling Pathway	106	11	PPP3CA, PFKP, FBP2, CALI	http://pathcards.genecards.org/card/glucagon_signaling_pathway
5.69	Deadenylation-dependent MRNA Decay	93	10	EXOSC9, EXOSC7, PFKP, C	http://pathcards.genecards.org/card/deadenylation-dependent_mrna_decay
5.67	Antigen Processing-Cross Presentation	121	12	CTSS, SNAP23, PSMB4, LM	http://pathcards.genecards.org/card/antigen_processing-cross_presentation
5.55	Regulation of Cholesterol Biosynthesis By SREBP (SREBF)	55	7	GPAM, SREBF1, SREBF2, I	http://pathcards.genecards.org/card/regulation_of_cholesterol_biosynthesis_by_srebp_(srebf)

Raw GeneAnalytics data – IR

Results

Score	SuperPath Name	# SuperPath Total Genes	# SuperPath Matched G	Matched Genes (Sy	Evidence URL
15.10	Translocation of ZAP-70 to Immunological Synapse	46	8	CSK, PTPN22, HLA-DQA	http://pathcards.genecards.org/card/translocation_of_zap-70_to_immunological_synapse
13.51	TCR Signaling (REACTOME)	122	12	CSK, PTPN22, HLA-DQA	http://pathcards.genecards.org/card/tcr_signaling_(reactome)
12.47	G-protein Signaling N-RAS Regulation Pathway	60	8	CTSS, ITPR1, HLA-DQA2	http://pathcards.genecards.org/card/g-protein_signaling_n-ras_regulation_pathway
11.62	Toxoplasmosis	141	12	FCGR2C, LAMC1, LAMC	http://pathcards.genecards.org/card/toxoplasmosis
10.67	Phagosome	152	12	FCGR2C, CTSS, STX7, ITI	http://pathcards.genecards.org/card/phagosome
10.31	Interferon Gamma Signaling	202	14	GBP2, KPNA3, GBP1, G	http://pathcards.genecards.org/card/interferon_gamma_signaling
10.23	Tuberculosis	180	13	CALML4, FCGR2C, CTSS	http://pathcards.genecards.org/card/tuberculosis
9.97	Nitrogen Metabolism	17	4	CA1, CA13, CA2, CA3	http://pathcards.genecards.org/card/nitrogen_metabolism
9.43	MHC Class II Antigen Presentation	103	9	CTSS, HLA-DQA2, HLA-D	http://pathcards.genecards.org/card/mhc_class_ii_antigen_presentation
9.13	CD28 Co-stimulation	86	8	CSK, HLA-DQA2, HLA-D	http://pathcards.genecards.org/card/cd28_co-stimulation
8.66	CTLA4 Signaling	71	7	PTPN22, HLA-DQA2, HL	http://pathcards.genecards.org/card/ctla4_signaling
8.45	Rheumatoid Arthritis	93	8	ITGB2, HLA-DQA2, HLA	http://pathcards.genecards.org/card/rheumatoid_arthritis
8.17	Staphylococcus Aureus Infection	96	8	FCGR2C, ITGB2, HLA-DC	http://pathcards.genecards.org/card/staphylococcus_aureus_infection
7.19	ICos-ICosL Pathway in T-Helper Cell	131	9	CSK, ITPR1, HLA-DQA2,	http://pathcards.genecards.org/card/icos-icosl_pathway_in_t-helper_cell
7.06	Pentose Phosphate Pathway	30	4	PFKP, FBP2, FBP1, PGD	http://pathcards.genecards.org/card/pentose_phosphate_pathway
6.13	Cell Adhesion Molecules (CAMs)	148	9	CLDN4, ITGB2, HLA-DQ	http://pathcards.genecards.org/card/cell_adhesion_molecules_(cams)
5.72	Peginterferon Alpha-2a/Peginterferon Alpha-2b Pathway (Hepatocyte), Pharmacodynamics	22	3	JAK1, PIAS1, MAVS	http://pathcards.genecards.org/card/peginterferon_alpha-2apeginterferon_alpha-2b_pathway_(hepatocyte)_pharmacodynamics
5.72	Regulation of Apoptosis By Parathyroid Hormone-related Protein	22	3	BCL2L13, BCL2L15, BID	http://pathcards.genecards.org/card/regulation_of_apoptosis_by_parathyroid_hormone-related_protein
5.40	HIF Repressors	24	3	SUMO3, ARNT, HIF1A	http://pathcards.genecards.org/card/hif_repressors
5.39	Th17 Cell Differentiation	162	9	JAK1, HLA-DQA2, HIF1	http://pathcards.genecards.org/card/th17_cell_differentiation

Results

Score	Name	# Genes	# Matched Genes	Matched Genes (S	Evidence URL	Ontology
15.90	Antigen Processing and Presentation of Peptide or Polysaccharide Antigen Via MHC Class II	12	5	HLA-DQA2, HLA-DQB2,	http://amigo.geneontology.org/amigo/term/GO:0002504	GoBiolProc
13.79	T Cell Receptor Signaling Pathway	179	15	BTNL3, CSK, PTPN22, I	http://amigo.geneontology.org/amigo/term/GO:0050852	GoBiolProc
13.10	Interferon-gamma-mediated Signaling Pathway	72	9	GBP2, GBP1, JAK1, HLA	http://amigo.geneontology.org/amigo/term/GO:0060333	GoBiolProc
12.66	Antigen Processing and Presentation	44	7	CTSS, HLA-DQA2, HLA-	http://amigo.geneontology.org/amigo/term/GO:0019882	GoBiolProc
11.83	Fructose 6-phosphate Metabolic Process	12	4	PFKP, FBP2, FBP1, GFP	http://amigo.geneontology.org/amigo/term/GO:0006002	GoBiolProc
10.85	Endoplasmic Reticulum Organization	39	6	CAV2, PEX5, LNPK, ATI	http://amigo.geneontology.org/amigo/term/GO:0007029	GoBiolProc
9.73	Negative Regulation of Coagulation	8	3	PROCR, ANXA5, ANXA	http://amigo.geneontology.org/amigo/term/GO:0050819	GoBiolProc
9.66	Sucrose Biosynthetic Process	2	2	FBP2, FBP1	http://amigo.geneontology.org/amigo/term/GO:0005986	GoBiolProc
9.66	Negative Regulation of Transposon Integration	2	2	ZNF91, ZNF93	http://amigo.geneontology.org/amigo/term/GO:0070895	GoBiolProc
9.25	Fructose 1,6-bisphosphate Metabolic Process	9	3	PFKP, FBP2, FBP1	http://amigo.geneontology.org/amigo/term/GO:0030388	GoBiolProc
8.52	Carnitine Transmembrane Transport	3	2	SLC22A4, SLC22A5	http://amigo.geneontology.org/amigo/term/GO:1902603	GoBiolProc
8.52	Positive Regulation of Mitotic Cell Cycle Phase Transition	3	2	KLHL18, TMOD3	http://amigo.geneontology.org/amigo/term/GO:1901992	GoBiolProc
8.52	Positive Regulation of Hormone Biosynthetic Process	3	2	ARNT, HIF1A	http://amigo.geneontology.org/amigo/term/GO:0046886	GoBiolProc
8.43	Peptidyl-proline Hydroxylation to 4-hydroxy-L-proline	11	3	P4HA1, P4HA2, ERO1A	http://amigo.geneontology.org/amigo/term/GO:0018401	GoBiolProc
8.39	Nucleotide-excision Repair, DNA Incision, 3'-to Lesion	23	4	GTF2H5, ERCC5, BIVM,	http://amigo.geneontology.org/amigo/term/GO:0006295	GoBiolProc
8.00	Antigen Processing and Presentation of Exogenous Peptide Antigen Via MHC Class II	98	8	CTSS, HLA-DQA2, HLA-	http://amigo.geneontology.org/amigo/term/GO:0019886	GoBiolProc
7.76	Regulation of Interferon-gamma-mediated Signaling Pathway	13	3	JAK1, PTPN2, PIAS1	http://amigo.geneontology.org/amigo/term/GO:0060334	GoBiolProc
7.71	Negative Regulation of Platelet-derived Growth Factor Receptor-beta Signaling Pathway	4	2	PTPN12, PTPN2	http://amigo.geneontology.org/amigo/term/GO:2000587	GoBiolProc
7.71	Immunoglobulin Production Involved in Immunoglobulin Mediated Immune Response	4	2	HLA-DQB1, HLA-DRB1	http://amigo.geneontology.org/amigo/term/GO:0002381	GoBiolProc
7.71	Quaternary Ammonium Group Transport	4	2	SLC22A4, SLC22A5	http://amigo.geneontology.org/amigo/term/GO:0015697	GoBiolProc
20.52	ER to Golgi Transport Vesicle Membrane	61	11	SREBF2, HLA-DQA2, SE	http://amigo.geneontology.org/amigo/term/GO:0012507	GoCellComp
17.37	Transport Vesicle Membrane	49	9	ITPR1, HLA-DQA2, HLA	http://amigo.geneontology.org/amigo/term/GO:0030658	GoCellComp
16.96	Trans-Golgi Network Membrane	95	12	MMP24, AP4B1, SCAM	http://amigo.geneontology.org/amigo/term/GO:0032588	GoCellComp
16.91	MHC Class II Protein Complex	18	6	HLA-DQA2, HLA-DQB1	http://amigo.geneontology.org/amigo/term/GO:0042613	GoCellComp
15.23	Clathrin-coated Endocytic Vesicle Membrane	33	7	HLA-DQA2, HLA-DQB1,	http://amigo.geneontology.org/amigo/term/GO:0030669	GoCellComp
13.12	Integral Component of Lumenal Side of Endoplasmic Reticulum Membrane	29	6	HLA-DQA2, HLA-DQB1	http://amigo.geneontology.org/amigo/term/GO:0071556	GoCellComp
12.18	Golgi Membrane	661	33	COPG2, CAV2, GBP2, G	http://amigo.geneontology.org/amigo/term/GO:0000139	GoCellComp
9.85	Lysosome	414	22	CXCR4, CTSS, ADA, SLC	http://amigo.geneontology.org/amigo/term/GO:0005764	GoCellComp
7.71	Tertiary Granule	4	2	STXBP3, STX7	http://amigo.geneontology.org/amigo/term/GO:0070820	GoCellComp
7.37	Golgi Apparatus	1366	50	PPP2R3C, COPG2, CAV	http://amigo.geneontology.org/amigo/term/GO:0005794	GoCellComp
7.34	Myofibril	45	5	LRRC10, PSMA6, TMOI	http://amigo.geneontology.org/amigo/term/GO:0030016	GoCellComp
7.09	Endoplasmic Reticulum Tubular Network Membrane	5	2	LNPK, ATL2	http://amigo.geneontology.org/amigo/term/GO:0098826	GoCellComp
7.09	Endolysosome Lumen	5	2	CTSS, CTSK	http://amigo.geneontology.org/amigo/term/GO:0036021	GoCellComp
6.92	Endocytic Vesicle Membrane	68	6	HLA-DQA2, HLA-DQB1	http://amigo.geneontology.org/amigo/term/GO:0030666	GoCellComp
6.24	Endosome Membrane	252	13	APPL2, ANTXR1, HLA-D	http://amigo.geneontology.org/amigo/term/GO:0010008	GoCellComp
6.02	Transport Vesicle	101	7	COPG2, CAV2, SCAMP	http://amigo.geneontology.org/amigo/term/GO:0030133	GoCellComp
5.88	PML Body	103	7	SUMO3, NR2C1, CALCO	http://amigo.geneontology.org/amigo/term/GO:0016605	GoCellComp
5.81	Extrinsic Component of Endosome Membrane	8	2	USP8, USP50	http://amigo.geneontology.org/amigo/term/GO:0031313	GoCellComp
5.32	Seh1-associated Complex	1	1	SEH1L	http://amigo.geneontology.org/amigo/term/GO:0035859	GoCellComp
5.32	Proximal Portion of Axoneme	1	1	DNAH11	http://amigo.geneontology.org/amigo/term/GO:0120134	GoCellComp
17.16	MHC Class II Receptor Activity	10	5	HLA-DQA2, HLA-DQB1,	http://amigo.geneontology.org/amigo/term/GO:0032395	GoMolecFunc
13.41	Non-membrane Spanning Protein Tyrosine Phosphatase Activity	9	4	ACP1, PTPN22, PTPN1	http://amigo.geneontology.org/amigo/term/GO:0004726	GoMolecFunc
10.92	Arylesterase Activity	6	3	CA1, APMAP, CA2	http://amigo.geneontology.org/amigo/term/GO:0004064	GoMolecFunc
10.29	Carbonate Dehydratase Activity	16	4	CA1, CA13, CA2, CA3	http://amigo.geneontology.org/amigo/term/GO:0004089	GoMolecFunc
10.28	Sialic Acid Transmembrane Transporter Activity	7	3	SLC17A3, SLC35A1, SLC	http://amigo.geneontology.org/amigo/term/GO:0015136	GoMolecFunc
9.66	Fructose 1,6-bisphosphate 1-phosphatase Activity	2	2	FBP2, FBP1	http://amigo.geneontology.org/amigo/term/GO:0042132	GoMolecFunc
9.11	Hydro-lyase Activity	20	4	CA1, CA13, CA2, CA3	http://amigo.geneontology.org/amigo/term/GO:0016836	GoMolecFunc
8.82	Phospholipase Inhibitor Activity	10	3	ANXA5, ANXA3, ANXA	http://amigo.geneontology.org/amigo/term/GO:0004859	GoMolecFunc
8.52	Amino Acid:proton Symporter Activity	3	2	SLC36A3, SLC36A2	http://amigo.geneontology.org/amigo/term/GO:0005280	GoMolecFunc
8.52	Carnitine Transmembrane Transporter Activity	3	2	SLC22A4, SLC22A5	http://amigo.geneontology.org/amigo/term/GO:0015226	GoMolecFunc
7.76	Endodeoxyribonuclease Activity	13	3	ERCC5, BIVM, BIVM-ER	http://amigo.geneontology.org/amigo/term/GO:0004520	GoMolecFunc
7.76	Mechanosensitive Ion Channel Activity	13	3	TMC8, TMC1, TMC6	http://amigo.geneontology.org/amigo/term/GO:0008381	GoMolecFunc
7.71	Quaternary Ammonium Group Transmembrane Transporter Activity	4	2	SLC22A4, SLC22A5	http://amigo.geneontology.org/amigo/term/GO:0015651	GoMolecFunc
7.47	AMP Binding	14	3	PFKP, FBP1, ACSS1	http://amigo.geneontology.org/amigo/term/GO:0016208	GoMolecFunc
7.09	Procollagen-proline 4-dioxygenase Activity	5	2	P4HA1, P4HA2	http://amigo.geneontology.org/amigo/term/GO:0004656	GoMolecFunc
6.59	Peptide Antigen Binding	33	4	HLA-DQB1, HLA-DRB1,	http://amigo.geneontology.org/amigo/term/GO:0042605	GoMolecFunc
6.51	E-box Binding	52	5	CLOCK, SREBF2, TCF12	http://amigo.geneontology.org/amigo/term/GO:0070888	GoMolecFunc
6.23	Phosphatidylinositol Binding	98	7	SNX19, APPL2, ITPR1,	http://amigo.geneontology.org/amigo/term/GO:0035091	GoMolecFunc
6.17	L-proline Transmembrane Transporter Activity	7	2	SLC36A3, SLC36A2	http://amigo.geneontology.org/amigo/term/GO:0015193	GoMolecFunc
6.17	L-alanine Transmembrane Transporter Activity	7	2	SLC36A3, SLC36A2	http://amigo.geneontology.org/amigo/term/GO:0015180	GoMolecFunc

Raw GeneAnalytics data - RA

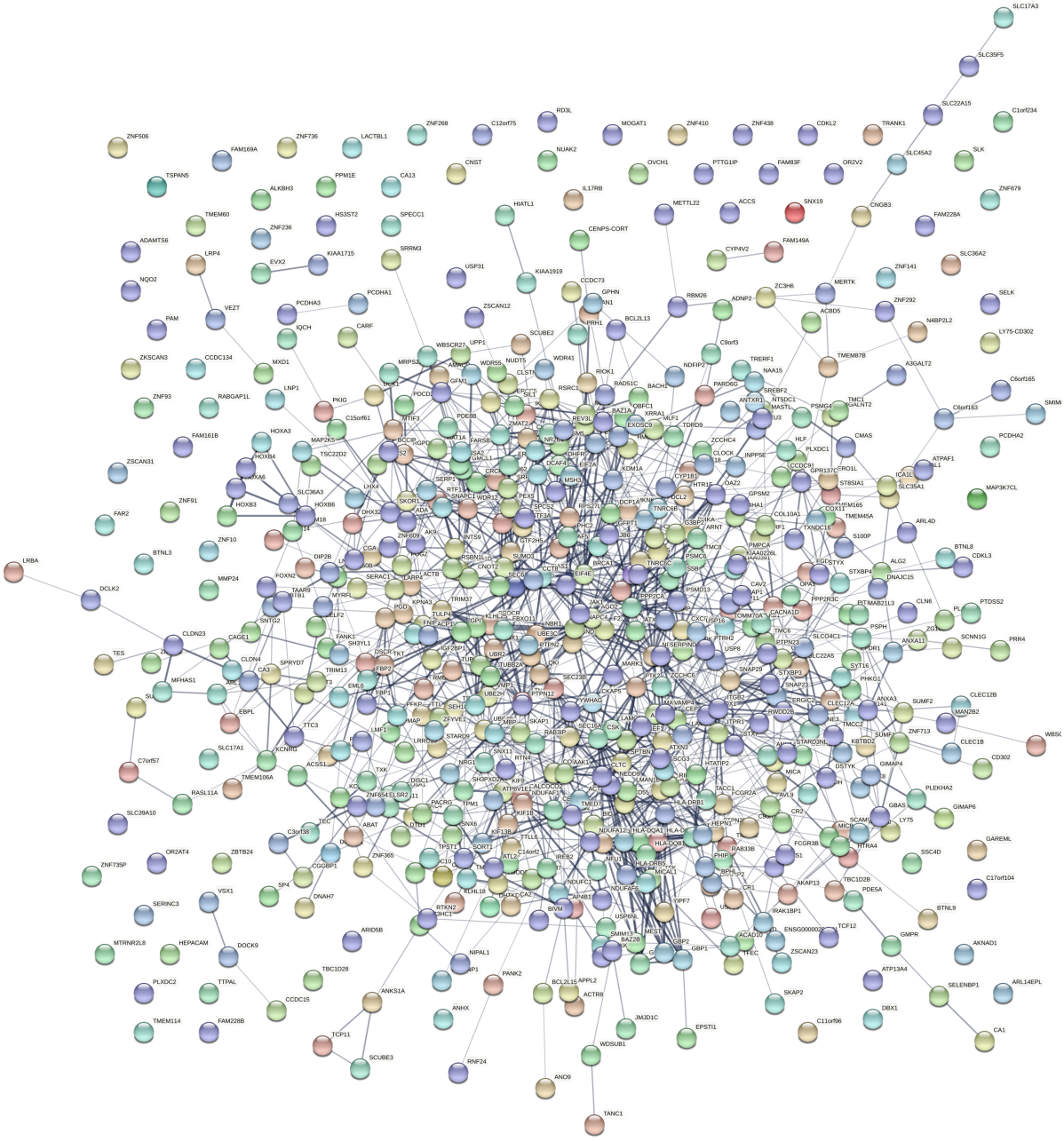
Results

Score	SuperPath Name	# SuperPath Total Genes	# SuperPath Matched Genes	Matched Genes (Sym	Evidence URL
9.92	Regulation of Apoptosis By Parathyroid Hormone-related Protein	22	4	BCL2L13, BCL2L1	PTHLH http://pathcards.genecards.org/card/regulation_of_apoptosis_by_parathyroid_hormone-related_protein
9.44	NAD Metabolism, Sirtuins and Aging	11	3	HIF1A, NFKB1, SIRT3	http://pathcards.genecards.org/card/nad_metabolism_sirtuins_and_aging
9.09	Uricosurics Pathway, Pharmacodynamics	12	3	SLC17A3, LGALS9,	SLC17A1 http://pathcards.genecards.org/card/uricosurics_pathway_pharmacodynamics
8.45	Aryl Hydrocarbon Receptor	48	5	RET, CYP1B1, ARNT,	PSRC http://pathcards.genecards.org/card/aryl_hydrocarbon_receptor
8.17	Mitotic Metaphase and Anaphase	180	10	ANAPC1, AHCTF1,	SEH1L http://pathcards.genecards.org/card/mitotic_metaphase_and_anaphase
7.99	Integrated Breast Cancer Pathway	154	9	MAP3K7CL, MSH6,	BACH http://pathcards.genecards.org/card/integrated_breast_cancer_pathway
7.85	Lysosome	128	8	CTSS, AP4B1, SUMF1,	GM http://pathcards.genecards.org/card/lysosome
7.69	Nitrogen Metabolism	17	3	CA1, CA2, CA3	http://pathcards.genecards.org/card/nitrogen_metabolism
7.28	Thyroxine (Thyroid Hormone) Production	6	2	TPO, CGA	http://pathcards.genecards.org/card/thyroxine_(thyroid_hormone)_production
6.88	MET Promotes Cell Motility	115	7	RAPGEF1, ACTN1,	LAMC1 http://pathcards.genecards.org/card/met_promotes_cell_motility
6.67	Peginterferon Alpha-2a/Peginterferon Alpha-2b Pathway (Hepatocyte), Pharmacodynamics	22	3	JAK1, PIAS1, MAVS	http://pathcards.genecards.org/card/peginterferon_alpha-2apeginterferon_alpha-2b_pathway_(hepatocyte)_pharmacodynamics
6.59	Tuberculosis	180	9	CALML4, CTSS, KSR1,	JAK http://pathcards.genecards.org/card/tuberculosis
6.41	Monoamine GPCRs	44	4	TAAR6, TAAR8, HTR1F,	HT http://pathcards.genecards.org/card/monoamine_gpcrs
6.41	Negative Regulation of MAPK Pathway	44	4	KSR1, BRAP, DUSP4,	MAF http://pathcards.genecards.org/card/negative_regulation_of_mapk_pathway
6.11	Osteoclast Differentiation	128	7	JAK1, CTSK, FCGR3A,	TGF http://pathcards.genecards.org/card/osteoclast_differentiation
5.88	Assembly of RNA Polymerase-I Initiation Complex	10	2	POLR1B, POLR1D	http://pathcards.genecards.org/card/assembly_of_rna_polymerase_i_initiation_complex
5.71	Hepatitis C and Hepatocellular Carcinoma	51	4	JAK1, HIF1A, TGFBR1,	NFK http://pathcards.genecards.org/card/hepatitis_c_and_hepatocellular_carcinoma
5.69	Herpes Simplex Virus 1 Infection	491	17	ZNF253, ZNF10, ZNF721,	http://pathcards.genecards.org/card/herpes_simplex_virus_1_infection
5.58	Pentose Phosphate Pathway	30	3	FBP2, FBP1, PGD	http://pathcards.genecards.org/card/pentose_phosphate_pathway
5.45	Innate Immune System	2124	55	TNFRSF13C, KLHL42,	TNF http://pathcards.genecards.org/card/innate_immune_system

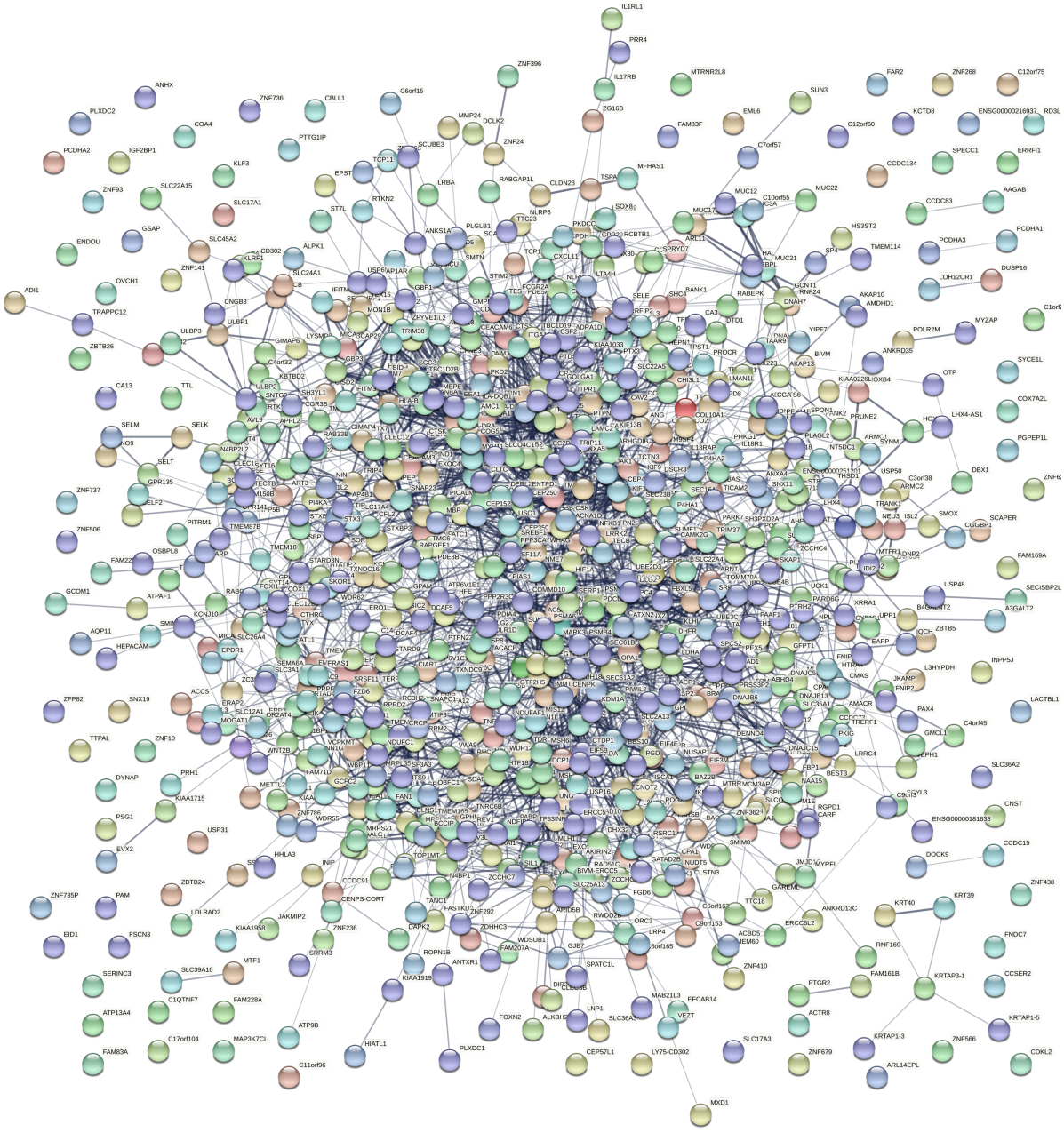
Results

Score	Name	# Genes	# Matched Genes	Matched Genes (S	Evidence URL	Ontology
12.73	Sialic Acid Transport	5	3	SLC17A2, SLC17A3,	SL(http://amigo.geneontology.org/amigo/term/GO:0015739	GoBiolProc
11.77	Dephosphorylation	179	12	PTPN23, FBP2, FBP1,	S http://amigo.geneontology.org/amigo/term/GO:0016311	GoBiolProc
10.37	Sucrose Biosynthetic Process	2	2	FBP2, FBP1	http://amigo.geneontology.org/amigo/term/GO:0005986	GoBiolProc
10.27	Toxin Metabolic Process	9	3	CYP1B1, N6AMT1,	PAI http://amigo.geneontology.org/amigo/term/GO:0009404	GoBiolProc
9.84	Peptidyl-tyrosine Dephosphorylation	102	8	PTPN23, DNAJC6,	ACP http://amigo.geneontology.org/amigo/term/GO:0035335	GoBiolProc
9.84	Response to Interferon-alpha	10	3	IFITM1, IFITM3,	IFITM http://amigo.geneontology.org/amigo/term/GO:0035455	GoBiolProc
9.84	Response to Interferon-beta	10	3	IFITM1, IFITM3,	IFITM2 http://amigo.geneontology.org/amigo/term/GO:0035456	GoBiolProc
9.46	Response to Interferon-gamma	24	4	IFITM1, IFITM3,	IFITM http://amigo.geneontology.org/amigo/term/GO:0034341	GoBiolProc
9.44	UDP-N-acetylglucosamine Biosynthetic Process	11	3	GNPNAT1, GNPDA2,	G http://amigo.geneontology.org/amigo/term/GO:0006048	GoBiolProc
9.22	Positive Regulation of Hormone Biosynthetic Process	3	2	ARNT, HIF1A	http://amigo.geneontology.org/amigo/term/GO:0046886	GoBiolProc
9.22	Positive Regulation of Mitotic Cell Cycle Phase Transition	3	2	KLHL18, TMOD3	http://amigo.geneontology.org/amigo/term/GO:1901992	GoBiolProc
9.22	Carnitine Transmembrane Transport	3	2	SLC22A4, SLC22A5	http://amigo.geneontology.org/amigo/term/GO:1902603	GoBiolProc
9.09	Fructose 6-phosphate Metabolic Process	12	3	FBP2, FBP1, GFPT1	http://amigo.geneontology.org/amigo/term/GO:0006002	GoBiolProc
8.76	Regulation of Interferon-gamma-mediated Signaling Pathway	13	3	JAK1, PTPN2, PIAS1	http://amigo.geneontology.org/amigo/term/GO:0060334	GoBiolProc
8.49	Intracellular Protein Transport	364	16	AP4B1, STX7, ARL11,	S http://amigo.geneontology.org/amigo/term/GO:0006886	GoBiolProc
8.41	Quaternary Ammonium Group Transport	4	2	SLC22A4, SLC22A5	http://amigo.geneontology.org/amigo/term/GO:0015697	GoBiolProc
8.41	Negative Regulation of Platelet-derived Growth Factor Receptor-beta Signaling Pathway	4	2	PTPN12, PTPN2	http://amigo.geneontology.org/amigo/term/GO:2000587	GoBiolProc
8.13	Regulation of Transcription From RNA Polymerase II Promoter in Response to Hypoxia	73	6	ARNT, BACH1, HIF1A,	http://amigo.geneontology.org/amigo/term/GO:0061418	GoBiolProc
7.94	Ras Protein Signal Transduction	75	6	RAPGEF6, KSR1, G3BP2,	http://amigo.geneontology.org/amigo/term/GO:0007265	GoBiolProc
7.93	Regulation of Signaling Receptor Activity	16	3	FCGR2B, CGA, MUC4	http://amigo.geneontology.org/amigo/term/GO:0010469	GoBiolProc
19.26	Cytosol	5213	146	FANK1, PTPN23, CA1,	http://amigo.geneontology.org/amigo/term/GO:0005829	GoCellComp
17.65	Nucleus	6708	176	FANK1, ERI1, PTPN23,	http://amigo.geneontology.org/amigo/term/GO:0005634	GoCellComp
16.09	Cytoplasm	6964	179	FANK1, ERI1, PTPN23,	http://amigo.geneontology.org/amigo/term/GO:0005737	GoCellComp
9.41	Endoplasmic Reticulum-Golgi Intermediate Compartment	83	7	STX17B, GNPNAT1, M	http://amigo.geneontology.org/amigo/term/GO:0005793	GoCellComp
8.85	Myofibril	45	5	LRRC10, PSMA6, TMO	http://amigo.geneontology.org/amigo/term/GO:0030016	GoCellComp
7.93	Lysosome	414	17	SLC17A2, CXCR4, CTSS	http://amigo.geneontology.org/amigo/term/GO:0005764	GoCellComp
7.78	Endolysosome Lumen	5	2	CTSS, CTSK	http://amigo.geneontology.org/amigo/term/GO:0036021	GoCellComp
7.10	Endoplasmic Reticulum	1562	45	VMP1, POMT1, MOGA	http://amigo.geneontology.org/amigo/term/GO:0005783	GoCellComp
7.07	Protein-containing Complex	743	25	PRPF3, RAPGEF1, IFITM	http://amigo.geneontology.org/amigo/term/GO:0032991	GoCellComp
6.85	Golgi Apparatus	1366	40	PPP2R3C, CCDC91, PE	http://amigo.geneontology.org/amigo/term/GO:0005794	GoCellComp
6.74	Sarcomere	64	5	ACTN1, MYBPHL, LRRC	http://amigo.geneontology.org/amigo/term/GO:0030017	GoCellComp
6.70	Perinuclear Region of Cytoplasm	720	24	RADS1C, RAPGEF1, PR	http://amigo.geneontology.org/amigo/term/GO:0048471	GoCellComp
6.69	Nucleoplasm	3743	93	FANK1, ERI1, PTPN23,	http://amigo.geneontology.org/amigo/term/GO:0005654	GoCellComp
6.48	Extrinsic Component of Endosome Membrane	8	2	USP8, USP50	http://amigo.geneontology.org/amigo/term/GO:0031313	GoCellComp
6.47	Extracellular Exosome	2173	58	PTPN23, CA1, ACP1,	FE http://amigo.geneontology.org/amigo/term/GO:0070062	GoCellComp
6.18	Preribosome, Large Subunit Precursor	25	3	NSA2, EIF6, WDR12	http://amigo.geneontology.org/amigo/term/GO:0030687	GoCellComp
5.90	Cytoskeleton	1384	39	FANK1, PTPN23, KLHL4	http://amigo.geneontology.org/amigo/term/GO:0005856	GoCellComp
5.88	Nuclear Pore Outer Ring	10	2	AHCTF1, SEH1L	http://amigo.geneontology.org/amigo/term/GO:0031080	GoCellComp
5.68	ERF1 Methyltransferase Complex	1	1	N6AMT1	http://amigo.geneontology.org/amigo/term/GO:0035657	GoCellComp
5.68	Rhabdomere	1	1	MERTK	http://amigo.geneontology.org/amigo/term/GO:0016028	GoCellComp
16.19	Sialic Acid Transmembrane Transporter Activity	7	4	SLC17A2, SLC17A3,	SL(http://amigo.geneontology.org/amigo/term/GO:0015136	GoMolecFunc
14.79	Non-membrane Spanning Protein Tyrosine Phosphatase Activity	9	4	ACP1, PTPN22, PTPN1	http://amigo.geneontology.org/amigo/term/GO:0004726	GoMolecFunc
14.40	Nucleotide Binding	1836	61	YME1L1, RAD51C, DH	http://amigo.geneontology.org/amigo/term/GO:0000166	GoMolecFunc
13.41	Phosphatase Activity	134	11	PTPN23, FBP2, FBP1,	S http://amigo.geneontology.org/amigo/term/GO:0016791	GoMolecFunc
10.37	Fructose 1,6-bisphosphate 1-phosphatase Activity	2	2	FBP2, FBP1	http://amigo.geneontology.org/amigo/term/GO:0042132	GoMolecFunc
10.33	Protein Binding	11207	254	KLHL18, FANK1, VMP1	http://amigo.geneontology.org/amigo/term/GO:0005515	GoMolecFunc
9.84	Sodium:phosphate Symporter Activity	10	3	SLC17A2, SLC17A3,	SL(http://amigo.geneontology.org/amigo/term/GO:0005436	GoMolecFunc
9.67	Protein Tyrosine Phosphatase Activity	104	8	PTPN23, DNAJC6,	ACP http://amigo.geneontology.org/amigo/term/GO:0004725	GoMolecFunc
9.44	IgG Binding	11	3	FCGR3A, FCGR2B,	FCGI http://amigo.geneontology.org/amigo/term/GO:0019864	GoMolecFunc
9.22	Amino Acid:proton Symporter Activity	3	2	SLC36A3, SLC36A2	http://amigo.geneontology.org/amigo/term/GO:0005280	GoMolecFunc
9.22	Carnitine Transmembrane Transporter Activity	3	2	SLC22A4, SLC22A5	http://amigo.geneontology.org/amigo/term/GO:0015226	GoMolecFunc
8.76	Mechanosensitive Ion Channel Activity	13	3	TMC8, TMC1, TMC6	http://amigo.geneontology.org/amigo/term/GO:0008381	GoMolecFunc
8.73	Hydrolase Activity	1674	50	ERI1, PTPN23, DNAJC6	http://amigo.geneontology.org/amigo/term/GO:0016787	GoMolecFunc
8.59	ATP Binding	1515	46	YME1L1, RAD51C, DH	http://amigo.geneontology.org/amigo/term/GO:0005524	GoMolecFunc
8.41	Quaternary Ammonium Group Transmembrane Transporter Activity	4	2	SLC22A4, SLC22A5	http://amigo.geneontology.org/amigo/term/GO:0015651	GoMolecFunc
8.41	ErbB-2 Class Receptor Binding	4	2	NRG1, MUC4	http://amigo.geneontology.org/amigo/term/GO:0005176	GoMolecFunc
8.41	Phosphoserine Phosphatase Activity	4	2	DUSP26, PSPH	http://amigo.geneontology.org/amigo/term/GO:0004647	GoMolecFunc
8.36	Phosphoprotein Phosphatase Activity	148	9	PTPN23, DNAJC6, ACP	http://amigo.geneontology.org/amigo/term/GO:0004721	GoMolecFunc
8.19	Xenobiotic Transmembrane Transporter Activity	15	3	SLC17A3, SLC22A5, AB	http://amigo.geneontology.org/amigo/term/GO:0042910	GoMolecFunc
7.97	E-box Binding	52	5	CLOCK, SREBF2, TCF12	http://amigo.geneontology.org/amigo/term/GO:0070888	GoMolecFunc

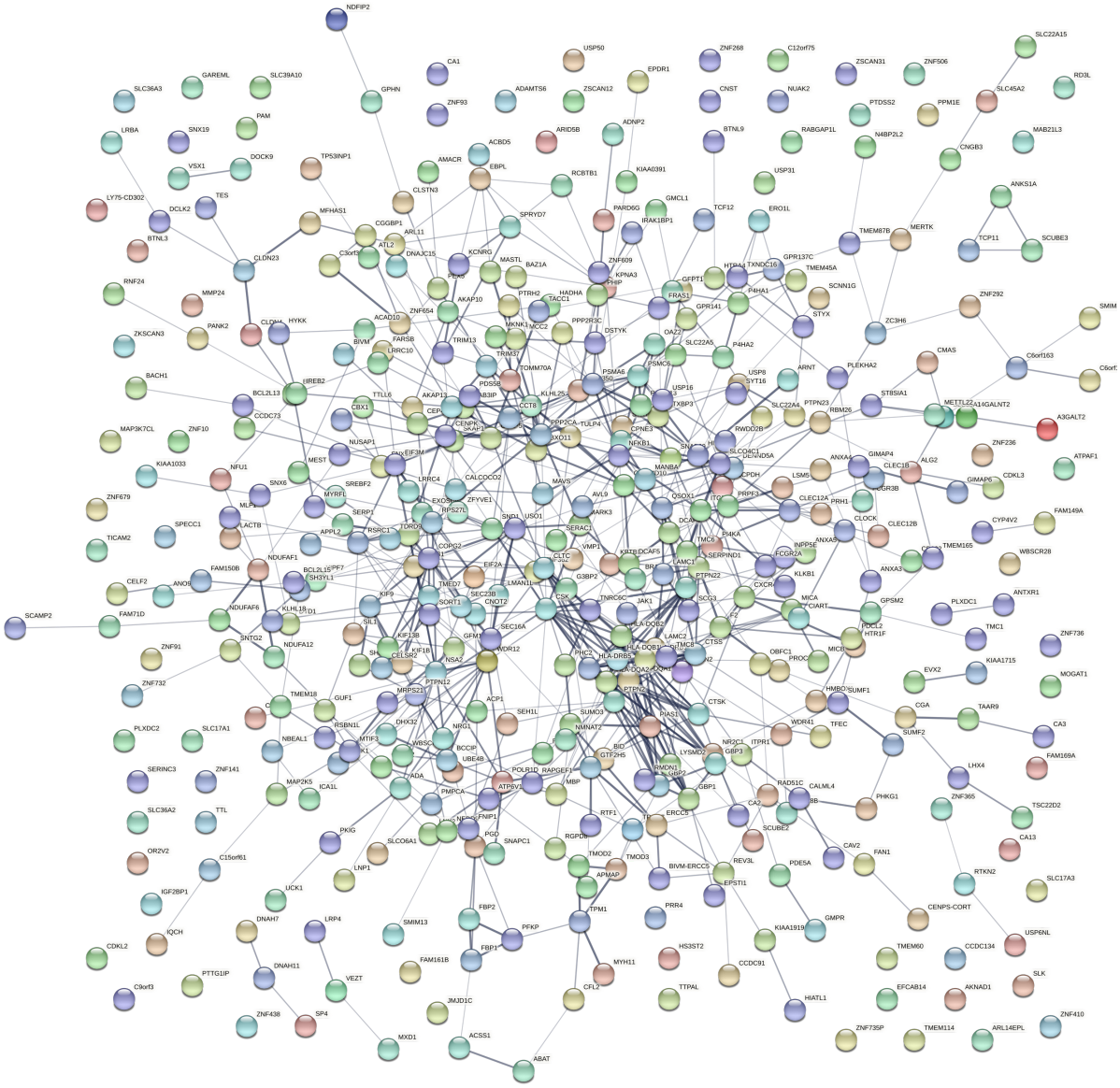
STRING network - R



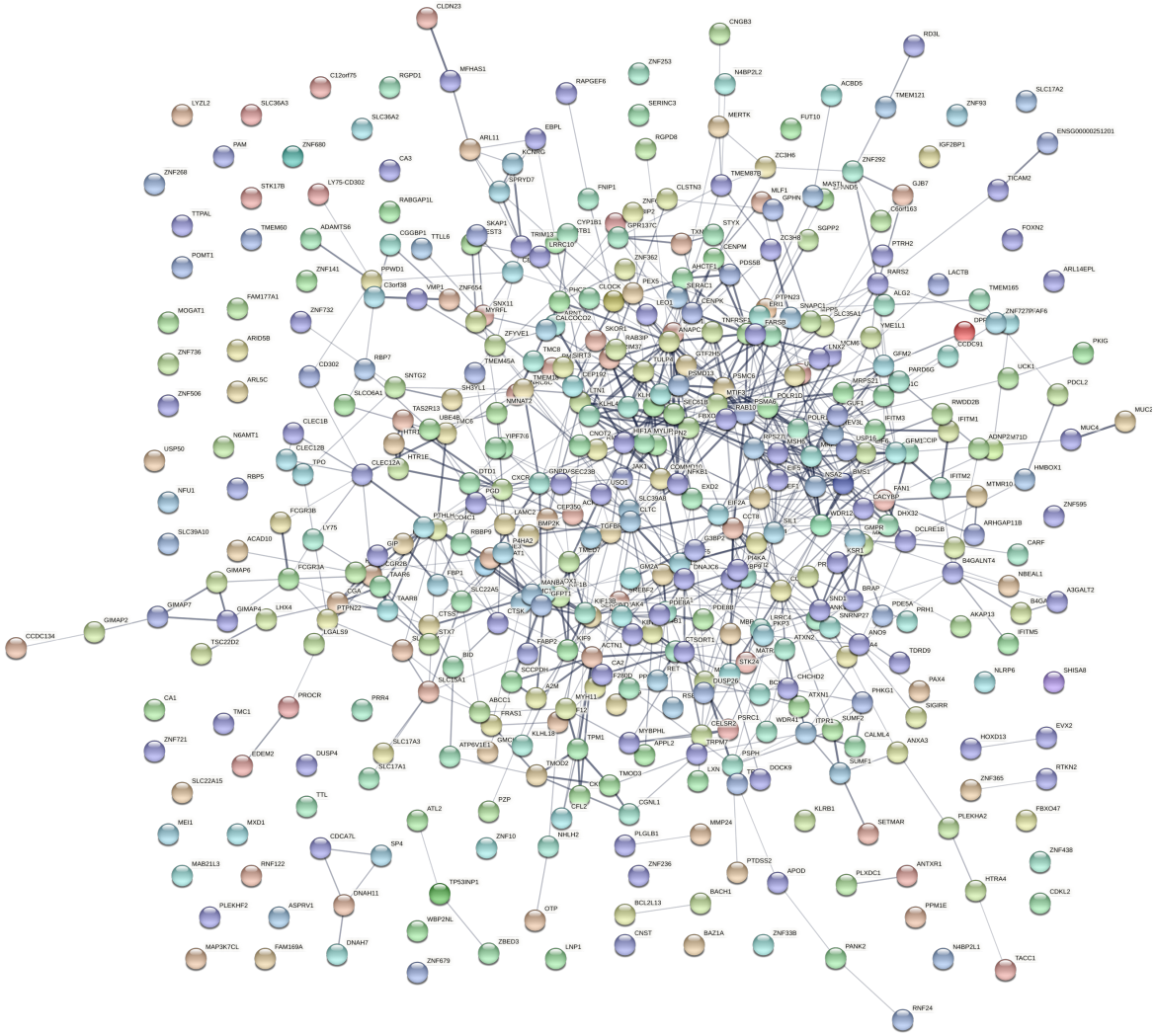
STRING network - NR



STRING network - IR



STRING network - RA



References

1. Gray SG. Epigenetic-based immune intervention for rheumatic diseases. Epigenomics. 2014 Apr;6(2):253–71.

2. Firestein GS, McInnes IB. Immunopathogenesis of Rheumatoid Arthritis. Immunity. Elsevier; 2017 Feb;46(2):183–96.

3. Tuncel J, Holmberg J, Haag S, Hopkins MH, Wester-Rosenlöf L, Carlsen S, et al. Self-reactive T cells induce and perpetuate chronic relapsing arthritis. Arthritis Research & Therapy. BioMed Central; 2020 Apr 28;22(1):95–12.

4. Symmons D, Turner G, Webb R, Asten P, Barrett E, Lunt M, et al. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology. 2002 Jul;41(7):793–800.

5. Pawłowska J, Smoleńska Z, Daca A, Witkowski JM, Bryl E. Older age of rheumatoid arthritis onset is associated with higher activation status of peripheral blood CD4(+) T cells and disease activity. Clinical & Experimental Immunology. John Wiley & Sons, Ltd; 2011 Feb;163(2):157–64.

6. Jia J, Li J, Yao X, Zhang Y, Yang X, Wang P, et al. Genetic architecture study of rheumatoid arthritis and juvenile idiopathic arthritis. PeerJ. PeerJ Inc; 2020;8(12):e8234.

7. Szekanecz Z, Meskó B, Poliska S, Váncsa A, Szamosi S, Végh E, et al. Pharmacogenetics and pharmacogenomics in rheumatology. Immunol Res. Springer-Verlag; 2013 Jul;56(2-3):325–33.

8. Littlejohn EA, Monrad SU. Early Diagnosis and Treatment of Rheumatoid Arthritis. Prim Care. 2018 Jun;45(2):237–55.

9. Montecucco F, Mach F. Common inflammatory mediators orchestrate pathophysiological processes in rheumatoid arthritis and atherosclerosis. Rheumatology (Oxford). 2009 Jan;48(1):11–22.

10. Gullick NJ, Scott DL. Co-morbidities in established rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2011 Aug;25(4):469–83.

11. Marrie RA, Walld R, Bolton JM, Sareen J, Walker JR, Patten SB, et al. Rising incidence of psychiatric disorders before diagnosis of immune-mediated inflammatory disease. Epidemiol Psychiatr Sci. Cambridge University Press; 2019 Jun;28(3):333–42.

12. Deb A, Dwibedi N, LeMasters T, Hornsby JA, Wei W, Sambamoorthi U. Burden of Depression among Working-Age Adults with Rheumatoid Arthritis. Arthritis. 2018;2018(6):8463632–11.

13. Scherrer JF, Virgo KS, Zeringue A, Bucholz KK, Jacob T, Johnson RG, et al. Depression increases risk of incident myocardial infarction among Veterans Administration patients with rheumatoid arthritis. Gen Hosp Psychiatry. 2009 Jul;31(4):353–9.

14. Matcham F, Norton S, Scott DL, Steer S, Hotopf M. Symptoms of depression and anxiety predict treatment response and long-term physical health outcomes in rheumatoid arthritis: secondary analysis of a randomized controlled trial. Rheumatology (Oxford). 2016 Feb;55(2):268–78.

15. Tobón GJ, Youinou P, Saraux A. The environment, geo-epidemiology, and autoimmune disease: Rheumatoid arthritis. J Autoimmun. 2010 Aug;35(1):10–4.

16. Viatte S, Plant D, Han B, Fu B, Yarwood A, Thomson W, et al. Association of HLA-DRB1 haplotypes with rheumatoid arthritis severity, mortality, and treatment response. JAMA. American Medical Association; 2015 Apr 28;313(16):1645–56.

17. Morgan AW, Robinson JI, Conaghan PG, Martin SG, Hensor EMA, Morgan MD, et al. Evaluation of the rheumatoid arthritis susceptibility loci HLA-DRB1, PTPN22, OLIG3/TNFAIP3, STAT4 and TRAF1/C5 in an inception cohort. Arthritis Research & Therapy. BioMed Central; 2010;12(2):R57–10.

18. Deane KD, Demoruelle MK, Kelmenson LB, Kuhn KA, Norris JM, Holers

VM. Genetic and environmental risk factors for rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2017 Feb;31(1):3–18.

19. Demoruelle MK, Deane KD, Holers VM. When and where does inflammation begin in rheumatoid arthritis? Curr Opin Rheumatol. 2014 Jan;26(1):64–71.

20. Chang K, Yang SM, Kim SH, Han KH, Park SJ, Shin JI. Smoking and rheumatoid arthritis. Int J Mol Sci. Multidisciplinary Digital Publishing Institute; 2014 Dec 3;15(12):22279–95.

21. Pryor WA, Stone K. Oxidants in cigarette smoke. Radicals, hydrogen peroxide, peroxynitrate, and peroxynitrite. Ann N Y Acad Sci. John Wiley & Sons, Ltd; 1993 May 28;686(1 Tobacco Smoki):12–27–discussion27–8.

22. Hyrich KL, Watson KD, Silman AJ, Symmons DPM, British Society for Rheumatology Biologics Register. Predictors of response to anti-TNF-alpha therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. Rheumatology. 2006 Dec;45(12):1558–65.

23. Mankia K, Cheng Z, Do T, Hunt L, Meade J, Kang J, et al. Prevalence of Periodontal Disease and Periodontopathic Bacteria in Anti-Cyclic Citrullinated Protein Antibody-Positive At-Risk Adults Without Arthritis. JAMA Netw Open. 2019 Jun 5;2(6):e195394.

24. Klein K, Gay S. Epigenetics in rheumatoid arthritis. Curr Opin Rheumatol. 2015 Jan;27(1):76–82.

25. Merlino LA, Curtis J, Mikuls TR, Cerhan JR, Criswell LA, Saag KG, et al. Vitamin D intake is inversely associated with rheumatoid arthritis: results from the Iowa Women's Health Study. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2004 Jan;50(1):72–7.

26. Sundström B, Johansson I, Rantapää-Dahlqvist S. Interaction between dietary sodium and smoking increases the risk for rheumatoid arthritis: results from a nested case-control study. Rheumatology (Oxford). 2015 Mar;54(3):487–93.

27. Hu Y, Costenbader KH, Gao X, Al-Daabil M, Sparks JA, Solomon DH, et al. Sugar-sweetened soda consumption and risk of developing rheumatoid arthritis in women. Am J Clin Nutr. 2014 Sep;100(3):959–67.

28. Meeuwisse CM, van der Linden MP, Rullmann TA, Allaart CF, Nelissen R, Huizinga TW, et al. Identification of CXCL13 as a marker for rheumatoid arthritis outcome using an in silico model of the rheumatic joint. Arthritis & Rheumatism. Wiley Subscription Services, Inc., A Wiley Company; 2011 Apr 27;63(5):1265–73.

29. Filipovic I, Walker D, Forster F, Curry AS. Quantifying the economic burden of productivity loss in rheumatoid arthritis. Rheumatology (Oxford). 2011 Jun;50(6):1083–90.

30. Sokka T. Work disability in early rheumatoid arthritis. Clin Exp Rheumatol. 2003 Sep;21(5 Suppl 31):S71–4.

31. Barrett EM, Scott DG, Wiles NJ, Symmons DP. The impact of rheumatoid arthritis on employment status in the early years of disease: a UK community-based study. Rheumatology. 2000 Dec;39(12):1403–9.

32. Strand V, Tundia N, Song Y, Macaulay D, Fuldeore M. Economic Burden of Patients with Inadequate Response to Targeted Immunomodulators for Rheumatoid Arthritis. J Manag Care Spec Pharm. Academy of Managed Care Pharmacy; 2018 Apr;24(4):344–52.

33. Lee A, Qiao Y, Grigoriev G, Chen J, Park-Min K-H, Park SH, et al. Tumor necrosis factor α induces sustained signaling and a prolonged and unremitting inflammatory response in rheumatoid arthritis synovial fibroblasts. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2013 Apr;65(4):928–38.

34. Popp J, Schicht M, Garreis F, Klinger P, Gelse K, Sesselmann S, et al. Human Synovia Contains Trefoil Factor Family (TFF) Peptides 1-3 Although Synovial Membrane Only Produces TFF3: Implications in Osteoarthritis and Rheumatoid Arthritis. Int J Mol Sci.

Multidisciplinary Digital Publishing Institute; 2019 Dec 3;20(23):6105.

35. Bartok B, Firestein GS. Fibroblast-like synoviocytes: key effector cells in rheumatoid arthritis. Immunol Rev. John Wiley & Sons, Ltd; 2010 Jan;233(1):233–55.

36. Kim GW, Lee NR, Pi RH, Lim YS, Lee YM, Lee JM, et al. IL-6 inhibitors for treatment of rheumatoid arthritis: past, present, and future. Arch Pharm Res. Pharmaceutical Society of Korea; 2015;38(5):575–84.

37. Feldmann M. Translating molecular insights in autoimmunity into effective therapy. Vol. 27, Annual review of immunology. Annual Reviews; 2009. 27 p.

38. Sarzi-Puttini P, Atzeni F, Shoenfeld Y, Ferraccioli G. TNF-alpha, rheumatoid arthritis, and heart failure: a rheumatological dilemma. Autoimmun Rev. 2005 Mar;4(3):153–61.

39. Taylor PC, Peters AM, Paleolog E, Chapman PT, Elliott MJ, McCloskey R, et al. Reduction of chemokine levels and leukocyte traffic to joints by tumor necrosis factor alpha blockade in patients with rheumatoid arthritis. Arthritis & Rheumatism. 2000 Jan;43(1):38–47.

40. Falconer J, Murphy AN, Young SP, Clark AR, Tiziani S, Guma M, et al. Review: Synovial Cell Metabolism and Chronic Inflammation in Rheumatoid Arthritis. Arthritis Rheumatol. John Wiley & Sons, Ltd; 2018 Jul;70(7):984–99.

41. Fearon U, Canavan M, Biniecka M, Veale DJ. Hypoxia, mitochondrial dysfunction and synovial invasiveness in rheumatoid arthritis. Nat Rev Rheumatol. Nature Publishing Group; 2016 Jul;12(7):385–97.

42. Hua S, Dias TH. Hypoxia-Inducible Factor (HIF) as a Target for Novel Therapies in Rheumatoid Arthritis. Front Pharmacol. Frontiers; 2016;7:184.

43. Guo X, Chen G. Hypoxia-Inducible Factor Is Critical for Pathogenesis and Regulation of Immune Cell Functions in Rheumatoid Arthritis. Frontiers in Immunology. Frontiers; 2020;11:1668.

44. Derksen VFAM, Huizinga TWJ, van der Woude D. The role of autoantibodies in the pathophysiology of rheumatoid arthritis. Semin Immunopathol. Springer Berlin Heidelberg; 2017 Jun;39(4):437–46.

45. Scott DL, Wolfe F, Huizinga TWJ. Rheumatoid arthritis. Lancet. 2010 Sep 25;376(9746):1094–108.

46. Fang Q, Zhou C, Nandakumar KS. Molecular and Cellular Pathways Contributing to Joint Damage in Rheumatoid Arthritis. Mediators of Inflammation. Hindawi; 2020;2020:3830212–20.

47. Baum R, Gravallese EM. Impact of inflammation on the osteoblast in rheumatic diseases. Curr Osteoporos Rep. Springer US; 2014 Mar;12(1):9–16.

48. Končarević S, Lößner C, Kuhn K, Prinz T, Pike I, Zucht H-D. In-depth profiling of the peripheral blood mononuclear cells proteome for clinical blood proteomics. Int J Proteomics. Hindawi; 2014;2014(11):129259–9.

49. Pourahmad J, Salimi A. Isolated Human Peripheral Blood Mononuclear Cell (PBMC), a Cost Effective Tool for Predicting Immunosuppressive Effects of Drugs and Xenobiotics. Iran J Pharm Res. Shahid Beheshti University of Medical Sciences; 2015;14(4):979.

50. Corkum CP, Ings DP, Burgess C, Karwowska S, Kroll W, Michalak TI. Immune cell subsets and their gene expression profiles from human PBMC isolated by Vacutainer Cell Preparation Tube (CPT™) and standard density gradient. BMC Immunol. 7 ed. BioMed Central; 2015 Aug 26;16(1):48–18.

51. Verhoeckx K, Cotter P, López-Expósito I, Kleiveland C, Lea T, Mackie A, et al. Peripheral Blood Mononuclear Cells. The Impact of Food Bioactives on Health. Cham (CH): Springer; 2015;129(Chapter 15):161–7.

52. Tseng C-C, Lin Y-Z, Lin C-H, Li R-N, Yen C-Y, Chan H-C, et al. Next-Generation Sequencing Profiles of the Methylome and Transcriptome in Peripheral Blood Mononuclear Cells of Rheumatoid Arthritis. J Clin

Med. Multidisciplinary Digital Publishing Institute; 2019 Aug 22;8(9):1284.

53. Kholodnyuk I, Kadisa A, Svirskis S, Gravelsina S, Studers P, Spaka I, et al. Proportion of the CD19-Positive and CD19-Negative Lymphocytes and Monocytes within the Peripheral Blood Mononuclear Cell Set is Characteristic for Rheumatoid Arthritis. Medicina (Kaunas). Multidisciplinary Digital Publishing Institute; 2019 Sep 24;55(10):630.

54. Auffray C, Sieweke MH, Geissmann F. Blood monocytes: development, heterogeneity, and relationship with dendritic cells. Annu Rev Immunol. Annual Reviews; 2009;27(1):669–92.

55. Rana AK, Li Y, Dang Q, Yang F. Monocytes in rheumatoid arthritis: Circulating precursors of macrophages and osteoclasts and, their heterogeneity and plasticity role in RA pathogenesis. International Immunopharmacology. 2018 Dec;65:348–59.

56. Hirose S, Lin Q, Ohtsuji M, Nishimura H, Verbeek JS. Monocyte subsets involved in the development of systemic lupus erythematosus and rheumatoid arthritis. Int Immunol. 2019 Oct 16;31(11):687–96.

57. Ruiz-Limon P, Ortega-Castro R, Barbarroja N, Perez-Sanchez C, Jamin C, Patiño-Trives AM, et al. Molecular Characterization of Monocyte Subsets Reveals Specific and Distinctive Molecular Signatures Associated With Cardiovascular Disease in Rheumatoid Arthritis. Frontiers in Immunology. 2019;10:1111.

58. Boyle WJ, Simonet WS, Lacey DL. Osteoclast differentiation and activation. Nature. Nature Publishing Group; 2003 May 15;423(6937):337–42.

59. Blair HA, Deeks ED. Abatacept: A Review in Rheumatoid Arthritis. Drugs. Springer International Publishing; 2017 Jul;77(11):1221–33.

60. Humby F, Bombardieri M, Manzo A, Kelly S, Blades MC, Kirkham B, et al. Ectopic lymphoid structures support ongoing production of class-switched autoantibodies in rheumatoid synovium. Huizinga T, editor. PLoS Med. Public Library of Science; 2009 Jan 13;6(1):e1.

61. Yasuda K, Takeuchi Y, Hirota K. Correction to: The pathogenicity of Th17 cells in autoimmune diseases. Semin Immunopathol. Springer Berlin Heidelberg; 2019 May;41(3):299–9.

62. Capone A, Volpe E. Transcriptional Regulators of T Helper 17 Cell Differentiation in Health and Autoimmune Diseases. Frontiers in Immunology. 2020;11:348.

63. Langrish CL, Chen Y, Blumenschein WM, Mattson J, Basham B, Sedgwick JD, et al. IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. J Exp Med. 2005 Jan 17;201(2):233–40.

64. Silverman GJ, Khanna S. B cell modulation in rheumatology. Curr Opin Pharmacol. 2007 Aug;7(4):426–33.

65. De Vita S, Zaja F, Sacco S, De Candia A, Fanin R, Ferraccioli G. Efficacy of selective B cell blockade in the treatment of rheumatoid arthritis: evidence for a pathogenetic role of B cells. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2002 Aug;46(8):2029–33.

66. Svensson L, Jirholt J, Holmdahl R, Jansson L. B cell-deficient mice do not develop type II collagen-induced arthritis (CIA). Clinical & Experimental Immunology. John Wiley & Sons, Ltd; 1998 Mar;111(3):521–6.

67. Litsiou E, Semitekolou M, Galani IE, Morianos I, Tsoutsa A, Kara P, et al. CXCL13 production in B cells via Toll-like receptor/lymphotoxin receptor signaling is involved in lymphoid neogenesis in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. American Thoracic Society; 2013 Jun 1;187(11):1194–202.

68. Bosello S, Fedele AL, Peluso G, Gremese E, Tolusso B, Ferraccioli G. Very early rheumatoid arthritis is the major predictor of major outcomes: clinical ACR remission and radiographic non-progression. Ann Rheum Dis. 2011 Jul;70(7):1292–5.

69. Kim D, Choi C-B, Lee J, Cho S-K, Won S, Bang S-Y, et al. Impact of early diagnosis on functional disability in rheumatoid arthritis. Korean

J Intern Med. 2017 Jul;32(4):738–46.

70. Cader MZ, Filer A, Hazlehurst J, de Pablo P, Buckley CD, Raza K. Performance of the 2010 ACR/EULAR criteria for rheumatoid arthritis: comparison with 1987 ACR criteria in a very early synovitis cohort. Ann Rheum Dis. 2011 Jun;70(6):949–55.

71. Zhao J, Su Y, Li R, Ye H, Zou Q, Fang Y, et al. Classification criteria of early rheumatoid arthritis and validation of its performance in a multi-centre cohort. Clin Exp Rheumatol. Clin Exp Rheumatol; 2014 Sep;32(5):667–73.

72. Sakellariou G, Scirè CA, Verstappen SMM, Montecucco C, Caporali R. In patients with early rheumatoid arthritis, the new ACR/EULAR definition of remission identifies patients with persistent absence of functional disability and suppression of ultrasonographic synovitis. Ann Rheum Dis. 2013 Feb;72(2):245–9.

73. Aletaha D, Martinez-Avila J, Kvien TK, Smolen JS. Definition of treatment response in rheumatoid arthritis based on the simplified and the clinical disease activity index. Ann Rheum Dis. 2012 Jul;71(7):1190–6.

74. Kay J, Upchurch KS. ACR/EULAR 2010 rheumatoid arthritis classification criteria. Rheumatology (Oxford). 2012 Dec;51 Suppl 6(suppl 6):vi5–9.

75. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. N Engl J Med. 2011 Dec 8;365(23):2205–19.

76. 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. Coles JA, editor. PLoS ONE. 2015;10(3):e0119683.

77. Sellam J, Marion-Thore S, Dumont F, Jacques S, Garchon H-J, Rouanet S, et al. Use of Whole-Blood Transcriptomic Profiling to Highlight Several Pathophysiologic Pathways Associated With Response to Rituximab in Patients With Rheumatoid Arthritis: Data

From a Randomized, Controlled, Open-Label Trial. Arthritis & Rheumatology. 2014 Jul 28;66(8):2015–25.

78. Braun J. Methotrexate: optimizing the efficacy in rheumatoid arthritis. Ther Adv Musculoskelet Dis. SAGE PublicationsSage UK: London, England; 2011 Jun;3(3):151–8.

79. Hyndman IJ. Rheumatoid arthritis: past, present and future approaches to treating the disease. Int J Rheum Dis. 2017 Apr;20(4):417–9.

80. Pieringer H, Studnicka-Benke A. What is causing my arthritis, doctor? A glimpse beyond the usual suspects in the pathogenesis of rheumatoid arthritis. QJM. 2013 Mar;106(3):219–28.

81. Stoffer MA, Schoels MM, Smolen JS, Aletaha D, Breedveld FC, Burmester G, et al. Evidence for treating rheumatoid arthritis to target: results of a systematic literature search update. Ann Rheum Dis. 2016 Jan;75(1):16–22.

82. Smolen JS, Landewé R, Breedveld FC, Buch M, Burmester G, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. 2014. pp. 492–509.

83. Solomon DH, Bitton A, Katz JN, Radner H, Brown EM, Fraenkel L. Review: treat to target in rheumatoid arthritis: fact, fiction, or hypothesis? Arthritis Rheumatol. John Wiley & Sons, Ltd; 2014 Apr;66(4):775–82.

84. Kerrigan SA, McInnes IB. Reflections on “older” drugs: learning new lessons in rheumatology. Nat Rev Rheumatol. Nature Publishing Group; 2020 Mar;16(3):179–83.

85. Mueller RB, Spaeth M, Restorff von C, Ackermann C, Schulze-Koops H, Kempis von J. Superiority of a Treat-to-Target Strategy over Conventional Treatment with Fixed csDMARD and Corticosteroids: A Multi-Center Randomized Controlled Trial in RA Patients with an Inadequate Response to Conventional Synthetic DMARDs, and New

Therapy with Certolizumab Pegol. J Clin Med. Multidisciplinary Digital Publishing Institute; 2019 Mar 3;8(3):302.

86. Durán J, Bockorny M, Dalal D, LaValley M, Felson DT. Methotrexate dosage as a source of bias in biological trials in rheumatoid arthritis: a systematic review. Ann Rheum Dis. 2016 Sep;75(9):1595–8.

87. Bedoui Y, Guillot X, Sélambarom J, Guiraud P, Giry C, Jaffar-Bandjee MC, et al. Methotrexate an Old Drug with New Tricks. Int J Mol Sci. Multidisciplinary Digital Publishing Institute; 2019 Oct 10;20(20):5023.

88. Friedman B, Cronstein B. Methotrexate mechanism in treatment of rheumatoid arthritis. Joint Bone Spine. 2018 Aug 3.

89. Visser K, van der Heijde D. Optimal dosage and route of administration of methotrexate in rheumatoid arthritis: a systematic review of the literature. Ann Rheum Dis. BMJ Publishing Group Ltd; 2009 Jul;68(7):1094–9.

90. Budzik GP, Colletti LM, Faltynek CR. Effects of methotrexate on nucleotide pools in normal human T cells and the CEM T cell line. Life Sci. 2000;66(23):2297–307.

91. Shea B, Swinden MV, Tanjong Ghogomu E, Ortiz Z, Katchamart W, Rader T, et al. Folic acid and folinic acid for reducing side effects in patients receiving methotrexate for rheumatoid arthritis. Cochrane Musculoskeletal Group, editor. Cochrane Database Syst Rev. John Wiley & Sons, Ltd; 2013 May 31;17(5):CD000951.

92. Pincus T, Gibson KA, Castrejón I. Update on methotrexate as the anchor drug for rheumatoid arthritis. Bull Hosp Jt Dis (2013). Bull Hosp Jt Dis (2013); 2013;71 Suppl 1:S9–19.

93. Weinblatt ME. Methotrexate in rheumatoid arthritis: a quarter century of development. Trans Am Clin Climatol Assoc. American Clinical and Climatological Association; 2013;124:16–25.

94. Davila L, Ranganathan P. Pharmacogenetics: implications for therapy in rheumatic diseases. Nat Rev Rheumatol. Nature Publishing Group;

2011 Aug 9;7(9):537–50.

95. Halilova KI, Brown EE, Morgan SL, Bridges SL, Hwang M-H, Arnett DK, et al. Markers of treatment response to methotrexate in rheumatoid arthritis: where do we stand? Int J Rheumatol. 2012;2012(12):978396–7.

96. Ćalasan MB, van den Bosch OFC, Creemers MCW, Custers M, Heurkens AHM, van Woerkom JM, et al. Prevalence of methotrexate intolerance in rheumatoid arthritis and psoriatic arthritis. Arthritis Research & Therapy. BioMed Central; 2013;15(6):R217–5.

97. Wang W, Zhou H, Liu L. Side effects of methotrexate therapy for rheumatoid arthritis: A systematic review. Eur J Med Chem. 2018 Oct 5;158:502–16.

98. Michaud K, Vrijens B, Tousset E, Pedro S, Schumacher R, Dasic G, et al. Real-World Adherence to Oral Methotrexate Measured Electronically in Patients With Established Rheumatoid Arthritis. ACR Open Rheumatol. John Wiley & Sons, Ltd; 2019 Nov;1(9):560–70.

99. Sergeant JC, Hyrich KL, Anderson J, Kopec-Harding K, Hope HF, Symmons DPM, et al. Prediction of primary non-response to methotrexate therapy using demographic, clinical and psychosocial variables: results from the UK Rheumatoid Arthritis Medication Study (RAMS). Arthritis Research & Therapy. BioMed Central; 2018 Jul 13;20(1):147.

100. Taylor PC, Balsa Criado A, Mongey A-B, Avouac J, Marotte H, Mueller RB. How to Get the Most from Methotrexate (MTX) Treatment for Your Rheumatoid Arthritis Patient?-MTX in the Treat-to-Target Strategy. J Clin Med. Multidisciplinary Digital Publishing Institute; 2019 Apr 15;8(4):515.

101. Relling MV, Gardner EE, Sandborn WJ, Schmiegelow K, Pui C-H, Yee SW, et al. Clinical pharmacogenetics implementation consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing: 2013 update. Vol. 93, Clinical pharmacology and therapeutics.

John Wiley & Sons, Ltd; 2013. pp. 324–5.

102. Oh K-T, Anis AH, Bae S-C. Pharmacoeconomic analysis of thiopurine methyltransferase polymorphism screening by polymerase chain reaction for treatment with azathioprine in Korea. Rheumatology. 2004 Feb;43(2):156–63.

103. Shippey EA, Wagler VD, Collamer AN. Hydroxychloroquine: An old drug with new relevance. Cleve Clin J Med. 2018 Jun;85(6):459–67.

104. Eugenia Schroeder M, Russo S, Costa C, Hori J, Tiscornia I, Bollati-Fogolín M, et al. Pro-inflammatory Ca++-activated K+ channels are inhibited by hydroxychloroquine. Scientific Reports. Nature Publishing Group; 2017 May 15;7(1):1892–7.

105. Torigoe M, Sakata K, Ishii A, Iwata S, Nakayamada S, Tanaka Y. Hydroxychloroquine efficiently suppresses inflammatory responses of human class-switched memory B cells via Toll-like receptor 9 inhibition. Clin Immunol. 2018 Oct;195:1–7.

106. Han J, Zhou Q, Li X, He J, Han Y, Jie H, et al. Novel function of hydroxychloroquine: Down regulation of T follicular helper cells in collagen-induced arthritis. Biomed Pharmacother. 2018 Jan;97:838–43.

107. Rempenault C, Combe B, Barnetche T, Gaujoux-Viala C, Lukas C, Morel J, et al. Clinical and Structural Efficacy of Hydroxychloroquine in Rheumatoid Arthritis: A Systematic Review. Arthritis Care & Research. 2020 Jan;72(1):36–40.

108. Shi N, Zhang S, Silverman G, Li M, Cai J, Niu H. Protective effect of hydroxychloroquine on rheumatoid arthritis-associated atherosclerosis. Animal Model Exp Med. 2019 Jun;2(2):98–106.

109. Hung Y-M, Wang Y-H, Lin L, Wang PYP, Chiou J-Y, Wei JC-C. Hydroxychloroquine may be associated with reduced risk of coronary artery diseases in patients with rheumatoid arthritis: A nationwide population-based cohort study. Int J Clin Pract. 2018 May;72(5):e13095.

110. Jorge A, Ung C, Young LH, Melles RB, Choi HK. Hydroxychloroquine retinopathy - implications of research advances for rheumatology care. Nat Rev Rheumatol. Nature Publishing Group; 2018 Dec;14(12):693–703.

111. MacMullan PA, Madigan AM, Paul N, Peace AJ, Alagha A, Nolan KB, et al. Sulfasalazine and its metabolites inhibit platelet function in patients with inflammatory arthritis. Clin Rheumatol. Springer London; 2016 Feb;35(2):447–55.

112. Wiese MD, Alotaibi N, O'Doherty C, Sorich MJ, Suppiah V, Cleland LG, et al. Pharmacogenomics of NAT2 and ABCG2 influence the toxicity and efficacy of sulphasalazine containing DMARD regimens in early rheumatoid arthritis. The Pharmacogenomics Journal. Nature Publishing Group; 2014 Aug;14(4):350–5.

113. Cildag S, Senturk T. Sulfasalazine-Related Hypersensitivity Reactions in Patients With Rheumatic Diseases. J Clin Rheumatol. 2017 Mar;23(2):77–9.

114. Volin MV, Harlow LA, Woods JM, Campbell PL, Amin MA, Tokuhira M, et al. Treatment with sulfasalazine or sulfapyridine, but not 5-aminosalicyclic acid, inhibits basic fibroblast growth factor-induced endothelial cell chemotaxis. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 1999 Sep;42(9):1927–35.

115. Kanerud L, Engström GN, Tarkowski A. Evidence for differential effects of sulphasalazine on systemic and mucosal immunity in rheumatoid arthritis. Ann Rheum Dis. BMJ Publishing Group Ltd; 1995 Apr;54(4):256–62.

116. Danis VA, Franic GM, Rathjen DA, Laurent RM, Brooks PM. Circulating cytokine levels in patients with rheumatoid arthritis: results of a double blind trial with sulphasalazine. Ann Rheum Dis. BMJ Publishing Group Ltd; 1992 Aug;51(8):946–50.

117. O'Dell JR, Leff R, Paulsen G, Haire C, Mallek J, Eckhoff PJ, et al. Treatment of rheumatoid arthritis with methotrexate and

hydroxychloroquine, methotrexate and sulfasalazine, or a combination of the three medications: results of a two-year, randomized, double-blind, placebo-controlled trial. Arthritis & Rheumatism. 2002 May;46(5):1164–70.

118. Peper SM, Lew R, Mikuls T, Brophy M, Rybin D, Wu H, et al. Rheumatoid Arthritis Treatment After Methotrexate: The Durability of Triple Therapy Versus Etanercept. Arthritis Care & Research. John Wiley & Sons, Ltd; 2017 Oct;69(10):1467–72.

119. Bansback N, Phibbs CS, Sun H, O'Dell JR, Brophy M, Keystone EC, et al. Triple Therapy Versus Biologic Therapy for Active Rheumatoid Arthritis: A Cost-Effectiveness Analysis. Ann Intern Med. 2017 Jul 4;167(1):8–16.

120. Quach LT, Chang B-H, Brophy MT, Soe Thwin S, Hannagan K, O'Dell JR. Rheumatoid arthritis triple therapy compared with etanercept: difference in infectious and gastrointestinal adverse events. Rheumatology (Oxford). 2017 Mar 1;56(3):378–83.

121. Van De Laar CJ, Oude Voshaar MAH, Fakhouri WKH, Zaremba-Pechmann L, De Leonardis F, La Torre De I, et al. Cost-Effectiveness of a JAK1/JAK2 Inhibitor vs a Biologic Disease-Modifying Antirheumatic Drug (bDMARD) in a Treat-to-Target Strategy for Rheumatoid Arthritis. Clinicoecon Outcomes Res. Dove Press; 2020;12:213–22.

122. Elliott MJ, Maini RN, Feldmann M, Long-Fox A, Charles P, Bijl H, et al. Repeated therapy with monoclonal antibody to tumour necrosis factor alpha (cA2) in patients with rheumatoid arthritis. Lancet. 1994 Oct 22;344(8930):1125–7.

123. Radner H, Aletaha D. Anti-TNF in rheumatoid arthritis: an overview. Wien Med Wochenschr. Springer Vienna; 2015 Jan;165(1-2):3–9.

124. Tavakolpour S, Alesaeidi S, Darvishi M, GhasemiAdl M, Darabi-Monadi S, Akhlaghdoust M, et al. A comprehensive review of rituximab therapy in rheumatoid arthritis patients. Clin Rheumatol. Springer

London; 2019 Nov;38(11):2977–94.

125. Buch MH, Smolen JS, Betteridge N, Breedveld FC, Burmester G, Dörner T, et al. Updated consensus statement on the use of rituximab in patients with rheumatoid arthritis. 2011. pp. 909–20.

126. Biggioggero M, Crotti C, Becciolini A, Favalli EG. Tocilizumab in the treatment of rheumatoid arthritis: an evidence-based review and patient selection. Drug Des Devel Ther. Dove Press; 2019;13:57–70.

127. Finckh A, Bansback N, Marra CA, Anis AH, Michaud K, Lubin S, et al. Treatment of very early rheumatoid arthritis with symptomatic therapy, disease-modifying antirheumatic drugs, or biologic agents: a cost-effectiveness analysis. Ann Intern Med. 2009 Nov 3;151(9):612–21.

128. Seif F, Khoshmirsafa M, Aazami H, Mohsenzadegan M, Sedighi G, Bahar M. The role of JAK-STAT signaling pathway and its regulators in the fate of T helper cells. Cell Commun Signal. BioMed Central; 2017 Jun 21;15(1):23–13.

129. Furumoto Y, Gadina M. The arrival of JAK inhibitors: advancing the treatment of immune and hematologic disorders. BioDrugs. Springer International Publishing; 2013 Oct;27(5):431–8.

130. Harigai M. Lymphoproliferative disorders in patients with rheumatoid arthritis in the era of widespread use of methotrexate: A review of the literature and current perspective. Mod Rheumatol. 4 ed. 2018 Jan;28(1):1–8.

131. Taylor PC, Keystone EC, van der Heijde D, Weinblatt ME, Del Carmen Morales L, Reyes Gonzaga J, et al. Baricitinib versus Placebo or Adalimumab in Rheumatoid Arthritis. N Engl J Med. Massachusetts Medical Society; 2017 Feb 16;376(7):652–62.

132. Hansen AS, Cattoglio C, Darzacq X, Tjian R. Recent evidence that TADs and chromatin loops are dynamic structures. Nucleus. Taylor & Francis; 2018 Jan 1;9(1):20–32.

133. de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. Genes Dev. Cold Spring Harbor Lab; 2012 Jan 1;26(1):11–24.

134. Magaña-Acosta M, Valadez-Graham V. Chromatin Remodelers in the 3D Nuclear Compartment. Front Genet. 2020;11:600615.

135. Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. Nature Publishing Group; 2016 Nov 1;17(11):661–78.

136. Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Clin Orthop Relat Res. 2007 Sep;462:3–5.

137. Giles KA, Gould CM, Du Q, Skvortsova K, Song JZ, Maddugoda MP, et al. Integrated epigenomic analysis stratifies chromatin remodellers into distinct functional groups. Epigenetics Chromatin. BioMed Central; 2019 Feb 12;12(1):12–9.

138. Rando OJ, Chang HY. Genome-wide views of chromatin structure. Annu Rev Biochem. Annual Reviews; 2009;78(1):245–71.

139. Björkegren C, Baranello L. DNA Supercoiling, Topoisomerases, and Cohesin: Partners in Regulating Chromatin Architecture? Int J Mol Sci. Multidisciplinary Digital Publishing Institute; 2018 Mar 16;19(3):884.

140. Lichter P, Cremer T, Borden J, Manuelidis L, Ward DC. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. Hum Genet. 2nd ed. Springer-Verlag; 1988 Nov;80(3):224–34.

141. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. American Association for the Advancement of Science; 2009 Oct 9;326(5950):289–93.

142. Araki Y, Mimura T. The Histone Modification Code in the Pathogenesis of Autoimmune Diseases. Mediators of Inflammation. Hindawi;

2017;2017(3):2608605–12.

143. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. An operational definition of epigenetics. Genes Dev. 2009 Apr 1;23(7):781–3.

144. Deans C, Maggert KA. What do you mean, "epigenetic"? Genetics. 2015 Apr;199(4):887–96.

145. Bird A. Perceptions of epigenetics. Nature. Nature Publishing Group; 2007 May 24;447(7143):396–8.

146. Deichmann U. Epigenetics: The origins and evolution of a fashionable topic. Dev Biol. 2016 Aug 1;416(1):249–54.

147. Jeffries MA, Sawalha AH. Autoimmune disease in the epigenetic era: how has epigenetics changed our understanding of disease and how can we expect the field to evolve? Expert Review of Clinical Immunology. Informa Healthcare; 2014 Dec 23;11(1):45–58.

148. Crutchley JL, Wang XQD, Ferraiuolo MA, Dostie J. Chromatin conformation signatures: ideal human disease biomarkers? Biomarkers in Medicine. Future Medicine Ltd London, UK; 2010 Aug;4(4):611–29.

149. Moosavi A, Motevalizadeh Ardekani A. Role of Epigenetics in Biology and Human Diseases. Iran Biomed J. Iranian Biomedical Journal; 2016 Nov;20(5):246–58.

150. Goulielmos GN, Zervou MI, Myrthianou E, Burska A, Niewold TB, Ponchel F. Genetic data: The new challenge of personalized medicine, insights for rheumatoid arthritis patients. Gene. 2016 Jun;583(2):90–101.

151. Zufferey F, Williams FMK, Spector TD. Epigenetics and methylation in the rheumatic diseases. Semin Arthritis Rheum. 2014 Apr;43(5):692–700.

152. Trenkmann M, Brock M, Ospelt C, allergy SGCRI, 2010. Epigenetics in rheumatoid arthritis. Springer

.

153. Dumitrescu RG. Early Epigenetic Markers for Precision Medicine. Methods Mol Biol. New York, NY: Springer New York; 2018;1856(3):3–17.

154. Glant TT, Mikecz K, Rauch TA. Epigenetics in the pathogenesis of rheumatoid arthritis. BMC Medicine. BioMed Central; 2014 Feb 26;12(1):860.

155. Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. Nat Rev Rheumatol. Nature Publishing Group; 2013 Mar 1;9(3):141–53.

156. Doody KM, Bottini N, Firestein GS. Epigenetic alterations in rheumatoid arthritis fibroblast-like synoviocytes. Epigenomics. Future Medicine Ltd London, UK; 2017 Apr;9(4):479–92.

157. Cribbs AP, Kennedy A, Penn H, Read JE, Amjadi P, Green P, 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 Rheumatol. John Wiley & Sons, Ltd; 2014 Sep 1;66(9):2344–54.

158. Strahl BD, Allis CD. The language of covalent histone modifications. Nature. Nature Publishing Group; 2000 Jan 6;403(6765):41–5.

159. Araki Y, Wang Z, Zang C, Wood WH, Schones D, Cui K, et al. Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8+ T cells. Immunity. 2009 Jun 19;30(6):912–25.

160. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences. 2010 Dec 14;107(50):21931–6.

161. Toussirot E, Abbas W, Khan KA, Tissot M, Jeudy A, Baud L, et al. Imbalance between HAT and HDAC activities in the PBMCs of patients with ankylosing spondylitis or rheumatoid arthritis and influence of HDAC inhibitors on TNF alpha production. Bobe P, editor. PLoS ONE.

Public Library of Science; 2013;8(8):e70939.

162. Rossetto D, Avvakumov N, Côté J. Histone phosphorylation: a chromatin modification involved in diverse nuclear events. Epigenetics. Taylor & Francis; 2012 Oct;7(10):1098–108.

163. Huber LC, Brock M, Hemmatazad H, Giger OT, Moritz F, Trenkmann M, et al. Histone deacetylase/acetylase activity in total synovial tissue derived from rheumatoid arthritis and osteoarthritis patients. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2007 Apr;56(4):1087–93.

164. Gillespie J, Savic S, Wong C, Hempshall A, Inman M, Emery P, et al. Histone deacetylases are dysregulated in rheumatoid arthritis and a novel histone deacetylase 3-selective inhibitor reduces interleukin-6 production by peripheral blood mononuclear cells from rheumatoid arthritis patients. Arthritis & Rheumatism. 2012 Feb;64(2):418–22.

165. Kadauke S, Blobel GA. Chromatin loops in gene regulation. Biochim Biophys Acta. 2009 Jan;1789(1):17–25.

166. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature. 2013 Oct 20;489:57.

167. Benyajati C, Worcel A. Isolation, characterization, and structure of the folded interphase genome of Drosophila melanogaster. Cell. 1976 Nov;9(3):393–407.

168. Williams A, Spilianakis CG, Flavell RA. Interchromosomal association and gene regulation in trans. Trends Genet. 2010 Apr;26(4):188–97.

169. Holwerda S, de Laat W. Chromatin loops, gene positioning, and gene expression. Front Genet. Frontiers; 2012;3:217.

170. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet. 2005 Jan;76(1):8–32.

171. Carter D, Chakalova L, Osborne CS, Dai Y-F, Fraser P. Long-range chromatin regulatory interactions in vivo. Nat Genet. Nature Publishing Group; 2002 Dec;32(4):623–6.

172. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations between alternatively expressed loci. Nature. Nature Publishing Group; 2005 Jun 2;435(7042):637–45.

173. Spilianakis CG, Flavell RA. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nat Immunol. Nature Publishing Group; 2004 Oct;5(10):1017–27.

174. Chetverina D, Fujioka M, Erokhin M, Georgiev P, Jaynes JB, Schedl P. Boundaries of loop domains (insulators): Determinants of chromosome form and function in multicellular eukaryotes. Bioessays. 2017 Mar;39(3):1600233.

175. Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell. 1999 Aug 6;98(3):387–96.

176. Fu Y, Sinha M, Peterson CL, Weng Z. The Insulator Binding Protein CTCF Positions 20 Nucleosomes around Its Binding Sites across the Human Genome. van Steensel B, editor. PLOS Genetics. Public Library of Science; 2008 Jul 25;4(7):e1000138.

177. Ghirlando R, Felsenfeld G. CTCF: making the right connections. Genes Dev. Cold Spring Harbor Lab; 2016 Apr 15;30(8):881–91.

178. Nakahashi H, Kieffer-Kwon K-R, Resch W, Vian L, Dose M, Stavreva D, et al. A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. 2013 May 30;3(5):1678–89.

179. Merkenschlager M, Nora EP. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu Rev Genomics Hum Genet. Annual Reviews; 2016 Aug 31;17(1):17–43.

180. Nasmyth K, Haering CH. Cohesin: its roles and mechanisms. Annu Rev Genet. Annual Reviews; 2009;43(1):525–58.

181. Degner SC, Wong TP, Jankevicius G, Feeney AJ. Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. J Immunol. American Association of Immunologists; 2009 Jan 1;182(1):44–8.

182. Constância M, Dean W, Lopes S, Moore T, Kelsey G, Reik W. Deletion of a silencer element in Igf2 results in loss of imprinting independent of H19. Nat Genet. Nature Publishing Group; 2000 Oct;26(2):203–6.

183. Tan-Wong SM, French JD, Proudfoot NJ, Brown MA. Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. Proc Natl Acad Sci USA. 2008 Apr 1;105(13):5160–5.

184. Misteli T, Soutoglou E. The emerging role of nuclear architecture in DNA repair and genome maintenance. Nature Reviews Molecular Cell Biology. Nature Publishing Group; 2009 Apr;10(4):243–54.

185. Hansen AS, Pustova I, Cattoglio C, Tjian R, Darzacq X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife. eLife Sciences Publications Limited; 2017 May 3;6:2848.

186. Ho Y, Shewchuk BM, Liebhaber SA, Cooke NE. Distinct chromatin configurations regulate the initiation and the maintenance of hGH gene expression. Molecular and Cellular Biology. 2013 May;33(9):1723–34.

187. Golubnitschaja O, Baban B, Boniolo G, Wang W, Bubnov R, Kapalla M, et al. Medicine in the early twenty-first century: paradigm and anticipation - EPMA position paper 2016. EPMA Journal. BioMed Central; 2016 Dec 1;7(1):1–13.

188. Relling MV, Evans WE. Pharmacogenomics in the clinic. Nature. Nature Publishing Group; 2015 Oct 15;526(7573):343–50.

189. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. Am J Hum Genet. 2018 May 3;102(5):717–30.

190. Ray P, Le Manach Y, Riou B, Houle TT. Statistical evaluation of a biomarker. Anesthesiology. 2010 Apr;112(4):1023–40.

191. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995 Oct 20;270(5235):467–70.

192. Hernández-Cabronero M, Blanes I, Pinho AJ, Marcellin MW, Serra-Sagristà J. Analysis-Driven Lossy Compression of DNA Microarray Images. IEEE Trans Med Imaging. 2016 Feb;35(2):654–64.

193. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proceedings of the National Academy of Sciences. 1994 May 24;91(11):5022–6.

194. Marzancola MG, Sedighi A, Li PCH. DNA Microarray-Based Diagnostics. Methods Mol Biol. New York, NY: Springer New York; 2016;1368(Chapter 12):161–78.

195. Wöhrle J, Krämer SD, Meyer PA, Rath C, Hügle M, Urban GA, et al. Digital DNA microarray generation on glass substrates. Scientific Reports. Nature Publishing Group; 2020 Apr 1;10(1):5770–9.

196. Anderson RC, Su X, Bogdan GJ, Fenton J. A miniature integrated device for automated multistep genetic assays. Nucleic Acids Res. 2000 Jun 15;28(12):E60–0.

197. Dekker J. Capturing Chromosome Conformation. Science. 2002 Feb 15;295(5558):1306–11.

198. Miele A, Dekker J. Mapping cis- and trans- chromatin interaction networks using chromosome conformation capture (3C). Methods Mol Biol. Totowa, NJ: Humana Press; 2009;464(Chapter 7):105–21.

199. Sati S, Cavalli G. Chromosome conformation capture technologies and their impact in understanding genome function. Chromosoma. Springer Berlin Heidelberg; 2017 Feb;126(1):33–44.

200. Dekker J. The three “C” s of chromosome conformation capture: controls, controls, controls. Nat Methods. Nature Publishing Group; 2006 Jan;3(1):17–21.

201. Yao X, Zha W, Song W, He H, Huang M, Jazrawi E, et al. Coordinated regulation of IL-4 and IL-13 expression in human T cells: 3C analysis for DNA looping. Biochem Biophys Res Commun. 2012 Jan 20;417(3):996–1001.

202. Hampel H, Goetzl EJ, Kapogiannis D, Lista S, Vergallo A. Biomarker-Drug and Liquid Biopsy Co-development for Disease Staging and Targeted Therapy: Cornerstones for Alzheimer's Precision Medicine and Pharmacology. Front Pharmacol. Frontiers; 2019;10:310.

203. Soda N, Rehm BHA, Sonar P, Nguyen N-T, Shiddiky MJA. Advanced liquid biopsy technologies for circulating biomarker detection. J Mater Chem B. The Royal Society of Chemistry; 2019 Nov 21;7(43):6670–704.

204. van Delft F, Koffijberg H, Retèl V, Heuvel MVD, IJzerman M. The Validity and Predictive Value of Blood-Based Biomarkers in Prediction of Response in the Treatment of Metastatic Non-Small Cell Lung Cancer: A Systematic Review. Cancers (Basel). Multidisciplinary Digital Publishing Institute; 2020 Apr 30;12(5):1120.

205. IJzerman MJ, Berghuis AMS, de Bono JS, Terstappen LWMM. Health economic impact of liquid biopsies in cancer management. Expert Rev Pharmacoecon Outcomes Res. 2018 Dec;18(6):593–9.

206. Coras R, Narasimhan R, Guma M. Liquid biopsies to guide therapeutic decisions in rheumatoid arthritis. Transl Res. 2018 Nov;201:1–12.

207. Raterman HG, Vosslamber S, de Ridder S, Nurmohamed MT, Lems WF, Boers M, et al. The interferon type I signature towards prediction of non-response to rituximab in rheumatoid arthritis patients. Arthritis Research & Therapy. BioMed Central; 2012 Apr 27;14(2):R95–10.

208. Scarsi M, Ziglioli T, Airò P. Decreased circulating CD28-negative T cells in patients with rheumatoid arthritis treated with abatacept are

correlated with clinical response. J Rheumatol. 2010 May;37(5):911–6.

209. Walsh AM, Whitaker JW, Huang CC, Cherkas Y, Lamberth SL, Brodmerkel C, et al. Integrative genomic deconvolution of rheumatoid arthritis GWAS loci into gene and cell type associations. Genome Biol. BioMed Central; 2016 Apr 30;17(1):2205.

210. You S, Koh JH, Leng L, Kim W-U, Bucala R. The Tumor-Like Phenotype of Rheumatoid Synovium: Molecular Profiling and Prospects for Precision Medicine. Arthritis Rheumatol. 2018 May;70(5):637–52.

211. De Groof A, Ducreux J, Humby F, Nzeusseu Toukap A, Badot V, Pitzalis C, et al. Higher expression of TNFα-induced genes in the synovium of patients with early rheumatoid arthritis correlates with disease activity, and predicts absence of response to first line therapy. Arthritis Research & Therapy. BioMed Central; 2016 Jan 20;18(1):19–12.

212. Humby F, Lewis M, Ramamoorthi N, Hackney JA, Barnes MR, Bombardieri M, et al. Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. Ann Rheum Dis. BMJ Publishing Group Ltd; 2019 Mar 16;:annrheumdis-2018–214539.

213. Ponchel F, Goëb V, Parmar R, El-Sherbiny Y, Boissinot M, Jawhari El J, et al. An immunological biomarker to predict MTX response in early RA. Ann Rheum Dis. 2014 Nov;73(11):2047–53.

214. Borry P, Bentzen HB, Budin-Ljøsne I, Cornel MC, Howard HC, Feeney O, et al. The challenges of the expanded availability of genomic information: an agenda-setting paper. J Community Genet. Springer Berlin Heidelberg; 2018 Apr;9(2):103–16.

215. Yan H, Hunter E, Akoulitchev A, Park P, Winchester DJ, Moo-Young TA, et al. Epigenetic chromatin conformation changes in peripheral blood can detect thyroid cancer. Surgery. 2019 Jan;165(1):44–9.

216. Jakub JW, Grotz TE, Jordan P, Hunter E, Pittelkow M, Ramadass A, et al. A pilot study of chromosomal aberrations and epigenetic changes in peripheral blood samples to identify patients with melanoma. Melanoma Research. 2015 Oct;25(5):406–11.

217. Bastonini E, Jeznach M, Field M, Juszczyk K, Corfield E, Dezfouli M, et al. Chromatin barcodes as biomarkers for melanoma. Pigment Cell & Melanoma Research. 2014 May 27;27(5):788–800.

218. Salter M, Corfield E, Ramadass A, Grand F, Green J, Westra J, et al. Initial Identification of a Blood-Based Chromosome Conformation Signature for Aiding in the Diagnosis of Amyotrophic Lateral Sclerosis. EBioMedicine. 2018 Jul;33:169–84.

219. Carini C, Hunter E, Scottish Early Rheumatoid Arthritis Inception cohort Investigators, Ramadass AS, Green J, Akoulitchev A, et al. Chromosome conformation signatures define predictive markers of inadequate response to methotrexate in early rheumatoid arthritis. J Transl Med. BioMed Central; 2018 Jan 29;16(1):18–11.

220. Pawlik A, Kotrych D, Malinowski D, Dziedziejko V, Czerewaty M, Safranow K. IL17A and IL17F gene polymorphisms in patients with rheumatoid arthritis. BMC Musculoskelet Disord. BioMed Central; 2016 May 11;17(1):208–6.

221. Ely LK, Fischer S, Garcia KC. Structural basis of receptor sharing by interleukin 17 cytokines. Nat Immunol. Nature Publishing Group; 2009 Dec;10(12):1245–51.

222. McKelvey KJ, Millier MJ, Doyle TC, Stamp LK, Highton J, Hessian PA. Co-expression of CD21L and IL17A defines a subset of rheumatoid synovia, characterised by large lymphoid aggregates and high inflammation. Zissel G, editor. PLoS ONE. 2018;13(8):e0202135.

223. van Hamburg JP, Tas SW. Molecular mechanisms underpinning T helper 17 cell heterogeneity and functions in rheumatoid arthritis. J Autoimmun. 2018 Feb;87:69–81.

224. Kirkham BW, Kavanaugh A, Reich K. Interleukin-17A: a unique

pathway in immune-mediated diseases: psoriasis, psoriatic arthritis and rheumatoid arthritis. Immunology. 2014 Feb;141(2):133–42.

225. Blanco FJ, Möricke R, Dokoupilova E, Codding C, Neal J, Andersson M, et al. Secukinumab in Active Rheumatoid Arthritis: A Phase III Randomized, Double-Blind, Active Comparator- and Placebo-Controlled Study. Arthritis Rheumatol. 2017 Jun;69(6):1144–53.

226. Ansel KM, Harris RBS, Cyster JG. CXCL13 is required for B1 cell homing, natural antibody production, and body cavity immunity. Immunity. 2002 Jan;16(1):67–76.

227. Jones JD, Hamilton BJ, Challener GJ, de Brum-Fernandes AJ, Cossette P, Liang P, et al. Serum C-X-C motif chemokine 13 is elevated in early and established rheumatoid arthritis and correlates with rheumatoid factor levels. Arthritis Research & Therapy. BioMed Central; 2014 Apr 25;16(2):R103–9.

228. Armas-González E, Domínguez-Luis MJ, Díaz-Martín A, Arce-Franco M, Castro-Hernández J, Danelon G, et al. Role of CXCL13 and CCL20 in the recruitment of B cells to inflammatory foci in chronic arthritis. Arthritis Research & Therapy. BioMed Central; 2018 Jun 7;20(1):114–12.

229. Takemura S, Braun A, Crowson C, Kurtin PJ, Cofield RH, O'Fallon WM, et al. Lymphoid neogenesis in rheumatoid synovitis. J Immunol. American Association of Immunologists; 2001 Jul 15;167(2):1072–80.

230. Yoshitomi H, Kobayashi S, Miyagawa-Hayashino A, Okahata A, Doi K, Nishitani K, et al. Human Sox4 facilitates the development of CXCL13-producing helper T cells in inflammatory environments. Nat Commun. Nature Publishing Group; 2018 Sep 19;9(1):3762–10.

231. Kobayashi S, Murata K, Shibuya H, Morita M, Ishikawa M, Furu M, et al. A distinct human CD4+ T cell subset that secretes CXCL13 in rheumatoid synovium. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2013 Dec;65(12):3063–72.

232. Moschovakis GL, Bubke A, Friedrichsen M, Falk CS, Feederle R,

Förster R. T cell specific Cxcr5 deficiency prevents rheumatoid arthritis. Scientific Reports. Nature Publishing Group; 2017 Aug 21;7(1):8933–13.

233. Greisen SR, Schelde KK, Rasmussen TK, Kragstrup TW, Stengaard-Pedersen K, Hetland ML, et al. CXCL13 predicts disease activity in early rheumatoid arthritis and could be an indicator of the therapeutic `window of opportunity'. Arthritis Research & Therapy. BioMed Central; 2014 Sep 24;16(5):434.

234. Greisen SR, Schelde KKM, Rasmussen TK, Kragstrup TW, Stengaard-Pedersen K, Hetland ML, et al. CXCL13 predicts disease activity in early rheumatoid arthritis and could be an indicator of the therapeutic `window of opportunity'. Arthritis Research & Therapy. BioMed Central; 2014 Sep 24;16(5):3381.

235. Han BK, Kuzin I, Gaughan JP, Olsen NJ, Bottaro A. Baseline CXCL10 and CXCL13 levels are predictive biomarkers for tumor necrosis factor inhibitor therapy in patients with moderate to severe rheumatoid arthritis: a pilot, prospective study. Arthritis Research & Therapy. BioMed Central; 2016 Apr 22;18(1):93–7.

236. Xing R, Jin Y, Sun L, Yang L, Li C, Li Z, et al. Interleukin-21 induces migration and invasion of fibroblast-like synoviocytes from patients with rheumatoid arthritis. Clinical & Experimental Immunology. 2016 Feb 15;184(2):147–58.

237. Marijnissen RJ, Roeleveld DM, Young D, Nickerson-Nutter C, Abdollahi-Roodsaz S, Garcia de Aquino S, et al. Interleukin-21 receptor deficiency increases the initial toll-like receptor 2 response but protects against joint pathology by reducing Th1 and Th17 cells during streptococcal cell wall arthritis. Arthritis Rheumatol. John Wiley & Sons, Ltd; 2014 Apr;66(4):886–95.

238. Campbell AL, Smith NC, Reilly JH, Kerr SC, Leach WJ, Fazzi UG, et al. IL-21 receptor expression in human tendinopathy. Mediators of Inflammation. Hindawi; 2014;2014(2):481206–7.

239. Dam EM, Maier AC, Hocking AM, Carlin J, Ng B, Buckner JH. Increased Binding of Specificity Protein 1 to the IL21R Promoter in B Cells Results in Enhanced B Cell Responses in Rheumatoid Arthritis. Frontiers in Immunology. 2018;9:1978.

240. Block KE, Huang H. The cellular source and target of IL-21 in K/BxN autoimmune arthritis. J Immunol. 2013 Sep 15;191(6):2948–55.

241. Jüngel A, Distler JHW, Kurowska-Stolarska M, Seemayer CA, Seibl R, Forster A, et al. Expression of interleukin-21 receptor, but not interleukin-21, in synovial fibroblasts and synovial macrophages of patients with rheumatoid arthritis. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2004 May;50(5):1468–76.

242. Yuan F-L, Hu W, Lu W-G, Li X, Li J-P, Xu R-S, et al. Targeting interleukin-21 in rheumatoid arthritis. Mol Biol Rep. Springer Netherlands; 2011 Mar;38(3):1717–21.

243. Young DA, Hegen M, Ma HLM, Whitters MJ, Albert LM, Lowe L, et al. Blockade of the interleukin-21/interleukin-21 receptor pathway ameliorates disease in animal models of rheumatoid arthritis. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2007 Apr;56(4):1152–63.

244. Dam EM, Maier AC, Hocking AM, Carlin J, Ng B, Buckner JH. Corrigendum: Increased Binding of Specificity Protein 1 to the IL21R Promoter in B Cells Results in Enhanced B Cell Responses in Rheumatoid Arthritis. Frontiers in Immunology. Frontiers; 2019;10:1122.

245. Du J, Wang X, Tan G, Liang Z, Zhang Z, Yu H. The association between genetic polymorphisms of interleukin 23 receptor gene and the risk of rheumatoid arthritis: An updated meta-analysis. Clin Immunol. 2020 Jan;210:108250.

246. Razawy W, van Driel M, Lubberts E. The role of IL-23 receptor signaling in inflammation-mediated erosive autoimmune arthritis and bone remodeling. Eur J Immunol. 2018 Feb;48(2):220–9.

247. Bunte K, Beikler T. Th17 Cells and the IL-23/IL-17 Axis in the Pathogenesis of Periodontitis and Immune-Mediated Inflammatory Diseases. Int J Mol Sci. Multidisciplinary Digital Publishing Institute; 2019 Jul 10;20(14):3394.

248. Lubberts E. The IL-23-IL-17 axis in inflammatory arthritis. Nat Rev Rheumatol. Nature Publishing Group; 2015 Oct;11(10):562–2.

249. Cua DJ, Sherlock J, Chen Y, Murphy CA, Joyce B, Seymour B, et al. Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. Nature. Nature Publishing Group; 2003 Feb 13;421(6924):744–8.

250. Zaky DSE, El-Nahrery EMA. Role of interleukin-23 as a biomarker in rheumatoid arthritis patients and its correlation with disease activity. International Immunopharmacology. 2016 Feb;31:105–8.

251. Cornelissen F, Asmawidjaja PS, Mus AMC, Corneth O, Kikly K, Lubberts E. IL-23 dependent and independent stages of experimental arthritis: no clinical effect of therapeutic IL-23p19 inhibition in collagen-induced arthritis. Moura IC, editor. PLoS ONE. Public Library of Science; 2013;8(2):e57553.

252. Smolen JS, Agarwal SK, Ilivanova E, Xu XL, Miao Y, Zhuang Y, et al. A randomised phase II study evaluating the efficacy and safety of subcutaneously administered ustekinumab and guselkumab in patients with active rheumatoid arthritis despite treatment with methotrexate. Ann Rheum Dis. 2017 May;76(5):831–9.

253. Lercher A, Bhattacharya A, Popa AM, Caldera M, Schlapansky MF, Baazim H, et al. Type I Interferon Signaling Disrupts the Hepatic Urea Cycle and Alters Systemic Metabolism to Suppress T Cell Function. Immunity. 2019 Dec 17;51(6):1074–9.

254. Ma H, Yang W, Zhang L, Liu S, Zhao M, Zhou G, et al. Interferon-alpha promotes immunosuppression through IFNAR1/STAT1 signalling in head and neck squamous cell carcinoma. Br J Cancer. Nature Publishing Group; 2019 Feb;120(3):317–30.

255. Harper MS, Guo K, Gibbert K, Lee EJ, Dillon SM, Barrett BS, et al. Interferon-α Subtypes in an Ex Vivo Model of Acute HIV-1 Infection: Expression, Potency and Effector Mechanisms. Emerman M, editor. PLoS Pathog. 2015;11(11):e1005254.

256. Agrawal H, Jacob N, Carreras E, Bajana S, Putterman C, Turner S, et al. Deficiency of type I IFN receptor in lupus-prone New Zealand mixed 2328 mice decreases dendritic cell numbers and activation and protects from disease. J Immunol. American Association of Immunologists; 2009 Nov 1;183(9):6021–9.

257. de Jong TD, Lübbers J, Turk S, Vosslamber S, Mantel E, Bontkes HJ, et al. The type I interferon signature in leukocyte subsets from peripheral blood of patients with early arthritis: a major contribution by granulocytes. Arthritis Research & Therapy. BioMed Central; 2016 Jul 13;18(1):165–10.

258. van der Pouw Kraan TCTM, Wijbrandts CA, van Baarsen LGM, Voskuyl AE, Rustenburg F, Baggen JM, et al. Rheumatoid arthritis subtypes identified by genomic profiling of peripheral blood cells: assignment of a type I interferon signature in a subpopulation of patients. Ann Rheum Dis. 2007 Aug;66(8):1008–14.

259. Lochhead RB, Sonderegger FL, Ma Y, Brewster JE, Cornwall D, Maylor-Hagen H, et al. Endothelial cells and fibroblasts amplify the arthritogenic type I IFN response in murine Lyme disease and are major sources of chemokines in Borrelia burgdorferi-infected joint tissue. J Immunol. 2012 Sep 1;189(5):2488–501.

260. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell. 2016 Nov 17;167(5):1369–1384.e19.

261. Feng Z, Prentice R, Srivastava S. Research issues and strategies for genomic and proteomic biomarker discovery and validation: a statistical perspective. Pharmacogenomics. 2004 Sep;5(6):709–19.

262. Bonassi S, Neri M, Puntoni R. Validation of biomarkers as early predictors of disease. Mutat Res. 2001 Sep 1;480-481:349–58.

263. Lin Y-J, Anzaghe M, Schülke S. Update on the Pathomechanism, Diagnosis, and Treatment Options for Rheumatoid Arthritis. Cells. Multidisciplinary Digital Publishing Institute; 2020 Apr 3;9(4):880.

264. Albrecht K, Müller-Ladner U. Side effects and management of side effects of methotrexate in rheumatoid arthritis. Clin Exp Rheumatol. Clin Exp Rheumatol; 2010 Sep;28(5 Suppl 61):S95–101.

265. Sotoudehmanesh R, Anvari B, Akhlaghi M, Shahraeeni S, Kolahdoozan S. Methotrexate hepatotoxicity in patients with rheumatoid arthritis. Middle East J Dig Dis. Iranian Society of Gastroenterology and Hepatology; 2010 Sep;2(2):104–9.

266. Onishi A, Kamitsuji S, Nishida M, Uemura Y, Takahashi M, Saito T, et al. Genetic and clinical prediction models for the efficacy and hepatotoxicity of methotrexate in patients with rheumatoid arthritis: a multicenter cohort study. The Pharmacogenomics Journal. Nature Publishing Group; 2020 Jun;20(3):433–42.

267. Teitsma XM, Jacobs JWG, Welsing PMJ, de Jong PHP, Hazes JMW, Weel AEAM, et al. Inadequate response to treat-to-target methotrexate therapy in patients with new-onset rheumatoid arthritis: development and validation of clinical predictors. Ann Rheum Dis. 2018 Sep;77(9):1261–7.

268. Maradit-Kremers H, Nicola PJ, Crowson CS, O'Fallon WM, Gabriel SE. Patient, disease, and therapy-related factors that influence discontinuation of disease-modifying antirheumatic drugs: a population-based incidence cohort of patients with rheumatoid arthritis. J Rheumatol. 2006 Feb;33(2):248–55.

269. Du M, Wang L. 3C-digital PCR for quantification of chromatin interactions. BMC Mol Biol. BioMed Central; 2016 Dec 6;17(1):23–9.

270. Hag ge HLN, Klous P, Braem C, Splinter E, Dekker J, Cathala G, et al. Quantitative analysis of chromosome conformation capture assays

(3C-qPCR). Nature Protocols. Nature Publishing Group; 2007 Jul;2(7):1722–33.

271. Würtele H, Chartrand P. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. Chromosome Research. Kluwer Academic Publishers; 2006;14(5):477–95.

272. Bustin S, Huggett J. qPCR primer design revisited. Biomol Detect Quantif. 2017 Dec;14:19–28.

273. Zahurak M, Parmigiani G, Yu W, Scharpf RB, Berman D, Schaeffer E, et al. Pre-processing Agilent microarray data. BMC Bioinformatics. BioMed Central; 2007 May 1;8(1):142–13.

274. Dombkowski AA, Thibodeau BJ, Starcevic SL, Novak RF. Gene-specific dye bias in microarray reference designs. FEBS Lett. 2004 Feb 27;560(1-3):120–4.

275. Dobbin KK, Kawasaki ES, Petersen DW, Simon RM. Characterizing dye bias in microarray experiments. Bioinformatics. 2005 May 15;21(10):2430–7.

276. Clénet M-L, Gagnon F, Moratalla AC, Viel EC, Arbour N. Peripheral human CD4+CD8+ T lymphocytes exhibit a memory phenotype and enhanced responses to IL-2, IL-7 and IL-15. Scientific Reports. Nature Publishing Group; 2017 Sep 14;7(1):11612–5.

277. Psifidi A, Dovas CI, Bramis G, Lazou T, Russel CL, Arsenos G, et al. Comparison of eleven methods for genomic DNA extraction suitable for large-scale whole-genome genotyping and long-term DNA banking using blood samples. PLoS ONE. Public Library of Science; 2015;10(1):e0115960.

278. Mardis E, McCombie WR. Library Quantification: Fluorometric Quantitation of Double-Stranded or Single-Stranded DNA Samples Using the Qubit System. Cold Spring Harb Protoc. 2017 Jun 1;2017(6):pdb.prot094730.

279. Dang J, Mendez P, Lee S, Kim JW, Yoon J-H, Kim TW, et al. Development of a robust DNA quality and quantity assessment qPCR assay for targeted next-generation sequencing library preparation. Int J Oncol. Spandidos Publications; 2016 Oct;49(4):1755–65.

280. Marwa OS, Kalthoum T, Wajih K, Kamel H. Association of IL17A and IL17F genes with rheumatoid arthritis disease and the impact of genetic polymorphisms on response to treatment. Immunology Letters. 2017 Mar;183:24–36.

281. Abou El Hassan M, Bremner R. A rapid simple approach to quantify chromosome conformation capture. Nucleic Acids Res. 2009 Apr;37(5):e35–5.

282. Chen SH, Lin CY, Cho CS, Lo CZ, Hsiung CA. Primer Design Assistant (PDA): A web-based primer design tool. Nucleic Acids Res. 2003 Jul 1;31(13):3751–4.

283. Spandidos A, Wang X, Wang H, Dragnev S, Thurber T, Seed B. A comprehensive collection of experimentally validated primers for Polymerase Chain Reaction quantitation of murine transcript abundance. BMC Genomics. BioMed Central; 2008 Dec 24;9(1):633–17.

284. Eslami G, Salehi R. Development of a chamber system for rapid, high yield and cost-effective purification of deoxyribonucleic acid fragments from agarose gel. Adv Biomed Res. 2014;3(1):78.

285. Gilbert MTP, Haselkorn T, Bunce M, Sanchez JJ, Lucas SB, Jewell LD, et al. The isolation of nucleic acids from fixed, paraffin-embedded tissues-which methods are useful when? Volff J-N, editor. PLoS ONE. 2007 Jun 20;2(6):e537.

286. Li K, Brownley A. Primer design for RT-PCR. Methods Mol Biol. Totowa, NJ: Humana Press; 2010;630(2):271–99.

287. Liu N, Zhou KI, Parisien M, Dai Q, Diatchenko L, Pan T. N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. Nucleic Acids Res. 2017 Jun 2;45(10):6051–63.

288. Nemtsova MV, Zaletaev DV, Bure IV, Mikhaylenko DS, Kuznetsova EB, Alekseeva EA, et al. Epigenetic Changes in the Pathogenesis of Rheumatoid Arthritis. Front Genet. Frontiers; 2019;10:570.

289. Mayeux R. Biomarkers: potential uses and limitations. NeuroRx. Springer-Verlag; 2004 Apr;1(2):182–8.

290. Selaas O, Nordal HH, Halse A-K, Brun JG, Jonsson R, Brokstad KA. Serum Markers in Rheumatoid Arthritis: A Longitudinal Study of Patients Undergoing Infliximab Treatment. Int J Rheumatol. Hindawi; 2015;2015(10):276815–9.

291. Redmond AM, Omarjee S, Chernukhin I, Le Romancer M, Carroll JS. Analysis of HER2 genomic binding in breast cancer cells identifies a global role in direct gene regulation. Mantovani R, editor. PLoS ONE. Public Library of Science; 2019;14(11):e0225180.

292. Tasaki S, Suzuki K, Kassai Y, Takeshita M, Murota A, Kondo Y, et al. Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. Nat Commun. Nature Publishing Group; 2018 Jul 16;9(1):2755.

293. Gerrard DL, Wang Y, Gaddis M, Zhou Y, Wang J, Witt H, et al. Three-dimensional analysis reveals altered chromatin interaction by enhancer inhibitors harbors TCF7L2-regulated cancer gene signature. J Cell Biochem. 2019 Mar;120(3):3056–70.

294. Rousseau M, Ferraiuolo MA, Crutchley JL, Wang XQ, Miura H, Blanchette M, et al. Classifying leukemia types with chromatin conformation data. Genome Biol. BioMed Central; 2014 Apr 30;15(4):R60–12.

295. Rousseau M, Fraser J, Ferraiuolo MA, Dostie J, Blanchette M. Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. BMC Bioinformatics. BioMed Central; 2011 Oct 25;12(1):414–6.

296. Opitz L, Salinas-Riester G, Grade M, Jung K, Jo P, Emons G, et al. Impact of RNA degradation on gene expression profiling. BMC Med

Genomics. BioMed Central; 2010 Aug 9;3(1):36–14.

297. Donohue DE, Gautam A, Miller S-A, Srinivasan S, Abu-Amara D, Campbell R, et al. Gene expression profiling of whole blood: A comparative assessment of RNA-stabilizing collection methods. Gerber AP, editor. PLoS ONE. Public Library of Science; 2019;14(10):e0223065.

298. Byamba D, Kim DY, Kim D-S, Kim T-G, Jee H, Kim SH, et al. Skin-penetrating methotrexate alleviates imiquimod-induced psoriasiform dermatitis via decreasing IL-17-producing gamma delta T cells. Exp Dermatol. John Wiley & Sons, Ltd; 2014 Jul;23(7):492–6.

299. Yu Q, Tong Y, Cui L, Zhang L, Gong Y, Diao H, et al. Efficacy and safety of etanercept combined plus methotrexate and comparison of expression of pro-inflammatory factors expression for the treatment of moderate-to-severe plaque psoriasis. International Immunopharmacology. 2019 Aug;73:442–50.

300. Zivojinovic SM, Pejnovic NN, Sefik-Bukilica MN, Kovacevic LV, Soldatovic II, Damjanov NS. Tumor necrosis factor blockade differentially affects innate inflammatory and Th17 cytokines in rheumatoid arthritis. J Rheumatol. The Journal of Rheumatology; 2012 Jan;39(1):18–21.

301. Jiang H, Mao H, Lu H, Lin P, Garry W, Lu H, et al. Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease. Int J Med Inform. 2021 Jan;145(6):104326.

302. Jacoby M, Gohrbandt S, Clausse V, Brons NH, Muller CP. Interindividual variability and co-regulation of DNA methylation differ among blood cell populations. Epigenetics. Taylor & Francis; 2012 Dec 1;7(12):1421–34.

303. Napoli C, Benincasa G, Donatelli F, Ambrosio G. Precision medicine in distinct heart failure phenotypes: Focus on clinical epigenetics. Am Heart J. 2020 Jun;224:113–28.

304. de Andres MC, Perez-Pampin E, Calaza M, Santaclara FJ, Ortea I, Gomez-Reino JJ, et al. Assessment of global DNA methylation in peripheral blood cell subpopulations of early rheumatoid arthritis before and after methotrexate. Arthritis Research & Therapy. BioMed Central; 2015 Aug 29;17(1):233–9.

305. Crawford GE, Holt IE, Mullikin JC, Tai D, Blakesley R, Bouffard G, et al. Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. Proceedings of the National Academy of Sciences. 2004 Jan 27;101(4):992–7.

306. Crawford GE, Davis S, Scacheri PC, Renaud G, Halawi MJ, Erdos MR, et al. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nat Methods. Nature Publishing Group; 2006 Jul;3(7):503–9.

307. Jou J, Gabdank I, Luo Y, Lin K, Sud P, Myers Z, et al. The ENCODE Portal as an Epigenomics Resource. Curr Protoc Bioinformatics. John Wiley & Sons, Ltd; 2019 Dec;68(1):e89.

308. Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, et al. BLUEPRINT to decode the epigenetic signature written in blood. Nat Biotechnol. Nature Publishing Group; 2012 Mar 7;30(3):224–6.

309. Julià A, Absher D, López-Lasanta M, Palau N, Pluma A, Waite Jones L, et al. Epigenome-wide association study of rheumatoid arthritis identifies differentially methylated loci in B cells. Hum Mol Genet. 2017 Jul 15;26(14):2803–11.

310. Ospelt C. Epigenetic biomarkers in rheumatology - the future? Swiss Medical Weekly. 2016;146:w14312.

311. Carlsen HS, Baekkevold ES, Morton HC, Haraldsen G, Brandtzaeg P. Monocyte-like and mature macrophages produce CXCL13 (B cell-attracting chemokine 1) in inflammatory lesions with lymphoid neogenesis. Blood. 2004 Nov 15;104(10):3021–7.

312. Wang Y, Lloyd KA, Melas I, Zhou D, Thyagarajan R, Lindqvist J, et al. Rheumatoid arthritis patients display B-cell dysregulation already in

the naïve repertoire consistent with defects in B-cell tolerance. Scientific Reports. Nature Publishing Group; 2019 Dec 27;9(1):19995–13.

313. Han S, Zhuang H, Lee PY, Li M, Yang L, Nigrovic PA, et al. Differential Responsiveness of Monocyte and Macrophage Subsets to Interferon. Arthritis Rheumatol. John Wiley & Sons, Ltd; 2020 Jan;72(1):100–13.

314. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods. Nature Publishing Group; 2007 Aug;4(8):651–7.

315. Spicuglia S, Vanhille L. Chromatin signatures of active enhancers. Nucleus. 2012 Mar 1;3(2):126–31.

316. Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc. 2010 Feb;2010(2):pdb.prot5384–4.

317. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. Nature Research; 2013 Dec;10(12):1213–8.

318. Mamberti S, Cardoso MC. Are the processes of DNA replication and DNA repair reading a common structural chromatin unit? Nucleus. Taylor & Francis; 2020 Dec;11(1):66–82.

319. Zhao L, Wang S, Cao Z, Ouyang W, Zhang Q, Xie L, et al. Chromatin loops associated with active genes and heterochromatin shape rice genome architecture for transcriptional regulation. Nat Commun. Nature Publishing Group; 2019 Aug 13;10(1):3640–13.

320. Tark-Dame M, van Driel R, Heermann DW. Chromatin folding--from biology to polymer models and back. J Cell Sci. The Company of Biologists Ltd; 2011 Mar 15;124(Pt 6):839–45.

321. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. Nature Publishing Group; 2008 Jul;40(7):897–903.

322. Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, et al. Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol. Nature Publishing Group; 2005 Jan;23(1):137–44.

323. Patel DD, Lee DM, Kolbinger F, Antoni C. Effect of IL-17A blockade with secukinumab in autoimmune diseases. Ann Rheum Dis. 2013 Apr;72 Suppl 2(suppl 2):ii116–23.

324. Mora A, Sandve GK, Gabrielsen OS, Eskeland R. In the loop: promoter-enhancer interactions and bioinformatics. Brief Bioinformatics. 2016 Nov;17(6):980–95.

325. Caterina JJ, Ciavatta DJ, Donze D, Behringer RR, Townes TM. Multiple elements in human beta-globin locus control region 5' HS 2 are involved in enhancer activity and position-independent, transgene expression. Nucleic Acids Res. 1994 Mar 25;22(6):1006–11.

326. Strauss EC, Orkin SH. In vivo protein-DNA interactions at hypersensitive site 3 of the human beta-globin locus control region. Proceedings of the National Academy of Sciences. 1992 Jul 1;89(13):5809–13.

327. Roberts CA, Dickinson AK, Taams LS. The Interplay Between Monocytes/Macrophages and CD4(+) T Cell Subsets in Rheumatoid Arthritis. Frontiers in Immunology. Frontiers; 2015;6(3):571.

328. Pogue SL, Preston BT, Stalder J, Bebbington CR, Cardarelli PM. The receptor for type I IFNs is highly expressed on peripheral blood B cells and monocytes and mediates a distinct profile of differentiation and activation of these cells. J Interferon Cytokine Res. Mary Ann Liebert, Inc; 2004 Feb;24(2):131–9.

329. Bolzoni M, Ronchetti D, Storti P, Donofrio G, Marchica V, Costa F, et

al. IL21R expressing CD14+CD16+ monocytes expand in multiple myeloma patients leading to increased osteoclasts. Haematologica. 2017 Apr;102(4):773–84.

330. Zhu Z, Wang X. Roles of cohesin in chromosome architecture and gene expression. Semin Cell Dev Biol. 2019 Jun;90:187–93.

331. Li Y, Huang W, Niu L, Umbach DM, Covo S, Li L. Characterization of constitutive CTCF/cohesin loci: a possible role in establishing topological domains in mammalian genomes. BMC Genomics. BioMed Central; 2013 Aug 14;14(1):553–12.

332. Domeier PP, Chodisetti SB, Schell SL, Kawasawa YI, Fasnacht MJ, Soni C, et al. B-Cell-Intrinsic Type 1 Interferon Signaling Is Crucial for Loss of Tolerance and the Development of Autoreactive B Cells. Cell Rep. 2018 Jul 10;24(2):406–18.

333. Silverman GJ, Carson DA. Roles of B cells in rheumatoid arthritis. Arthritis Research & Therapy. BioMed Central; 2003;5 Suppl 4(Suppl 4):S1–6.

334. Marston B, Palanichamy A, Anolik JH. B cells in the pathogenesis and treatment of rheumatoid arthritis. Curr Opin Rheumatol. 2010 May;22(3):307–15.

335. Görtz B, Hayer S, Redlich K, Zwerina J, Tohidast-Akrad M, Tuerk B, et al. Arthritis induces lymphocytic bone marrow inflammation and endosteal bone formation. J Bone Miner Res. 2004 Jun;19(6):990–8.

336. Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, et al. A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res. Cold Spring Harbor Lab; 2010 May;20(5):578–88.

337. Medvedovic J, Ebert A, Tagoh H, Busslinger M. Pax5: a master regulator of B cell development and leukemogenesis. Adv Immunol. Elsevier; 2011;111:179–206.

338. Siersbæk R, Madsen JGS, Javierre BM, Nielsen R, Bagge EK, Cairns J,

et al. Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. Mol Cell. 2017 May 4;66(3):420–5.

339. Downes DJ, Beagrie RA, Gosden ME, Telenius J, Carpenter SJ, Nussbaum L, et al. High-resolution targeted 3C interrogation of cis-regulatory element organization at genome-wide scale. Nat Commun. Nature Publishing Group; 2021 Jan 22;12(1):531–15.

340. Wang J, Dai X, Berry LD, Cogan JD, Liu Q, Shyr Y. HACER: an atlas of human active enhancers to interpret regulatory variants. Nucleic Acids Res. 2019 Jan 8;47(D1):D106–12.

341. Choy M-K, Javierre BM, Williams SG, Baross SL, Liu Y, Wingett SW, et al. Author Correction: Promoter interactome of human embryonic stem cell-derived cardiomyocytes connects GWAS regions to cardiac gene networks. Nat Commun. Nature Publishing Group; 2018 Nov 12;9(1):4792–1.

342. Kapellos TS, Iqbal AJ. Epigenetic Control of Macrophage Polarisation and Soluble Mediator Gene Expression during Inflammation. Mediators of Inflammation. Hindawi; 2016;2016(3):6591703–15.

343. Gordon RA, Grigoriev G, Lee A, Kalliolias GD, Ivashkiv LB. The interferon signature and STAT1 expression in rheumatoid arthritis synovial fluid macrophages are induced by tumor necrosis factor α and counter-regulated by the synovial fluid microenvironment. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2012 Oct;64(10):3119–28.

344. Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet. Nature Publishing Group; 2012 Dec;13(12):840–52.

345. Qu W, Tsukahara T, Nakamura R, Yurino H, Hashimoto S-I, Tsuji S, et al. Assessing Cell-to-Cell DNA Methylation Variability on Individual Long Reads. Scientific Reports. Nature Publishing Group; 2016 Feb 18;6(1):21317–7.

346. Beck D, Brandl MB, Boelen L, Unnikrishnan A, Pimanda JE, Wong JWH.

Signal analysis for genome-wide maps of histone modifications measured by ChIP-seq. Bioinformatics. 2012 Apr 15;28(8):1062–9.

347. Karlić R, Chung H-R, Lasserre J, Vlahovicek K, Vingron M. Histone modification levels are predictive for gene expression. Proc Natl Acad Sci USA. 2010 Feb 16;107(7):2926–31.

348. Wampler Muskardin TL, Fan W, Jin Z, Jensen MA, Dorschner JM, Ghodke-Puranik Y, et al. Distinct Single Cell Gene Expression in Peripheral Blood Monocytes Correlates With Tumor Necrosis Factor Inhibitor Treatment Response Groups Defined by Type I Interferon in Rheumatoid Arthritis. Frontiers in Immunology. Frontiers; 2020;11:1384.

349. Sims NA, Romas E. Is RANKL inhibition both anti-resorptive and anabolic in rheumatoid arthritis? Arthritis Research & Therapy. BioMed Central; 2015 Nov 17;17(1):328–2.

350. Andersson AK, Feldmann M, Brennan FM. Neutralizing IL-21 and IL-15 inhibits pro-inflammatory cytokine production in rheumatoid arthritis. Scand J Immunol. John Wiley & Sons, Ltd; 2008 Jul;68(1):103–11.

351. Blair JPM, Bager C, Platt A, Karsdal M, Bay-Jensen A-C. Identification of pathological RA endotypes using blood-based biomarkers reflecting tissue metabolism. A retrospective and explorative analysis of two phase III RA studies. Gualillo O, editor. PLoS ONE. Public Library of Science; 2019;14(7):e0219980.

352. Lewis MJ, Barnes MR, Blighe K, Goldmann K, Rana S, Hackney JA, et al. Molecular Portraits of Early Rheumatoid Arthritis Identify Clinical and Treatment Response Phenotypes. Cell Rep. 2019 Aug 27;28(9):2455–5.

353. Hunter E, McCord R, Ramadass AS, Green J, Westra JW, Mundt K, et al. Comparative molecular cell-of-origin classification of diffuse large B-cell lymphoma based on liquid and tissue biopsies. transl med commun. BioMed Central; 2020 Dec 1;5(1):1–13.

354. Courvoisier DS, Alpizar-Rodriguez D, Gottenberg JE, Hernandez MV,

Iannone F, Lie E, et al. Rheumatoid Arthritis Patients after Initiation of a New Biologic Agent: Trajectories of Disease Activity in a Large Multinational Cohort Study. EBioMedicine. 2016 Sep;11:302–6.

355. Siemons L, Klooster ten PM, Vonkeman HE, Glas CAW, van de Laar MAFJ. Distinct Trajectories of Disease Activity Over the First Year in Early Rheumatoid Arthritis Patients Following a Treat-to-Target Strategy. Arthritis Care & Research. John Wiley & Sons, Ltd; 2014 Mar 26;66(4):625–30.

356. Barnabe C, Sun Y, Boire G, Hitchon CA, Haraoui B, Thorne JC, et al. Heterogeneous Disease Trajectories Explain Variable Radiographic, Function and Quality of Life Outcomes in the Canadian Early Arthritis Cohort (CATCH). Kuwana M, editor. PLoS ONE. Public Library of Science; 2015;10(8):e0135327.

357. Drouin J, Haraoui B, 3e Initiative Group. Predictors of clinical response and radiographic progression in patients with rheumatoid arthritis treated with methotrexate monotherapy. J Rheumatol. The Journal of Rheumatology; 2010 Jul;37(7):1405–10.

358. Aletaha D, Funovits J, Keystone EC, Smolen JS. Disease activity early in the course of treatment predicts response to therapy after one year in rheumatoid arthritis patients. Arthritis & Rheumatism. Wiley Subscription Services, Inc., A Wiley Company; 2007 Oct;56(10):3226–35.

359. Katchamart W, Johnson S, Lin H-JL, Phumethum V, Salliot C, Bombardier C. Predictors for remission in rheumatoid arthritis patients: A systematic review. Arthritis Care & Research. John Wiley & Sons, Ltd; 2010 Aug;62(8):1128–43.

360. Verstappen SMM, van Albada-Kuipers GA, Bijlsma JWJ, Blaauw AAM, Schenk Y, Haanen HCM, et al. A good response to early DMARD treatment of patients with rheumatoid arthritis in the first year predicts remission during follow up. Ann Rheum Dis. 2005 Jan;64(1):38–43.

361. Drossaers-Bakker KW, Zwinderman AH, Vliet Vlieland TPM, Van Zeben D, Vos K, Breedveld FC, et al. Long-term outcome in rheumatoid arthritis: a simple algorithm of baseline parameters can predict radiographic damage, disability, and disease course at 12-year followup. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2002 Aug;47(4):383–90.

362. Saevarsdottir S, Wallin H, Seddighzadeh M, Ernestam S, Geborek P, Petersson IF, et al. Predictors of response to methotrexate in early DMARD naive rheumatoid arthritis: results from the initial open-label phase of the SWEFOT trial. Ann Rheum Dis. 2011 Mar;70(3):469–75.

363. Michelsen B, Kristianslund EK, Sexton J, Hammer HB, Fagerli KM, Lie E, et al. Do depression and anxiety reduce the likelihood of remission in rheumatoid arthritis and psoriatic arthritis? Data from the prospective multicentre NOR-DMARD study. Ann Rheum Dis. BMJ Publishing Group Ltd; 2017 Nov;76(11):1906–10.

364. Pope JE, Rampakakis E, Movahedi M, Cesta A, Li X, Couto S, et al. Treatment patterns in rheumatoid arthritis after discontinuation of methotrexate: data from the Ontario Best Practices Research Initiative (OBRI). Clin Exp Rheumatol. Clin Exp Rheumatol; 2018 Mar;36(2):215–22.

365. Rigby WFC, Lampl K, Low JM, Furst DE. Review of Routine Laboratory Monitoring for Patients with Rheumatoid Arthritis Receiving Biologic or Nonbiologic DMARDs. Int J Rheumatol. Hindawi; 2017;2017(2):9614241–15.

366. Capell HA, Madhok R, Porter DR, Munro RAL, McInnes IB, Hunter JA, et al. Combination therapy with sulfasalazine and methotrexate is more effective than either drug alone in patients with rheumatoid arthritis with a suboptimal response to sulfasalazine: results from the double-blind placebo-controlled MASCOT study. Ann Rheum Dis. BMJ Publishing Group Ltd; 2007 Feb;66(2):235–41.

367. Shaw Y, Metes ID, Michaud K, Donohue JM, Roberts MS, Levesque MC, et al. Rheumatoid Arthritis Patients' Motivations for Accepting or

Resisting Disease-Modifying Antirheumatic Drug Treatment Regimens. Arthritis Care & Research. 3rd ed. 2018 Apr;70(4):533–41.

368. Fraenkel L, Seng EK, Cunningham M, Mattocks K. Understanding how patients (vs physicians) approach the decision to escalate treatment: a proposed conceptual model. Rheumatology (Oxford). 2015 Feb;54(2):278–85.

369. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, et al. A comparison of background correction methods for two-colour microarrays. Bioinformatics. 2007 Oct 15;23(20):2700–7.

370. Akey JM, Biswas S, Leek JT, Storey JD. On the design and analysis of gene expression studies in human populations. Nat Genet. Nature Publishing Group; 2007 Jul;39(7):807–8–authorreply808–9.

371. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47–7.

372. Del Carratore F, Jankevics A, Eisinga R, Heskes T, Hong F, Breitling R. RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets. Bar-Joseph Z, editor. Bioinformatics. 2017 Sep 1;33(17):2774–5.

373. Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett. 2004 Aug 27;573(1-3):83–92.

374. Plantinga TS, Fransen J, Takahashi N, Stienstra R, van Riel PL, van den Berg WB, et al. Functional consequences of DECTIN-1 early stop codon polymorphism Y238X in rheumatoid arthritis. Arthritis Research & Therapy. BioMed Central; 2010;12(1):R26–10.

375. Shi M, Wang J, Xiao Y, Wang C, Qiu Q, Lao M, et al. Glycogen Metabolism and Rheumatoid Arthritis: The Role of Glycogen Synthase 1 in Regulation of Synovial Inflammation via Blocking AMP-Activated Protein Kinase Activation. Frontiers in Immunology. Frontiers;

2018;9:1714.

376. Gaber T, Dziurla R, Tripmacher R, Burmester GR, Buttgereit F. Hypoxia inducible factor (HIF) in rheumatology: low O2! See what HIF can do! Ann Rheum Dis. BMJ Publishing Group Ltd; 2005 Jul;64(7):971–80.

377. Weyand CM, Goronzy JJ. Immunometabolism in the development of rheumatoid arthritis. Immunol Rev. John Wiley & Sons, Ltd; 2020 Mar;294(1):177–87.

378. Pucino V, Certo M, Varricchi G, Marone G, Ursini F, Rossi FW, et al. Metabolic Checkpoints in Rheumatoid Arthritis. Front Physiol. Frontiers; 2020 Apr 17;11:707.

379. Tan W, Sunahori K, Zhao J, Deng Y, Kaufman KM, Kelly JA, et al. Association of PPP2CA polymorphisms with systemic lupus erythematosus susceptibility in multiple ethnic groups. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2011 Sep;63(9):2755–63.

380. Zhang J, Meng Y, Wu H, Wu Y, Yang B, Wang L. Association between PPP2CA polymorphisms and clinical features in southwest Chinese systemic lupus erythematosus patients. Medicine (Baltimore). 2018 Jul;97(27):e11451.

381. López-Mejías R, Corrales A, Vicente E, Robustillo-Villarino M, González-Juanatey C, Llorca J, et al. Influence of coronary artery disease and subclinical atherosclerosis related polymorphisms on the risk of atherosclerosis in rheumatoid arthritis. Scientific Reports. Nature Publishing Group; 2017 Jan 6;7(1):40303–6.

382. Garcia AI, Buisson M, Bertrand P, Rimokh R, Rouleau E, Lopez BS, et al. Down-regulation of BRCA1 expression by miR-146a and miR-146b-5p in triple negative sporadic breast cancers. EMBO Mol Med. John Wiley & Sons, Ltd; 2011 May;3(5):279–90.

383. Chang S, Sharan SK. BRCA1 and microRNAs: emerging networks and potential therapeutic targets. Mol Cells. 2012 Nov;34(5):425–32.

384. Pauley KM, Cha S. miRNA-146a in rheumatoid arthritis: a new therapeutic strategy. Immunotherapy. Future Medicine Ltd London, UK; 2011 Jul;3(7):829–31.

385. Peralta-Leal A, Rodríguez-Vargas JM, Aguilar-Quesada R, Rodríguez MI, Linares JL, de Almodóvar MR, et al. PARP inhibitors: new partners in the therapy of cancer and inflammatory diseases. Free Radic Biol Med. 2009 Jul 1;47(1):13–26.

386. Pierer M, Wagner U, Rossol M, Ibrahim S. Toll-like receptor 4 is involved in inflammatory and joint destructive pathways in collagen-induced arthritis in DBA1J mice. Feghali-Bostwick C, editor. PLoS ONE. Public Library of Science; 2011;6(8):e23539.

387. Samarpita S, Kim JY, Rasool MK, Kim KS. Investigation of toll-like receptor (TLR) 4 inhibitor TAK-242 as a new potential anti-rheumatoid arthritis drug. Arthritis Research & Therapy. BioMed Central; 2020 Jan 23;22(1):16–10.

388. Choi IY, Gerlag DM, Herenius MJ, Thurlings RM, Wijbrandts CA, Foell D, et al. MRP8/14 serum levels as a strong predictor of response to biological treatments in patients with rheumatoid arthritis. Ann Rheum Dis. BMJ Publishing Group Ltd; 2015 Mar;74(3):499–505.

389. Cavalli M, Baltzer N, Umer HM, Grau J, Lemnian I, Pan G, et al. Allele specific chromatin signals, 3D interactions, and motif predictions for immune and B cell related diseases. Scientific Reports. Nature Publishing Group; 2019 Feb 25;9(1):2695–14.

390. Fraser J, Rousseau M, Shenker S, Ferraiuolo MA, Hayashizaki Y, Blanchette M, et al. Chromatin conformation signatures of cellular differentiation. Genome Biol. BioMed Central; 2009;10(4):R37–18.

391. Liu H-W, Lin H-L, Yen J-H, Tsai W-C, Chiou S-S, Chang J-G, et al. Demethylation within the proximal promoter region of human estrogen receptor alpha gene correlates with its enhanced expression: Implications for female bias in lupus. Mol Immunol. 2014 Sep;61(1):28–37.

392. Algate K, Haynes D, Fitzsimmons T, Romeo O, Wagner F, Holson E, et al. Histone deacetylases 1 and 2 inhibition suppresses cytokine production and osteoclast bone resorption in vitro. J Cell Biochem. John Wiley & Sons, Ltd; 2020 Jan;121(1):244–58.

393. van Heemst J, Hensvold AH, Jiang X, van Steenbergen H, Klareskog L, Huizinga TWJ, et al. Protective effect of HLA-DRB1*13 alleles during specific phases in the development of ACPA-positive RA. Ann Rheum Dis. BMJ Publishing Group Ltd; 2016 Oct 1;75(10):1891–8.

394. Kim K-W, Kim H-R, Park J-Y, Park J-S, Oh H-J, Woo Y-J, et al. Interleukin-22 promotes osteoclastogenesis in rheumatoid arthritis through induction of RANKL in human synovial fibroblasts. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2012 Apr;64(4):1015–23.

395. Pörings A-S, Lowin T, Dufner B, Grifka J, Straub RH. A thyroid hormone network exists in synovial fibroblasts of rheumatoid arthritis and osteoarthritis patients. Scientific Reports. Nature Publishing Group; 2019 Sep 13;9(1):13235–14.

396. CLARK WS, WATKINS AL, TONNING HO, BAUER W. The effects of resistance exercises on the nitrogen, phosphorus, and calcium metabolism of patients with rheumatoid arthritis. J Clin Invest. American Society for Clinical Investigation; 1954 Apr;33(4):505–9.

397. Boland K, Flanagan L, McCawley N, Pabari R, Kay EW, McNamara DA, et al. Targeting the 19S proteasomal subunit, Rpt4, for the treatment of colon cancer. Eur J Pharmacol. 2016 Jun 5;780(1 Suppl.):53–64.

398. Sjakste T, Paramonova N, Rumba-Rozenfelde I, Trapina I, Sugoka O, Sjakste N. Juvenile idiopathic arthritis subtype- and sex-specific associations with genetic variants in the PSMA6/PSMC6/PSMA3 gene cluster. Pediatr Neonatol. 2014 Oct;55(5):393–403.

399. Men L-J, Liu J-Z, Chen H-Y, Zhang L, Chen S-F, Xiao T-W, et al. Down regulation of G protein-coupled receptor 137 expression inhibits proliferation and promotes apoptosis in leukemia cells. Cancer Cell Int. BioMed Central; 2018;18(1):13–10.

400. Di Giammartino DC, Polyzos A, Apostolou E. Transcription factors: building hubs in the 3D space. Cell Cycle. Taylor & Francis; 2020 Oct;19(19):2395–410.

401. Petrovic J, Zhou Y, Fasolino M, Goldman N, Schwartz GW, Mumbach MR, et al. Oncogenic Notch Promotes Long-Range Regulatory Interactions within Hyperconnected 3D Cliques. Mol Cell. 2019 Mar 21;73(6):1174–1190.e12.

402. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012 Jan 20;148(1-2):84–98.

403. Brenner M, Laragione T, Gulko PS. Analyses of synovial tissues from arthritic and protected congenic rat strains reveal a new core set of genes associated with disease severity. Physiol Genomics. 2013 Nov 15;45(22):1109–22.

404. Negi S, Juyal G, Senapati S, Prasad P, Gupta A, Singh S, et al. A genome-wide association study reveals ARL15, a novel non-HLA susceptibility gene for rheumatoid arthritis in North Indians. Arthritis & Rheumatism. John Wiley & Sons, Ltd; 2013 Dec;65(12):3026–35.

405. Kar S, Islam C. New approaches for biomarker stability determination in regulated bioanalysis: trending, bridging and incurred samples. Bioanalysis. 2019 Oct;11(20):1837–44.

406. Yang Y, Huang N, Hao L, Kong W. A clustering-based approach for efficient identification of microRNA combinatorial biomarkers. BMC Genomics. BioMed Central; 2017 Mar 14;18(Suppl 2):210–4.

407. Olson MT, Epstein JA, Sackett DL, Yergey AL. Production of reliable MALDI spectra with quality threshold clustering of replicates. J Am Soc Mass Spectrom. Springer-VerlagNew York; 2011 Jun;22(6):969–75.

408. Matthijssen XME, Niemantsverdriet E, Huizinga TWJ, van der Helm-van Mil AHM. Enhanced treatment strategies and distinct disease outcomes among autoantibody-positive and -negative rheumatoid

arthritis patients over 25 years: A longitudinal cohort study in the Netherlands. Montecucco C, editor. PLoS Med. Public Library of Science; 2020 Sep;17(9):e1003296.

409. Gomez-Cabrero D, Almgren M, Sjöholm LK, Hensvold AH, Ringh MV, Tryggvadottir R, et al. High-specificity bioinformatics framework for epigenomic profiling of discordant twins reveals specific and shared markers for ACPA and ACPA-positive rheumatoid arthritis. Genome Med. BioMed Central; 2016 Nov 22;8(1):124–15.

410. Davis KD, Aghaeepour N, Ahn AH, Angst MS, Borsook D, Brenton A, et al. Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. Nat Rev Neurol. 2020 Jul;16(7):381–400.

411. Kowalski LP, Coletta RD, Salo T, Maschietto M, Chojniak R, Lima JM, et al. Head and neck cancer: Emerging concepts in biomarker discovery and opportunities for clinical translation. Clin Transl Med. John Wiley & Sons, Ltd; 2020 Nov;10(7):e209.

412. Silberring J, Ciborowski P. Biomarker discovery and clinical proteomics. Trends Analyt Chem. 2010 Feb 1;29(2):128–40.

413. Ren AH, Fiala CA, Diamandis EP, Kulasingam V. Pitfalls in Cancer Biomarker Discovery and Validation with Emphasis on Circulating Tumor DNA. Cancer Epidemiol Biomarkers Prev. 2020 Dec;29(12):2568–74.

414. Yekula A, Muralidharan K, Kang KM, Wang L, Balaj L, Carter BS. From laboratory to clinic: Translation of extracellular vesicle based cancer biomarkers. Methods. 2020 May 1;177:58–66.

415. Oh D-J. A long journey for acute kidney injury biomarkers. Ren Fail. Taylor & Francis; 2020 Nov;42(1):154–65.

416. Chaikijurajai T, Tang WHW. Reappraisal of Inflammatory Biomarkers in Heart Failure. Curr Heart Fail Rep. 2020 Feb;17(1):9–19.

417. McGranaghan P, Düngen H-D, Saxena A, Rubens M, Salami J,

Radenkovic J, et al. Incremental prognostic value of a novel metabolite-based biomarker score in congestive heart failure patients. ESC Heart Fail. 2020 Oct;7(5):3029–39.

418. Luo Y, Wu J-Y, Lu M-H, Shi Z, Na N, Di J-M. Carvacrol Alleviates Prostate Cancer Cell Proliferation, Migration, and Invasion through Regulation of PI3K/Akt and MAPK Signaling Pathways. Oxid Med Cell Longev. Hindawi; 2016;2016(5):1469693–11.

419. Saura C, Roda D, Roselló S, Oliveira M, Macarulla T, Pérez-Fidalgo JA, et al. A First-in-Human Phase I Study of the ATP-Competitive AKT Inhibitor Ipatasertib Demonstrates Robust and Safe Targeting of AKT in Patients with Solid Tumors. Cancer Discov. American Association for Cancer Research; 2017 Jan;7(1):102–13.

420. Grimison PS, Stockler MR, Chatfield M, Thomson DB, Gebski V, Friedlander M, et al. Accelerated BEP for metastatic germ cell tumours: a multicenter phase II trial by the Australian and New Zealand Urogenital and Prostate Cancer Trials Group (ANZUP). Ann Oncol. 2014 Jan;25(1):143–8.

421. Liu C, Armstrong C, Zhu Y, Lou W, Gao AC. Niclosamide enhances abiraterone treatment via inhibition of androgen receptor variants in castration resistant prostate cancer. Oncotarget. Impact Journals; 2016 May 31;7(22):32210–20.

422. Spiliopoulou A, Colombo M, Plant D, Nair N, Cui J, Coenen MJ, et al. Association of response to TNF inhibitors in rheumatoid arthritis with quantitative trait loci for CD40 and CD39. Ann Rheum Dis. 2019 Aug;78(8):1055–61.

423. Tao W, Concepcion AN, Vianen M, Marijnissen ACA, Lafeber FPGJ, Radstake TRDJ, et al. Multiomics and Machine Learning Accurately Predict Clinical Response to Adalimumab and Etanercept Therapy in Patients With Rheumatoid Arthritis. Arthritis Rheumatol. John Wiley & Sons, Ltd; 2020 Sep 10;10:3310.

424. Roodenrijs NMT, Hamar A, Kedves M, Nagy G, van Laar JM, van der

Heijde D, et al. Pharmacological and non-pharmacological therapeutic strategies in difficult-to-treat rheumatoid arthritis: a systematic literature review informing the EULAR recommendations for the management of difficult-to-treat rheumatoid arthritis. RMD Open. BMJ Specialist Journals; 2021 Jan;7(1):e001512.

425. Wilkinson MF, Shyu AB. Multifunctional regulatory proteins that control gene expression in both the nucleus and the cytoplasm. Bioessays. John Wiley & Sons, Ltd; 2001 Sep;23(9):775–87.

426. Bitetto G, Di Fonzo A. Nucleo-cytoplasmic transport defects and protein aggregates in neurodegeneration. Transl Neurodegener. BioMed Central; 2020 Jul 3;9(1):25–16.

427. Shi C-X, Kortüm KM, Zhu YX, Bruins LA, Jedlowski P, Votruba PG, et al. CRISPR Genome-Wide Screening Identifies Dependence on the Proteasome Subunit PSMC6 for Bortezomib Sensitivity in Multiple Myeloma. Mol Cancer Ther. 2017 Dec;16(12):2862–70.

428. Zemeckiene Z, Sitkauskiene B, Gasiuniene E, Paramonova N, Tamasauskiene L, Vitkauskiene A, et al. Evaluation of proteasomal gene polymorphisms in Lithuanian patients with asthma. J Asthma. Informa Healthcare; 2015 Jun;52(5):447–52.

429. Deng J, Tan H, Hu J, Su G, Cao Q, Huang X, et al. Genetic aspects of idiopathic paediatric uveitis and juvenile idiopathic arthritis associated uveitis in Chinese Han. Br J Ophthalmol. BMJ Publishing Group Ltd; 2020 Mar;104(3):443–7.

430. Albert PS, Schisterman EF. Novel statistical methodology for analyzing longitudinal biomarker data. Schisterman EF, Albert PS, editors. Stat Med. John Wiley & Sons, Ltd; 2012 Sep 28;31(22):2457–60.

431. Ho D, Quake SR, McCabe ERB, Chng WJ, Chow EK, Ding X, et al. Enabling Technologies for Personalized and Precision Medicine. Trends Biotechnol. 2020 May;38(5):497–518.

432. Westerman K, Reaver A, Roy C, Ploch M, Sharoni E, Nogal B, et al.

Publisher Correction: Longitudinal analysis of biomarker data from a personalized nutrition platform in healthy subjects. Scientific Reports. Nature Publishing Group; 2018 Nov 8;8(1):16778–1.

433. Lakhanpal A, Smith MH, Donlin LT. Rheumatology in the era of precision medicine: synovial tissue molecular patterns and treatment response in rheumatoid arthritis. Curr Opin Rheumatol. 2021 Jan;33(1):58–63.

434. Kane D, Jensen LE, Grehan S, Whitehead AS, Bresnihan B, Fitzgerald O. Quantitation of metalloproteinase gene expression in rheumatoid and psoriatic arthritis synovial tissue distal and proximal to the cartilage-pannus junction. J Rheumatol. J Rheumatol; 2004 Jul;31(7):1274–80.

435. de Hair MJH, Harty LC, Gerlag DM, Pitzalis C, Veale DJ, Tak PP. Synovial tissue analysis for the discovery of diagnostic and prognostic biomarkers in patients with early arthritis. J Rheumatol. The Journal of Rheumatology; 2011 Sep;38(9):2068–72.

436. Nehar-Belaid D, Hong S, Marches R, Chen G, Bolisetty M, Baisch J, et al. Mapping systemic lupus erythematosus heterogeneity at the single-cell level. Nat Immunol. Nature Publishing Group; 2020 Sep;21(9):1094–106.

437. Yu W, Uzun Y, Zhu Q, Chen C, Tan K. scATAC-pro: a comprehensive workbench for single-cell chromatin accessibility sequencing data. Genome Biol. BioMed Central; 2020 Apr 20;21(1):94–17.

438. Glossop JR, Emes RD, Nixon NB, Packham JC, Fryer AA, Mattey DL, et al. Genome-wide profiling in treatment-naive early rheumatoid arthritis reveals DNA methylome changes in T and B lymphocytes. Epigenomics. Future Medicine Ltd London, UK; 2016 Feb;8(2):209–24.

439. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol. Nature Publishing Group; 2013 Feb;31(2):142–7.

440. Hirata S, Dirven L, Shen Y, Centola M, Cavet G, Lems WF, et al. A multi-biomarker score measures rheumatoid arthritis disease activity in the BeSt study. Rheumatology (Oxford). 2013 Jul;52(7):1202–7.

441. Milman N, Karsh J, Booth RA. Correlation of a multi-cytokine panel with clinical disease activity in patients with rheumatoid arthritis. Clin Biochem. 2010 Nov;43(16-17):1309–14.

442. Tomita M. Whole-cell simulation: a grand challenge of the 21st century. Trends Biotechnol. 2001 Jun;19(6):205–10.

443. Damerau A, Gaber T. Modeling Rheumatoid Arthritis In Vitro: From Experimental Feasibility to Physiological Proximity. Int J Mol Sci. Multidisciplinary Digital Publishing Institute; 2020 Oct 25;21(21):7916.

444. Chen H-J, Li Yim AYF, Griffith GR, de Jonge WJ, Mannens MMAM, Ferrero E, et al. Meta-Analysis of in vitro-Differentiated Macrophages Identifies Transcriptomic Signatures That Classify Disease Macrophages in vivo. Frontiers in Immunology. Frontiers; 2019;10:2887.

445. Pagani S, Torricelli P, Veronesi F, Salamanna F, Cepollaro S, Fini M. An advanced tri-culture model to evaluate the dynamic interplay among osteoblasts, osteoclasts, and endothelial cells. Journal of Cellular Physiology. John Wiley & Sons, Ltd; 2018 Jan;233(1):291–301.

446. Asadipour M, Hassan-Zadeh V, Aryaeian N, Shahram F, Mahmoudi M. Histone variants expression in peripheral blood mononuclear cells of patients with rheumatoid arthritis. Int J Rheum Dis. John Wiley & Sons, Ltd; 2018 Oct;21(10):1831–7.

447. Yan F, Powell DR, Curtis DJ, Wong NC. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. Genome Biol. BioMed Central; 2020 Feb 3;21(1):22–16.

448. Gate RE, Cheng CS, Aiden AP, Siba A, Tabaka M, Lituiev D, et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. Nat Genet. Nature Publishing Group; 2018 Aug;50(8):1140–50.

449. Hall ECR, Murgatroyd C, Stebbings GK, Cunniffe B, Harle L, Salter M, et al. The Prospective Study of Epigenetic Regulatory Profiles in Sport and Exercise Monitored Through Chromosome Conformation Signatures. Genes (Basel). Multidisciplinary Digital Publishing Institute; 2020 Aug 7;11(8):905.