Predicting Response to Disease Modifying Treatment in Multiple Sclerosis

PhD Thesis

Arie R Gafson

Imperial College London

Division of Brain Sciences, Department of Medicine

2017

Supervisors:

Professor Paul Matthews
Division of Brain Sciences, Department of Medicine, Centre for Neurotechnology and UK Dementia Research Institute, Imperial College London

Professor Gavin Giovannoni
Blizard Institute, Barts and The London School of Medicine and Dentistry Queen Mary University London

Abstract

Multiple sclerosis (MS) is an autoimmune disease affecting the central nervous system (CNS) that most commonly begins with a relapsing-remitting course (RRMS). Many disease modifying treatments now are available, but none have efficacy in all patients, all are expensive and all are associated with possible adverse events. Stratifying patients to the best tolerated and most efficacious treatment either prior to or soon after commencing treatment would enhance relative benefits and reduce harm. Effective stratification depends on an understanding of relevant aspects of a drug's mechanism of action, characterisation of key pharmacodynamic effects and being able to monitor disease activity over time. In this study, I set out to determine whether multi-omics profiling (transcriptome, cytokines, lipoproteins and metabolome) can fulfil these three requirements for one of the newer, oral treatments for RRMS, dimethyl fumarate (DMF).

Chapter 1 provides an introduction to MS and explores the need for a stratified approach to treatment. Chapter 2 outlines the materials and methods used in this study including a discussion of modelling approaches that are used for data reduction.

In Chapter 3, I aimed to discriminate MS patients from healthy controls using multi-omics profiling. The RRMS patients showed greater expression of immune pathway genes, as well as raised concentrations of lipids within lipoprotein sub-fractions, relative to healthy controls. The lipid measures were predictive of disability as measured using the Expanded Disability Status Scale (EDSS) when combined in a multivariate regression model.

In Chapter 4, I tested whether multi-omics profiling could further elucidate the pharmacodynamic actions of dimethyl fumarate (DMF), a disease modifying treatment for RRMS. Comparisons of patient samples pre- and 6 weeks post-initiation of DMF revealed transcriptome changes enriched for activation of nuclear factor (erythroid-derived 2)-like 2 (*Nrf2*) and inhibition of nuclear factor κB (NFκB). Metabolomics profiling defined elevated levels of tricarboxylic acid metabolites, fumarate, succinate, succinyl-carnitine and methyl-succinylcarnitine.

In Chapter 5, I used my prospective longitudinal data to test whether gene expression and metabolite changes associated with drug action in the blood mononuclear cell fraction at 6 weeks are associated with clinical and radiological responses at 15 months. Patients responding to treatment (measured using the composite outcome measure 'no evidence of disease activity') showed robust transcriptome changes between baseline and 6-weeks that were not present in non-responders. They also showed a relative stabilisation of gene expression over the remaining study period.

My study thus provides evidence that multi-omics profiling could be a useful tool for stratified medicine in MS. It promises to elucidate differences that exist between disease and healthy states, further understanding of the pharmacodynamics of treatments and can provide longitudinal measures of response for monitoring the impact of a medicine. The latter could be used to optimise treatment choice for individual patients. If these methods were reduced to practice they could increase the chances of better clinical outcomes whilst avoiding otherwise unnecessary adverse events.

Contents

Declaration of Originality

The data presented in this thesis is my own original work. I would like to acknowledge the following for their contribution.

The Clinical Phenome Centre, Imperial College London for performing the NMR and Mass-Spectrometry based metabolomic profiling including MMF quantification assay. For providing Methodology for these assays.

Metabolon for performing Mass-Spectrometry based metabolomics profiling. For providing Methodology for these assays.

Genewiz for performing the library preparation and sequencing for the next-generation sequencing data.

Imperial College BRC Genomics Facility for performing the library preparation and sequencing for the pilot next-generation sequencing data.

Copyright Declaration

Publications arising from this work

Raffel J, Gafson AR, Malik O, Nicholas R. Anti-JC virus antibody titres increase over time with natalizumab treatment**. *Mult Scler*. 2015 Dec;21(14):1833-8. PMID: 26449743

Gafson A, Craner MJ, Matthews PM. Personalised Medicine for Multiple Sclerosis Care. *Mult Scler*. 2016; Sep 26. PMID: 27672137

Raffel J, Gafson AR, Malik O, Nicholas R. Inflammatory Activity on Natalizumab Predicts Short-term but not Long-term Disability in Multiple Sclerosis. *PLOS One,* 2017; accepted Jan 04, 2017 (In Press) PMID: 28081190.

Gafson AR, Thorne T, McKechnie CIJ, Jimenez B, Nicholas R, Matthews PM. Lipoprotein markers associated with disability from multiple sclerosis. *Mult Scler*, 2017 (*Submitted for publication*).

Gafson AR, Kim K, Cencioni MT, Nicholas R, Baranzini S, Matthews PM. Blood mononuclear cell transcriptome changes associated with dimethyl fumarate treatment of multiple sclerosis. *Neurology, 2017* (*submitted for publication*).

Abstracts

Gafson AR, Raffel J, Malik O, Nicholas R. Anti-JC virus antibody titres increase over time with natalizumab treatment**. *Association of British Neurologists Annual Meeting*; 19th – 22nd May 2015; *J Neurol Neurosurg Psychiatry* 2015;86:e4. *31st Congress of the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS)*. 7 – 10 October 2015, Barcelona, Spain.

Gafson AR, Raffel J, Malik O, Nicholas R. Early prediction of long-term response to natalizumab in multiple sclerosis. *Association of British Neurologists Annual Meeting*; 19th – 22nd May 2015; *J Neurol Neurosurg Psychiatry* 2015;86:e4. *31st Congress of the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS)*. 7 – 10 October 2015, Barcelona, Spain.

Gafson AR, Porter J, Wang Yi-Fang, Khadayate S, Nicholas R, Matthews PM. Tecfidera modulates transcriptional expression of genes involved in regulation of the immune response but not anti-oxidant pathway. *The 68th Annual AAN Annual Meeting*; April 15-21 2016, Vancouver, Canada.

Porter J, Gafson AR, Nicholas R, Matthews PM. Transcriptome profiles from relapsing-remitting multiple sclerosis patients show aberrant immunological phenotypes compared with healthy controls. *The 68th Annual AAN Annual Meeting*; April 15-21 2016, Vancouver, Canada.

Gafson AR, Nicholas R, Giovannoni G, Matthews PM. Plasma cytokine concentration changes in multiple sclerosis patients after treatment with dimethyl

fumarate. *32nd Congress of the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS)*. 14 – 17 September 2016, London, UK.

Savva C, Gafson AR, McKecnnie CIJ, Matthews PM. Tecfidera Modulates short term metabolomics signatures in multiple sclerosis patients. *The 69th Annual AAN Annual Meeting*; April 2017, Boston, USA.

McKechnie CIJ, Gafson AR, Savva C, Matthews PM. Lipoprotein markers of disability for patients with multiple sclerosis. *The 69th Annual AAN Annual Meeting*; April 2017, Boston, USA.

List of Figures

Fig. 17. Comparative correlation heatmap demonstrating differentially coexpressed modules pre- and post- treatment with DMF

Fig 18: Box plots pre (1) and post (2) treatment for cytokines IL-4 and IL-13.

Fig. 19. Sample representation using the first 2 latent variables from sPLS-DA (50 metabolites selected) for RRMS patients samples pre- and post- treatment.

Fig. 20. Comparative correlation heatmap demonstrating differentially coexpressed modules pre- and post- treatment with DMF

Figure 21(a-d). Boxplots of TCA metabolites (a: succinyl-carnitine, b: methyl succinyl-carnitine, c: fumarate, d: succinate) at baseline and 6 weeks in controls (untreated) and patients (treated).

Figure 22: Random Forest plot displaying 30 most discriminatory variables distinguishing pre- and post-treatment samples and corresponding mean decrease accuracy values.

Fig 23. Correlation plot of MMF concentration (ng/ml) and raw counts of methyl succinyl-carnitine.

Fig 24a. Correlation plot of MMF concentrations (ng/mL) and raw counts of 1-methylnicotinamide.

Fig 24b. Correlation plot of MMF concentrations (ng/mL) and raw counts of 1-methylnicotinamide (technical replicate).

Fig 25. Integrated metabolomics and transcriptomics analysis of RRMS patients samples pre- and post- treatment with DMF.

Fig 26 (a-d). Boxplots of Nrf2 related genes at baseline and 6 weeks in responders, non-responders and healthy controls

Fig 27 (a-d). Boxplots of NFkB related genes at baseline and 6 weeks in responders, non-responders and healthy controls

Fig 28a. Principal component analysis of non-responders at 15 months
Fig 28b. Principal component analysis on non-responder group A (green) and group B (red) at 15 months post-treatment.

Fig 29. Venn Diagrams of differentially expressed genes in responder (a) and non-responder (b) groups compared with controls at baseline, 6 weeks and 15 months.

List of tables

Table 17a. 10 most significantly altered metabolites in the flusher group.

Table 17b. 10 most significantly altered metabolites in the non-flusher group.

Table 18. Metabolites significantly increased 6-weeks post- treatment with DMF and also associated with DMF associated gastrointestinal adverse events

Acknowledgements

I would like to first and foremost thank Paul for everything he has done for me during the last three years. I first met Paul before starting my Academic Foundation Programme to discuss potential projects that I could undertake during my four-month placement. A week later, I met him again with the other academic trainees and he didn't remember me… This was disappointing as I thought I had left a lasting impression.

After a successful four-month placement and as I approached the end of my FY2 year as a junior doctor, I met with Paul, who remembered me this time, to discuss future career plans. We discussed my reservations about immediately commencing Core Medical Training and Paul mentioned that there might be a way of funding a clinical research fellowship for a project in multiple sclerosis. This was the start of my PhD journey…

Over the last three years, I have learnt a lot from and about Paul. His attention to detail, his vast scientific knowledge and his willingness to share and collaborate are skills I hope to emulate in my own future career. Paul has been an incredibly kind, supportive and patient supervisor. He has provided me with some brilliant opportunities, not least my visit to UCSF earlier this year to work with the Baranzini lab in UCSF. Paul will always encourage enthusiasm and fosters an environment for open enquiry with scientific rigor. He has taught me when to use a semi-colon; when not to use a hyphen - and also never to use the word 'interestingly'. Perhaps most fundamentally, Paul has demonstrated to me what one can achieve with hard work. I

am incredibly grateful for the opportunity he has given me and I thank him sincerely for this.

I would also like to thank Gavin, the only MS consultant in London who responded to my emails in 2009 asking to undertake a special study module in their department! Gavin was the first to foster my interest in MS and encouraged me to pursue this. Following on from two very successful summer placements at the Royal London, I was more enthusiastic than ever and I am very grateful to have him as a mentor and role-model.

There are so many people at Imperial to whom I am grateful for helping me over the last three years. Rehiana and MT were able to show me the ropes in the lab, David helped me with my protocol, and Keerthi at the BRC was always patient when I came to ask for help with bioinformatics. Tom and Chris have also been incredibly useful resources as I struggled getting to grips with 'R'. And how could I possibly forget Tabitha, who has provided excellent company during numerous coffee and lunch breaks!

Dedication

To Galia, Anthony, Irene, Leonie, Jeremy and Rosie

Abbreviations

AE	Adverse event
AG	Arie Gafson
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
AQP4	Aquaporin 4
AR-BVL	Annualised rate of brain volume loss
BBB	Blood-Brain-Barrier
BMI	Body Mass Index
BRC	Imperial College Biomedical Research Centre
CSF	Cerebrospinal fluid
DMARDS	Disease-modifying anti-rheumatic drugs
DMF	Dimethyl Fumarate
DMT	Disease Modifying Treatment
DNA	Deoxyribonucleic acid
DRAGEN	Dynamic Read Analysis for Genomics
EAE	Experimental Autoimmune Encephalomyelitis
EBV	Epstein-Barr virus
EDSS	Expanded Disability Status Score
EDTA	Ethylenediaminetetraacetic acid
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GA	Glatiramer Acetate
Gd+	Gadolinium-enhancing
HDL	High Density Lipoprotein
HILIC	Hydrophilic interaction chromatography
HLA	Human leucocyte antigen
IDL	Intermediate Density Lipoprotein
IFN- β	Interferon Beta
IL	Interleukin
IPA	Ingenuity Pathway Analysis
JCV	Latent polyomavirus JC
KA	Kynurenic Acid
LASSO	Least absolute shrinkage and selector operator
LDL	Low Density Lipoprotein
LFC	Log Fold Change
LOD	Limit of Detection
LOOCV	Leave-one-out cross validation
LOQ	Limit of quantification
MCS	Mental Component Score
MDA	Mean decrease accuracy
MHC	Major Histocompatibility Complex
miRBAse	Micro RNA database
MiRNA	Micro-RNA
MMF	Monomethyl fumarate
MMP	Matrix metalloproteinases
MRI	Magnetic resonance imaging
mRNA	messenger ribonucleic acid
MS	Multiple Sclerosis

MSFC	Multiple Sclerosis Functional Composite
NEDA	No evidence of disease activity
Nf-L	Neurofilament light chain
NFκB	Nuclear factor κB
NGS	Next Generation Sequencing
NMOSD	Neuromyelitis optica spectrum disorders
NMR	Nuclear magnetic resonance
NPC	National Phenome Centre
NRF2	Nuclear factor (erythroid-derived 2)-like 2
OOB	Out-of-bag
Padj	Adjusted p-value
PBMC	Peripheral blood mononuclear cell
PCA	Principal Component Analysis
PCS	Physical Component Score
PLS-DA	Partial least squares discriminant analysis
PML	Progressive Multifocal Leukoencephalopathy
QA	Quinolinic Acid
QOL	Quality of Life
RA	Rheumatoid Arthritis
RNA	Ribonucleic acid
RRMS	Relapsing-remitting multiple sclerosis
rRNA	ribosomal ribonucleic acid
S1P	Sphingosine-1-phosphate
SNP	Single nucleotide polymorphism
sPLS	Sparse partial least squares
SPMS	Secondary progressive multiple sclerosis
SSE	Sum of Squared errors
STAR	Spliced Transcripts Alignment to a Reference
STK11	Serine/threonine kinase 11
SVD	Singular value decomposition
TCA	Tricarboxylic acid cycle
Th	T-helper cell
TNF	Tumor Necrosis Factor
UPLC	Ultra-performance liquid chromatography
VLDL	Very Low Density Lipoprotein
VMA	Vanillylmandelic acid
WGCNA	Weighted gene correlation network analysis

Genes

ASPH	Aspartate beta-hydroxylase
ATF4	Activating transcription factor 4
BACH1	BTB domain and CNC homolog 1
C21orf33	Chromosome 21 Open Reading Frame 33
CD36	CD36 molecule
CD83	CD83 molecule

CEBPB	CCAAT/enhancer binding protein beta
DCHS1	Protein dachsous homolog
FOSB	FBJ murine osteosarcoma viral oncogene homolog B
FOSL1	FOS like 1, AP-1 transcription factor subunit
FTH1	Ferritin heavy chain 1
FTL	Ferritin light chain
GCLC	Glutamate-cysteine ligase catalytic subunit
GPX1	Glutathione peroxidase 1
GRASP	General receptor for phosphoinositides 1 associated scaffold protein
GSR	Glutathione-disulfide reductase
GSR	Glutathione-disulfide reductase
GSTP1	Glutathione S-transferase pi 1
ICAM1	Intercellular adhesion molecule 1
ID1	DNA-binding protein inhibitor ID-1
JUN	Proto-Oncogene C-Jun
KIR3DL2	Killer cell immunoglobulin-like receptor 3DL2
MAFG	MAF bZIP transcription factor G
MAFG	MAF bZIP transcription factor G
MGST1	Microsomal glutathione S-transferase 1
MITF	Melanogenesis associated transcription factor
MRPL55	39S ribosomal protein L55, mitochondrial
N4BP3	NEDD4 binding protein 3
NCSTN	Nicastrin
NFKBIA	NFKB inhibitor alpha
NFKBIE	NFKB inhibitor epsilon
NQO1	NAD(P)H quinone dehydrogenase 1
NR4A2	Nuclear receptor related 1 protein
PDGFC	Platelet derived growth factor C
PNISR	PNN interacting serine and arginine rich protein
PRDX1	Peroxiredoxin 1
PTGS1	Prostaglandin-endoperoxide synthase 1
RNF19B	Ring finger protein 19B
RPL10P9	Ribosomal Protein L10 Pseudogene 9
RUNX3	Runt related transcription factor 3
SBDSP1	SBDS, Ribosome Maturation Factor Pseudogene 1
SIGLEC6	Sialic acid binding Ig like lectin 6
SSPN	Sarcospan
TLR9	Toll like receptor 9
TOB1	Transducer Of ERBB2
TUBB2A	Tubulin Beta 2A Class IIa
ZNF547	Zinc finger protein 547
ZNF727	Zinc finger protein 727

Chapter 1

Introduction

Multiple Sclerosis (MS) is a 'syndrome' with great variation in disease course, presentation and response to treatments(1). It is now crucial to unpick this heterogeneity given the increasing number of available treatments and a more informed patient population. Furthermore, there is significant heterogeneity of treatment response which is currently difficult to predict. Many of these medicines are expensive and therefore better ways of demonstrating efficacy are required (Fig. 1). This is important as treatments have different mechanisms of action, efficacy and relative risk of adverse events. In conjunction with their neurologist, a newly diagnosed patient must now make complex decisions in deciding to initiate treatment, treatment choice and treatment escalation.

Multiple Sclerosis Subtypes

MS is a neuroinflammatory disorder characterised by demyelination, gliosis and axonal injury. Patients commonly present in young adulthood (between the ages of 20-40), there is a greater preponderance in females (typically a 2:1 ratio) and there is often significant heterogeneity in clinical symptomatology. Most patients will initially follow a relapsing-remitting course (RRMS) characterised by symptomatic episodes from which there is either partial or complete recovery. The frequency of such relapses tends to be highly variable, although on average will occur once every one to two years(2). This clinical course is typically superseded by a progressive stage of disease where relapses become less clearly defined and are replaced by a steady worsening of symptoms (referred to as secondary-progressive MS (SPMS). Whilst

this may be the most common trajectory, patients have highly variable disease courses with respect to frequency of relapses and time to disease progression. Despite a substantial literature on the natural history of MS disease course, the future severity of the disease in an individual patient cannot be predicted well. A minority of patients (<10%) will follow a progressive course from the outset (termed primary progressive MS - PPMS) and a similar proportion will have mild relapses with very limited accrual of disability over time (controversially termed benign MS)(3). In reality, the majority of this group are not 'spared' from motor or cognitive impairment; it just appears to be somewhat delayed(4-6).

Pathology of MS

Neuropathological studies support the hypothesis that MS is both an inflammatory and neurodegenerative disorder (7). Associated neuropathological features seem inextricably linked. For example, post-mortem studies reveal that dense meningeal infiltrates are associated with demyelination, neuronal damage and cortical atrophy(8, 9). Furthermore, brain biopsies from patients early in disease course have demonstrated that inflammatory cortical demyelination often precedes the appearance of white matter plaques, the latter being associated with neurodegenerative changes such as reactive astrocytosis and neuroaxonal injury(10, 11).

Diagnosis of MS

Patients are diagnosed with MS based on clinical and radiological findings according to the McDonald Criteria, for which there must be evidence of dissemination in time and space(12). Diagnosis couples clinical histories with conventional laboratory tests and imaging (and, in some cases, cerebrospinal fluid (CSF) examinations)(12). This is

necessary in order to ensure a more accurate diagnosis and to potentially identify

subtypes of MS such as progressive forms of the disease. Auto-antibody testing can

improve diagnostic accuracy. For example, neuromyelitis optica spectrum disorders

(NMOSD), identified by antibodies to aquaporin4 (AQP4-IgG), have a distinct

clinical course and may worsen with IFN treatment(13). Specificity of AQP4-IgG is

high(14, 15). However, cautions arise from recognition of the potential insensitivity

of AQP-4 assays for NMOSD(16). Patients with MOG antibody-associated

demyelination also have a unique course characterized by a unique clinical,

radiological, and therapeutic profile(17).

Treatments for MS

The last twenty years have seen remarkable developments in the therapeutic

management of MS. This has almost exclusively benefited patients with RRMS; a

large unmet clinical need still exists for those in the progressive phase of the disease

for which there are no currently licensed treatments. The recent results of the

ocrelizumab trial in progressive MS have provided some hope and this currently

awaits licensing in the UK. The first available treatments for RRMS were interferon-β

(IFN- β) and glatiramer acetate (GA), both of which were shown in clinical trials to

reduce annual relapses rates by approximately 30% and to have an impact on

radiological measures(18). Long-term follow-up of patients from these early trials has

demonstrated that a subset experienced reductions in long-term disease severity,

progression(19) and risk of mortality(20). These latter findings of heterogeneity in

long term clinical outcomes accurately represent the unmet need in MS management;

the inability to accurately predict who will respond (or not respond) to a specific

treatment.

A second generation of MS treatments became available at the turn of the century: natalizumab in 2003 and fingolimod in 2009(21-23). Natalizumab is a monoclonal antibody targeting integrin-α and preventing lymphocyte egress across the blood brain barrier (BBB). Initially with proof of principle in the rodent model of MS(24), experimental autoimmune encephalomyelitis (EAE), it progressed through pivotal clinical trials to boast a greater than 65% reduction in relapse rate and >90% attenuation of new MRI lesions(25, 26). The widespread excitement associated with this highly effective disease modifying treatment (DMT) was tempered somewhat by the diagnosis of progressive multifocal leukoencephalopathy (PML) in two patients from the trial(27). PML is a serious adverse event associated with natalizumab that affects 1 in 1000 patients taking this DMT. Thought to be caused by reactivation of latent polyomavirus JC (JCV); the mechanism by which natalizumab causes PML remains unknown(28). Risk factors for PML include prior use of other DMT, duration of time on the medication and JCV serostatus(29).

Fingolimod, a sphingosine-1-phosphate (S1P) analogue exerts its action by downregulating the S1P receptor 1 on leukocytes and the endothelium thus preventing naïve and central memory T lymphocytes from migrating from lymph nodes to the peripheral vasculature. Reductions in relapse rates of over 50% and reductions in new MRI lesions were observed in clinical trials of this medication(21, 23); however the drug is associated with symptomatic bradycardia, lymphopenia, retinal oedema and rarely fulminant hepatic failure(30).

Most recently, a number of further DMTs have become available for the management of RRMS. These include two oral medications, teriflunomide and dimethyl fumarate

(DMF) (31-33) and alemtuzumab, a monoclonal antibody to CD52 which is reserved for more aggressive forms of the disease but is associated with secondary autoimmune disorders such as thyroid disease and immune thrombocytopenic purpura(34-36). The emergence of a number of new medications for MS, with similar efficacy but a diversity of adverse event types, severity and frequency has driven the need for a personalised approach to therapy, to ensure a patient is on the most effective medication for them in order to positively impact on disease course and to avoid unwanted side effects and adverse events.

Dimethyl Fumarate

Dimethyl fumarate (BG-12; Tecfidera) (DMF) is a fumaric acid ester licensed as a DMT for RRMS. The primary therapeutic mechanism of action still is debated, but has been proposed to involve activation of the transcription factor nuclear factor (erythroid-derived 2)-like 2 (Nrf2)(37, 38), inhibition of nuclear factor κB (NFκB)(39) or agonism of the hydroxylic acid receptor 2(40). Because DMF is believed to act on circulating, as well as tissue-resident immune cells, may trigger anti-oxidant pathways through modulation of transcription factors, and is metabolized to the intermediate metabolite fumaric acid, its mechanism of action lends itself well to characterisation using gene expression studies and metabolomic phenotyping(41).

The pharmacokinetic properties of DMF are well characterized. After oral administration, DMF is rapidly hydrolysed by esterases to its principle metabolite monomethyl fumarate (MMF)(42). MMF is highly bioavailable, has a half-life of 12 hours and reaches serum peak concentrations of approximately 20 μM. MMF is

ultimately hydrolysed inside cells to fumaric acid entering the tricarboxylic acid cycle (TCA)(43, 44).

Initial studies in murine models of MS showed beneficial effects of this treatment on clinical parameters and this was attributed to its anti-inflammatory properties including induction of IL-10 and reduced macrophage inflammation(45). DMF was subsequently tested for efficacy in a phase I clinical trial, which demonstrated that the treatment was generally well tolerated and adverse events (AE) (which included flushing and gastrointestinal symptoms) were mild and reversible. The trial was also able to show a significant reduction in Gd+ enhancing lesions(46). The phase II follow up study conducted by Kappos and colleagues reiterated initial findings, reporting a 60% reduction in new Gd+ enhancing lesions, as well as a reduction in annualised relapse rates by 32%(47). The phase III DEFINE study demonstrated a reduction in annualised relapse rates by 0.19 with a relative risk reduction in relapses of 53%(31). Further AEs reported included reduction in lymphocyte count and elevated liver aminotransferase liver levels.

Anti-Oxidant Properties

DMF has been shown to activate the Nrf2 transcriptional pathway which reduces oxidative stress(37, 38). In support of this postulated mechanism of action, other groups have shown that pre-incubation of astrocytes with DMF can prevent dopamine mediated neurotoxicity by increasing the activities of intracellular antioxidant enzymes(48). Others have argued these antioxidant properties may be exerted by astrocytes rather than neurons(49). The anti-oxidant properties of DMF have also been shown to be important for T cell activation and differentiation and the 'healthy'

functioning of the peripheral immune system preventing release of pro-inflammatory cytokines and ensuring effective intracellular signalling networks(50).

Anti-Inflammatory Properties

It has been well described that DMF suppresses nuclear factor κB-dependent transcription with consequent preferential expression of anti-inflammatory chemokines and cytokines(51, 52). Ockenfels and others have shown that DMF can diminish IL-6 and TGF-alpha secretion shifting the immune response from a Th1 to a Th2 profile(53, 54). These anti-inflammatory properties of DMF have also been well characterised in peripheral blood mononuclear cells (PBMCs)(55).

Personalising treatment in MS

Stratification of baseline risk

In order to optimise the management of MS, it is important to be able to accurately prognosticate for individual patients. Because all treatments carry costs – both financially and in risks to future health- defining the *net* benefit of treatment involves balancing the expected natural history of an individual's disease against the risks of treatment(56) (Fig. 1a and 1b). Patients differ widely in their baseline risk of untreated disease progression, however we are not good at predicting this.
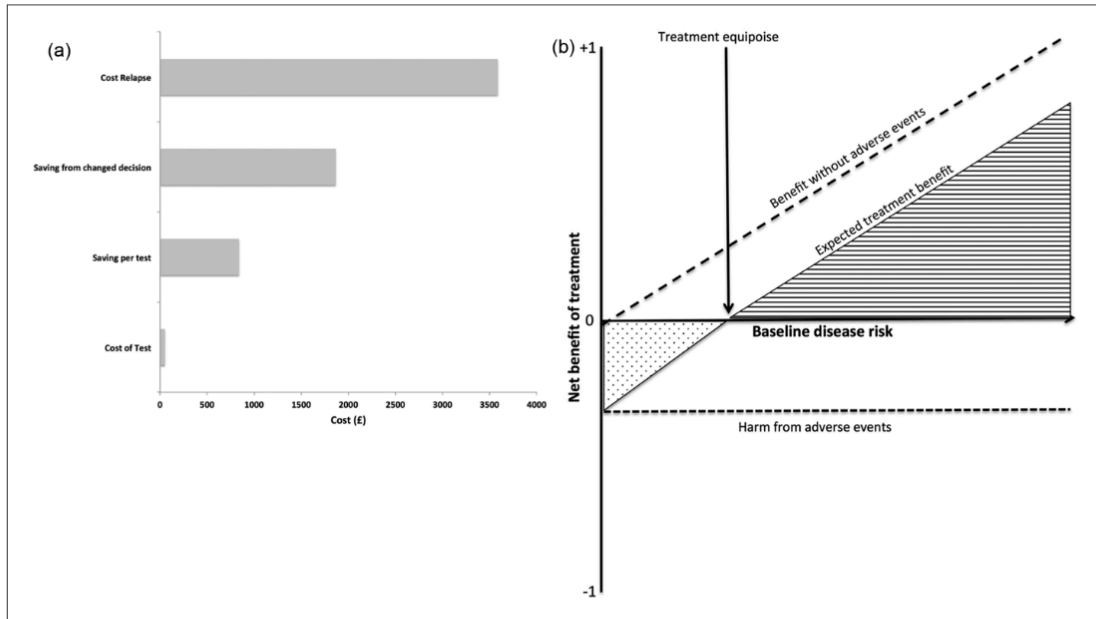
Figure 1. (a) A simplified illustration of the potential cost-effectiveness of treatment monitoring for routine interferon neutralising antibody (NAb) testing of MS patients on IFN treatment. The mean cost of relapse treated in the UK National Health System is £3586.[4] The cost of NAb testing is about £50. It was assumed that the NAb test provides an actionable index discriminating potential efficacy of the IFN treatment from lack of any impact and that patients recognised to have NAb would be switched to another, equally effective disease modifying treatment of similar cost. Recognition of the NAb therefore leads to savings accruing from avoidance of potential relapse costs associated with loss of efficacy of the IFN (change in relapse rates with loss of efficacy × cost of relapse = 0.52 × 3586 = £1864).[5] The potential savings per test is the product of savings from avoidance of relapse and probability of a positive test (1864 × 0.45 = £840). (b) A graphical illustration of the dependence of net benefit of any treatment on baseline risk for a patient. The benefit of treatment was assumed to be proportional to the risk of disease progression (relapse). An 'ideal treatment' – one in which there is no cost or risk of adverse events – should benefit all patients, with an absolute benefit (measured, for example, in adjusted quality of life years) that increases with disease risk. However, in practice, treatments also can have a negative impact on patients, either as cost or adverse events. Here, for simplicity, the negative impact is assumed to arise only from adverse effects of treatment and the risk and impact are assumed to be the same for all patients. Estimation of the *net* benefit to a patient includes consideration both of the impact of treatment to reduce disease activity and the impact of adverse events associated with medicine use. While patients with higher baseline risk from the untreated disease will receive greater net benefit (area defined by grid lines to the right), those at lower risk from their disease may experience net harm (punctate area marked to the left). A goal of personalised medicine is to match the benefit/risk profile of a medicine and the baseline disease risk to optimise the net benefit for a patient.

(Figure from Gafson et al, 2017, MS Journal(57))

Considerable community effort has been addressed to the identification of prognostic factors for estimation of the baseline risk, but the precision with which this can be done still is limited. For example, MRI measures of disease activity (by gadolinium contrast enhancement) or T2-hyperintense lesion load are important prognostic measures for prediction of risk of clinically definite disease after first symptoms(58). Large, single clinical centre-based studies additionally have highlighted interactions of MRI measures with age and sex in determining risk of progression(59). Lesion distribution or clinical presentation appear to be independent predictors of medium term prognosis(60). Epidemiological studies suggest that other phenotypic (e.g.,

24

obesity, serum vitamin D), exposure (e.g., sunlight) and lifestyle factors (e.g., smoking) impact on prognosis(61). However, I did not find models that define their quantitative interactions with individual susceptibilities. For example, is the impact of smoking, low vitamin D and obesity meaningfully higher in people carrying the DRB1501 allele, or with early presentations of disease? Genotype alone has not yet been shown to contribute significantly to disease severity risk(62). In the absence of single, highly predictive markers, future personalisation will depend on clusters of markers in multivariate models.

Stratification of treatment

There still are few specific indices to guide the timing of treatment beyond evidence from trials that early treatment delays short-term clinical progression(63). The choice of initial treatment also does not have a good evidence base, other than the lack of benefit (or worsening) that has been found with IFN and other conventional DMTs(64) in patients with progressive onset disease or NMOSD. While not fully evidence-based, patients with a higher baseline risk likely will probably receive greater *net* benefit from any treatment. Information concerning the relative efficacy of medicines is limited due to the small number of head-to-head clinical trials and the limitations of inference even when comparisons of the pivotal trials of individual agent efficacies are made using formalized meta-analytic structures(65). One of the most promising approaches to gathering evidence concerning relative clinical effectiveness is through real-life data aggregation in multi-centre consortia, such as MSBase (66). Generally, choices regarding medicine use in clinical practice are framed in terms of a hierarchy of efficacy and risk for treatments based on data from their pivotal clinical trials (which were intended to demonstrate efficacy, rather than

comparative effectiveness). Decision-making then represents the balancing of these data against estimates of relative disease severity for any given patient. Patient and neurologist-specific factors of preference and access also play a role. However, while there may be general guidelines, there is not a general consensus regarding the criteria and methods for arriving at the balance of evidence for an individual patient.

Currently, in the absence of strongly predictive prospective markers, treatment *monitoring* plays a major role. In fact, unquestionably the currently best developed example of personalized medicine in MS is for safety monitoring of natalizumab treatment(29). This model combines titres of anti-JC virus antibody, treatment duration and previous history of immunosuppressive therapy in order to stratify patient risk of PML. Baseline risk assessment and monitoring with treatment rapidly became the standard of practice as the manufacturer and regulators worked together to define a way of keeping this powerful treatment available once PML was recognized as a complication. An international pharmacovigilance effort developed by the manufacturer rapidly led to validation of a clinically practical approach to personalization of risk and subsequent monitoring.

Monitoring for effectiveness is more challenging, in no small part because the target outcome (ultimately, the accrual of fixed disability) is less easily defined. Nonetheless, there are examples that are widely, if not universally, employed. Neutralising antibody levels for IFN(67) and for natalizumab(68) explain a major proportion of poorer efficacy of these medications. More generally, T2 lesion increases and brain volume reduction on treatment are predictive of longer term clinical efficacy, at least at a group level(69). The latter, combined with clinical

measures of disease activity (relapse frequency or progression of fixed disability) already are incorporated into treatment escalation decisions in many clinical centres, although specific criteria for a switch in treatment are not generally agreed. While the availability of large datasets is a major confound, this also is a consequence of lack of standardization of MRI field strength, criteria for lesion identification and software for brain volume change measures. Possibly even more sensitive markers of sub-clinical disease activity are emerging, e.g., with monitoring of CSF or serum neurofilament light chain (Nf-L) concentrations(70). In a 15-year follow up study, higher levels of CSF Nf-L at baseline were associated with greater disability progression in RRMS patients(71). Changes in concentrations while on treatment also have been linked to treatment response(72). Additional markers are being explored actively, but most of the profusion of reports based on studies with smaller populations have later failed replication; Kroksveen and colleagues recently reported that from 188 proposed CSF MS biomarkers, only 10 (5%) have been successfully validated(73).

The importance of real world data

One of the reasons that no predictive biomarkers of response to medication have emerged in MS may lie in the fact that the largest studies with the best power are undertaken in the context of clinical trials; where patients are not necessarily representative of the real world populations. For example, patients enrolled into such studies are restricted by their age and normally have high baseline clinical activity. Furthermore, because the medications are not usually tested against one another, it is difficult to ascertain response markers unique to the treatment but also associated with specific individuals(74). For these reasons, we are more likely to advance discovery

of personalised treatment strategies using real world populations in observational studies. These can proceed for a longer duration than clinical trials, patients on different treatments can be directly compared and prognostic subgroups can be characterised. Despite these advantages, a number of inherent biases must be taken into account for such studies, including selection bias and detection bias. Real world observational studies have been helpful in showing that escalation of treatment to a more aggressive treatment after failure on a DMT can have benefits with respect to relapses and the risk of disability progression(75, 76). Furthermore, natalizumab may be preferable to fingolimod in patients with more active disease(77).

Omics as a stratification tool

Deconstructing the heterogeneity of a disease like MS involves understanding the differences in populations (responders vs non responders, high disease activity vs low disease activity). Given the importance of both genetics and the environment in the susceptibility to MS, profiling patients genomes, transcriptomes, proteomes and metabolomes is a promising approach to deconstruct this heterogeneity. There are a number of ways such data can be used for this purpose. Firstly, one can take a data driven approach which is unbiased but relies on good quality data (and samples) in order to derive meaningful conclusions. This is an important approach given that there is much we still do not understand about the causes of MS or the reasons for differing responses to medication or disease course. One can also take a knowledge-driven approach but this may ignore other candidates that are not known about a priori. A good combined approach may involve an initial pilot study which identifies candidates of interest followed by a validation using a knowledge driven approach.

Once data is collected using the data-driven approach, there are two approaches to data analysis. One can either perform unsupervised classification (discussed in greater detail in the Methods section) in which sub-groups of samples are separated based on their similarity and then differences in their biological behavior are interrogated. The second approach involves supervised classification were samples are assigned to relevant groups (e.g. control and patients) and those variables that are best able to discriminate them are identified. This approach can be validated using train and test subsets to test the predictive value of such a model. In the context of using such an approach for identifying predictive markers of treatment response, a number of factors must be borne in mind. Firstly, not responding to a medication may result from a number of different factors and thus there may not be a specific signature common to all those that do not respond to the drug. Secondly, in order to reduce the possibility of spurious results, large sample sizes are required. Lastly, any such signature must be validated in a separate cohort in order to progress to something that will be clinically meaningful.

Pharmacogenomics has been championed as a necessary foundation for personalised medicine, based particularly on examples drawn from cancer and rare inherited metabolic diseases(78). Individualized, genetic based diagnoses have had considerable impact, e.g., in licensing of imatinib for Philadelphia chromosome positive chronic myeloid leukaemia and trastuzumab in HER2 positive breast cancer(79).

Part of the reason for the success of this approach in cancer is that the underlying pathology can in large part be explained by genetic dysfunction. Understanding the

mutations that can arise in specific cancers has led to the development of molecular targeted therapy(80-82). Treatments can then be individualised to those with the specific molecular defect(83). As such, the ultimate goal is that 'each patient receives the right drug at the right time at the right dose for the right disease'(84). Despite promising developments in this field, there are drawbacks, for example, on and off target side effects(85-87) and the acquisition of resistance to a so-called targeted therapy(88). The latter may arise from alterations in the drug target(89-91) or target amplification(92, 93). A further role for pharmacogenomics may be in determining drug dosage; indeed it is well documented that at a certain dose, a drug may be below a therapeutic level in some individuals, but may exert toxicity in others. The most common example of this is in the use of warfarin(94). In the context of MS, individual characteristics can determine drug metabolism (e.g., hepatic and renal function for elimination of IFN) and this may guide choice of drug in less common situations in patients with comorbidities(95). Differences in ethnicity may alter drug absorption or metabolism(96).

Extending 'omics to MS

A greater understanding of the underlying genetic contribution to MS risk in recent years has raised the prospect of applying pharmacogenomics to the personalised treatment of MS. Risk susceptibility genes are thought to contribute modestly to disease risk heritability governed by complex gene-environment interactions. The first genetic association in MS was reported in 1972 for HLA Class I antigens(97). This field has now developed such that there are now a number of MHC related alleles which confer an increased risk of susceptibility for and severity of MS. Over 100 further genes of interest became apparent following the first genome wide association

study in MS in 2007, most significantly, IL-2 and IL-7 receptor alpha genes(98). Cytotoxic T lymphocyte antigen 4 has also been implicated(99-101), however results for this have not always proved conclusive and effect sizes are often modest.

Gene expression is a quantitative phenotype that can be used as an endophenotype for characterising disease risk genes(102, 103). It is now clear that most of the genome is transcribed into RNA, however only a small proportion (1-2%) codes for protein(104). Messenger RNA (mRNA) contains the coding region but typically has untranslated sequences at both its 5' and 3' ends that may play a role in regulatory function(105). There are also many non-coding RNA species with roles in intracellular signalling and transcriptional modification(106). It appears that mRNA does not have a linear relationship with its downstream protein-coding counterpart suggesting that a significant regulation may take place before translation occurs(107).

In the context of MS, gene expression studies have been applied in an attempt to gain greater insight into the disease(108-111) and also to identify markers of disease activity or treatment response(112, 113). These studies have principally relied on microarray technology which is cheaper, easier to use and computationally less intensive to analyse. Gene expression can be studied in a network context using gene co-expression networks. This approach has been used for studying neurological disorders with considerable success(114). Here, genes that have similar expression patterns are grouped together on the assumption they may be functionally related(115). Similarity in gene expression is calculated using a number of approaches, including correlation, regression or Bayesian methods. On occasion these can be combined to create more robust results(116). Despite promising observations,

these associations have not been translated into clinically useful stratification tools or robust biomarkers.

Identifying and quantifying the full set of transcripts including as-yet unidentified genes and RNA subtypes in a biological sample has become possible using next-generation sequencing (NGS), known as RNA-Seq. This technique is a multi-stage process that begins with fragmentation of RNA species and conversion by reverse transcription to cDNA fragments containing sequence adaptors(117). Depending on protocol, these fragments will sometimes undergo PCR amplification before high throughput sequencing to generate millions of short reads from one or both ends of the fragment (single-end and paired-end sequencing respectively). These reads are then reconstructed to their original RNA assembly and are then quantified to give an estimation of their expression levels. There are many advantages to RNA-Seq over traditional sequencing platforms. These include the ability to detect as yet unidentified RNA species, SNPS and transcription boundaries(117, 118).

Proteomics

One approach to validating gene expression studies is to look for correlation with downstream proteins of interest. Proteomics has been widely utilised in MS for biomarker discovery. Candidates of interest include matrix metalloproteinases (MMPs) which are involved in regulating pro-inflammatory cytokine release, myelin breakdown and axonal damage. Levels of MMP-9 specifically have been correlated with relapses and MRI lesion activity(119-121). A further candidate is ostepontin, levels of which are raised in MS patients compared to controls(122, 123) and fall following treatment with natalizumab or GA(124).

Quantifying cytokines can also be helpful in identifying markers of disease activity. In the context of MS, IL-17, IL-12 and IL-23 have been extensively explored. IL-17 appears to have a strong association with MS, correlating well with worsening disease activity. This finding has led to a trial of an anti-IL-17A monoclonal antibody(125) in patients with MS (NCT00882999).

Metabolomics

In combination with transcriptomics and proteomics, one can derive yet more information by looking at the end products of physiological and pathological processes in the form of metabolomics. Indeed, this has already provided information into the pathogenesis of disease(126-129). In the context of a multi-omics experiment, metabolites represent the downstream output of the genome. They can now be used for patient diagnosis and disease monitoring(130, 131).

In order to conduct an effective metabolomics experiment it is essential to consider sample size and to match for co-variates that may affect metabolites of interest (e.g. age and sex). It is also important to consider whether to undertake a targeted or untargeted approach to the experiment; the latter having the disadvantage that many metabolites are unidentifiable. Metabolomics experiments should also take into account factors that may affect specific metabolites, for example timing of dose with respect to last meal, timing of medications and time of day when blood is collected. The main platforms for metabolomics are NMR spectroscopy and mass spectrometry. Details of the analysis of such data will be provided in the Methods section.

Tools used for metabolomics profiling already provide precision medicine in the context of newborn screening. Indeed, these tools are used for diagnosing and predicting disease and also for determining optimal therapy (enzyme supplementation of dietary restriction) and dosing regime(132, 133). Further examples in adult life include measurement of glucose levels to identify pre-diabetes, iron levels in haemochromatosis and vitamin levels to diagnose dietary deficiencies.

Pharmacometabolomics has also advanced as a clinically useful tool(134). For example, in kidney transplantation, MS-based monitoring of immunosuppressants such as mycophenolate and tacrolimus help to optimise patient dosing. This is important given the serious adverse advents associated with such medications including reduced white cell counts, anaemia and thrombosis(135). Similarly, in Alzheimer's disease where adherence to medication may be problematic, there is some promise(136).

A number of metabolomics studies using CSF, serum and urine samples have been undertaken in MS patients. With regards to CSF, lactate has been identified as increased in patients with active MRI lesions(137), however follow-up validation experiments have failed to corroborate findings(138). In serum samples, an NMR-based approach using partial least squares discriminant analysis (PLS-DA) was able to separate RRMS from SPMS subjects(139) with a high degree of accuracy. Other studies have also identified metabolites that can accurately predict MS patients vs controls(140) and NMO patients(141).

One of the advantages of assaying metabolomic profiles in MS patients is the dynamic interplay between environmental exposures and their effect on the metabolome. Equally, gene-environment interactions can also be studied to gain further insight into the disease(142). Environmental exposures seem of particular importance in the pathogenesis of MS; the most investigated being Epstein-Barr virus infection (EBV), levels of vitamin D and smoking(143, 144). Others include adolescent obesity and nocturnal shift work. Despite strong evidence supporting these as causative factors, there are a number of biases to be borne in mind. These include residual confounders, reverse causation and selection bias. Heritability studies reveal that a sibling of a patient with MS has a 7-fold increased risk of the condition. Whilst this risk provides some evidence for genetic predisposition, it highlights the importance of environmental factors(145, 146). The steady rise in incidence of MS amongst women in the last decade is suggestive of environmental triggers. These include changes in lifestyle, for example, increases in smoking, obesity and changes in reproductive behaviour(147-150).

Different 'omics can be combined using correlation-based or pathways analysis approaches. In a complex disease such as MS, there is continual interplay between genes, proteins, metabolites and the environment(151). Integrating these at an individual level will enable personalised management. The current approach to integrating such data sets is by mathematical modelling(152). The first step involves compiling the current knowledge through a literature search in order to establish pathways that are affected by the disease. These are then combined to form a network which can be evaluated using experimental data. Approaches that provide large amounts of information (i.e. omics) are generally more reliable for testing such

networks(153, 154). If the network is reliable, a mechanistic model can be created which may predict an individuals response to a treatment.

How might personalised medicine work in MS?

As outlined above, MS requires a personalised approach to its diagnosis, prognostication and therapeutic strategy. This will require greater understanding of the disease itself and the subtypes within it. Whilst all of the current DMTs that are available target the immune system, they target different cell types, their movement and their ability to function. I envisage an approach to personalised medicine that incorporates a more accurate endophenotype of patients based on genetic and environmental information that can be objectively measured using combinations of genetics, transcriptomics, metabolomics and proteomics. Furthermore, patients will be stratified to an appropriate medication based on short-term pharmacodynamic responses measured using the same omic tools. Such tools may determine response to treatment and risk of adverse events. Most optimistically, detailed profiling may be able to prognosticate patients and predict their future disease course which has implications for social, medical and psychological impacts related to the disease itself.

Summary and Aims of My Thesis Work

MS is a clinically heterogeneous condition. It is sometimes difficult to diagnose and even harder to predict prognosis, disease course or response to a treatment. This is important as the medications are expensive and have hazardous risk profiles.

Once a patient is accurately diagnosed, it is important to prioritise initiation of treatment as ongoing inflammatory activity leads to accumulation of disability. There is currently no algorithm in place to match patients to an appropriate treatment. This relies on a 'trial-and-error' approach, usually combining a clinicians' expertise with patient values. Escalation of treatment usually occurs when a patient fails to respond to their current medication; however classifying 'response' to a medication is also poorly defined.

Given that a number of treatments are now available for RRMS, it is essential to be able to stratify patients to a treatment that is most efficacious. One approach to achieving this is through pharmacogenomics. Pharmacogenomics has emerged as a key player in personalisation of treatment for cancer. Harnessing this approach for MS is an attractive option as there is a genetic predisposition to the disease.

Despite there being some genetic component to MS there are widely recognised environmental factors that seem to be associated with MS. Their effects can be captured by looking at gene-environment interactions. Additionally, one can derive information about environmental risk factors through exploring the metabolome. Ultimately, a hierarchical pathway approach combining genetic, transcriptomic,

proteomic and metabolomics will yield the 'bigger picture'. Currently, tools to achieve this are very much in their infancy.

A commonly prescribed first line treatment for RRMS is DMF. DMF has been shown to be effective at both reducing relapses and MRI lesion load. The therapeutic mechanism of action of this drug is not known, however, it is believed to have anti-oxidant and anti-inflammatory properties. These properties make it an ideal candidate for identifying multi-omic treatment response signatures.

Hypothesis

Changes in gene expression in PBMC and metabolomics profiles in plasma within 6 weeks of initiation of DMF for RR-MS can associate with clinical and radiological response at 15 months.

Aims

To determine;

1) Whether gene expression changes in the blood mononuclear cell fraction and metabolite changes in plasma 6-weeks post treatment initiation can further elucidate the pharmacodynamic actions of DMF.

2) Whether, multi-omics profiling can discriminate treatment-naïve MS patients and controls.

3) Whether, in MS patients starting DMT, gene expression and metabolite changes associated with drug action in the blood mononuclear cell fraction at 6 weeks can be associated with clinical and radiological response at 15 months.

Chapter 2. Materials and Methods

Ethical Approval

The study was approved by the NRES Committee London – Camden and Islington – 14/LO/1896.

Study Design

This longitudinal cohort study intended to explore whether changes in transcriptome, cytokine and metabolomic signatures could help predict response to DMF in RRMS patients. 36 patients were recruited to the study and 10 healthy age- and sex- matched controls. Subjects and controls attended a screening visit and two further study visits over the course of 15 months. Patients were recruited from the MS clinics at Imperial College Healthcare NHS Trust and consent for the study was obtained. Healthy volunteers were recruited by advertisement. Inclusion and Exclusion criteria for patients outlined below.

Inclusion Criteria

- Male or female between 18 and 65 years of age inclusive, at the time of signing the informed consent.
- Clinical or clinical and laboratory supported diagnosis of MS (Revised McDonald criteria, 2010).
- Recently prescribed, but yet to commence Tecfidera for RRMS.
- Able to lie comfortably on back for up to 60 minutes at a time.
- Capable of giving written informed consent, which includes compliance with the requirements and restrictions listed in the consent form.

Exclusion Criteria

A subject will not be eligible for inclusion in this study if any of the following criteria apply:

- Any clinical significant medical conditions that in the opinion of the investigator would compromise subjects' safety or compliance with study procedures.

- Unwillingness or inability to follow the procedures outlined in the protocol.

- Subject is mentally or legally incapacitated.

- Presence of a cardiac pacemaker or other electronic device or ferromagnetic metal foreign bodies as assessed by a standard pre-MRI questionnaire.

- Claustrophobia limiting tolerance of MRI

Study Days

For the patients with MS, clinical assessments were performed at baseline, 6 weeks (+/- 3 weeks) and 15 months (+/- 8 weeks). Patients underwent magnetic resonance imaging (MRI) at 6 weeks (+/- 3 weeks) and 15 months (+/- 8 weeks), and detailed clinical and patient-centred histories were taken to determine whether a patient was responding to the drug. Clinical assessment included MS Functional Composite (MSFC) at all timepoints as well as a quality of life questionnaire (SF-36) at baseline and 6 weeks (+/- 3 weeks). Blood samples were taken on entry to the study, and 6

weeks (+/- 3 weeks) and 15 months (+/- 8 weeks) post treatment onset. Blood from healthy volunteers was taken at the same timepoints. A schema for the project is outlined in Fig 2.
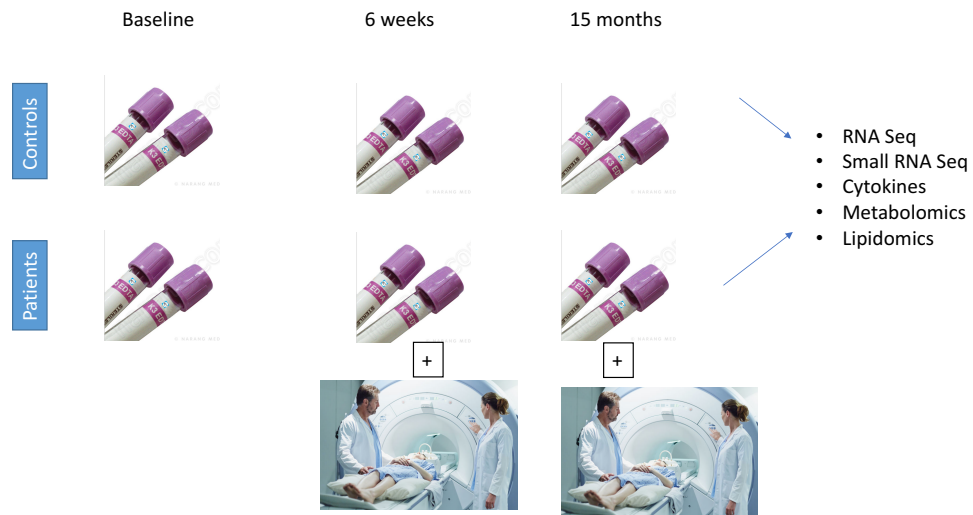
Fig 2. Schematic of study visits for RRMS patients and healthy controls. Images taken from Google Images.

Patient and Control Demographics

The study cohort included 36 patients with RRMS (RRMS; median EDSS, 2.5, range 1- 6.5) diagnosed by McDonald criteria(12), who were recruited from the Imperial College Healthcare NHS Trust and who consented for participation in the study. Patients recruited were aged between 18-65 and treatment-free (DMT and steroids) for at least 3 months. Three patients experienced regular migraines, 2 had asthma, 2 psoriasis and 2 had autoimmune thyroid disorders. Three patients had thalassemia trait and one patient had hypercholesterolaemia. Within the patient cohort, there were 13 men and 25 women. Mean age was 42.9 (+/- 12. 1), average disease duration from diagnosis was 6 years (+/- 5 years) and average disease duration from symptom onset was 12 years (+/- 9 years). 19 patients were treatment naïve and 16 patients had been

on previous treatments. 3 patients were current smokers. Amongst those patients who had been on previous treatments, 9 had been on a β-Interferon, 6 on copaxone and 1 on azathioprine. Table 1 summarises patient demographics.

10 age- and sex- matched healthy controls were recruited by local advertising. 4 were men and 6 were women. 1 patient was a current smoker. There were no comorbidities in this cohort. Table 1 summarises control demographics.

	MS Patients	Controls
Gender	13 Men and 23 Women (n =36)	4 Men 6 Women (n = 10)
Mean Age (years)	42.9 $\pm$ 12.1	37.3 $\pm$ 11.0
Average disease duration from diagnosis (years)	6 $\pm$ 5	N/A
Average disease duration from first symptom (years)	12 $\pm$ 9	N/A
EDSS (median, range)	2.5 (1 – 6.5)	N/A
Treatment Naïve patients	15	N/A
Current Smoker	3	1

Table 1a. Patient and Control Demographics for full cohort. Values quoted as mean $\pm$ standard deviation if not indicated otherwise.

With regards to follow-up, at the 6-week visit, 5 patients did not continue in the study. 4 were unable to take the medication due to severe abdominal symptoms and 1 was unable to attend for follow-up. At the 15 month timepoint, a further 5 patients did not continue in the study. 3 had terminated treatment due to side effects and 2 had stopped the medication due to disease progression.

Sample size for this study was calculated based on previous findings from RNA-Seq experiments, feasibility and the limited timeframe of the study(155). A depth coverage of 50 million reads and a CV of 0.4 powered the study at 90% when looking for expression changes between groups with fold change > 0.5 fold.

Peripheral Blood Mononuclear Cell Extraction

Whole blood samples were taken at time points described above using EDTA tubes (40 mls total). PBMC were extracted from fresh whole blood using Ficoll technique. EDTA tubes containing blood samples were initially spun at 1400rpm for 10 minutes (Acceleration (A) 7, Deceleration (D) 7) at room temperature. Following centrifugation, overlying plasma supernatant was collected into 2ml cryovials and stored at -80 $^{\circ}$C. Remaining blood was transferred to 50 ml Falcon tubes (maximum 12mls per Falcon tube) and diluted in Dulbecco's Phosphate-buffered saline (1:2 ratio). Equivalent number of 50 ml Falcon tubes were prepared with 15mls histopaque at room temperature (Histopaque-1077, Sigma Life Science). Diluted blood was then overlaid onto Ficoll and tubes centrifuged at 1600 rpm for 30 minutes (A:5, D:2). Buffy coat containing PBMC was aspirated using sterile Pasteur pipettes and washed with sterile PBS. Two wash cycles were performed followed by re-suspension of cell pellet in 10mls PBS. Cells were then counted using Trypan Blue and 5-10 million cells placed into 2 sterile eppendorf tubes for duplicate RNA extraction. Remaining cells were stored in a PBS/DMSO (9:1 ratio) solution at -80 $^{\circ}$C.

RNA extraction

RNA extraction was performed in duplicate on fresh pellet directly following PBMC extraction by Ficoll. RNA extraction was performed using Qiagen RNeasy kit (Qiagen, Hilden, Germany). The kit uses the selective-binding properties of a silica-based membrane to bind up to 100 micrograms of RNA. The cell pellet was lysed and homogenised using buffer RLT – a denaturing guanidine-thiocyanate-containing buffer that inactivates RNases. Ethanol was then added and the sample inserted into a RNeasy Mini spin column for RNA purification and removal of contaminants. The sample underwent three buffer washes (buffer RPE and RW1) and an on-column DNase digestion step using DNase1 stock solution. The duplicate assay yielded 2 50Ml samples. RNA quantity was measured using a spectrophotometer and the ratio of absorbance at 260nm and 280nm was calculated to assess the purity of RNA (a ratio of 2.0 is generally accepted as pure for RNA). The ratio of absorbance at 260/230 was also calculated as a secondary measure of nucleic acid purity (a ratio of 2.0-2.2 is generally accepted as pure for RNA).

Cytokine Assay

Cytokine and inflammatory markers were measured using the Meso Scale Discovery (MSD) v-PLEX Neuroinflammation kit. (Meso Scale Discovery, Maryland, USA). This kit consists of 5 microplates pre-coated with antibodies to 40 neuroinflammatory markers. The panels are separated as follows. Pro-inflammatory panel (IFN- , IL-1 , IL-2, IL-4, IL-6, IL-8, IL-10, IL-13 and TNF-a), cytokine panel (IL-1, IL-5, IL-7, IL-12/IL-23p40, IL-15, IL-16, IL-17A, TNF-a, VEGF), chemokine panel (Eotaxin, MIP-1, Eotaxin-3, TARC, IP-10, MIP-1, MCP-1, MDC, MCP-4), angiogenesis panel

(VEGF-C, VEGF-D, Tie-2, Flt-1, PlGF, bFGF) and vascular injury panel (SAA, CRP, VCAM-1, ICAM-1).

Calibration curves were prepared in the supplied assay diluent. Arrays were preincubated with 25 µL per well of assay diluent for 30 minutes. After the preincubation, 25 µL sample or calibrator was added in duplicate to the appropriate wells. The array was then incubated at room temperature for 2 hours. The array was washed with PBS plus 0.05% Tween 20, and 25 µL detection antibody reagent was added. After 2 hours of incubation at room temperature, the array was washed and the detection buffer was added. Results were read with a MSD Sector Imager 6000. Sample cytokine concentrations were determined with Softmax Pro Version 4.6 software using curve fit models.

RNA-Seq protocol

Samples were sent to Genewiz (Genewiz, New Jersey, USA) for library preparation and sequencing for mRNA and small RNA. The first step in this process was to run a QC on RNA samples and to ensure sufficient quantity and quality. Initial quantities assayed using Nanodrop and Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and RNA integrity using Tapestation (Agilent Technologies, Palo Alto, Ca, USA). Of all samples sent to Genewiz, all had 260/280 ratio of >2. RNA integration number was also > 9 for all samples. Two samples had insufficient material for RNA sequencing and were therefore omitted from further analysis. A table of this QC is provided in Appendix 1. Following QC, library preparation was performed in the same facility (Genewiz, New Jersey, USA).

This used the NEBNext Ultra RNA Library Preparation Kit from Illumina following manufacturer's recommendations (NEB, Ipswich, MA, USA). mRNA was enriched with Oligod(T) beads and enriched mRNAs were fragmented for 15 minutes at 94°c. cDNA for first and second strand were then synthesized, end-paired and adenylated at 3' ends. The universal adapter was then ligated to cDNA fragments along with the index sequence and the library enriched with limited cycle PCR. Sequencing libraries were validated using Agilent Tapestation and quantified using Qubit 2.0 Fluorometer and quantitative PCR (Applied Biosystems, Carlsbad, Ca, USA).

Small RNA sequencing was prepared using the Illumina Small RNA Library Prep kit (Illumina, San Diego, Ca). Here, Illumina's 3' and 5' adapter was added to RNA molecules with a 5'-phosphate and a 3'-hydroxyl group sequentially. Reverse transcription was then performed to create single stranded cDNA. cDNA was PCR amplified with a common primer and a primer containing the index sequence. The amplified cDNA construct was purified using polyacrylamide gel electrophoresis, the correct band was excised from the gel, eluted with water and concentrated by ethanol precipitation. This final library was then quality controlled using Qubit 2.0 Fluoremeter and Agilent TapeStation (Applied Biosystems, Carlsbad, Ca, USA).

Sequencing for RNA (Illumina Hi-Seq platform 2 x 150bp PE configuration) and small RNA (Illumina HiSeq 2500, 1 x 50bp SR configuration) was performed. The sequencing libraries were multiplexed and clustered on two flowcells and these were loaded onto the Illumina HiSeq platform. The small RNA libraries were multiplexed and clustered on 3 lanes of a flowcell.

Raw sequence data was quality controlled, initially by determining the Phred quality score, the most common metric used to assess the accuracy of a sequencing platform(156). Q scores are defined as a property logarithmically related to base calling error probabilities (P)[2].

Q = -10 log$_{10}$P

As an example, if Phred assigns a Q score of 30 (Q30), the probability of an incorrect score is 1 per 1000. In my dataset, the percentage of >Q30 bases was >96% for all samples. An initial FASTQC was then performed on all samples. FASTQC is a software that checks the per base sequence quality, the per base GC content, Kmer content and adapter content of a sample(157). An example of a FASTQC report is provided in Appendix 1.

 The second QC step involved clipping the adapters from the raw sequences. This was performed using TrimGalore software. Trimgalore is a multi-step software tool that removes low-quality base calls and removes adapters from raw sequences(158). Following this step, a further FASTQC was performed on all samples to ensure that data was of sufficient quality.

Following adapter clipping, sequence reads (in the form of fastq files) were aligned to a reference genome and raw counts for genes were derived. In order to ensure greater reproducibility of findings, alignment was performed using two different pipelines.

Method 1: DRAGEN alignment and Counts using HTSeq

DRAGEN (Dynamic Read Analysis for Genomics) is a commercially available software that is capable of aligning raw sequence reads from the Illumina HiSeq platform (http://www.edicogenome.com/) to a reference genome. The Fastq files containing the raw sequences were inputted into the DRAGEN pipeline and aligned using reference genome GRCh38.p10. Alignment accuracy for this software is reportedly superior to existing open-source alignment software such as TopHat and Star(159). Gene hit counts were calculated from the output BAM files derived from DRAGEN software using HTSeq-count(160). HTSeq-count is a python library that counts aligned reads overlapping exons for each gene. Only reads mapping unambiguously to a single gene are counted using the software; and reads possibly mapping to more than one gene are discarded. HTSeq count is a reliable counting tool for RNA-Seq data(161).

Method 2: STAR alignment and Counts using StringTie and Ballgown suite

Alignment and Count derivation were also performed using a different pipeline to ensure robustness of analysis and to increase the probability of reproducible findings. Alignment was performed using STAR software (Spliced Transcripts Alignment to a Reference)(162). STAR is an open-source pipeline that utilises a seed finding phase to search for a Maximal Mappable Prefix; a concept used by large-scale genome alignment tools. MMP involves sequential searching for a given read sequence within the reference genome sequences. The longest substring that matches the reference genome is found using the maximum mappable length of the sequence. Alignments are then built using clustering and stitching and alignments are scored based on penalties for mismatches, insertions, deletions and splice junction gaps.

Small RNA alignment and count derivation

Sequence reads were trimmed to remove possible adapter sequences at the 3' end.

Reads with 17 to 35 nucleotides were retained and mapped against micro RNAs for

Homo Sapiens in the micro RNA database(163) (miRBase 21) using the CLC

Genomics workbench v.9.0.1 (Qiagen, Hilden, Germany) . Mapping statistics were

generated using QualiMap v.2.2.1(164).

Metabolomics

A pilot plasma sample set was sent for global metabolomic profiling to Metabolon

Inc. (Metabolon, Durham, NC, USA). This included 15 MS patients and 10 healthy

controls at baseline and 6 weeks.

	MS Patients	Controls
Gender	8 Men and 7 Women (n = 15)	4 Men 6 Women
Mean Age (years)	37.3 $\pm$ 11.3	38.3 $\pm$ 9.7
Average disease duration from diagnosis (years)	5 $\pm$ 4	N/A
Average disease duration from first symptom (years)	7 $\pm$ 6	N/A
EDSS (median, range)	1.5 (1 – 6.5)	N/A
Treatment Naïve patients	8	N/A
Current Smoker	0	1

Table 1b. Patient and healthy volunteer demographic data for Metabolon Inc. samples. Values quoted as mean $\pm$ standard deviation if not indicated otherwise.

Metabolon Mass Spectrometry Analysis

Mass Spectroscopy (Discovery cohort)

Samples were sent to Metabolon Inc. (Durham, NC) for untargeted metabolomics

analysis. Samples were precipitated with methanol followed by centrifugation prior to

QC. QC involved addition of several control samples to aid chromatographic alignment. This included pooled matrix samples, technical replicates (derived from a pool of well characterized human plasma), process blanks and within-sample spiking of endogenous compounds. Experimental samples were randomized across the platform and run with the QC samples.

Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS).

All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) (Waters, MA, USA) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer (Thermo Fisher Scientific Inc., MA, USA) interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. Samples were analysed using acidic positive ion conditions (optimized both for hydrophilic and hydrophobic compounds), basic negative ion conditions and a negative ionization following elution from a HILIC column. The scan range covered 70-1000 m/z.

Metabolite Identification

Raw data was extracted and peaks identified using the Metabolon library. Biochemical identifications are based on three criteria: retention index (RI) within a narrow RI window of the proposed identification (within 150 RI units ~10s) , accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards(165). Peaks were quantified using area-under-the-curve.

National Phenome Centre Metabolomics

All baseline and 6 week plasma samples underwent metabolomic profiling at the National Phenome Centre (NPC) at Imperial College London. This included global mass spectrometry analysis and a NMR-based lipoprotein analysis.

Samples were formatted into 96-well polypropylene plates as previously described(166) and diluted in a 5:1 ratio (v/v) with an aqueous internal standard (IS) solution (sample:IS). The sample was further diluted by addition of acetonitrile in a 1:3 ratio (sample+IS:ACN, v/v). The plate was heat sealed using Thermo-Seal foil sheets and an ALPS 50 V-Manual Heat Sealer (Thermo Fisher Scientific Inc., MA, USA) prior to mixing for two minutes using a MixMate (Eppendorf) operating at 1400 rpm and subsequent 2 h incubation, both at 4°C. All samples were then centrifuged at 4°C for 10 minutes at 3486 × g to separate precipitated protein and other particulate material from the supernatant, which was aspirated and dispensed to a separate 96-well plate for UPLC-MS analysis.

A single HILIC UPLC-MS analysis was performed on an Acquity UPLC instrument coupled to a Xevo G2-S oaTOF mass spectrometer (Waters Corp., Manchester, UK) via a Zspray electrospray ionization (ESI) source operating in the positive ion mode. Details of the UPLC-MS system configuration and HILIC analytical method used for profiling have been reported previously(166). Briefly, a 2 µl full loop injection (with 5× overfill) of prepared sample was made to a 2.1 × 150 mm BEH HILIC column (Waters Corp., Milford, MA, USA) thermostatted at 40 °C. Gradient elution was performed using acetonitrile + 0.1% formic acid (A) and 20mM ammonium formate in water + 0.1% formic acid (B) solvents with a flow rate of 0.6 mL/min. Details of the gradient can be found in the supporting information of the provided reference.

Feature extraction and data processing were performed using Progenesis QI 2.1 software (Waters Corp., Manchester, UK) as previously described(166). Metabolites of interest from the discovery cohort analysis were located either by retention time and accurate mass match to an authentic reference standard or by accurate mass and interpretation of the MS/MS fragmentation pattern (specifically for methyl succinyl-carnitine, as no reference standard was commercially available).

Targeted Quantitative Analysis of Monomethyl Fumaric acid (MMF)

All samples were subjected to targeted analysis for the absolute quantification of MMF. Briefly, samples were prepared by dilution with three volumes of acetonitrile + 0.1% formic acid containing100 mg/mL heavy labeled MMF (mono-methyl-^{13}C,d_3 fumarate, Sigma-Aldrich). The samples were mixed and centrifuged as described above prior to solid phase extraction using OSTRO sample preparation plates (Waters Corp., Milford, MA, USA) operated by vacuum manifold for two minutes. The product sample was dried overnight under a continuous flow of nitrogen gas and reconstituted using an amount of ultra-pure water equal to the original volume of plasma used (150 μl).

Sample analysis was performed using an Acquity UPLC instrument coupled to a Xevo TQ-S tandem quadrupole mass spectrometer (Waters Corp., Manchester, UK) via a Zspray electrospray ionization (ESI) source operating in the negative ion mode. MMF was identified using the National Phenome Centre reference library as being well retained by the reversed-phase chromatographic method described previously(166) and that method was therefore validated with the following parameters: limit of detection (LOD) = 0.5 ng/mL; limit of quantification (LOQ) =

5ng/mL; linear range = 0.5-100ng/mL; dynamic range = 0.5-2000ng/mL; sensitivity = 0.99+/-0.023. Within-run precision was measured by seven repeated analyses of samples at the low, medium, and high range of the method (% relative standard deviation) = 3.3, 1.6, and 2.4 respectively. Matrix effects and absolute recovery covered a chosen low, medium and high range within the dynamic range (40, 400 and 800ng/mL). Matrix effects indicated negligible ion suppression with values above 95% with no ion enhancement. Absolute recovery was within acceptable range of 77 to 118% in accordance to stated GLP and GMP Bioanalytical method validation guidelines.). Peak integration and calculation of final MMF concentration was performed using TargetLynx software (Waters Corp., Milford, MA, USA).

NMR Spectroscopy

Plasma samples were centrifuged for 5 minutes at 4ºC at 13,000 rpm to remove solid particles in suspension. Bruker 600 Avance III spectrometer was used to acquire 1D NMR general profiling and 2D J-res spectra(167). Spectra were processed, phased and baseline corrected in automation using TopSpin software (v3.2, Bruker BioSpin, Rheinstetten, Germany). The signal from the anomeric proton of the glucose at 5.23 ppm was used to calibrate the plasma spectra.

Lipoprotein and cytokine analyses

Lipoprotein quantification was performed with the Bruker B.I.-LISA (Bruker IVDr Lipoprotein Subclass Analysis) platform using the -CH3 and CH2 resonances in the 1H-general profile NMR spectrum (0.88 and 1.29 ppm, respectively). These broad resonances were bucketed and fitted against a Partial Least Square (PLS2) regression model. The model has been validated against direct assays after plasma

ultracentrifugation(168). For each sample, the method estimates concentrations of lipids (cholesterol, free cholesterol, phospholipids and triglycerides) within the main lipoprotein classes (VLDL, IDL, LDL, HDL), subdivided according to increasing density and decreasing size (VLDL-1 - VLDL-6, LDL-1 - LDL-6, HDL-1 - HDL-4). 105 lipoprotein sub-fractions were analysed in each sample using this method. A test-retest comparison of sub-fractions samples measured in a single healthy control from samples taken on 5 consecutive days showed that mean lipoprotein sub-fraction concentrations varied by < 5%.

Modelling approaches for data reduction

Omics approaches yield large datasets and statistical tools have been developed to interpret them (9,10). These include dimension reduction techniques that rationalise the date into a limited number of variables (called components) that account for the greatest variance in the dataset. For any given omics experiment (X) there are variables (also referred to as features) (x) and samples (n). This can be represented by:

$$X = (x_1 + x_2,....x_p)$$

Dimension reduction identifies a set of new variables or components(**f**) through a linear combination of the original variables. Coefficients (q) will make up the components and these are also known as loadings. Dimension reduction finds a set of q's that take into account the greatest possible variance of f components.

$$f = Xq$$

Principal Component Analysis

Perhaps the most widely used dimension reduction approach is principal component analysis (PCA)(169). This can be performed in a number of ways including Eigen analysis, latent variable analysis, factor analysis, singular value decomposition (SVD) or linear regression(170). The method reduces the dimensions of a dataset to a few principal components. The first principal component is the mathematical combination of measurements accounting for the largest amount of variability in the data(171). The second component finds the coefficients that maximise variance orthogonally to the first component. The total variance is defined as the sum of variances of the predicted values for each component. In general, the selection of components is subjective, however it will sometimes be the case that the number of components is decided upon based on a cumulative proportion of variance(172).

Partial least squares discriminant analysis

The major criticism of principal component analysis is that the components themselves may be modelling variation in x-variables that are of little or no relevance to the groups they belong to (y-variables). Partial least squares (PLS) (173, 174) attempts to overcome this problem by also calculating a set of latent variables, but uses a criterion other than maximum variance for the decomposition step. The criterion is a normalised weight vector which is calculated as the covariance between the y-variable and the x-variables. In simple terms, the model selects the components which describe the greatest amount of variance based on knowledge of group membership of samples. Each component is then checked for predictive power

through cross-validation(175, 176). In this way, PLS can be used as a supervised classification method and the response variable can be classified using a binary vector. This is also known as PLS-DA. PLS-DA is regularly applied to gene expression data (Datta 2001).

Sparse partial-least squares discriminant analysis (sPLS-DA)

sPLS-DA combines integration and simultaneous variable selection. It reduces data further than PLS-DA using a Lasso penalization combined with singular value decomposition (SVD)(177). The advantage of this approach is that the number of dimensions and components can be selected in addition to the number of variables to select for each dimension. This can then be tested using cross-validation or leave-one-out analysis(178, 179).

Random Forest Analysis

This is a supervised classification technique utilising an ensemble of decision trees(180). For a given decision tree, a randomised set of the data with class identifiers is selected to build the tree (this is known as the 'bootstrap sample' or 'training set'. The remaining data not used in this sample is known as the 'out-of-bag (OOB)' variables and these are passed down the tree to provide class predictions for each sample. This process if repeated up to thousands of times to produce a forest. Each sample is then classified as belonging to a group depending on the class prediction frequency derived from the forest. The method is unbiased as the prediction for each sample is based on trees that were built without that sample. Once all class predictions are made, the OOB error rate is calculated as a measure of prediction accuracy. The advantages of random forest include the fact is does not

make parametric assumptions and it does not overfit the data. One can also derive the variables that make the largest contribution to the classification using the 'mean decrease accuracy' (MDA). The MDA is determined by running the decision trees with and without each variable to determine if this has an effect on the predictive accuracy of the mode. If a variable is important, the prediction accuracy drops and this is recorded as the MDA.

Multivariate regression and penalisation

Simple regression consists of one independent variable, X, and one dependent variable Y. For a given value of X, one can estimate a value for Y. When performing simple regression, one can calculate the sum of squared errors (SSE) which is the sum of squared deviations of the data from the predicted values; representing variation in the data that is not explained by the regression model.

When analysing big data sets, the number of dependent variables will be greater than 1. In this context, multivariate linear regression can be used. In this model, R^2 represents the amount of variation that is explained by the dependent variables. One of the disadvantages of multivariate linear regression is that often the dependent variables are correlated[181] or there are too many of them and this causes variation in the regression slope and intercept that make the regression model unstable. It also increases the standard error of the estimated regression coefficients and can cause overfitting of the data[182].

One approach to remedy overfitting data is to constrain the magnitude of the parameters or 'budget' them. This can be done a number of ways but the classic approach is through ridge regression. This approach allows one to choose a value for

λ; a weight that balances minimising SSE whilst limiting model complexity. The best value for λ is normally chosen using cross-validation which evaluates the model with a validation data set(183). A further regularisation method is the least absolute shrinkage and selector operator (LASSO) regression(184). This model works by removing variables as a variable selection method. There are a number of weaknesses to this approach, not least that it is not robust to collinearity of variables. Furthermore, it will not choose more variables than the number of samples so it a poor application for 'omics datasets. One regression method that blends both of these approaches is elastic net regression. This approach allows evaluation of number of variables to be used through cross-validation. The advantages of this approach is that it selects more variables than LASSO but also shrinks nonzero parameters like ridge regression(185). A final approach to reduce variable number but maximise variance accounted for is Akaike Information Criterion(186). AIC allows us to perform model selection to derive a preferred model based on a trade-off between goodness of fit and model complexity. Given a set of candidate models, this approach will give AIC values for each model; the lowest value representing the best preferred model. When increasing parameters are added, the AIC fixes a penalty and this can discourage overfitting.

Chapter 3. Discriminating disease activity using a multi-omics comparative approach between MS patients and healthy controls

Introduction

Current markers of disease activity in MS are principally surrogates of accumulated disease activity over time. They provide a picture of a patient's history with MS rather than capturing the current disease status. This is well exemplified by MRI measures such as lesion load and size and by the EDSS score.

An approach to mitigate this is by comparing MS patients to healthy controls and then correlating any discriminating features with existing measures of disability.

Gene expression profiling is a method of measuring the approximate biological activity of a cell or tissue. It has been utilised in the field of MS to further our understanding of the genetic component of the disease. Early proof of concept studies have established the MS transcriptome in blood[108] and brain tissue[109]. A more recent study compared the transcriptome of PBMCs in RRMS, SPMS and PPMS patients with healthy controls uncovering a signature of 380 transcripts that differentiated MS groups from controls. As with many such studies, the effect sizes were only moderate (0.7 – 2.29) but the transcripts identified were concordant with genetic susceptibility studies in MS[187, 188]. A further approach has been to perform gene expression profiling in post-mortem tissues from MS patients[189]. However, little has been done comparing healthy controls to MS patients at post-mortem. MicroRNAs (MiRNA) are a novel approach to addressing genetic dysregulation in MS patients relative to controls. A number of studies have demonstrated either up- or down- regulated miRNAs in cell subsets of MS

patients(190). The most robust miRNAs include miR-15a, miR-16-1, miR-20b, miR-27b, miR-29b, miR-128, miR-155 and miR-320a(190). These findings have led to a limited number of studies testing the utility of such miRNAs as biomarkers in the diagnosis of MS(191, 192). Nevertheless, poor replication of such studies has meant that no clinically validated micro-RNAs exist to aid in the diagnosis of MS.

The role of metabolomics in distinguishing MS patients from healthy controls is another approach that also ties in environmental influences. Dickens and colleagues recently performed a metabolomics analysis of MS patients and controls. Using a PLS-DA, they were able to reliably identify MS patients from controls with high sensitivity and specificity. Interestingly, the two most discriminant metabolites were glucose and phosphocholine(139). Most recently, Lim and colleagues identified metabolites of the kynurenine pathway, known to be involved in chronic inflammation and progression of neurodegenerative disease, as dysregulated in MS patients compared with controls. A decision tree method was able to discriminate MS subtypes with high sensitivity (91%), maintained in a validation cohort(193).

A potential role for lipids in MS

Disease progression biomarkers may be related to previously identified factors associated with the expression of co-morbidities of MS(194). Cardiovascular disease is one of the most frequent co-morbidities of MS and other systemic autoimmune disorders such as rheumatoid arthritis, systemic lupus erythematosus and psoriasis (195-197). While the underlying causes of vascular comorbidity may be multi-factorial, elevated plasma lipids, lipoproteins (the carrier molecules of lipids in blood e.g. LDL) and oxidised lipids are found with both cardiovascular disease and MS, at

least in part because of a direct effect of pro-inflammatory cytokines on hepatic lipid metabolism(197, 198).

Previous studies have suggested an association between dyslipidaemia, dyslipoproteinaemia and MS disease activity (new MRI lesions(199) or worsening EDSS(199-201)). Recently, indirect evidence consistent with a causal role for dyslipidaemia in disease progression was provided by results from MS-STAT, a Phase IIa study showing that high dose simvastatin reduced rates of brain atrophy and disability progression in patients with secondary progressive MS(202).

Lipids and lipoproteins can be measured using a number of techniques including ultracentrifugation, high performance liquid chromatography and nuclear magnetic resonance (NMR) spectroscopy(203). A novel NMR method has recently been developed that can measure lipid concentrations within plasma lipoprotein sub-fractions (lipoproteins subdivided based on density and size). This method has the advantages of being amenable to high throughput use, is highly reproducible and can provide simultaneous class-specific information on both lipoproteins and their constituent lipids; both of which have been associated with MS disease activity(168). It promises increased sensitivity to changes in circulating lipids related to systemic inflammatory states.

A role for cytokines

Measuring cytokines in MS patients has identified that Th2 cytokines are reduced in MS patients (despite possibly being protective) and that Th1 cytokines (traditionally seen as pro-inflammatory) are associated with a more active disease course.

Furthermore, Th17 cytokines are known to be elevated in MS as well as IL-1β and IL-6. Further studies have identified that cytokines can be altered by treatments such as interferon, natalizumab and DMF(204).

In this study, I utilised baseline samples from untreated RRMS patients and healthy controls to determine whether multi-omics approaches can elucidate markers that are associated with disease and that may correlate with existing measures of disease disability. I performed the multi-omics analyses in sequential biological order starting with gene expression, moving on to proteomics (cytokines) and lipidomics and finally to the end products of metabolism (metabolomics). Each modality was analysed using statistically appropriate tools that have previously been validated for this type of 'omics. I also correlated these markers with existing markers of disease activity.

Methods

All methods for sample collection, assays and basic statistics are provided in the Methods chapter. Specific methods for this chapter are outlined below.

Transcriptomics

Statistical Analyses

Differential expression analysis was performed on count data derived from StringTie and HT-Seq software using DESEq2(205). Deseq2 performs differential expression analysis by first performing a regularised logarithm transformation followed by detection and correction of dispersion estimates that are too low through modelling using average expression strength over all samples. An assumption made by this software is that genes of similar average expression have similar dispersion. Where counts are low or dispersion is high for a specific gene, DESEq2 shrinks the log fold change (LFC) towards zero. The software provides a log fold change value across conditions as well as an adjusted p-value corrected for multiple comparisons using Benjamini and Hochberg(206).

I repeated the differential expression analysis using Ballgown, an open-source software that uses FPKM values in a linear model(207). The model log transforms the count data and applies linear models to test for differential expression at the gene, transcript, exon or junction level. Contrasts between patients and controls were performed controlling for variables age and gender. Adjusted p-value for significance (Padj) was set at (Padj < 0.05). Fold change cut offs were thresholded at log2-fold change of ±0.3.

Following DE analysis, principal component analysis was performed. Description of this technique is outlined in Methods section above. Differentially expressed genes with adjusted p-value < 0.05 were identified and submitted to G:profiler functional annotation tool for gene enrichment. Downstream pathway analysis was also performed using the Ingenuity Pathway Analysis (IPA) software; commericially licensed by Qiagen.

Metabolomics

Full lipoprotein and mass spectrometry profiling were performed on all patients and all control samples at both the National Phenome Centre and Metabolon.

Statistical Analyses

For lipoprotein analyses, contrasts between patients and controls were performed using ANOVA (F test). If the ANOVA showed a statistically significant difference (p < 0.05), a *post hoc* analysis was performed using Holm-Bonferroni to correct for multiple comparisons. The desired level of significance was set at p < 0.05 after correction.

Very Low Density Lipoprotein (VLDL) sub-fractions that had statistically significantly different concentrations between patients and controls with a fold change > 1.30-fold were taken forward for further analyses. Pearson's correlation coefficients between each pair of VLDL sub-fractions were calculated and used to create the matrix of pairwise correlations (Fig 9a,9b).

Correlations between EDSS and plasma lipoproteins and MS patient characteristics were analysed using multivariate linear regression models. I included well recognised

risk factors of age, gender, disease duration and body mass index (BMI)(208). EDSS was defined as the dependent variable, with the patient characteristics and lipid sub-fraction measures as predictive variables:

$$EDSS = \alpha_1 VLDL \text{ (sub-fraction)} + \alpha_2 Age + \alpha_3 Gender + \alpha_4 Disease\ Duration + \alpha_5 BMI + \alpha_6$$

where coefficients $\alpha_{1,...,}\alpha_5$ determine the contributions of each of the predictor variables to EDSS and α_6 expresses the residual error. I used analysis of residuals to check the required assumptions of normally distributed errors with constant variance.

Due to the high collinearity between the VLDL sub-fractions, which would have led to variance inflation in a model with all of the lipid sub-fraction measures(209), each lipid sub-fraction was analysed separately using the regression model. These multivariate regression models then were evaluated from their R^2 values, residual standard errors and the overall p-values. The level of significance was set at $p < 0.05$, corrected for multiple comparisons using the Holm-Bonferroni method. The relative significance of contributions of individual predictive variables was assessed by averaging sequential sums of squares obtained from all possible orderings of the predictors using the LMG method(210). Statistics analysed using 'relaimpo' package in R(211).

I used the Akaike Information Criterion (AIC) to determine the most parsimonious model with the predictor variables available(212). AIC allows us to perform model selection to derive a preferred model based on a trade off between goodness of fit and model complexity. These statistics were analysed using 'mass' package in R(213). In

order to test if the resulting model was unstable, I performed a leave-one-out cross validation.

To explore associations between cytokines and each of the individual VLDL sub-fractions increased in patients with MS relative to healthy volunteers, I created regression models using each of the VLDL sub-types included in the parsimonious model derived from AIC and input these into a generalised linear model via penalised maximum likelihood. To improve the interpretability and accuracy given the large number cytokine variables, I used a lasso regression based cross–validation(214). Analyses were performed using the glmnet package in R(215).

For metabolomic data provided by Metabolon, two way ANOVA was initially performed to determine whether any metabolites were significantly different between patients and controls. These values were corrected by false discovery rate and a threshold q-value of <0.05 was set for level of significant. sPLS-DA was then used to identify whether patient samples could be distinguished from controls. The sPLS-DA analysis was run for 4 components using 10-fold repeated cross-validation and 50 variables selected for each component. The output of this analysis is an R^2 value, which expresses the variance in the dataset that can be explained by each component(178). Statistics were analysed using the 'mixomics' package in R(216).

In order to derive the most discriminatory variables between patients and controls, I used the Random Forest classification(180). I used an 80:20 ratio (train-test subsets) and 1000 trees to build the model. A variable importance measure was computed

based on the mean decrease accuracy metric. In this analysis, the number of discriminatory variables was limited to 30.

Cytokines

Cytokine and inflammatory markers were measured using the MSD V-PLEX kit at baseline (pre-treatment) in both the patient and control groups.

Statistical Analysis

Comparisons between patients and controls was performed using Student's t-test corrected for multiple comparisons with a significance threshold of p-value < 0.05.

Integrated Analyses

Transcriptome and metabolome data were integrated using two approaches. Firstly, pairwise correlation analyses were performed using Pearson's correlation coefficient between differentially expressed genes at baseline and differentially expressed metabolites. Cut-off for significant correlation was $-0.7 > r > 0.7$. I also conducted a pathway-based integration using Ingenuity Pathway Analysis (Qiagen, Helden, Germany).

Results

Transcriptomic analyses

MS patients show variable gene expression compared to healthy controls

A differential expression analysis between untreated RRMS patients and healthy controls was initially performed. The fold change results of the DE analysis underwent variance-stabilising transformation and a principal component analysis was performed on this data. There was a reasonable visual separation of samples by disease phenotype in the first two principal components (Fig 3). Despite this apparent separation, a sample distance heatmap which plots the Euclidean distance between samples was also generated which showed very limited separation between patients and healthy controls. This is demonstrated by the mixed clustering of subjects with no obvious separation between the two groups (Fig 4).
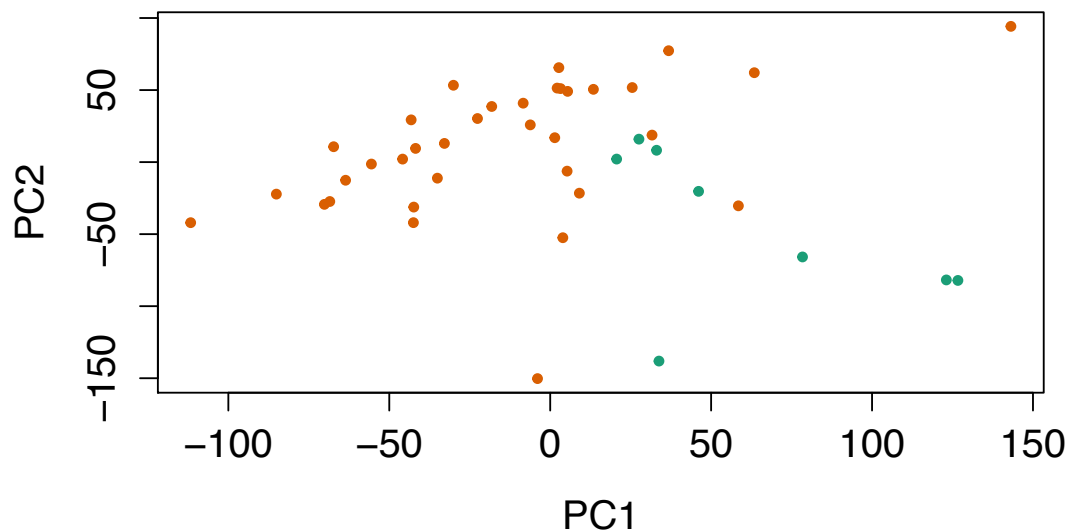
Figure 3: RNA samples from PBMCs of RRMS patients and healthy controls show moderate separation. PCA plot of RRMS patients at baseline (no treatment) and healthy controls. Red dots represent RRMS samples. Green dots represent control samples.
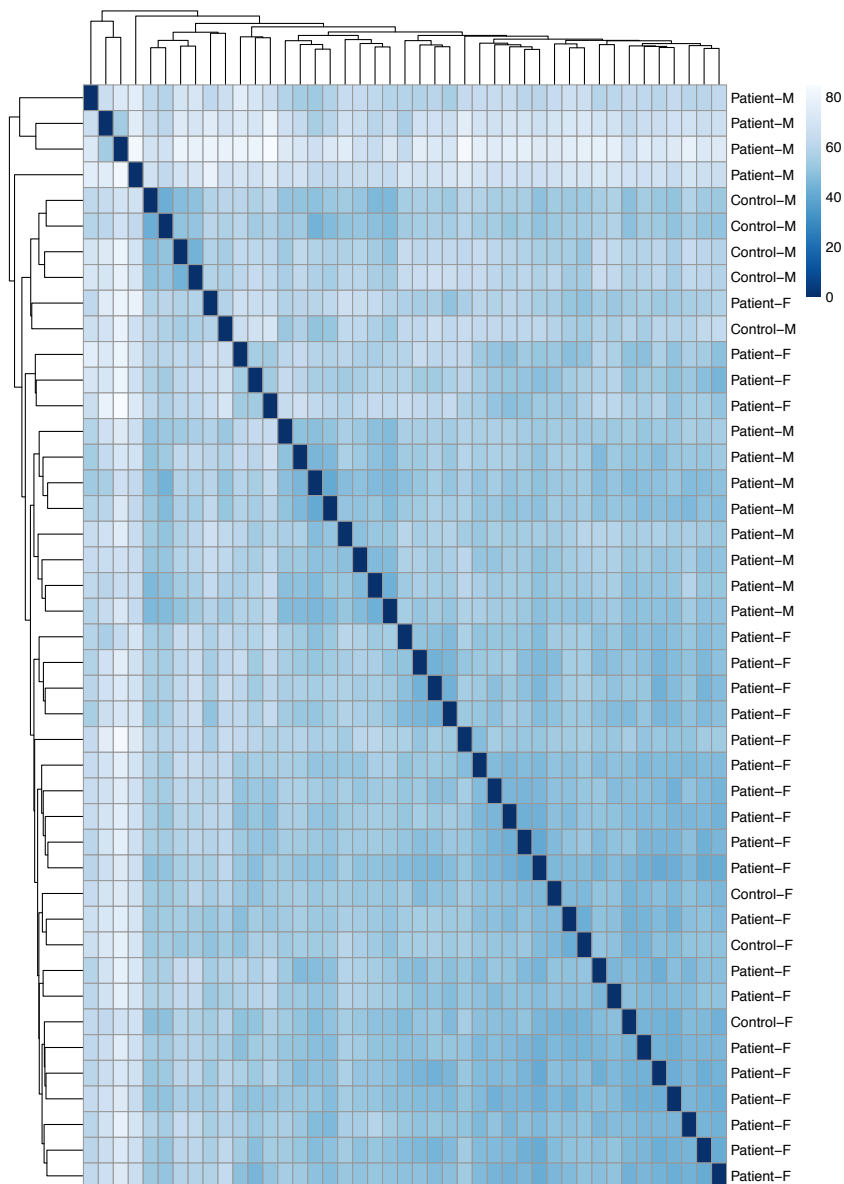
Fig. 4. RNA samples from RRMS patients and health controls show very limited separation using unsupervised clustering analysis. Sample distance heatmap between RRMS patients and control as calculated from the regularized log transformation. Samples which demonstrate similarity in Euclidean space are clustered close to each other.

Controlling for age and gender, I found 522 genes that were differentially expressed (DE) between patients and controls (Padj < 0.05) (Fig 5). Of these, 254 were downregulated in patients and 268 were upregulated. Fold changes ranged from (-2.6)

to (+2.15). Two genes were differentially upregulated with a fold change >2 (padj <0.05) and 5 genes were differentially downregulated with a fold change < 0.5 (padj < 0.05). A table of these genes (Table 2) along with boxplots of specific genes are displayed (Fig 6). Killer cell immunoglobulin-like receptor 3DL2 (KIR3DL2) which was significantly differentially expressed in my dataset has previously been implicated in the pathogenesis of MS due to its interaction with HLA-A3(217, 218). I found significant DE of TOB1 (fold change -0.49, padj < 0.005) which has also demonstrated a strong association with MS(111, 219-221).
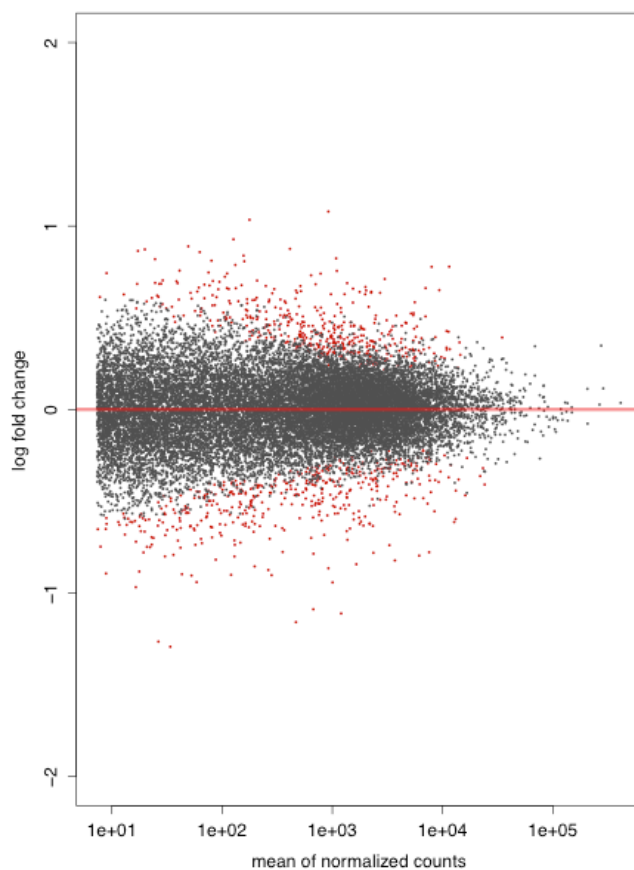
Fig. 5. 522 genes are differentially expressed between PBMCs of RRMS patients compared to healthy controls. MA plot of log$_2$fold change of normalised counts of RRMS patients (untreated) compared to healthy controls. Each dot represents a gene. Black dots are genes with Padj > 0.05. Red dots are genes with Padj < 0.05.

Gene	Log2FoldChange	Adjusted P-Value
ID1	-1.29	1.04E-06
KIR3DL2	-1.27	3.57E-06
NR4A2	-1.16	1.07E-07
FOSB	-1.11	1.79E-05
GRASP	-1.09	4.23E-14
C21orf33	1.03	9.75E-05
RPL10P9	1.08	0.0001

Table 2: Some genes are highly differentially expressed in the PBMCs of RRMS patients compared to healthy controls. The 7 most differentially expressed genes between MS patients and healthy controls (log$_2$fold change > 1 or < - 1, padj < 0.05). Abbreviations described in Abbreviation section.
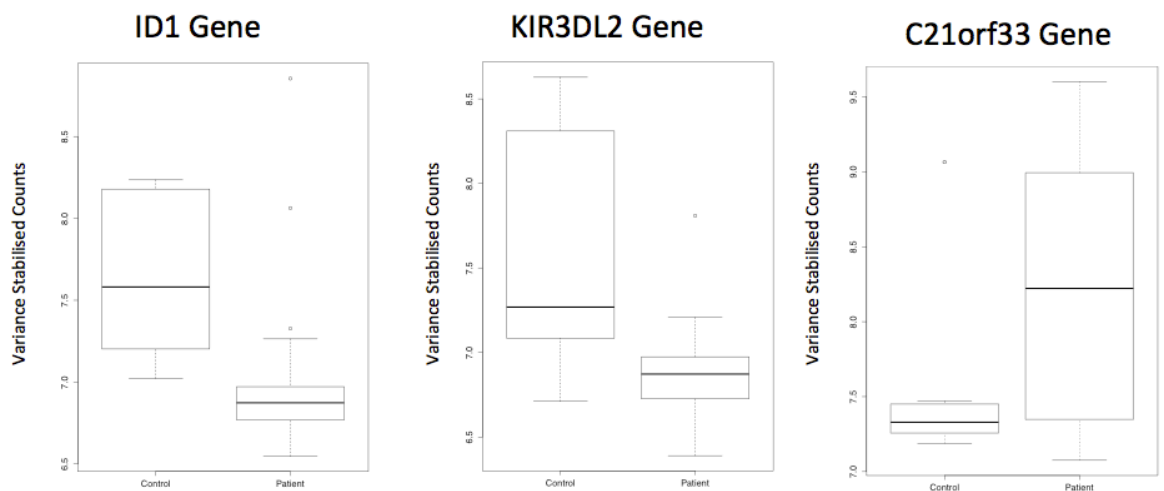
Fig 6. Some genes are highly differentially expressed in the PBMCs of RRMS patients compared to healthy controls. Boxplots of the three most differentially expressed genes between MS patients and controls are visualized. Padj < 0.05 for all ID1, KIR3DL2 and C21orf33.

Pathway Analysis reveals a role for immune-based mechanisms in MS

I performed an enrichment based analysis using those genes that showed the greatest amount of DE between patients and controls (padj < 0.05) using G:Profiler. Kegg pathways enriched in the dataset included 'B Cell activation involved in the immune response' and 'TNF signaling pathway' (p < 0.001).

Pathway analysis demonstrated that canonical pathways Nrf2-mediated oxidative stress, Th2 signaling, IL-9 signaling, Toll-like receptor signaling and Th1 and Th2 activation were enriched in the differentially expressed genes in patients relative to healthy controls (Fig 7).
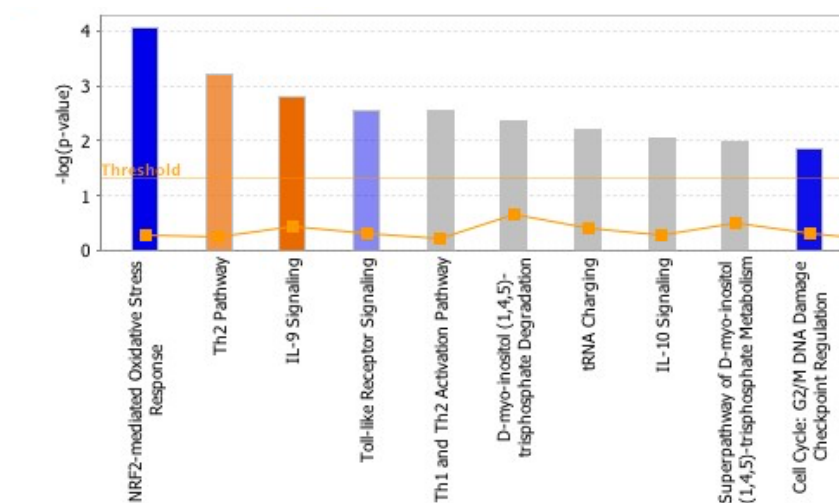
Fig. 7. Differentially expressed genes in RRMS patients are enriched for immune pathways. Pathway analysis of differentially expressed genes between MS patients and controls. All pathways above the orange line are statistically significant (p < 0.005).

I performed a correlation analysis in patients between transcript levels for all genes and EDSS. No gene was strongly positively correlated with EDSS (r > 0.6) and no gene was strongly negatively correlated with EDSS (r < 0.6).

Plasma IL-2 is raised in MS Patients

Given the enrichment of immune mediated pathways such as Th1 and Th2, I investigated whether there were corresponding differences in plasma cytokines in RRMS patients compared to healthy controls. Patients had higher plasma concentrations of IL-2 than healthy volunteers ([0.56 pg/ml vs 0.29 pg/ml], 1.9-fold difference, $p < 0.05$) (Fig. 8). I also tested whether there was a difference in concentrations of Th2 cytokines IL-4 and IL-13, but these were non-significant ($p > 0.05$).
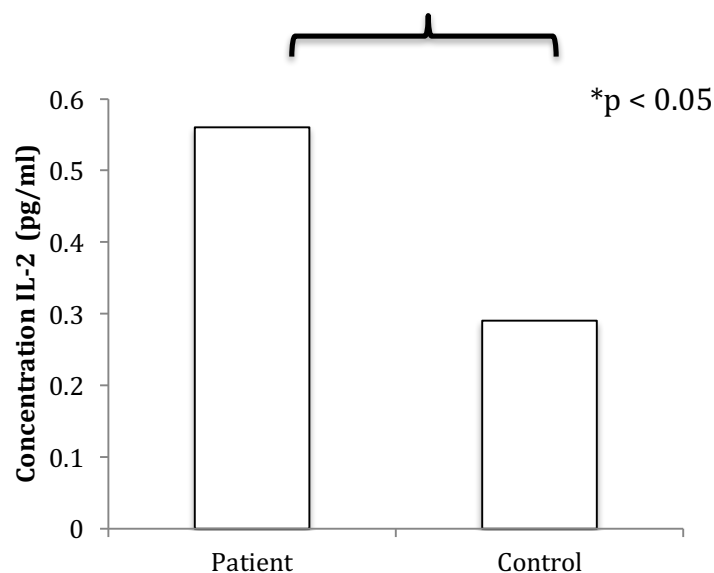
Fig. 8. RRMS patients have higher concentrations of IL-2 than healthy controls. Bar chart demonstrating concentrations of IL-2 in RRMS patients and healthy controls. Concentration provided in pg/ml.

NMR based lipoprotein analysis

VLDL sub-fractions are increased in MS patients compared to healthy controls

A global lipoprotein sub-fraction survey was performed comparing the MS patient
and healthy volunteer groups. I found 13 lipid sub-fractions with increased
concentrations in the patient group relative to the healthy volunteers (Table 3). 12 of
these were VLDL sub-fractions and 1 was a HDL sub-fraction. 8/12 of the VLDL
sub-fractions had changes >1.30-fold in patients relative to volunteers (range, 1.38-
2.3-fold)(Table 3). I also found elevated concentrations of plasma triglycerides in the
patient group ($p < 0.0001$). Strong pairwise correlations were observed between the
relative concentrations of VLDL lipid sub-fractions in the patient and healthy
volunteer groups ($r > 0.5$) (Fig 9).

Lipid Sub-Fraction	Concentration (mg/dL)		Relative Change (patients/controls)	F Test (corrected)
	Patient	**Control**		
HDL-1				
Free Cholesterol	34.39 (± 19.10)	23.53 (± 8.36)	1.46	P = 0.03
VLDL-1				
Triglycerides	31.92 (± 27.24)	24.58 (± 27.32)	1.29	p = 0.002
Free Cholesterol	2.04 (± 2.13)	0.88 (± 0.78)	2.30	p = 0.002
Cholesterol	5.94 (± 5.87)	3.87 (± 1.81)	1.53	p = 0.0001
Phospholipids	5.26 (± 4.53)	3.63 (± 1.78)	1.45	p = 0.004
VLDL-2				
Free Cholesterol	1.27 (± 1.32)	0.72 (± 0.37)	1.76	p = 0.00009
Triglycerides	13.66 (± 9.99)	8.92 (± 3.81)	1.53	p = 0.003
Phospholipids	3.21 (± 2.46)	2.11 (± 0.94)	1.52	p = 0.003
Cholesterol	3.02 (± 2.96)	1.98 (± 0.83)	1.53	p = 0.00009
VLDL-3				
Triglycerides	11.40 (± 8.78)	8.23 (± 3.47)	1.38	p = 0.004
Free Cholesterol	1.49 (± 1.50)	1.19 (± 0.50)	1.25	p = 0.0002
Cholesterol	3.67 (± 3.60)	2.87 (± 1.30)	1.28	p = 0.0006
Phospholipids	3.39 (± 2.80)	2.62 (± 1.15)	1.29	p = 0.002

**Table 3: Lipid sub-fractions that were significantly increased in MS patients relative to healthy
volunteers. Concentrations are quoted in mg/dL ± standard deviation.**
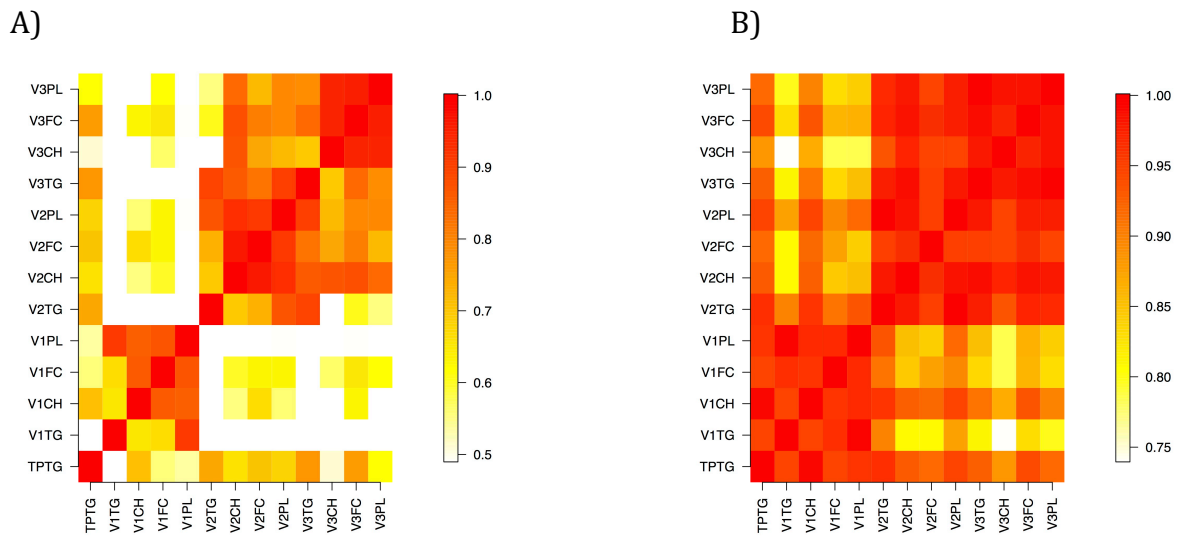
Fig 9 a,b. Matrices illustrating pairwise correlations between VLDL sub-fraction concentrations for healthy volunteers (a) and MS patients (b). Abbreviations: TPTG = total plasma triglycerides, V1TG = VLDL-1; triglycerides, V1CH = VLDL-1; cholesterol, V1FC = VLDL-1; free cholesterol, V1PL = VLDL-1; phospholipids, V2TG = VLDL-2; triglycerides, V2CH = VLDL-2; cholesterol, V2FC = VLDL-2; free cholesterol, V2PL = VLDL-2; phospholipids, V3TG = VLDL-3; triglycerides, V3CH = VLDL-3; cholesterol, V3FC = VLDL-3; free cholesterol, V3PL = VLDL-3, phospholipids. Bar represents r-value for pairwise correlation.

VLDL sub-fraction concentrations are correlated with disability in people with MS

I wanted to test whether VLDL sub-fractions elevated in my cohort were correlated with MS disease disability. Multivariate regression models showed that each of the VLDL sub-fractions found to be significantly increased in the patients relative to the healthy volunteers was correlated with EDSS (Table 4).

Lipid Sub-Fraction	Model R^2	Model P-value	Lipid P-value
VLDL-1			
Free Cholesterol	0.35	p = 0.0006	p = 0.02
Cholesterol	0.38	p = 0.0003	p = 0.005
Phospholipids	0.33	p = 0.001	p = 0.04
VLDL-2			
Free Cholesterol	0.40	p = 0.0001	p = 0.002
Triglycerides	0.39	p = 0.0002	p = 0.004
Phospholipids	0.37	p = 0.0003	p = 0.0007
Cholesterol	0.39	p = 0.0002	p = 0.003
VLDL-3			
Triglycerides	0.40	p = 0.0001	p = 0.002

Table 4: VLDL sub-fractions are associated with EDSS. Statistical measures of fit for each VLDL sub-fraction with corresponding p-values for the regression model and for each VLDL predictor coefficient.

I assessed the relative importance of each predictive variable for each model using a residual sum of squares approach. For two of the VLDL-2 models (free cholesterol and cholesterol), the VLDL sub-fractions accounted for the greatest proportions of variance (0.39 and 0.38, respectively) (Fig. 10 a, b). For the remaining models, the VLDL sub-fractions accounted for the second greatest proportion of variance after age.

A)

B)
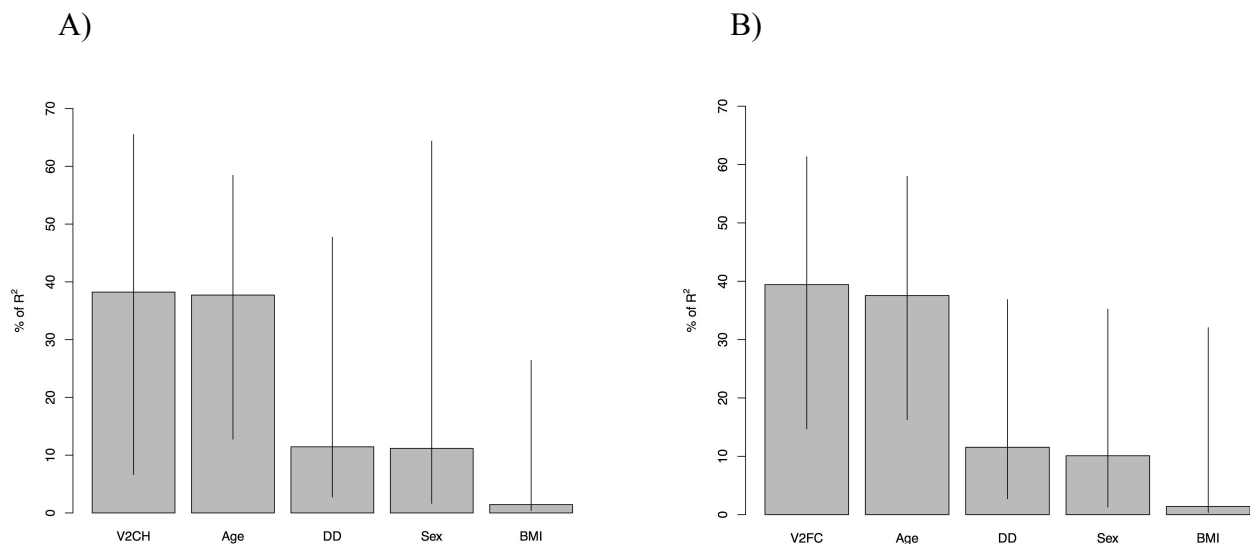
Fig 10 a, b. Relative weighting of variables to multivariate regression models for VLDL-2 (free cholesterol and cholesterol). Values expressed as percentage of R^2 with 95% confidence intervals. Abbreviations: DD = Disease Duration; V2CH = VLDL-2; cholesterol, V2FC = VLDL-2; free cholesterol.

I selected an optimal model based on Akaike Information Criterion (AIC) after incorporating all VLDL sub-fraction concentrations and clinical characteristics. The best quality model excluded BMI, disease duration and five of the lipoprotein sub-fractions, leaving VLDL-1 (phospholipids) and VLDL-2 (triglycerides and phospholipids), age and sex as explanatory variables:

$$\text{EDSS} = \beta_1 \text{VLDL-1 (phospholipids)} + \beta_2 \text{VLDL-2 (triglycerides)} + \beta_3 \text{VLDL-2 (phospholipids)} + \beta_4 \text{Gender} + \beta_5 \text{Age} + \beta_6$$

This multivariate regression model described a moderately strong correlation between EDSS and the lipid sub-fraction measures after accounting for age and sex ($R^2 = 0.48$, $p < 0.0005$). All 3 lipid sub-fractions in this model were statistically significant ($p = 0.01$, $p = 0.003$ and $p = 0.01$, respectively). I tested stability of the model using leave-one-out cross validation (LOOCV). For each leave-one-out cross validation, the model was statistically significant and the lipid sub-fractions were correlated with EDSS score ($R^2 > 0.46$) (Supplementary Lipid Table – Appendix 1).

Integrated analysis reveals that cytokines are associated with relative increases in VLDL sub-fractions in MS patients

Generalised linear regression models were created using the VLDL sub-fractions in the optimal model derived from AIC to test for an association between VLDL sub-fractions and plasma cytokine levels. Consistently strong correlations between VLDL sub-fraction increases and both CCL-17 and IL-7 concentrations were found ($R^2 = 0.78$, $p < 0.0001$).

Mass Spectrometry based Metabolomics

Increased triacylglycerols in MS patients

Of over 1600 metabolites analysed in the mass spectrometry analysis, 14 were
significantly different in patients compared to controls using two-way ANOVA (q-
value < 0.05). 7 were overexpressed in MS patients and 7 were underexpressed in MS
patients (Table 5). The over-expressed metabolites were mainly triacylglycerols (6 of
seven over-expressed, q > 0.05).

Metabolite	Q-Value	Fold Change
TAG54:7-FA20:4	0.0094	1.76
TAG58:8-FA20:4	0.0080	1.79
TAG58:9-FA20:4	0.0060	1.85
TAG56:9-FA20:4	0.0076	1.88
TAG58:8-FA20:3	0.0069	1.88
TAG58:10-FA20:4	0.0014	2.06
dihomo-linolenoyl-choline	0.0017	2.35
4-acetylphenol sulfate	0.0017	0.39
isoeugenol sulfate	0.0008	0.22
5alpha-pregnan-3beta,20alpha-diol monosulfate	0.0006	0.20
pregnanolone/allopregnanolone sulfate	0.0007	0.23
4-ethylphenylsulfate	0.0003	0.23
5alpha-pregnan-3beta,20beta-diol monosulfate	0.0003	0.25
pregnanediol-3-glucuronide	0.0002	0.25

**Table 5: 14 statistically significant metabolites different between MS patients and healthy
controls. Q-value is corrected result from two-way ANOVA. Fold changes provided.**

**RRMS patients can be distinguished from healthy controls with high predictive
accuracy based on metabolomics**

Next, I wanted to determine whether patient and control groups could be separated
based on condition state using sPLS-DA. The model was developed using 100-fold
cross-validation. The estimated error rates stabilised after four dimensions for any

number of selected variables. The optimal number of variables to select for the model were determined as 50, 30, 4 and 90 using the model developed by 100-fold cross-validation. Visual inspection showed that the model could distinguish DMF treatment state accurately (Fig 11). The metabolites most associated with the first principal component are seen in Fig. 12. These included pregnanolone (in agreement with the two-way ANOVA), vanillylmandelate (VMA) and citrulline. I next tested the predictive ability of sPLSDA model using a train and test protocol (80:20 respectively). I re-built the model using 100-fold cross-validation and 4 components. The model had an 80% positive predictive accuracy.
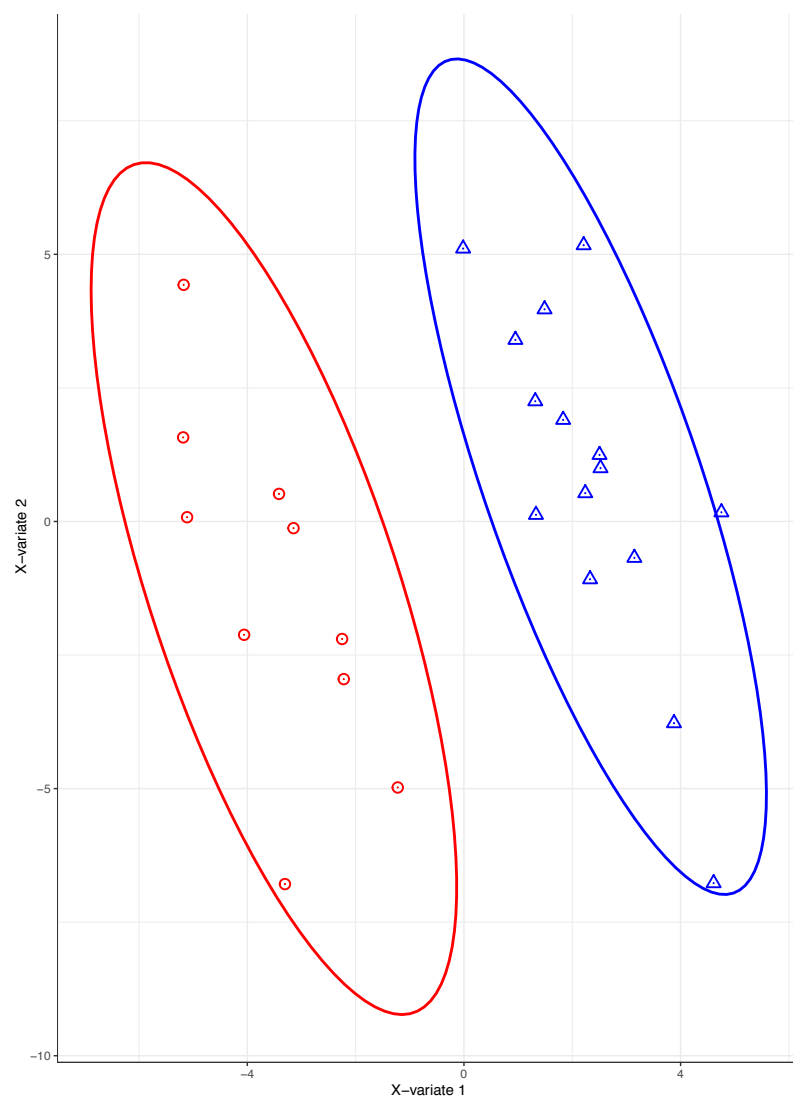
Fig 11. RRMS patients can be visually distinguished from healthy controls using supervised sPLS-DA analysis. Sample representation using the first 2 latent variables from sPLS-DA (50 metabolites selected). RRMS samples (blue triangles) and control samples (red circles) are displayed. X-variate 1 is the first latent variable (horizontal axis). X-variate 2 is the second latent variable (vertical axis).
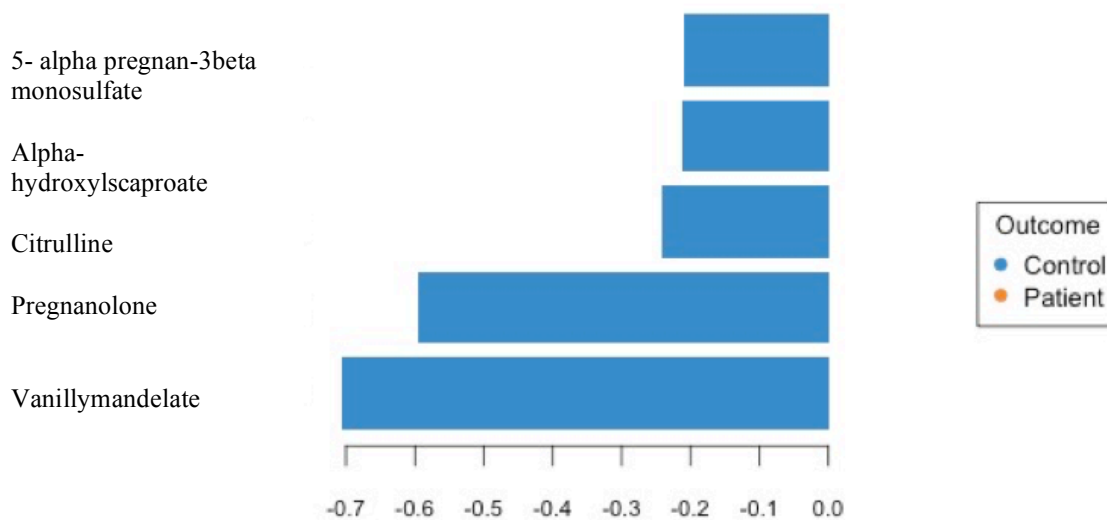
Fig 12. The most important 5 variables for the first principal component ranked from the bottom of the graph. Negative signs indicate the correlation structure between variables and underlying group. The negative correlation between these metabolites and healthy controls (represented by the blue bars) represents a reciprocal positive correlation between these metabolites and RRMS patients.

Finally, I ran a Random Forest calculation on the dataset to both look for individual discriminatory variables and also to look for predictive accuracy of the model. Consistent with my findings from sPLS-DA, the most discriminatory variable was VMA (Fig. 13). I also found that kynurenate was a discriminatory variable. The predictive accuracy of the random forest model was 76%
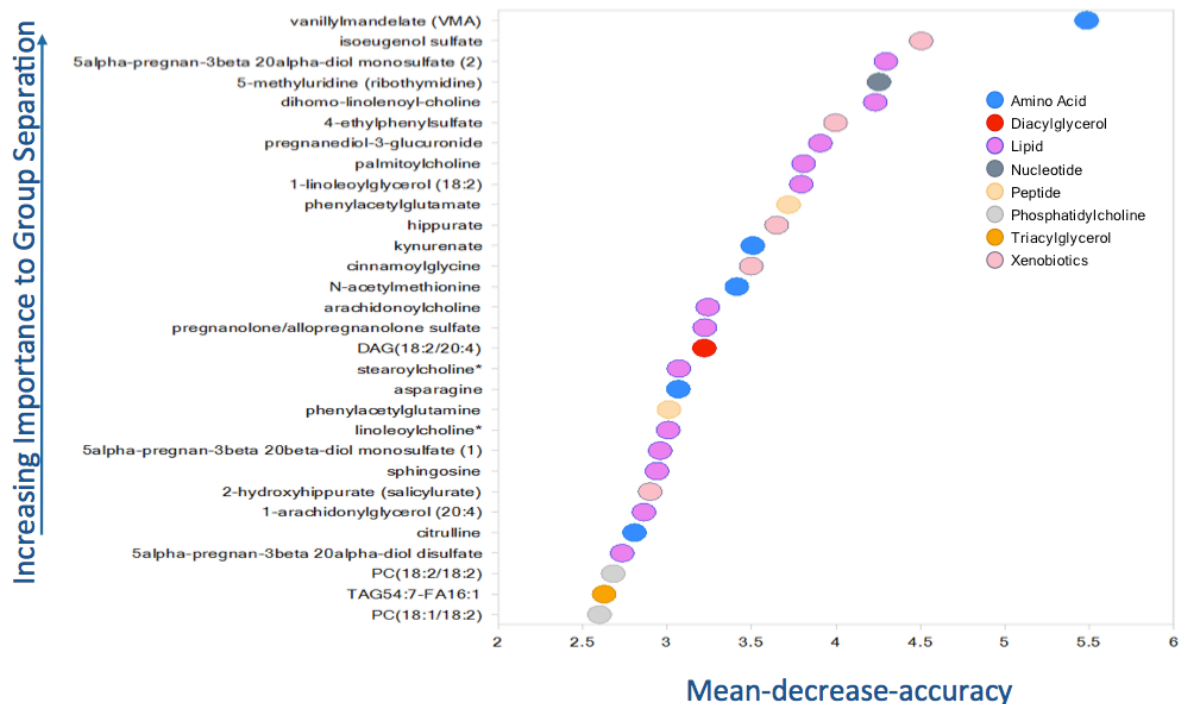
Fig. 13. VMA is the most highly discriminant metabolite comparing RRMS patients to healthy controls. Random Forest analysis demonstrating the most important variables in a comparison between RRMS patients and controls. Importance of variables expressed as mean-decrease-accuracy.

Two recent papers have described metabolomics signatures associated with MS. The first, by Lim and colleagues(193) described an aberrant kynuerine pathway in MS. Using the same statistical analyses, I found that levels of kynurenic acid (KA) in controls was almost two-fold higher in controls than in patients (fold-change 1.78, $p < 0.00005$); the *reverse* of Lim and colleagues findings. Furthermore, I found that the Kynuerine/Tryptophan ratio and Quinolinic acid /KA ratio between patients and controls was not significant ($p > 0.05$ for both comparisons). I validated reduced tryptophan levels in MS patients compared with controls although the fold-change was modest (fold-change 0.83l, $p < 0.0005$). A further recent paper of interest by

Villoslada and colleagues identified a number of discriminatory metabolites in RRMS patients compared with controls also using PLS-DA methods(222). These included sphingomyelin and lysophosphatidylethanolamine. In my cohort, neither sphingomyelin nor lysophosphatidylethanolamine were significantly different in patients compared to controls (q > 0.05). I also performed a correlation analysis with EDSS for all discriminatory variables found by Villoslada and colleagues which I was also able to detect in my cohort (namely glutamic acid, tryptophan and eiosapentaenoic acid). Consistent with this groups findings, I found a strong negative correlation between eicosapentaenoic acid and EDSS ($r = -0.59$, $p < 0.05$). However, other discriminatory variables were non-significant.

Integrated analyses implicate multiple inflammatory pathways in the pathogenesis of MS

I performed a pairwise correlation of all differentially expressed genes in the RRMS cohort with differentially expressed metabolites. For pregnanolone, I found 5 differentially expressed genes that were highly correlated with pregnanolone ($r > 0.7$) (Table 6). For other differentially expressed metabolites, only ZNF727 was highly correlated with 4-ethylphenylsulfate.

Gene	Correlation
SBDSP1	0.75
MRPL55	0.71
JUN	0.70
DCHS1	0.70
TUBB2A	0.70

Table 6: Correlation between metabolite pregnanolone and differentially expressed genes in RRMS patients (untreated) compared to healthy controls.

An integrated pathway analysis approach undertaken on differentially expressed genes in MS patients compared to controls and metabolites identified a number of inflammatory pathways significantly altered in the MS population. These included TNF and Serine/threonine kinase 11 (STK11) signalling, CD4+ proliferation and inflammation of the central nervous system (Fig 14).

A)
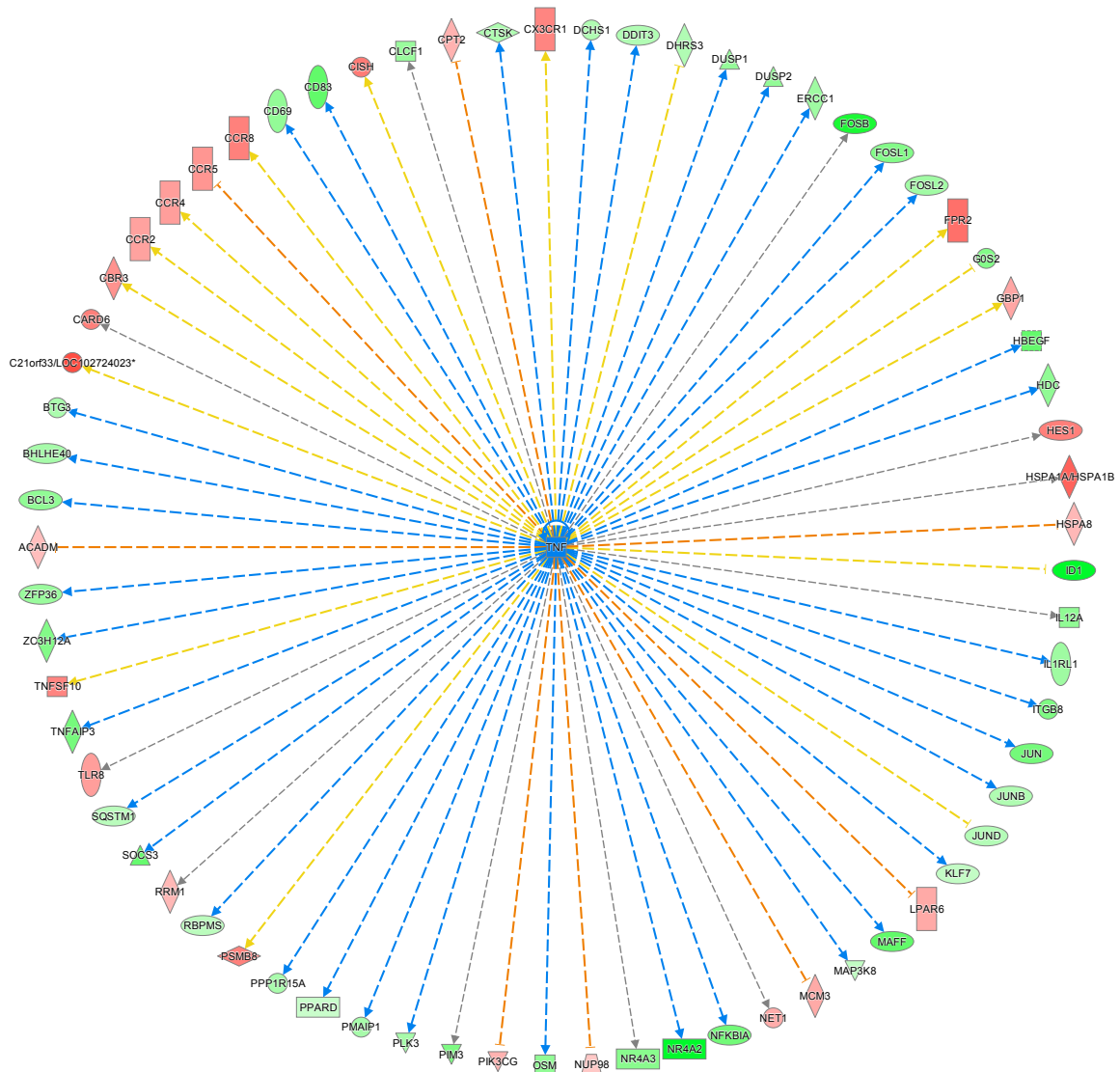
B)
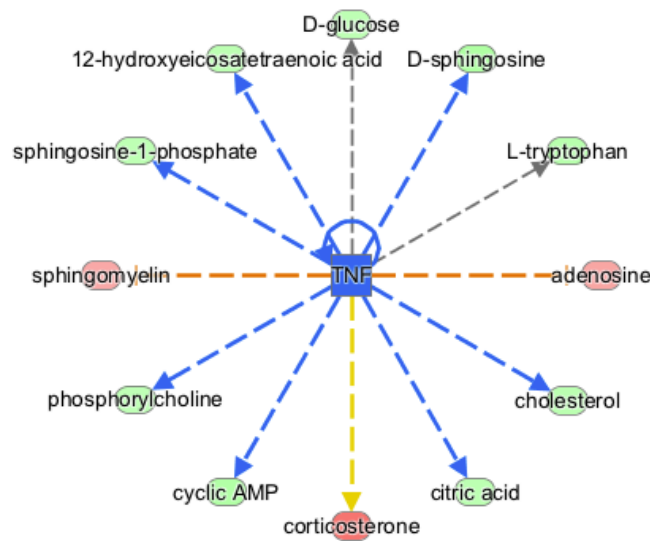
Fig 14a and 14b. TNF associated transcripts (a) and metabolites (b) differentially expressed in RRMS patients compared to controls.

C)
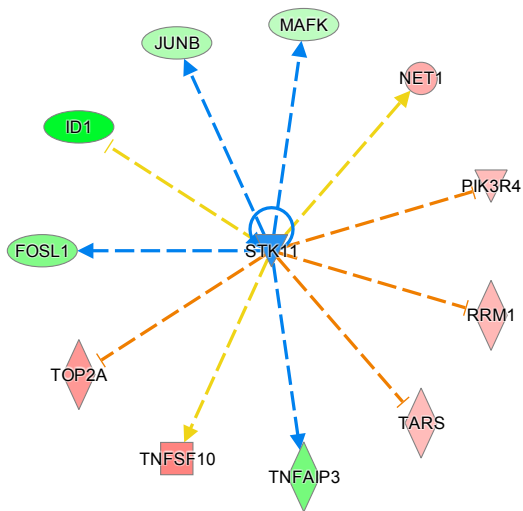
D)
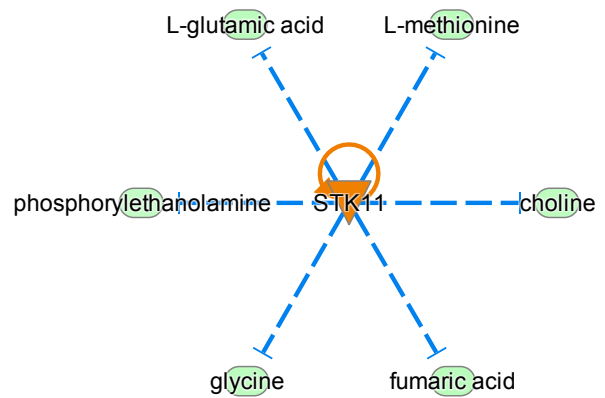
Fig 14c and 14d. STK-11 associated transcripts (c) and metabolites (d) differentially expressed in RRMS patients compared to controls.

E)
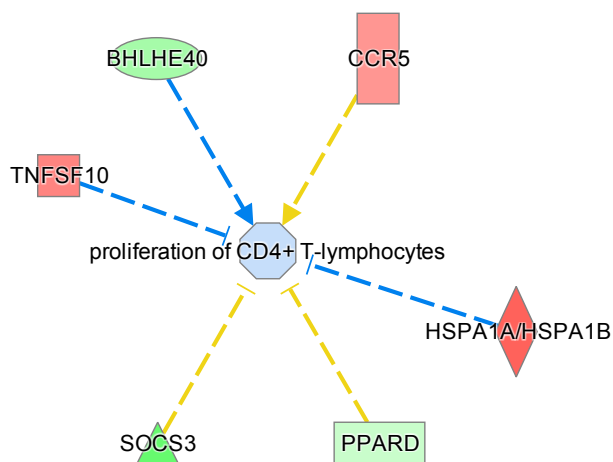
F)
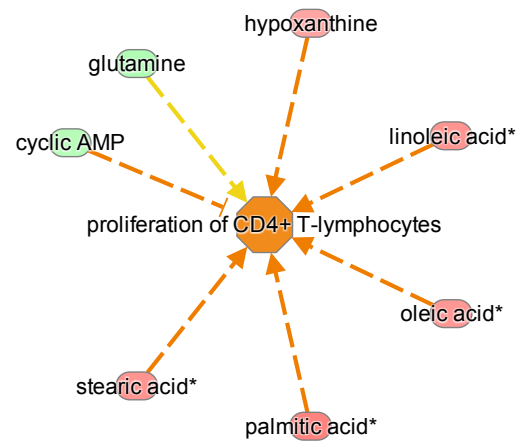
Fig 14e and 14f. Proliferation of CD4+ lymphocyte associated transcripts (e) and metabolites (f) differentially expressed in RRMS patients compared to controls.

Results Summary

- **522 genes and 1 small RNA were differentially expressed between RRMS patients and healthy controls. PCA analysis showed very limited separation between the two groups based on disease state, but very clear separation based on gender.**

- **An enrichment analysis demonstrated that the differentially expressed genes between RRMS patients and controls were involved in B-Cell activation and TNF-signalling.**

- **A cytokine analysis demonstrated that patients had higher concentrations of IL-2 than healthy controls.**

- **VLDL lipoprotein sub-fractions were elevated in RRMS patients compared to controls. These demonstrated strong correlation with disability as measured by EDSS score. A multivariate regression model combining VLDL sub-fractions with gender and age described moderately strong correlation with EDSS.**

- **Data reduction techniques applied to mass spectrometry metabolomics demonstrated good separation between RRMS patients and controls. The most discriminant variables were VMA, pregnanolone and citrulline.**

- **An integrated pathway analysis approach using the metabolomics and transcriptomic datasets identified alterations in TNF signalling and CD4+ proliferation in the MS cohort.**

Discussion

Gene expression profiling is a developing field that has been utilised in MS to try and understand the functional consequences of genetic variants that predispose to the condition(113). A number of studies have looked at gene expression in MS patients and compared this to controls, however these studies have exclusively been conducted using microarrays rather than NGS. This current study identified a large number of differentially expressed genes between RRMS patients and healthy controls that were associated with inflammatory disease. Previous studies have also identified similar differences between these groups(223) as well as specifically in PBMCs(224), however other results have been more modest, in part due to the use of microarray technology which identifies a much smaller number of RNA species(108). I was able to confirm that within the differentially expressed genes, there was enrichment for inflammatory pathways and signalling and this result was corroborated in part by the finding of raised IL-2 in RRMS patients compared to controls.

Given that dyslipidaemia has previously been associated with MS disease activity I also investigated using metabolomics approaches whether lipoprotein sub-fractions are associated with disease and disability level in people with RRMS. Using a novel NMR analytical methodology that provides information on lipid concentrations within lipoprotein sub-classes, I showed that some lipoprotein sub-fractions are elevated selectively in the plasma of the MS patients. Because of the disproportionately high number of elevated VLDL sub-fractions, I tested for a relationship between these and measures of clinical disability (EDSS). For two of my models, selected VLDL concentrations individually were the strongest explanatory variables for EDSS. A model combining selected VLDL sub-fraction concentrations (VLDL-1

(phospholipids) and VLDL-2 (triglycerides and phospholipids)) with age and sex was highly associated with disability.

By evaluating correlations of VLDL sub-fractions using different multivariate models and then applying penalised regression methods, I derived a subset of VLDL sub-fractions that was most explanatory of disease disability in this cohort. My models showed that EDSS score could be partially explained by a combination of VLDL sub-fractions, age and sex. The latter two variables previously were associated independently with a poorer prognosis in MS(225). In this study, VLDL-2 (free cholesterol and cholesterol) sub-fractions independently explained as much of the variance in disability as age.

Prior studies also have reported that plasma lipids are raised in MS and that these correlate with disease disability(194). In this study, I identified novel associations between VLDL sub-fractions, which transport cholesterol and triglycerides in the blood, and MS disability. More than one potential mechanistic relationship has been hypothesised. Lipoproteins may be involved in demyelination(226) and may exacerbate the blood-brain barrier disruption associated with inflammation(227). Increases in VLDL levels(228) also have been reported during other inflammatory disorders . In non-alcoholic steatohepatitis, VLDL has been specifically associated with hepatic inflammation(229).

However, mechanisms contributing to VLDL increases in MS are unknown. They are associated with monocyte activation in the periphery(230) and a shift to a pro-inflammatory cytokine secretion profile(231). VLDLs play a role in transport of lipids

and proteins in the blood(232). I postulate a possible mechanistic role for VLDL in disease progression in MS: VLDL may transport cytokines from the periphery to the CNS. A recent study in rheumatoid arthritis demonstrated that variants of the same lipoproteins may carry more pro-inflammatory cytokine cargo resulting in greater inflammation(233). In the context of this study, I observed an association between VLDL and both CCL-17 and IL-7. CCL-17 has been shown to play a role in T cell maturation. A common polymorphism in the IL-7 receptor gene is a known genetic risk factor for MS(98).

Using mass spectrometry based metabolomics I was able to further investigate the use of this approach in discriminating MS patients from healthy controls. This type of analysis is in relative infancy, however, a number of studies have already been performed in this field and have demonstrated a reasonable degree of accuracy in discriminating MS patients from controls(139, 141). Consistent with my NMR lipoprotein findings, I also found raised levels of triacylglycerols in RRMS patients. Furthermore, using Random Forest, ANOVA and sPLSDA I identified that VMA, pregnanolone and citrulline were highly discriminant for my RRMS cohort compared to healthy controls. Dysregulation of pregnanolone in MS patients has already been identified; it may be involved in neuroinflammatory mechanisms or neurodegeneration(234). Furthermore, using an integrated correlation-based analysis, I identified high correlation between pregnanolone and c-JUN gene expression. C-JUN is known to be activated by steroid hormones such as preganolone(235) and has previously been implicated in the pathogenesis of MS(236).

Using integrated pathway analysis, I was able to show interacting pathways amongst those over-represented in both the metabolomics and transcriptomic analyses. These included pathways and cells well known to be implicated in MS such as TNF signalling and CD4+ T lymphocytes but also lesser known variants such as STK11, a variant of which has recently been identified as a risk factor for MS(237).

A sPLSDA model using train and test subsets was moderately effective at discriminating MS patients from healthy controls. A recent study identified abnormal kynurenine pathway activity in MS. I demonstrated opposing effects to those found in this study; namely lower levels of KA in RRMS patients. Notably, this study also found lower levels in secondary progressive and primary-progressive patients suggesting this discrepancy may have resulted from different levels of disability in the RRMS cohorts respectively(193).

The main limitation of my study was the small sample size and the lack of a validation cohort. These factors limit any conclusions that can be made on the generalisability of my findings to the broader RRMS population or whether they may only describe a subset. Nevertheless, the broad range of EDSS scores, ages and disease durations in the cohort may increase the likelihood these findings are reproducible. My findings, particularly with respect to the NMR lipoprotein analysis, were highly significant. Additional validation studies are needed.

In conclusion, I have provided evidence that increased concentrations of VLDL sub-fractions in plasma correlate with disability in RRMS patients. I also further highlight gene expression and metabolite changes in RRMS patients compared with controls

that are associated with inflammatory pathways. My mass spectrometry metabolomics data confirmed differences in plasma concentrations of a number of previously known metabolites associated with MS and identified differences in concentrations of additional metabolites such as VMA, pregnanolone and citrulline. My findings highlight relationships between systemic metabolism and MS that provide insights into possible mechanisms responsible for some of the common co-morbidities of MS. However, these findings need to be replicated in larger populations to test for generalisability. They should be explored prospectively in a cohort to explore how metabolomic and transcriptomic changes evolve during disease course and progressive disability.

Chapter 4. Pharmacodynamics of DMF using multi-omic profiling

Introduction

The mechanism of action of a medication should be thought of as a complex pathway that begins from ingestion and ends in metabolism and excretion of the drug or its metabolites. The processes involved in between can be measured using a number of omics approaches in order to help understand a treatment's efficacy.

DMF is a relatively newly licensed treatment for RRMS. Its precise mechanism of action is unknown. The study of DMF's therapeutic effect lends itself well to a hierarchical signalling approach. After oral administration, DMF is rapidly hydrolysed by esterases to its principle metabolite monomethyl fumarate (MMF)(42). MMF is highly bioavailable, has a half-life of 12 hours and reaches a mean plasma peak concentration of approximately 20 μM. MMF is ultimately hydrolysed inside cells to fumaric acid, which enters the tricarboxylic acid cycle (TCA)(43, 44). The primary therapeutic mechanism of action still is debated, but is believed to involve activation of the transcription factor nuclear factor (erythroid-derived 2)-like 2 (Nrf2)(37, 38), inhibition of nuclear factor κB (NFκB)(39) and/or agonism of the hydroxylic acid receptor 2(40).

Downstream effects of DMF have been explored using gene expression studies. Vandermeeren and colleagues used northern blotting techniques to show that DMF reduces levels of ICAM-1, VCAM-1 and E-selectin mRNA arguing this supported evidence of inhibition of NF-κB by DMF(51). Further gene expression studies using northern blotting have supported this assertion(52, 238) and also the downregulation

of endothelial adhesion molecule mRNA by DMF(239). This was followed up by another study demonstrating reduced levels of TGF-alpha and Th1 cytokines with an increase in the Th2 cytokine IL-10(240). Consistent with an anti-oxidant role for DMF, Lee and colleagues found raised levels of Nrf2 expression in the CNS of DMF treated mice(241). Gene expression studies have also provided evidence for an antioxidant role of DMF. Zhao and colleagues described inhibition of hypoxia-inducible factor-1α by DMF which, in turn, resulted in activation of RACK1 and the Nrf2 anti-oxidant pathway. Most recently, Brennan and colleagues also identified activation of Nrf2 through upregulation of OSGIN1 resulting in cell-cylce inhibition and cell protection against oxidative challenge(242, 243).

Metabolomics provides an additional approach for drug target discovery and understanding pharmacodynamics(244). Study of patients with MS also may suggest novel cerebrospinal fluid (CSF), serum and urine biomarkers of disease, disease activity or disability (245). Because DMF triggers anti-oxidant pathways and is metabolized to the intermediate metabolite fumaric acid, its mechanism of action lends itself well to characterisation using metabolomics(41).

Here, I took an unbiased biologically-based sequential omics approach (transcriptomics, cyotokine and metabolomics) to investigate the short-term pharmacodynamics effects of DMF in patients with RRMS. I performed the multi-omics analyses in sequential biological order starting with gene expression, moving on to proteomics (cytokines) and lipidomics and finally to the end products of metabolism (metabolomics). Each modality was analysed using statistically appropriate tools that have previously been validated for this type of 'omics.

Methods

Transcriptomics

Patients attended at baseline and 6 weeks following onset of treatment where RNA was extracted from PBMCs as described in the protocol outlined in Methods chapter.

Differential expression analysis was performed on count data derived from StringTie and HT-Seq software using DESEq2(205). Deseq2 performs differential expression analysis by first performing a regularised logarithm transformation followed by detection and correction of dispersion estimates that are too low through modelling using average expression strength over all samples. An assumption made by this software is that genes of similar average expression have similar dispersion. Where counts are low or dispersion is high for a specific gene, DESEq2 shrinks the log fold change (LFC) towards zero. The software provides a log fold change value across conditions as well as an adjusted p-value corrected for multiple comparisons using Benjamini and Hochberg(206).

Contrasts between patients pre- and post- treatment were performed using a pairwise approach controlling for intra-individual differences. Adjusted p-value for significance (Padj) was set at (Padj < 0.05). Following DE analysis, principal component analysis was performed. Description of this technique is outlined in Methods section above. Differentially expressed genes with adjusted p-value < 0.05 were identified and submitted to G:profiler functional annotation tool for gene enrichment. Downstream pathway analysis was also performed using the Ingenuity Pathway Analysis (IPA) software (Qiagen, Helden, Germany).

I identified modules of highly correlated genes using a weighted correlation network analysis (WGCNA). This analysis involved the formation of an adjacency matrix representing the correlation changes that occurred pre- and post- treatment. The method groups genes together when their correlations to the same sets of genes change between different conditions(246). The output is a list of differentially co-expressed modules with their associated genes and an estimate of the statistical significance of the difference in co-expression derived from a comparison of the dispersion index values of each module with the null distribution obtained from permuted data(115) .

To validate my RNA-Seq findings, I repeated sequencing pre- and post- treatment for a subset of patients (n=9) in a different facility (Imperial College BRC Genomics Facility). TruSeq stranded mRNA libraries were multiplexed and sequenced with an average of 35 million cDNA fragments per sample (100 bp paired-end reads) (Illumina, San Diego, USA). Quality control was performed using FastQC software (version 0.11.2)(157). Sequencing reads were aligned to GRCh37 reference human genome by Tophat tool (version 2.0.10)(247) using the set of known genes provided by Ensembl database (release 75) with an average alignment rate of 85%. rRNA contamination (< 4%) and transcript integrity (median ratio of coverage at 5' end vs 3' end for 1000 most highly expressed transcripts > 0.6) were monitored with Picard tool (version 1.85). The raw number of read pairs mapped to each Ensembl gene was calculated with HTSeq (version 0.6.0) in 'union' mode. Reads (or read pairs) that overlap more than one gene or mapped to multiple locations were discarded. The raw counts data set was normalized with DESeq2 statistical tool(205). Sample's pairing

was taken into account while analysing patients at baseline and after treatment using DESeq2.

Cytokines

31 patients were recruited from Imperial College Healthcare NHS Trust prior to commencing DMF. Full demographic details as well as sample collection and processing are available in the Methods section. Cytokine and inflammatory markers were measured using the MesoScale Discovery (MDF) V-PLEX kit at baseline (pre-treatment) and 6-weeks post treatment in the patient cohort. The same was done for 10 age- and sex- matched healthy controls and two timepoints separated by 6 weeks.

Statistical Analysis

Paired comparisons between patients pre- and post- treatment and controls at equivalent timepoints was performed using nonparametric Wilcoxon 2-sample test.

Metabolomics

Patients were separated into an initial discovery cohort and a validation cohort to test for generalisability of results across groups. The discovery cohort included 15 patients with RRMS and the validation cohort included 12 patients with RRMS (median EDSS 3.0, range 1 – 7). Samples were analysed using mass spectrometry. Discovery and validation cohort samples were sent to independent laboratories (Metabolon and Phenome Centre respectively). The reason for this was that Metabolon provide a library of known metabolites which was unavailable at the Phenome Centre. As such I could test metabolites of interest derived from the discovery cohort using a targeted approach at the Phenome Centre. Furthermore, being able to validate findings in a

different laboratory would increase the chances of future successful replication by other groups. It is well known that inter-laboratory sampling can reduce the chances of successful replication of findings(248). Samples from the discovery cohort were also profiled at the Phenome Centre as technical replicates. Mass Spectrometry assay outlined in the Methods section.

Sparse Partial Least Squares Discriminant Analysis (sPLS-DA) was used to initially identify whether patient samples pre- and post- treatment could be distinguished on the basis of metabolites as a treatment effect. sPLS-DA is a PLS regression method in which variables are selected from the dataset in a supervised framework, i.e. with respect to the different categories of samples, which in my case was conditions pre- and post-treatment. The sPLS-DA analysis was run for 4 components using 10-fold repeated cross-validation and 50 variables selected for each component. The output of this analysis is an R^2 value, which expresses the variance in the dataset that can be explained by each component(178). Statistics were analysed using the 'mixomics' package in R(216).

I identified modules of highly correlated metabolites using a weighted correlation network analysis (WGCNA). This analysis involved the formation of an adjacency matrix representing the correlation changes that occurred pre- and post- treatment. The method groups metabolites together when their correlations to the same sets of metabolites change between different conditions(246). The output is a list of differentially co-expressed modules with their associated metabolites and an estimate of the statistical significance of the difference in co-expression derived from a

comparison of the dispersion index values of each module with the null distribution obtained from permuted data(115)

The significance of changes in specific metabolites pre- and post- treatment were estimated relative to equivalent comparisons in the healthy control population using one-way ANOVA. All changes were corrected for multiple comparisons using the false discovery rate. Statistical significance was set at a q-value < 0.05. Correlation between discriminant variables of interest and concentrations of MMF was performed using Pearson's correlation coefficient. Quantification of MMF is outlined in Methods section.

In order to derive the most discriminatory variables from the dataset pre- and post-treatment, I used the Random Forest classification(180). I used an 80:20 ratio (train-test subsets) and 1000 trees to build the model. A variable importance measure was computed based on the mean decrease accuracy metric. The Mean Decrease Accuracy (MDA) is the number of observations that are incorrectly classified by removing the metabolite in question from the model. In this analysis, the number of discriminatory variables was limited to 30. Change in lipoprotein sub-fraction levels pre- and post-treatment determined using paired Student's T-test (corrected for multiple comparisons using Benjamini-Hochberg correction).

Integrated Analyses

Transcriptome and metabolome data were integrated using a pathway-based integration method from Ingenuity Pathway Analysis (Qiagen, Helden, Germany).

Results

Transcriptomics

DMF is associated with gene expression changes in PBMCs

I initially performed a differential expression analysis comparing patient samples before and 6 weeks post-treatment. PCA analysis of the variance-stabilised data demonstrated very little separation between samples by treatment state (Fig 15). 542 genes were differentially expressed 6-weeks post treatment with DMF (Padj < 0.05). 329 were upregulated and 213 were downregulated. The 10 most significantly altered genes are listed in Table 7.
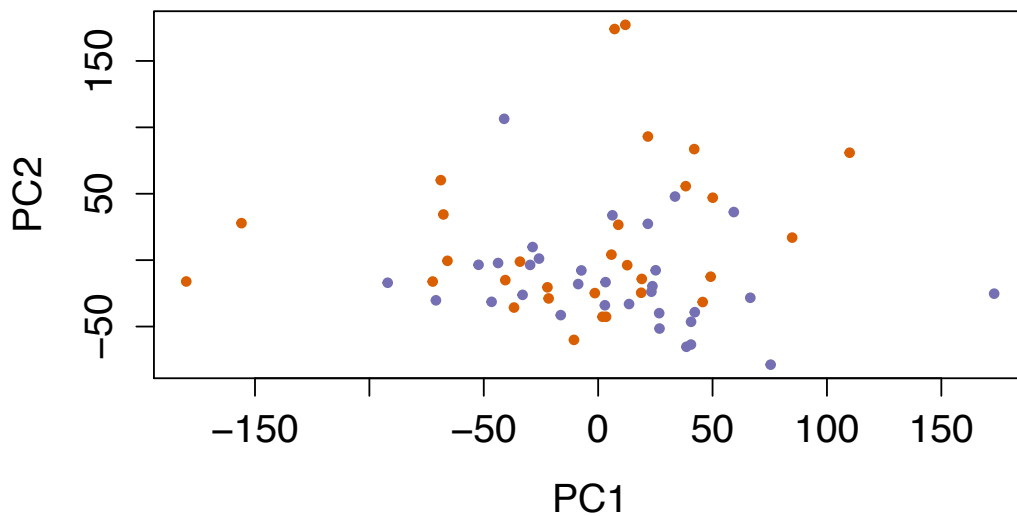
Fig 15. RNA samples from RRMS patients pre- and post- treatment cannot accurately discriminate treatment state. PCA plot of RRMS patients at baseline (no treatment) and 6 weeks post-treatment with DMF. Red dots represent RRMS samples pre-treatment. Lilac dots represent RRMS samples post-treatment. Principal component 1 on horizontal axis, principal component 2 on vertical axis.

Gene	Log2Fold Change	Padj
MITF	-0.39	1.36E-06
TRNQ	-0.34	3.88E-06
LOC107987261	-0.32	0.0017
PDGFC	-0.32	0.0016
TRNW	-0.31	1.54E-05
CD36	-0.30	0.0008
N4BP3	0.28	0.0038
SSPN	0.28	0.0120
SIGLEC6	0.28	0.0045
ZNF547	0.29	0.0111

Table 7: 10 most differentially expressed genes 6 weeks post treatment with DMF (padj <

0.05). Fold changes expressed as Log$_2$Fold Change. P-values expressed as adjusted p-

values (corrected by Benjamini and Hochberg. Abbreviations described in abbreviation

section.

I performed an enrichment based analysis using those genes that showed the

significant differential expression pre- and post- treatment in the full sample set (padj

< 0.05) using G:Profiler. Kegg pathways enriched in the dataset included 'positive

response to oxidative stress' and 'cellular oxidant detoxification ($p < 0.001$). In this

latter pathway, genes involved in this pathway and also differentially expressed post

treatment included CD36, GPX1, MGST1, NQO1, PTGS1, GSR and PRDX1.

Pathway analysis revealed that the most highly enriched canonical pathways were

Nrf2-mediated oxidative stress response ($p < 0.005$) and glutathione redox reactions

($p < 0.005$). A table and figure of the Nrf2 genes that were DE post-treatment are

highlighted in Table 8 and Figure 16.

Gene	Log2Fold Change
BACH1	-0.17
FOSL1	-0.252
FTH1	-0.166
FTL	-0.186
GCLC	-0.17
GSR	-0.14
GSTP1	-0.155
MAFG	-0.175
MGST1	-0.242
NQO1	-0.236
PRDX1	-0.142
TLR9	0.222

Table 8: Nrf2 related differentially expressed genes 6-weeks post treatment with DMF (Padj <

0.05). Fold changes expressed as log₂fold-change.
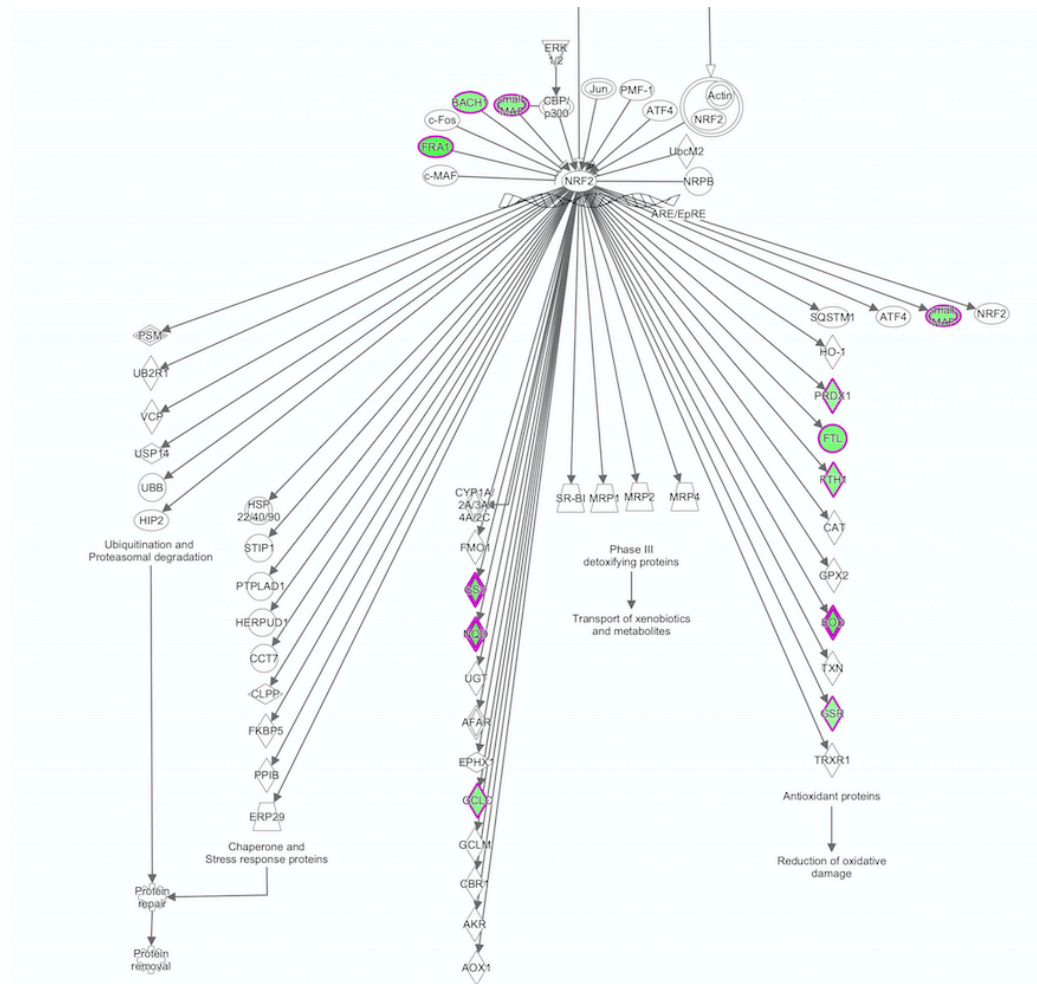
Fig 16. Nrf2 pathway with associated differentially expressed genes 6-weeks post treatment with

DMF. Green genes are those which were differentially expressed. A purple border denotes

association with multiple sclerosis.

Modules of differentially expressed genes after treatment with DMF

I identified 4 differentially co-expressed modules comprising a total of 2,444 genes (which I arbitrarily described as yellow, salmon, tan and brown) (p-value < 0001 for each of four modules). 2 of these modules were more highly correlated after treatment (salmon pink and brown, which included 288 and 854 genes, respectively) and 2 were less strongly correlated after treatment (yellow and tan, which included 358 and 944 genes, respectively) (Fig 17). Enrichment analysis demonstrated that the yellow and brown modules were enriched for KEGG pathways 'regulation of metabolic process' (p = 0.005 and p = 0.001 respectively), the salmon module was enriched for KEGG pathway 'positive regulation of innate immune response' (p = 0.001) and 'immune system process' (p = 0.001) and the tan module was enriched for KEGG pathway MHC class 1 antigen processing and presentation' (p = 0.0005). The hub genes for the brown, salmon, tan and yellow modules were PNISR, NCSTN, KIAA1429 and ZNF294 respectively.
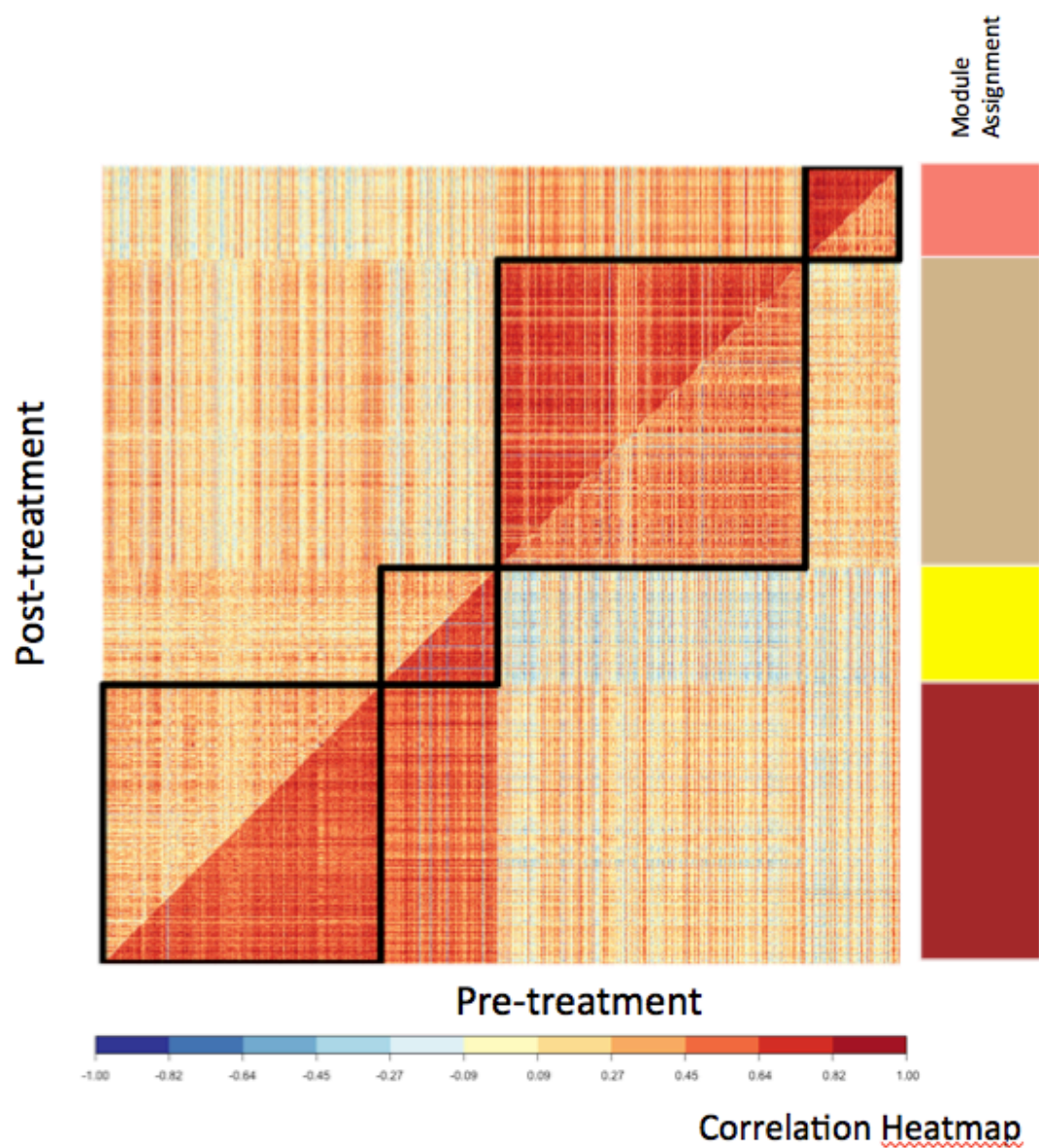
Fig. 17. Comparative correlation heatmap demonstrating differentially coexpressed modules pre- and post- treatment with DMF. The lower diagonal shows pairwise correlation of metabolites pre-treatment. The upper diagonal shows pairwise correlation of metabolites post-treatment. Module assignments are identified by black squares and by the colour bar (salmon, tan, yellow and brown). Correlation heatmap bar represented (range -1.00 (blue) to +1.00 (red)).

Validation of findings through re-sequencing in a subset of patients

I performed a separate differential expression analysis in a subset of 9 patients from the overall study whose samples had been sequenced in duplicate at two separate facilities (Imperial BRC Genomics Facility and Genewiz, NJ, USA)) using the same Illumina sequencing machine model (Illumina Hi-Seq Platform, 2 x 150bp PE configuration) but different alignment software (Tophat2(247) vs DRAGEN(159) respectively). For these 9 patients, data derived from Genewiz yielded 595 differentially expressed genes post-treatment ($p < 0.05$). In the same subset of patients, sequenced at a different facility (BRC) but undergoing the same differential expression analysis using DESEQ2 software, 498 genes were differentially expressed post treatment ($p < 0.05$). Of these 498 genes found from BRC data, 217 (44%) were differentially expressed in the Genewiz dataset. To distinguish differences related to sequencing runs from those related to the use of different alignment software, I repeated the alignment using TopHat2 with the raw sequence data from both the BRC and Genewiz. I then performed identical count derivation using HTSeq. In this analysis, there was 498 and 523 DE genes respectively representing 54% overlap of statistically significant DE genes. Of the 10 most differentially expressed genes, 8 overlapped. I investigated whether those genes that did not overlap had lower average counts than those which were differentially expressed consistently across both sequencing facilities. For those overlapping, mean counts were 804 +/- 619 genes, whilst for non-overlapping genes mean counts were 278 +/- 208. The distribution of mean counts between overlapping and non-overlapping genes were significantly different ($p = 0.01$).

Effect of DMF on small-RNA species

I also performed a differential expression analysis on small RNA count data using software package DESEQ2(205). Controlling for intra-individual variability, 17 small RNAs were differentially expressed post-treatment. A similar comparison in the control group at the same timepoints yielded no differentially expressed small RNA. Expression of 8 of the small RNAs was down-regulated by the treatment (log_2fold change -0.21 to – 0.4) and 9 were upregulated by treatment (0.18 to 0.31) (Table 9).

Gene	Log2Fold Change	Padj
mir-31	0.31	0.002
mir-769	0.28	0.001
mir-192	0.26	0.014
mir-342	0.24	0.001
mir-146b	0.22	0.05
let-7i	0.18	0.02
mir-140	0.18	0.001
mir-361	0.18	0.006
mir-339	0.16	0.01
mir-130a	-0.22	0.05
mir-628	-0.24	0.03
mir-654	-0.26	0.006
mir-493	-0.26	0.016
mir-6810	-0.29	0.025
mir-1908	-0.30	0.01
mir-3908	-0.30	0.01
mir-410	-0.41	0.0003

Table 9: Differentially expressed small RNAs 6-weeks post treatment with DMF in RRMS patients (padj < 0.05 – corrected using Benjamini Hochberg). Changes in expression expressed as log_2fold change.

A shift towards a Th2 cytokine profile after 6-weeks of DMF treatment

Given the difference in cytokines observed when comparing untreated RRMS patients to healthy controls, I also analysed differences in cytokines as a response to DMF.

Increases in plasma IL-4 (36%) [0.05 pg/ml (pre), 0.07 pg/ml (post)] and IL-13 (21%) [1.5 pg/ml (pre), 1.8 pg/ml (post)] cytokine concentrations were found after treatment with DMF (p = 0.03 and p = 0.001 respectively) (Fig. 18). Similar findings were not found in the control cohort.

The influence of DMF on Th1 and Th17 cytokines was investigated. There was no significant change pre- and post- treatment in Th1 or Th17 cytokines (Table 10). In my previous chapter, I identified altered levels of IL-2 in MS patients relative to healthy controls. I was unable to show a treatment-mediated change in IL-2 6-weeks post-treatment with DMF.

Th17 Cytokines	Pre-drug concn (pg/ml)	Post-drug concn (pg/ml)	p-value
IL-17	3.75 (1.63 – 6.64)	3.92 (0.15 – 7.11)	0.06
GM-CSF	0.23 (0.06 – 0.72)	0.22 (0.02 – 0.84)	0.46
IL-22	0.24 (0.03 – 3.3)	0.2 (0.01 – 2.98)	0.31
Th1 Cytokines	Pre-drug concn (pg/ml)	Post-drug concn (pg/ml)	p-value
IFN-γ	13.6 (4.3 – 22.6)	12.3 (9.7 – 19.6)	0.21

Table 10. Median concentration of Th17 and Th1 cytokines pre- and post- DMF. Values expressed as median (range). P-values as defined.
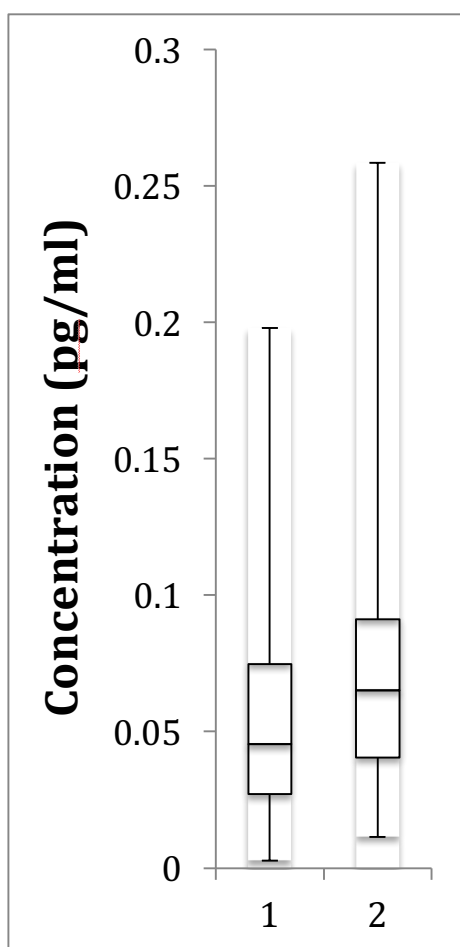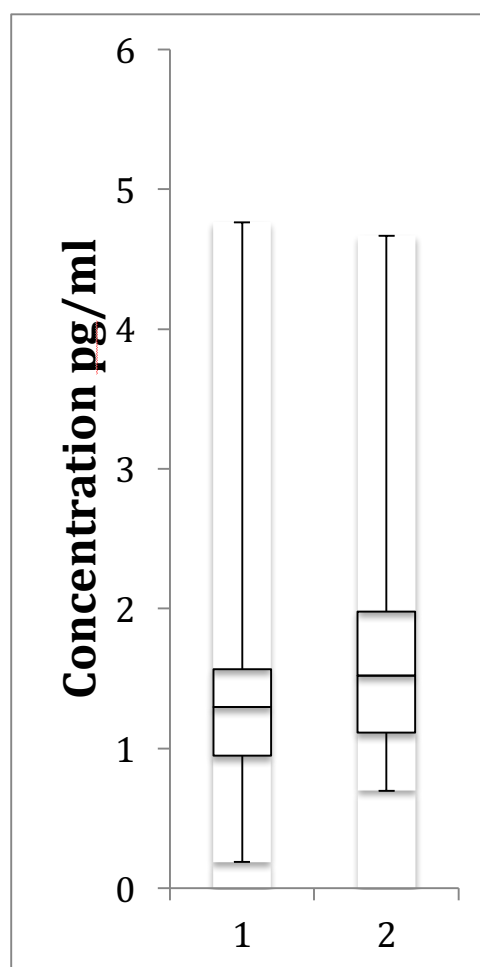
Fig 18: Box plots pre (1) and post (2) treatment for cytokines IL-4 (left) and IL-13 (right). Differences pre- and post- treatment were significant (p < 0.05).

Plasma metabolites change after DMF treatment

A sPLS-DA analysis was performed on the pre- and post- treatment discovery datasets. The estimated error rates stabilised after four dimensions for any number of selected variables. The R^2 values for these 4 components were 0.1, 0.36, 0.42 and 0.44, respectively (Fig 19).
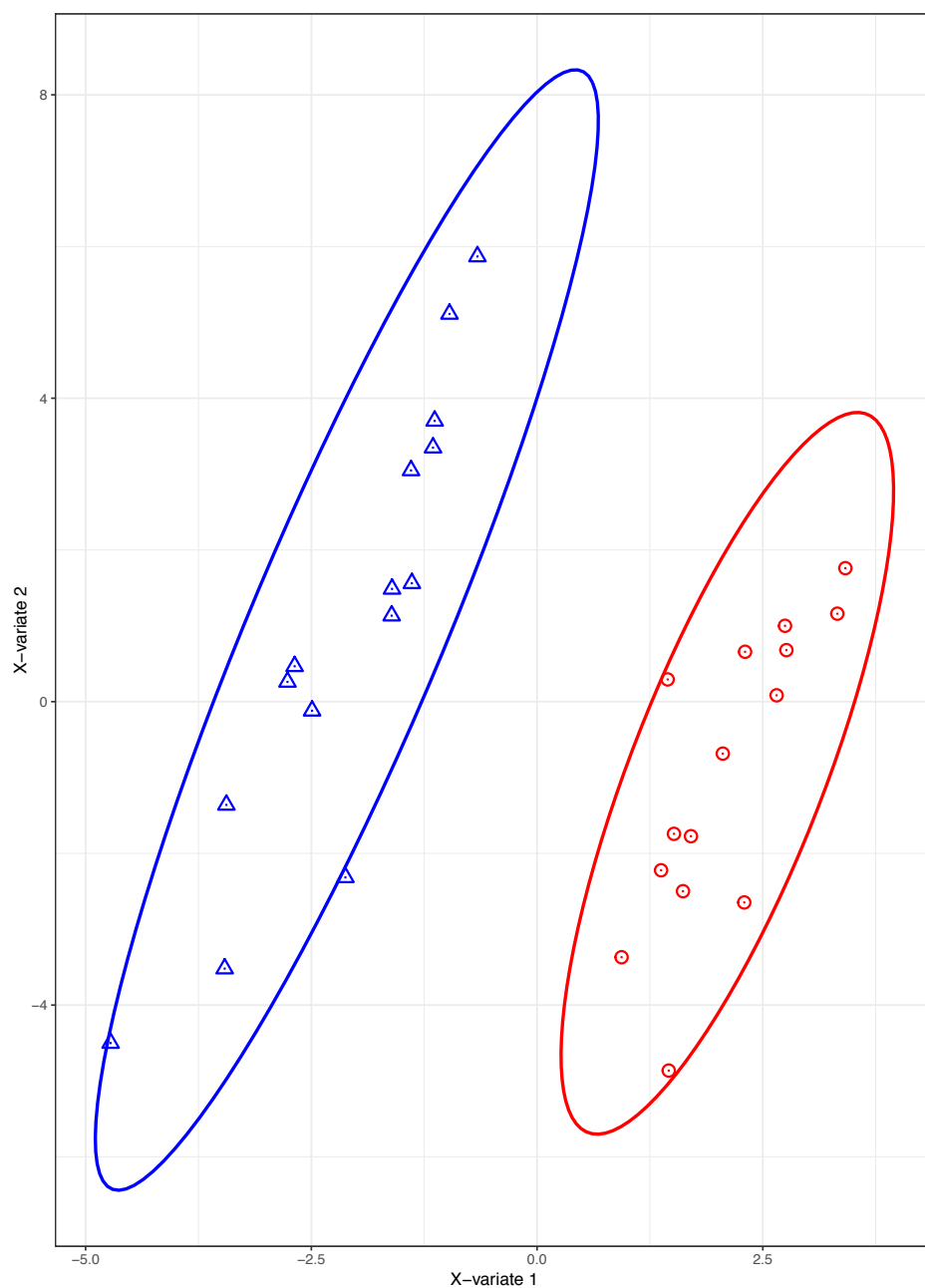
Fig. 19. Pre- and post- treatment samples can be accurately distinguished using supervised sPLS-DA analysis. Sample representation using the first 2 latent variables from sPLS-DA (50 metabolites selected). Pre-treatment samples (blue triangles) and post-treatment samples (red circles) are displayed. First latent variable represented on horizontal axis. Second latent variable represented on vertical axis.

Modules of differentially expressed metabolites after treatment with DMF

I identified 5 differentially co-expressed modules comprising a total of 924 metabolites. 4 of these were statistically significant ($p < 0.001$ for orange, red, green and blue modules). 2 of these modules were significantly more highly correlated after treatment (orange and green, which included 114 and 38 metabolites, respectively) and 2 were less strongly correlated after treatment (blue and red, which included 48 and 29 metabolites, respectively) (Fig 20). A high proportion of the orange module was comprised of carnitine species (13%) and phosphatidycholines (11%), lysophosphatidylcholines (18%) and lysophosphatidylethanolamines (20%) were over-represented in the green module.
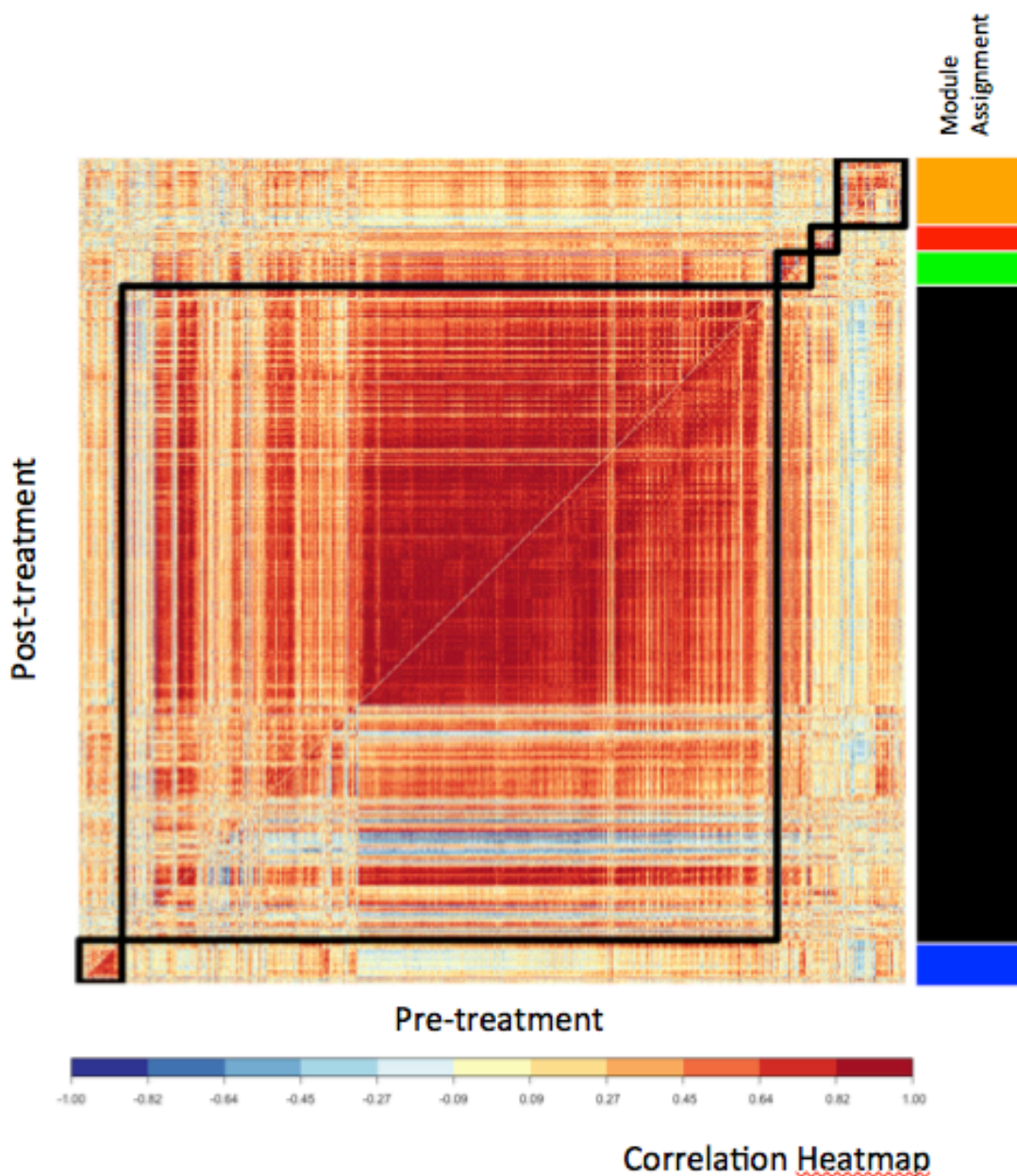
Fig. 20. Comparative correlation heatmap demonstrating differentially coexpressed modules pre- and post- treatment with DMF. The lower diagonal shows pairwise correlation of metabolites pre-treatment. The upper diagonal shows pairwise correlation of metabolites post-treatment. Modules are identified by black squares and by the Module Assignment bar (orange, red, green, black and blue). Correlation displayed using correlation heatmap bar (range -1.00 (blue) to +1.00 (red)).

Exploratory Analyses

Tricarboxylic Acid intermediates are increased by DMF

As DMF is ultimately metabolised to fumarate, I tested whether TCA cycle intermediates were altered with treatment. Both fumarate and succinate were significantly increased 6 weeks after the start of treatment. Furthermore, succinyl-carnitine and methyl succinyl-carnitine, which can be derived from succinyl-CoA, also were increased significantly ($q < 0.05$) (Table 11) (Fig 21 a-d). Comparison of baseline and 6-week samples in controls did not show similar changes. Of all metabolites assayed, a random forest analysis confirmed that methyl succinyl-carnitine, fumarate and succinyl-carnitine were the most discriminatory variables in my dataset (Table 11) (Fig 22).

Metabolite	Fold Change	p-value	q-value	Mean Decrease Accuracy
Fumarate	1.58	0.00004	0.00500	8.6
Succinate	1.18	0.00300	0.04060	2.7
Succinyl-Carnitine	1.74	0.00120	0.02670	6.0
Methyl Succinyl-Carnitine	39.24	0.000001	0.000001	10.2

Table 11: TCA metabolites significantly increased in MS patients post-treatment with DMF. Fold change, P-values and Q values (corrected using false discovery rate) provided. Mean Decrease Accuracy values from Random Forest also provided.

(a)

(b)

(c)

(d)
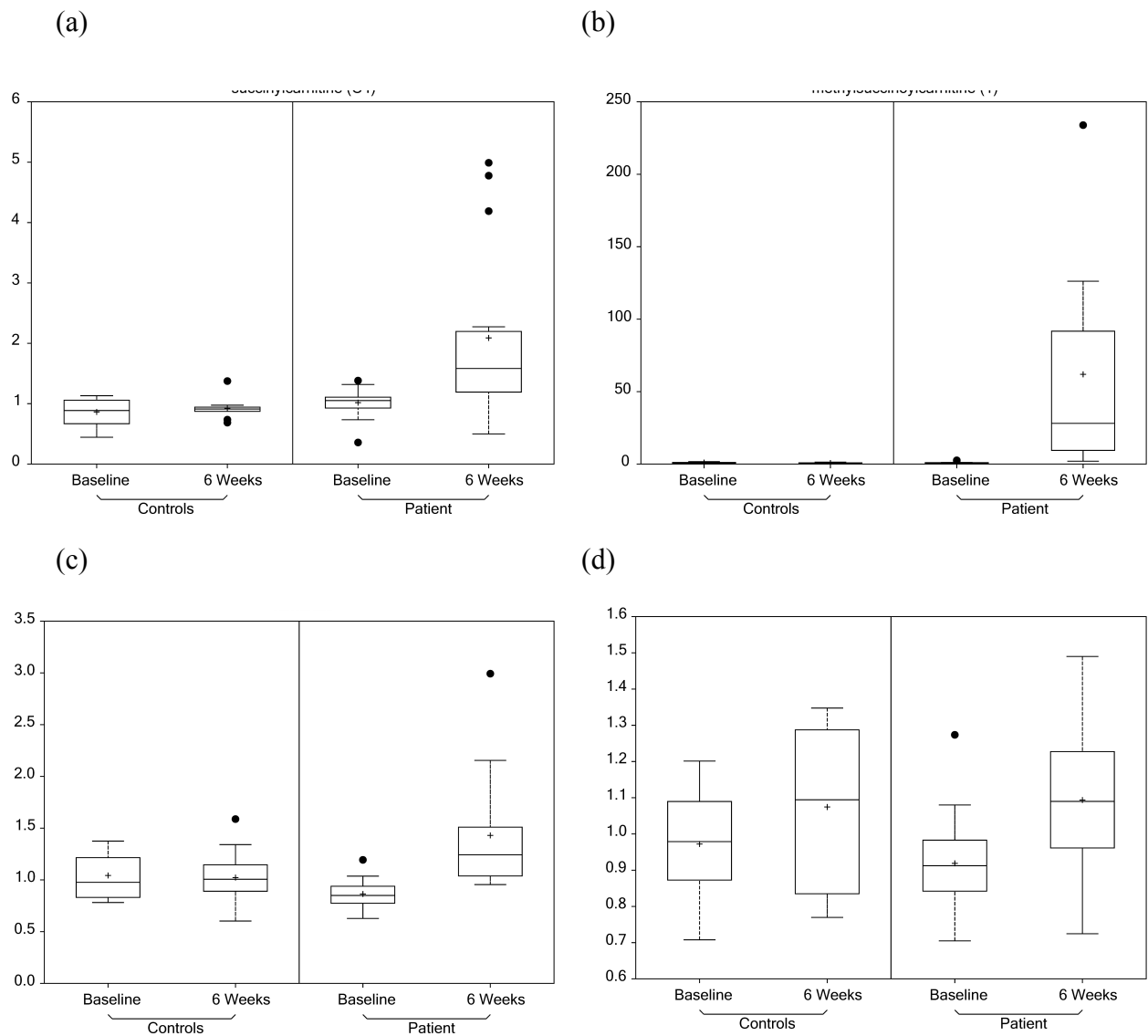
Figure 21(a-d). TCA metabolites are significantly increased 6-weeks post treatment with DMF. Boxplots of TCA metabolites (a: succinyl-carnitine, b: methyl succinyl-carnitine, c: fumarate, d: succinate) at baseline and 6 weeks in controls (untreated) and patients (treated). Metabolites quantified by raw counts. All 4 metabolites statistically increased in the patient cohort after correction for multiple comparison. Dots represent outlier values.
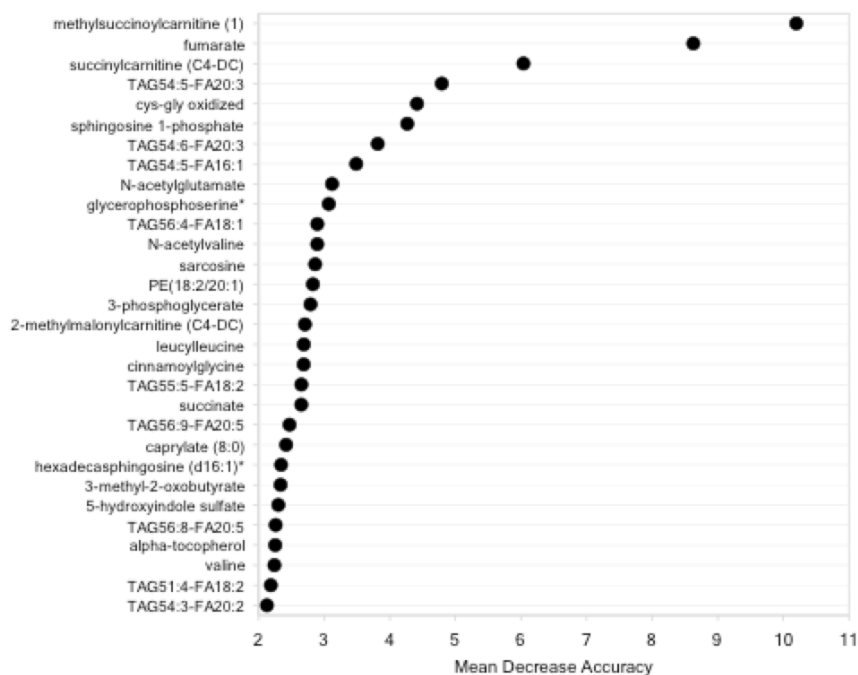
Figure 22: Pre- and post- treatment metabolites highly discriminatory for treatment state. **Random Forest plot displaying 30 most discriminatory variables distinguishing pre- and post-treatment samples and corresponding mean decrease accuracy values**.

DMF effects on anti-oxidant metabolites

I investigated whether DMF had any effect on metabolites of glutathione, a peptide associated with antioxidant pathways (38). Glutathione was not detected in this study, but its metabolites glycine and 5-oxoproline were significantly increased by 6 weeks after initiation of treatment with DMF (q < 0.05) (Table 12). However, concentrations of cysteinylglycine and cysteine (also metabolites of glutathione) in plasma were not significantly altered.

Metabolite	Fold Change	p-value	q-value
Cysteinylglycine	1.25	0.01	0.07
Cysteine	0.94	0.31	0.34
Glycine	1.19	0.00002	0.005
5-oxoproline	1.18	0.002	0.033

Table 12. Glutathione metabolites in MS patients post-treatment with DMF. Fold change, P-values and Q values (corrected using false discovery rate) provided.

Validation of methyl Succinyl-Carnitine as a marker discriminating pre- from post-treatment samples

Given the over-representation of carnitines in the green co-expression module, the highly significant increases of succinyl carnitine and methyl succinyl carnitine after treatment and the results of the exploratory random forest classification, I sought to confirm their association with the pharmacodynamic response in a separate validation cohort. The most significantly increased metabolite in the validation cohort was also methyl succinyl carnitine (mass/charge ratio 4.71_276.1448), which changed by 145-fold ($p < 0.005$). I was also able to validate a rise in glycine post treatment (fold change 1.40 ($p < 0.005$) but findings for 5-oxoproline were non-significant (fold change 0.98, $p > 0.05$). I was unable to identify a peak for succinyl carnitine using the different mass spectroscopy platform employed for the validation dataset.

Levels of MMF correlate with levels of methyl succinyl-carnitine

I was able to measure plasma concentrations of MMF; the most abundant, pharmacologically active metabolite of DMF. Samples seemed to either show low concentrations of MMF (<20 ng/mL) or high concentrations of MMF (range 95 – 592 ng/mL) (Table 13). Pearson's correlation showed a strong relationship between levels of methyl succinyl-carnitine and MMF ($r = 0.66$) (Fig 23). Levels of the active drug did not correlate with reported timing of last dose ($r = -0.19$) suggesting either a previously uncharacterised extreme pharmacokinetic variation of the medication or poor adherence to medication.

Subject Number	MMF Concentration (ng/mL)	Reported Timing of last dose (hours)
1	<5	6
2	7.3	2
3	11	2
4	<5	4
5	<5	1
6	26.4	3
7	<5	3
8	<5	2
9	<5	13
10	<5	4
11	<5	15
12	<5	3
13	11.9	3
14	<5	15
15	<5	1
16	95	4
17	486.2	2
18	105.3	4
19	243.2	4
20	432.2	4
21	165.9	3
22	574.7	8
23	358.3	1
24	154.3	3
25	591.5	3
26	242.5	6
27	386	4

Table 13. Concentrations of MMF in RRMS subjects 6-weeks post treatment with DMF. Values expresses as ng/mL. Timing of reported dose also stated in hours.
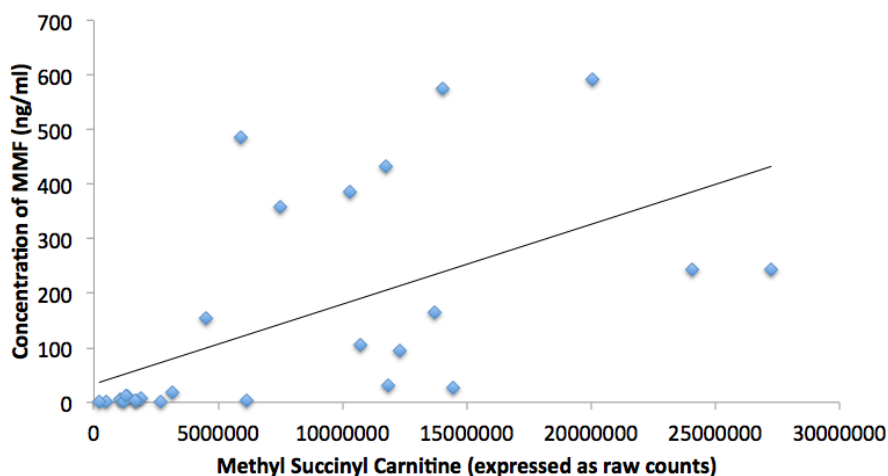
Fig 23. MMF is highly correlated with methyl succinyl-carnitine. Correlation plot of MMF concentration (ng/ml) and raw counts of methyl succinyl-carnitine.

Levels of MMF correlate with levels of 1-methylnicotinamide

I undertook an unbiased correlation analysis to investigate which metabolites correlated most strongly with MMF concentrations in my discovery cohort. The most highly correlated metabolite was 1-methylnicotinamide (r = 0.76) (Fig 24a). Because I undertook full metabolomic profiling on all samples at the National Phenome Centre (including technical replicates) I was able to cross-validate this finding in the discovery cohort based on the equivalent feature in that dataset (m/z ratio 137.0710, retention time 4.05). In this analysis, I found an even stronger correlation between drug concentration and levels of 1-methylnicotinamide (r = 0.85) (Fig 24b). In order to validate this finding, I repeated the correlation analysis in the validation cohort. The correlation between MMF levels and 1-methylnicotinamide were maintained (r = 0.68).

High levels of 1-methylnicotinamide are associated with corresponding reductions in levels of the methyl-donor betaine(249). In my cohort, there was a strong inverse correlation between betaine and both MMF and 1-methylnicotinamide (r = -0.6 and -0.77 respectively).
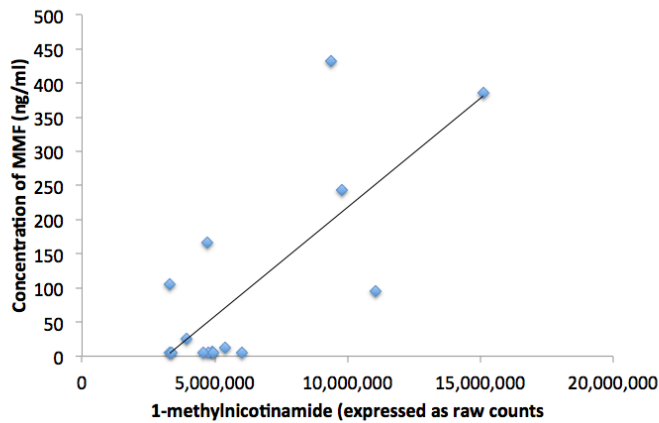
Fig 24a. MMF is highly correlated with 1-methylnicotinamide. Correlation plot of MMF concentrations (ng/mL) and raw counts of 1-methylnicotinamide.
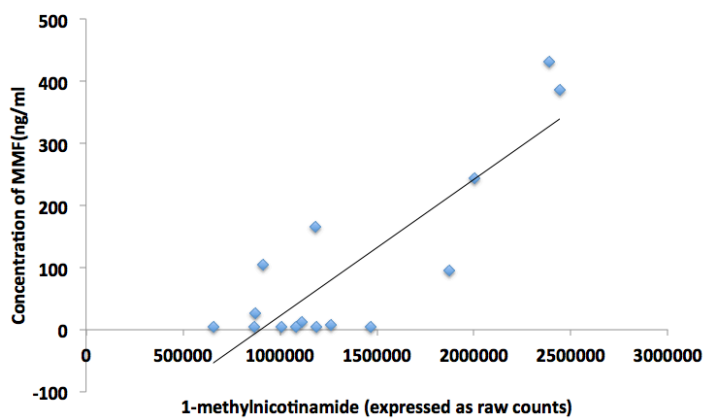
**Fig 24b. MMF is highly correlated with 1-methylnicotinamide. Correlation plot of MMF concentrations (ng/mL) and raw counts of 1-methylnicotinamide (technical replicate).**

Effect of DMF on lipoprotein sub-fractions

In chapter 3, I reported an increase in VLDL lipoprotein subfractions in MS patients. I sought to evaluate whether 6 weeks of DMF would affect the levels of lipoprotein subfractions in my patient cohort. Concentrations of VLDL subfractions were not affected by 6 weeks of DMF treatment. However, I was able to identify 5 LDL species (L2-Cholesterol, L2-phospholipid, L2-apolipoprotein, L2PN, L2-FreeCholesterol) that were all significantly reduced 6 weeks post treatment with DMF ($p = 0.02$, $p= 0.03$, $p = 0.03$, $p=0.04$, $p=0.04$ respectively), corrected for multiple comparisons.

Integration of transcriptome, small RNA and metabolomics suggest a role for IL-4 in mediating pharmacodynamic effects of DMF.

I combined RNA-Seq and metabolomic data pre- and post- treatment in patients using an integrated pathway analysis tool from Ingenuity Pathway Analysis software (Qiagen, Hilden, Germany). In my previous chapter, I identified TNF signalling as an over-represented and over-activated pathway in patients when compared to my healthy volunteer population. I was able to see a down-regulation of this pathway following treatment. The IL-4 pathway was significantly over-represented 6-weeks post treatment with DMF (Fig 25). This confirmed my previous result showing an increase in IL-4 cytokine concentration following onset of treatment.

(a)
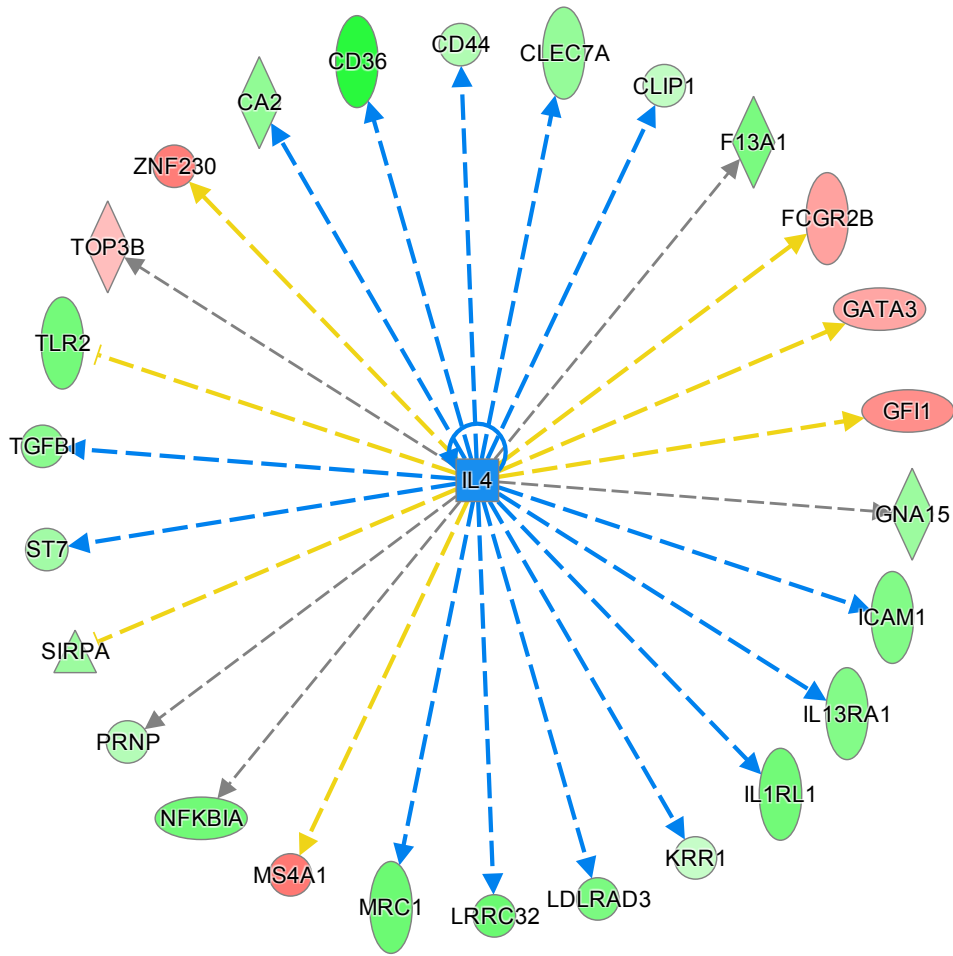
(b)
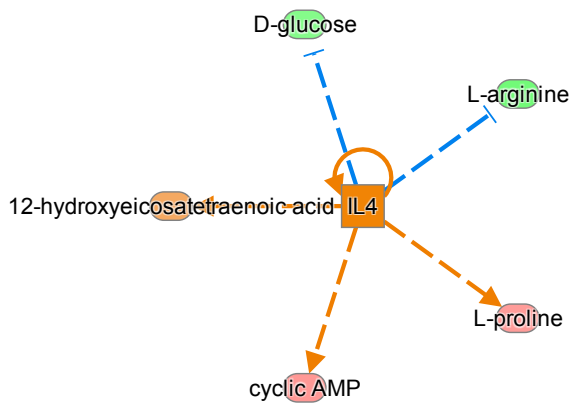
Fig 25a and 25b. IL-4 associated transcripts (a) and metabolites (b) differentially expressed post-treatment with DMF.

Results Summary

- **Data reduction techniques could accurately separate MS metabolomic samples based on treatment state (pre- and post-treatment). Differential coexpression identified modules of metabolites that showed high intra-modular correlation.**

- **TCA cycle intermediates were elevated after treatment with DMF. Methyl-succinylcarnitine was the most discriminatory metabolite as determined by basic statistical analyses and data reduction methods.**

- **The increase in methyl-succinylcarnitine was replicated in the validation cohort and showed strong correlation with concentrations of the active component of DMF; monomethyl fumarate. The most highly correlated metabolite with MMF was 1-methylnicotinamide. This metabolite was correlated with common adverse effects associated with DMF.**

- **DMF caused an increase in Th2 cytokines IL-4 and IL-13. Using integrated pathway analysis approached, I confirmed a role for IL-4 in the short-term pharmacodynamic effect of DMF in metabolomics and transcriptomic datasets.**

- **Short term administration of DMF caused a robust alteration in the gene expression signature of MS patients, however the effect sizes were modest. Differential coexpression analysis identified an interesting module of differentially expressed genes that were enriched for immune system processes. The hub gene for this module was nicastrin.**

Discussion

The primary mechanism of action of DMF is not well understood. Here I investigated whether an integrated 'omics approach (transcriptomics, metabolomics, cytokine analysis) can further elucidate pharmacodynamic effects. Using NGS, I was able to demonstrate that just over 500 genes appeared to be differentially expressed after short-term administration of DMF and that genes encoding proteins involved in the Nrf2 activation pathway were over-represented. Other gene expression studies have identified genes related to anti-oxidant pathways(242, 250-252), anti-inflammatory pathways(53, 253) and NFκB(52, 254) that may be related to DMF's therapeutic effect. My results confirm in-vivo that Nrf2 related gene expression may be implicated in the effect of DMF. Furthermore, I demonstrated that genes altered by DMF could be partitioned into differential co-expression modules enriched for immune response protein transcripts.

I investigated the reproducibility of differential expression estimates to see if they were consistent across different alignment software and sequencing runs at different facilities. Consistency improved when alignment software was comparable leaving sequencing facilities as the only source of discrepancy. Sequencing methods are imperfect and there are a number of potential sources of experimental error that can lead to thousands of false positive variants in a fully sequenced human transcriptome(255). Potential sources of discrepancy could have occurred at the library preparation or sequencing stage. With regards to the former, this may have been at the PCR amplification stage(256) which is subject to unmeasured biases or the use of barcodes(257, 258) which can result in decreased read quality. With respect to sequencing, Illumina is known to have a sequence specific error profile where

inverted repeats exists or there are repetitive GGC sequences(259). I also explained part of the discrepancy on the basis that those genes which did not overlap were likely to have fewer mean counts and therefore were less likely to show sufficient variance to make differences statistically significant.

Using mass spectrometry, I observed that short-term administration of DMF could only accurately separate samples using supervised classification techniques such as sPLS-DA. While some of the changes detected post-treatment could be related to anti-oxidant pathways, the greatest changes (as measured by Random Forest) were seen in TCA cycle metabolites fumarate, succinyl-carnitine and methyl succinyl-carnitine. The finding of elevated methyl succinyl-carnitine was validated in samples from a separate group of patients using a different mass spectrometry platform.

One prior study has investigated the effect of DMF on oligodendrocytes using a metabolomics approach(41). This report also identified increases in succinate and fumarate with treatment, but studied much shorter time intervals (24 and 72 hours). Increases in carnitine were observed but not succinyl-carnitines. This difference is likely due to succinylation of carnitines in peripheral organs such as the liver, which would not have occurred in an in-vitro experiment in isolated oligodendrocytes(260). My results confirm an increase in TCA cycle intermediates *in vivo* 6 weeks after initiation of treatment with DMF. Furthermore, I identify two novel carnitine species (succinyl-carnitine and methyl succinyl-carnitine) that are significantly increased post-treatment.

The isolated increase in succinate without increases in intermediates involved early in the forward direction of the cycle (malate, citrate and α-ketoglutarate) led us to hypothesise that fumarate generated from metabolism of DMF(261) may be reduced by succinate dehydrogenase to succinate (resulting in corresponding oxidation of NADH to NAD+). High levels of succinate and hence succinyl-CoA can result in raised levels of succinylation(262, 263) of carnitines and other species, in turn (264, 265). For example, patients with mutations in succinate-CoA ligase, which normally catalyses conversion of succinyl-CoA to succinate, characteristically show increased concentrations of both succinyl-CoA and succinylcarnitine(265-268).

L-carnitine and acetyl-l-carnitine have both been demonstrated to activate anti-oxidant pathways mediated by Nrf2(269-271). In a rat model of interstitial nephropathy, reduction in two acyl-carnitines (stearoyl-carnitine and palmitoyl-carnitine) resulted in impaired Nrf2 pathways and activation of NF-κB; the latter known to be inhibited by DMF(272). Carnitine species can increase fatty acid flux through acyltransferase and can accumulate to accept acyl groups from CoA in β-oxidation(273-275). This represents an alternative pathway that would produce a similar intra-cellular effect as agonism of the HCA2 receptor during ketogenesis, the latter being an alternative proposed mechanism of action of DMF(276, 277).

My finding of a high correlation between MMF and 1-methylnicotinamide was replicated using both technical replicates and in a separate validation cohort Furthermore, I was able to demonstrate a corresponding inverse correlation between 1-methylnicotinamide and betaine which has already been identified(249). The increase in NAD+ through reversal of succinate dehydrogenase (outlined above)

would result in raised 1-methylnicotinamide by the enzyme nicotinamide n-methyltransferase(278). I hypothesise that this is the mechanism by which I observed a high correlation between MMF and 1-methylnicotinamide. Furthermore, I hypothesise that 1-methylnicotinamide may play a role in the mechanism of action of DMF. 1-methylnicotinamide has been widely explored for its anti-inflammatory(279), anti-oxidant(280) and neuroprotective properties(281). However, its potential therapeutic effects for MS have yet to be explored. 1-methylnicotinamide has already been explored in other murine models of human disease(282) and its pharmacokinetic profile has been well characterised(283).

 With respect to cytokine changes following short-term administration of DMF, I found an increase in Th2 cytokines IL-4 and IL-13 but no corresponding dynamic changes in Th1 or Th17 related cytokines. Pathway analyses of both transcripts and metabolites suggest a role for IL-4 in the action of DMF. My finding of increases in IL-4 is consistent with a recent study in PBMCs of patients with psoriasis treated with DMF(253), which was also able to find alterations in Th1 and Th17 cytokines. The difference in results may arise from use of cell subsets (PBMCs) in the latter study, rather than measurements directly from plasma.

In conclusion, I have provided evidence for a robust transcriptomic response to DMF in the short-term which involves up regulation of Nrf2 and enrichment of anti-inflammatory and anti-oxidant pathways. Furthermore, I have provided evidence that TCA cycle intermediates are significantly modulated in patients receiving DMF. The most significantly altered metabolites are derivatives of TCA cycle intermediates: succinyl-carnitine and methyl succinyl-carnitine. The levels of methyl succinyl-

carnitine and 1-methylnicotinamide correlate well with concentrations of the active drug MMF in sample plasma. Despite previous evidence for the anti-inflammatory effects of 1-methylnicotinamide, this is the first evidence for a link between this molecule and DMF. I hypothesise that 1-methylnicotinamide and methyl succinyl-carnitine may be the downstream effectors of DMF's known mechanisms of action. In future work, this could be tested directly in vitro or in pre-clinical models.

Chapter 5. Identifying markers associated with response to DMF and its adverse events.

Introduction

Accurately stratifying RRMS patients to the best treatment based on their personal characteristics is a therapeutic ideal as it would avoid the use of expensive and potentially harmful medications that may yield little benefit and will not prevent accrual of disability(57). While an optimal stratification might occur before dosing, an alternative would be to assess response only a short time after initiation of treatment. The main criticism of current response measures (both clinical and radiological) is that they demand longer periods of observation (on the order of a year or more, typically). There thus is a strong rationale for searching for short-term biomarkers associated with longer-term responses to treatment.

The current major approach to ensuring patients are receiving the most efficacious medication for them involves disease activity monitoring over 1-2 years. A recent approach to this has been the introduction of the composite outcome measure termed 'no evidence of disease activity' (NEDA) which is increasingly being evaluated for measuring treatment response(284-286). NEDA evolved from the term 'freedom from disease activity' and first appeared in the MS literature in a 2009 *post hoc* analysis of natalizumab(284). The most recent definition; 'NEDA-4', combines four measures of disease activity to increase sensitivity: no relapses, no EDSS progression, no new MRI activity (Gd+ lesions or new/newly enlarging T2 lesions) and no worsening brain volume loss(287, 288). The most practical limitation of this approach is the time it takes for such measures of disease activity to appear. As such, there is a clinically

unmet need for identifying short-term predictive markers of these longer-term outcome measures.

A recent post-hoc analysis revealed that only 26% of patients in the DMF DEFINE/CONFIRM trials achieved NEDA at two years(289). This suggests that there are definable responder and non-responder populations. The hypothesised mechanism of action involving activity on transcription factor-encoding genes nuclear factor (erythroid-derived 2)-like 2 (*Nrf2*)(37, 38) and nuclear factor κB (NFκB)(39) within immune cells further suggested to us that the pharmacodynamic effects on the transcriptome of PBMC could be used to predict treatment response(290). To test this, I used next-generation RNA sequencing and plasma metabolomic profiling to identify short-term changes in gene expression associated with medium-term treatment response defined by the composite outcome measure NEDA-4 at 15 months post-treatment.

Methods

Patients attended for three study visits; at baseline (pre-drug), at 6-weeks after commencing treatment, and at 15 months post-treatment. Full details of assessments are outlined in the Methods section.

Outcome Measures

MRI Scans

The RRMS patients underwent MRI scanning at the Imperial College Clinical Imaging Facility at six weeks and 15 months after the start of treatment (Siemens Verio 3T; 32-channel head coil; T1- and T2-weighted structural scans). Scans were analysed using MSmetrix, a scanner independent software developed by Icometrix to extract whole brain atrophy, lesion volume changes and the number of new lesions between two timepoints(291, 292). The longitudinal analysis performed using MSmetrix incorporates both spatial and temporal information for accurate and consistent lesion segmentation based on Markov Random Field modelling and difference imaging across the two time points (293).

Clinical Outcomes

Clinical outcomes of patients were assessed by a single trained physician (AG). This included detailed patient histories to establish whether any new clinical relapses had occurred during the study period, a full EDSS assessment, MS Functional Composite scoring (MSFC) and SF-36 quality of life questionnaire.

The main outcome measure used to designate patients as responders or non-responders was the 15-month NEDA-4 outcome. NEDA-4 was defined as no

evidence of relapses, active MRI lesions (both new or enlarged T2 lesions), 6-month confirmed disability progression (CDP) (defined as an increase in EDSS score of 1.5 points from a baseline score of 0, of 1.0 point from a baseline score of 1.0 or more or 0.5 points from baseline greater 5.0) or mean annualised rate of brain volume loss (AR-BVL) of more than 0.4%.

The secondary outcomes were MSFC Z-score comparisons between baseline and 15 months. The MSFC is a multidimensional outcome tool that comprises quantitative tests of 3 neurological domains; upper limb coordination (9-Hole Peg Test), ambulation (25-Foot Walking test), and cognitive ability (Paced Auditory Serial Addition Test-4). The raw scores for these components are standardized as Z scores (SD units from mean baseline score in the population) with an increase in MSFC scores indicating improvement(294). Z-score comparisons pre- and 18 months post-treatment were calculated using a paired student's T-test. Disease progression was defined as a greater than 20% increase in 9-Hole Peg Test or 25-foot walk compared between baseline and the 15-month study visit using either of these measures(295).

Patients were also assessed using a quality of life questionnaire; SF36(296). The SF36 is scored in 8 domains of QOL and simplified into two scores, a physical summary score (PCS) and a mental component score (MCS). These are expressed as an adjusted score range between 0 and 100 (0 being most impaired and 100 being least impaired with regards to QOL). Comparisons between baseline and 15-months were performed using a paired student's T-test.

Statistical Analysis

Treatment 'responder' or 'non-responder' groups were analysed separately with respect to identification of treatment response biomarkers. Significant of change in MSFC was assessed using Student's T-test.

Transcriptomics

A series of differential expression analyses were performed on count data derived from HT-Seq software using DESEq2(205) in parallel in the 'responder' and 'non-responder' groups. A full description of Deseq2 can be found in Chapters 3&4. Here, the statistical tests are described in the context of treatment responders, however the same analyses were also performed in the non-responder group.

Time-course comparisons were performed between baseline and six week samples and then between six weeks and 15-month samples controlling for intra-individual variation. Cross-sectional contrasts between patients and controls were performed controlling for age as a covariate. Adjusted p-value for significance (Padj) was set at (Padj < 0.05). The threshold for fold change cut offs was a log2-fold change of ±0.3.

Treatment response genes were defined as those that were significantly different in the responder vs control group at baseline but that were no longer differentially expressed in the responder group at 6 weeks and 18 months; i.e. these genes had 'normalised' in the responder population. Additionally, these genes had to be continuously different in the non-responder group at all timepoints.

A further set of treatment response genes were defined as those that were differentially expressed in the responder group at both post-treatment timepoints and that were also different in the treatment responder group vs controls at baseline but not at 6 weeks or 18 months; i.e. genes that were persistently altered by treatment and were also different in patients compared to controls at baseline but not following treatment.

Downstream pathway analysis was performed using the Ingenuity Pathway Analysis (IPA) software (Qiagen, Helden, Germany). Genes of interest imputed into IPA had Padj < 0.05 and log2-fold change of ±0.3. Enrichment analysis was performed using g:Profiler isolated to KEGG pathways (p-value < 0.0005).

Comparison between atrophy rates in non-responder groups was calculated using Fisher's exact test (p < 0.05). Permutation analysis was performed using DESEq2 (100-fold) with random selection of RRMS subjects (n = 8).

Contrasts were also made in the small RNA dataset. Treatment response small RNAs were defined as those that were different in the responder vs control group at baseline but not at 6 weeks or 18 months (i.e. normalised) but remained different in the non-responder group at these timepoints.

In order to investigate any possible interactions between the mRNAs and small RNAs of interest I identified target mRNAs for the mi-RNAs of interest using TargetScan v7.0. All of the target genes with conserved sites were selected. I then identified genes

within this target set that were previously defined as treatment response genes in my preliminary analysis.

Treatment responders and non-responders were also compared with reference to known MS risk genes. There are currently 200 known genes with variants that increase the risk of MS (IMSGC, unpublished data). I also investigated differential expression of MS disease risk variants in the subset of genes where there was adequate coverage in the RNA-Seq data to accurately identify these variants. Variants were derived using GATK software(297) and differential expression was calculated using a paired students T-test ($p < 0.01$).

Metabolomics

In order to evaluate whether the gene transcripts or microRNAs of interest were associated with metabolites, I undertook a pairwise correlation analysis for each of the transcripts and microRNAs of interest and all metabolites within both the responder and non-responder groups using Pearson's correlation coefficient. Metabolites with high correlations ($-0.7 > r$ or $r > 0.7$) were treated as significant. Metabolite enrichment analysis was performed using Metabolite Set Enrichment Analysis (MSEA), an open source tool to identify pathways enriched for a given set of metabolites(298).

Results

DMF is an effective DMT in a subset of RRMS patients

8/24 patients (33%) achieved NEDA-4 over the 15 month period after initiating treatment with DMF. 12 patients (50%) had an annualised brain volume loss (AR-BVL) greater than -0.4% (range, -0.44% to -2.19%). Enlarging or new lesions occurred in 9 patients (38% and 4 of these had an AR-BVL < -0.4%). 3 patients experienced relapses and 6-month confirmed disability progression occurred in 4 patients (2/4 of whom also experienced relapses).

The mean change in MSFC from baseline to 15 months for the whole cohort was +0.27 (range, -0.27 to 1.33) (p < 0.005). The SF-36 physical summary score increased from 58.4 ± 25.8 to 62.8 ± 25.0 and the mental component score increased from 53.6 ± 18.9 to 59.9 ± 18.0, but these changes were not statistically significant (p =0.24 and p = 0.1 respectively).

Short-term dynamic effects of DMF

I assessed pharmacodynamic effects independently in the clinical responder and non-responder groups. In the responder group, there were 478 differentially expressed genes 6 weeks after the start of treatment with DMF relative to baseline (padj < 0.05). These differences showed enrichment of transcripts related to the Nrf2 pathway (p < 0.0005) (Fig 26) and increased expression of those associated with down-regulation of NFκB associated responses (overlap p < 0.0005) (Fig 27). In the non- responder group, no differentially expressed genes were identified 6 weeks after the start of treatment relative to baseline (Table 14).
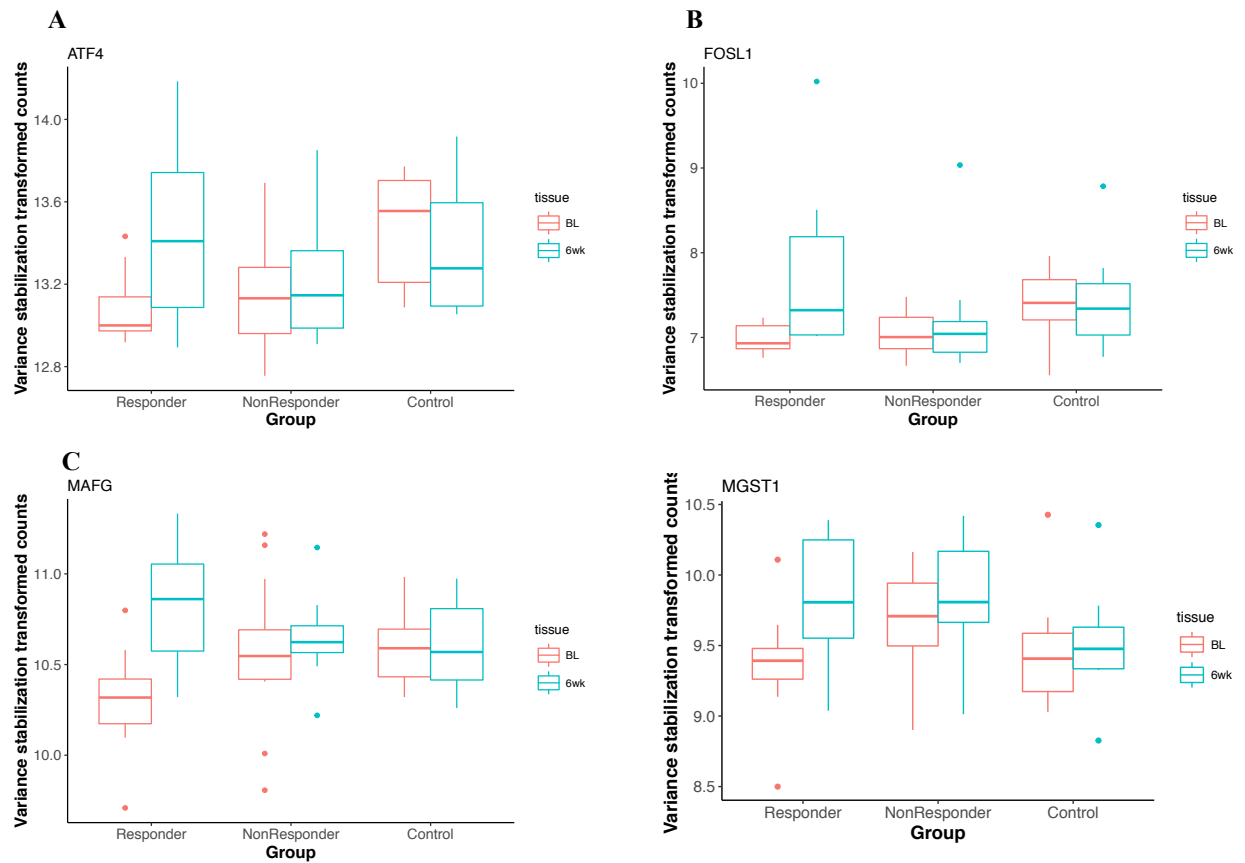
Fig 26 (a-d). Nrf2 related transcripts are increased 6 weeks post treatment in responders but not in non-responders or healthy controls. Boxplots represent variance stabilised transformed counts for transcripts (A) ATF4, (B) FOSL1, (C) MAFG, (D) MGST1 at baseline and 6 weeks in responders, non-responders and healthy controls. FOSL1 = fos-related antigen 1; ATF4 = activating transcription factor 4; MAFG = transcription factor MafG; MGST1 = microsomal glutathione S-transferase 1.

	Responders	Non-Responders	Controls
Baseline vs 6 weeks	478	0	7
6 weeks vs 15 months	0	1264	180

Table 14. Number of differentially expressed genes between baseline and 6 weeks (short-term), and 6 weeks and 15 months (medium-term) in responders, non-responders and controls.
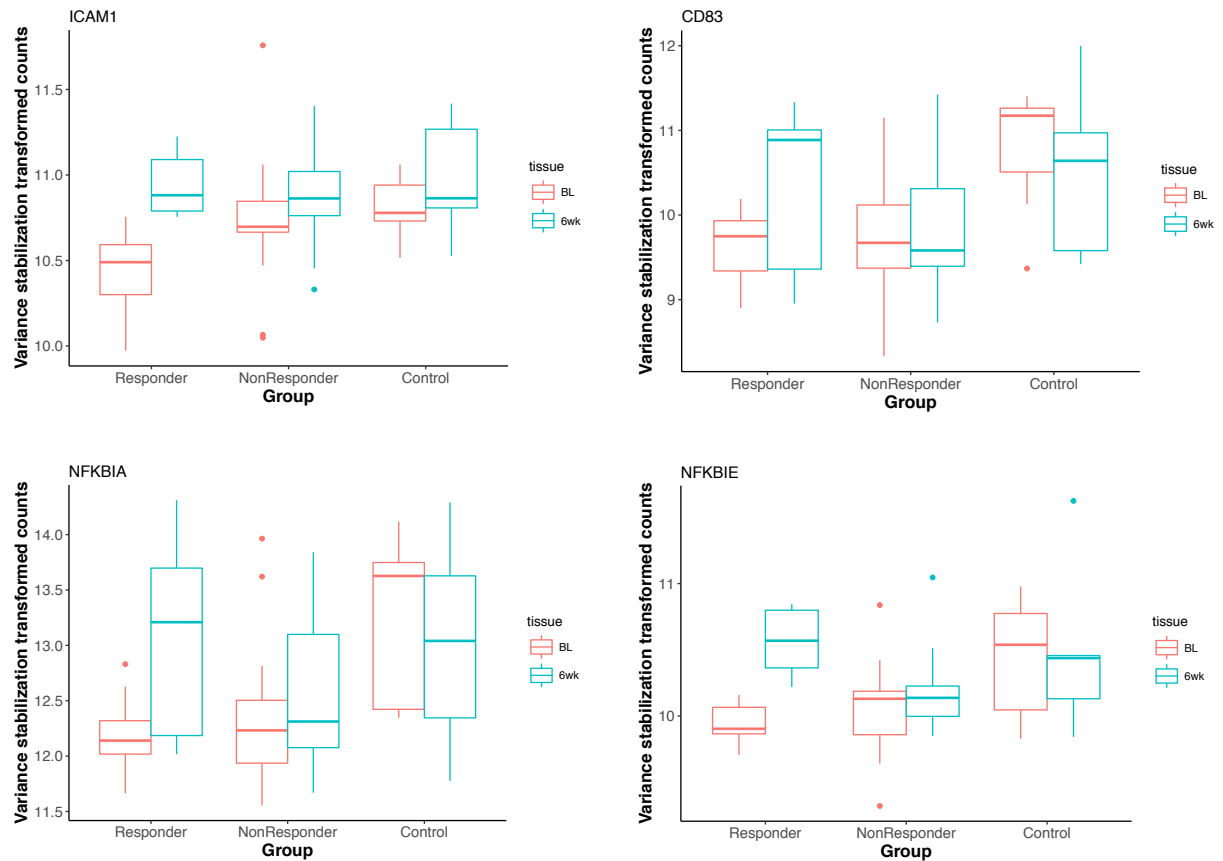
Fig 27 (a-d). NFκB related transcripts are increased 6 weeks post treatment in responders but not in non-responders or healthy controls. Boxplots represent variance stabilised transformed counts for transcripts (A) ICAM1, (B) CD83, (C) NFκBIA, (D) NFκBIE at baseline and 6 weeks in responders, non-responders and healthy controls. CD83 =cluster of differentiation 83; ICAM1 = Intercellular adhesion molecule 1; NFκBIA = nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha; NFκBIE = nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, epsilon.

I confirmed the significance of this difference between responder and non-responder groups by testing for effects of outlier values using leave-one-out cross validation (LOOCV). The median numbers of differentially expressed genes post-treatment were 404 and 0 in the responder and non-responder groups, respectively (p < 0.0005). I also assessed RNA-Seq data from untreated healthy controls (n=7). Comparison from

baseline to the end of a 6-week period without any intervention showed only 7 differentially expressed genes (padj < 0.05) (Table 14).

In the responder group, I found 10 genes altered by treatment (at both 6 weeks and 15 months) and also differentially expressed between patients and controls. These were designated as a set of potential treatment response markers: ZNF594, ZEB2, SERTAD3, ZFP36, PIM4, RNF19B, PPP1R15A, LOC105378248, GCNt2, HBEF.

Treatment response is associated with a stable pattern of gene expression

Between 6 weeks and 15 months, 0 and 1264 differentially expressed genes were detected in the responder and non-responder groups, respectively (Table 14). I further confirmed a difference between the two groups using a 100-fold permutation analysis in randomly selected combinations of 8 RRMS patients. The median number of differentially expressed genes in this analysis was 702 (range, 31 – 3230). In healthy controls (n = 7), who were not given any intervention and who were followed over the same time period, there were 180 differentially expressed genes (Table 14).

The large number of differentially expressed genes found between 6 weeks and 15 months in the non-responder group prompted us to test for response heterogeneity within this group. I first tested for individual outliers. Based on PCA of the 16 non-responders at 15 months, responses in 2 patients were outliers. Data from these subjects therefore were excluded from further analysis (Fig 28a). Following removal of these two outliers, two distinct non-responder groups (arbitrarily called groups A and B) were identified in a subsequent round of PCA (Fig 28b). I then independently assessed differentially expressed genes between 6 weeks and 15 months in these

groups: 560 differentially expressed genes were found for group A and 648 for group B (117 [11%] of these differentially expressed genes overlapped between the two groups). I tested for the significance of these differences relative to the stable expression pattern in the responder group using LOOCV. The median numbers of differentially expressed genes with LOOCV were 270 and 497 for groups A and B, significantly different to the equivalent analysis in the responder group where the median number of differentially expressed genes was 0 (p = 0.03).
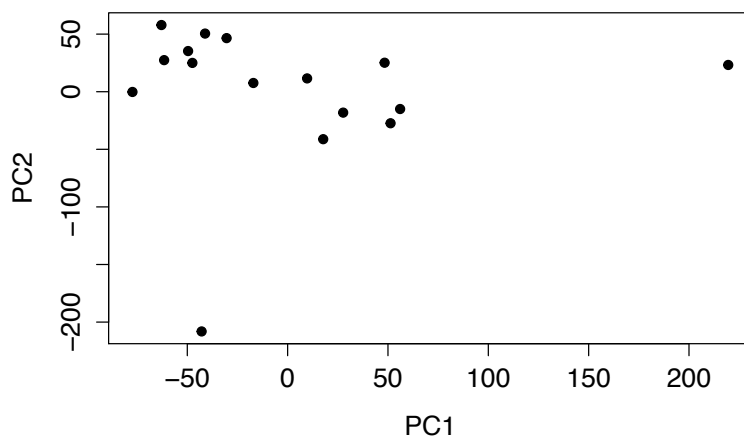
Fig 28a. Two subjects appear as outliers in 15 month samples within the non-responder group. Principal component analysis of non-responders at 15 months. Each dot represents a subject's RNA sample.
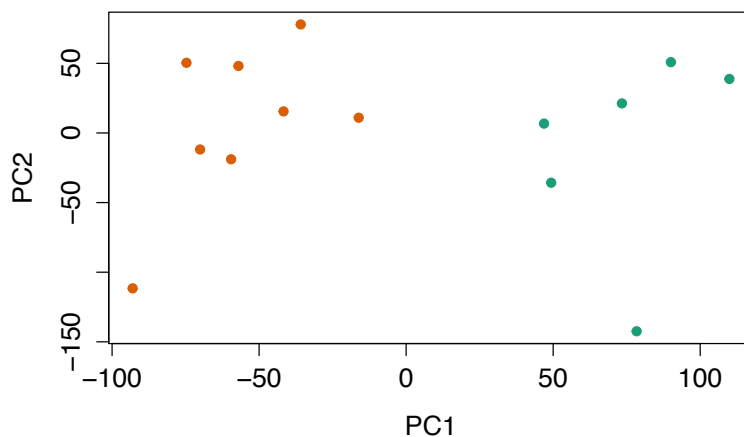
Fig 28b. Two distinct non-responder groups are identified at 15 months. Principal component analysis of non-responder group A (green) and group B (red) at 15 months. Each dot represents a subject's RNA sample.

In Group A, the most enriched pathways were involved with Th1 and Th2 activation and T cell receptor signalling (p < 0.0001). The differentially expressed genes showed enrichment for the KEGG pathway 'T cell receptor signalling' (p < 0.0001). In Group B, there were no significantly enriched canonical or KEGG pathways. Given the enrichment of immune-related genes in Group A, I investigated whether this correlated with more adverse clinic outcomes. Within Group A, there appeared to be a higher rate of grey matter atrophy (100% vs 62%) and white matter atrophy (100% vs 75%) (defined as greater than -0.4% change in annualised atrophy for both measures), however neither of these differences in number of subjects exceeding the pathological threshold set were statistically significant (p = 0.2 and 0.48, respectively).

DMF is associated with a short term 'pseudo-normalisation' of gene expression in responders

After controlling for gender, 668 differentially expressed genes were found between patients in the responder group and healthy controls at baseline (padj < 0.05). However, 6 weeks after the start of treatment, only 3 genes were differentially expressed between these patients and healthy controls (Table 15). At 15 months, there were 85 differentially expressed genes between these patients, although only 14 genes (2%) overlapped with the differentially expressed genes found at baseline (Fig 29a).

478 differentially expressed genes were found between patients in the non-responder group and healthy controls at baseline (padj < 0.05) (Table 15). 202 (21%) overlapped with those in the equivalent analysis in the responder group. At 6 weeks after the start of treatment, 18 genes were differentially expressed between the same patients and

healthy controls, 8 of which (44%) also had been identified baseline (Fig 29b). At 15 months, 391 differentially expressed genes were found in non-responder group A and 340 in non-responder group B (Table 15).

	Responders vs Control	Non-Responders vs Control
Baseline	668	478
6 weeks	3	18
15 months	85	391 (Group A) 340 (Group B)

Table 15. Number of differentially expressed genes in cross-sectional analysis between RRMS patients and controls at baseline, 6 weeks and 15 months.

(a)   **(b)**
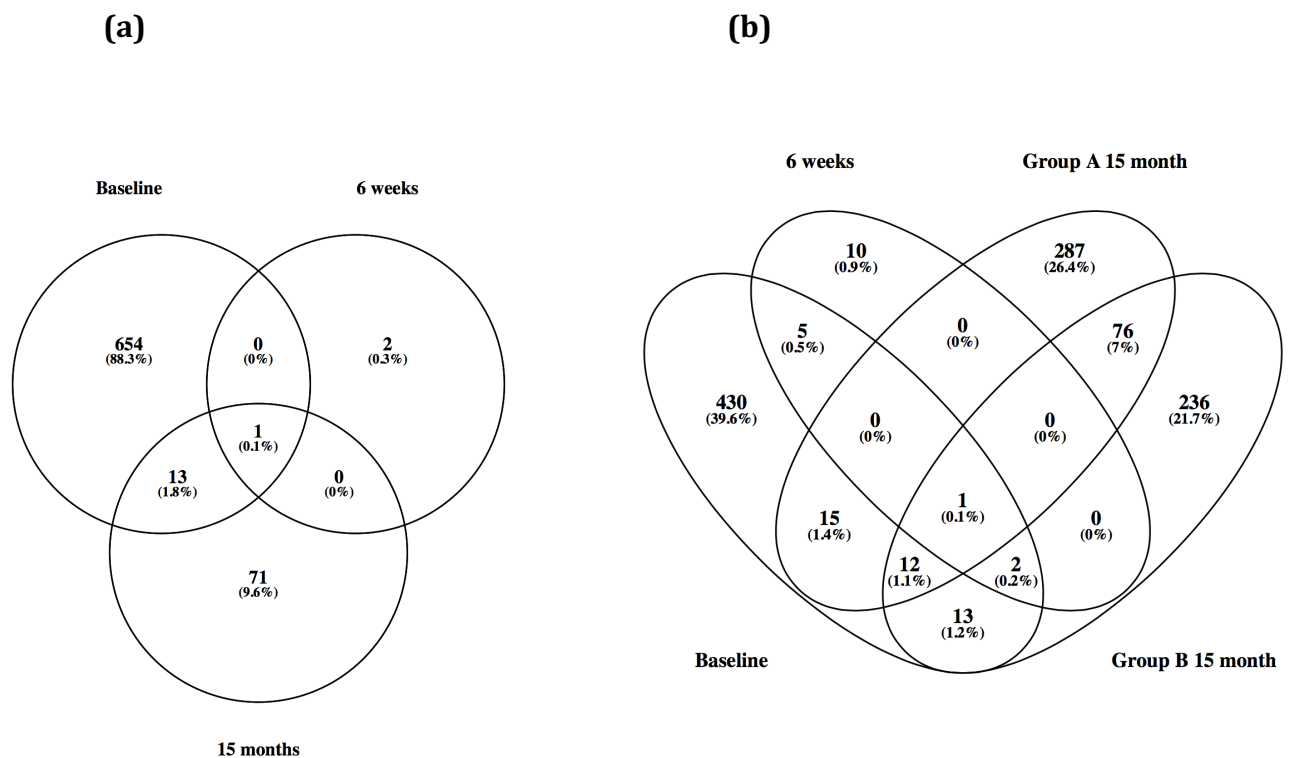
Fig 29. DMF treatment is associated with a relative normalisation of gene expression in responders but not in non-responders. Venn Diagrams represent the number of DEG in responder (A) and non-responder (B) groups compared with controls at baseline, 6 weeks and 15 months.

Nrf2 and NFκB related genes are persistently altered by DMF in responders but not in non-responders

Given the postulated roles of Nrf2 activation and NFκB inhibition in the mechanism of action of DMF I sought to identify downstream genes known to be transcriptionally regulated by Nrf2(299) and NFκB(300) that may be affected differently in the responder and non-responder group. For Nrf2, in the responder group, 14/75 (19%) Nrf2 regulated genes were differentially expressed 6 weeks post-treatment. 6 of these remained differentially expressed at 15 months. In the 'non-responder' group, there were no Nrf2-related genes differentially expressed at 6 weeks. CCAAT/enhancer-binding protein beta (CEBPB) was the only gene that was persistently changed in the responder group and which did not change in the non-responder group at either timepoints. 2/392 genes known to be regulated by NFκB(300) were persistently differentially expressed in the responder group: ferritin heavy chain (FTH1) and aspartyl beta-hydroxylase gene (ASPH). FTH1 was differentially expressed in the non-responder group at 15 months leaving ASPH as the only NFκB regulated gene to be persistently differentially expressed in the responder group but not in the non-responder group over the study duration.

Small RNA as a treatment response marker in RRMS patients

I performed similar analyses using data derived from small RNA. In the responder group two miRNAs differentially expressed at baseline normalised at 6 weeks and 15 months (mir-423 and mir-6718). In the non-responders mir-423 was also differentially expressed at baseline but remained so at 6 weeks demonstrating no significant normalisation of this transcript post-treatment. Mir-423 is therefore a candidate marker of drug response.

Correlations between RNA and small RNA treatment response markers

In order to look for correlations between miRNAs and RNAs in the dataset I identified all RNAs that are regulated by miRNA of interest mir-423 using TargetScan v7.0. I compared this list of 249 conserved RNAs with those RNAs I found to be persistently changed at 6 weeks and 15 months in the responder group. RNF19B is a gene, known to be regulated by mir-423, that was persistently altered in the responder group and also differentially expressed in responders vs controls at baseline but not in an equivalent comparison at 6-weeeks and 15 months.

Transcripts of genes associated with MS susceptibility are altered by DMF

There are currently 200 genes identified by genome wide association studies containing variants that confer an increased risk of MS(301). In the responder group, at 6 weeks, 16 of these genes were differentially expressed post-treatment. In the non-responder group, no MS risk genes were differentially expressed 6 weeks post-treatment. A leave-one-out cross validation was performed in the responder and non-responder groups respectively. In the responder group, the median number of differentially expressed susceptibility genes was 9 (range 0 – 21). In an equivalent analysis in the non-responder group the median was 0 (range 0 -1). The difference between these distributions was highly significant ($p < 0.0001$).

Using GATK software, I was able to look at differential expression of genetic risk variant alleles pre- and post- treatment selectively in patients heterozygous for a risk allele. 7 MS risk genes had sufficient depth for this analysis. 1 gene (RUNX3)

showed statistically significant reduction in the risk allele (p < 0.05). For this gene, the reciprocal healthy variant also showed a significant reduction, representing overall reduction in the expression of this gene post treatment rather than a reduction in the disease-specific variant itself.

Metabolic correlates of transcriptomic biomarkers

Next, I attempted to correlate the treatment response RNAs and small RNAs (KIR3DL2, RNF19B and mir-423) with metabolomic profiles in the individual subjects. Firstly, I correlated MMF concentrations with the three putative transcriptomic response markers. Samples showed either low concentrations of MMF (<20 ng/mL) or high concentrations of MMF (range 95 – 592 ng/mL). Using all samples, I found no correlation between the active drug and the three transcriptome biomarkers. A previously identified correlated of MMF (1-methylnicotinamide – see Chapter 4) also showed no correlation. When performing the same analysis only in those patients with therapeutic concentrations of MMF (range 95 – 592 ng/mL), I found a high inverse correlation between mir-423 and levels of MMF (r = -0.68) and 1-methylnicotinamide (r = -0.85).

I also undertook an unbiased analysis by performing a pairwise correlation between each of my transcriptome markers of interest and all metabolites assayed by metabolomic profiling. There were 13 metabolites highly correlated with levels of RNF19B (r > 0.7) (Table 16). For KIR3DL2 and mir-423, there were no highly correlated or inversely correlated metabolites.

Metabolite	Correlation
LPC(18:1)	0.80
CER(24:1)	0.78
uridine	0.77
1-methylhistidine	0.77
Total LPC	0.77
LPC(16:0)	0.75
N-acetylisoleucine	0.74
Gamma-glutamylvaline	0.72
LPC(17:0)	0.72
PC(17:0/22:5)	0.72
N-methylalanine	0.71
LPC(20:2)	0.71
LPE(22:5)	0.71

Table 16. Metabolites demonstrating high correlation with gene transcript RNF19B. LPC = lipophosphatidylcholine, CER=Ceramide, PC = phosphatidylcholine, LPE = lipophosphatidylethanolamine

I repeated this analysis looking only in patients with therapeutic concentrations of MMF. There were 249 metabolites that were highly correlated with RNF19B (r > 0.7). including succinylcarnitine (r = 0.87), methylsuccinylcarnitine (r = 0.85) and succinate (r = 0.84) all of which were noted to be significantly elevated 6-weeks post treatment (Supplementary Correlation Table – Appendix 1). For KIR3DL2, there were 94 highly correlated metabolites (r > 0.7). For mir-423, I found 56 highly inversely correlated metabolites, the most significant was 1-methylnicotinamide (r = -0.85) (Supplementary Correlation Table).

Metabolomic correlates of DMF-induced flushing

One of the most common side effects of DMF is flushing. At 6 weeks follow-up post-treatment with DMF, 18/35 patients reported flushing at least once since commencing DMF. I separated patients out by this symptom and investigated corresponding metabolite changes pre- and post- treatment, restricting the analysis to those with therapeutic concentrations of the active component of the drug, MMF. 73 metabolites were significantly altered post-treatment in the symptomatic group ($p < 0.05$). The 10 showing greatest changes are highlighted in Table 17a. 7 of these metabolites were over-expressed post-treatment (fold-change 1.26 – 1.87). In the group which did not experience symptoms, 200 metabolites were significantly different; 170 of these were triacylglycerols that were all under-expressed post-treatment with DMF. For the 10 most significantly reduced triacylglycerols the mean fold-change was 0.53 (s.d. 0.18) (Fig 17b)

Metabolite	Fold change	Standard Deviation	Adjusted p-value
cys-gly, oxidized	1.39	0.24	0.0006
leucylleucine	0.48	0.38	0.0039
sebacate (decanedioate)	1.66	1.28	0.0054
adipoylcarnitine (C6-DC)	1.87	0.57	0.0056
creatinine	0.44	0.04	0.0069
phenylalanylglycine	1.34	0.54	0.0075
PE(18:1/18:2)	1.26	0.47	0.0076
LPC(18:2)	1.19	0.28	0.0076
PE(18:2/18:2)	1.56	0.67	0.0092
PE(P-18:2/18:2)	0.47	0.34	0.0096

Table 17a. Patients experiencing flushing demonstrated significant differences in metabolites. 10 most significantly altered metabolites in the flusher group. Values displayed as fold changes pre- and post- treatment with associated standard deviations and adjusted p-values.

Metabolite	Fold Change	Standard Deviation	Adjusted p-value
TAG54:6-FA18:1	0.55	0.15	0.0001
TAG52:4-FA20:0	0.65	0.06	0.0003
TAG50:4-FA18:2	0.55	0.13	0.0004
TAG58:8-FA22:5	0.75	0.06	0.0004
TAG54:6-FA18:3	0.53	0.09	0.0004
TAG52:6-FA16:1	0.49	0.21	0.0005
TAG52:6-FA18:3	0.51	0.15	0.0006
TAG58:7-FA18:1	0.69	0.09	0.0007
TAG54:5-FA18:3	0.59	0.13	0.0009
TAG52:6-FA18:2	0.49	0.15	0.0009

Table 17b. Patients not experiencing flushing demonstrated significant differences in metabolites. 10 most significantly altered metabolites in the non-flusher group. Values displayed as fold changes pre- and post- treatment with associated standard deviations and adjusted p-values.

Metabolomic correlates of DMF-induced gastrointestinal symptoms

Similarly, I repeated this analysis for gastrointestinal symptoms. In my cohort, 20/35 patients reported gastrointestinal symptoms at least once since commencing DMF. In those patients reporting abdominal symptoms, 54 metabolites were significantly altered ($p < 0.05$). These included pregnanolone which I previously identified as being significantly reduced in MS patients as well as 1-methylnicotinamide, succinylcarnitine and methyl-succinylcarnitine; all of which were also significantly increased 6 weeks post- treatment with DMF (Table 18).

Finally, I tested whether concentrations of MMF were associated with side effects of the medication. In the group with therapeutic concentrations of MMF (range 95 – 592 ng/mL), 67% experienced flushing and 67% experienced gastrointestinal (GI)

symptoms. In the group with low plasma concentrations, 28% reported flushing and 36% experienced GI symptoms (p > 0.05, Fisher's exact test).

Metabolite	Fold Change	St Deviation	Adjusted p-value
Palmitoylcholine	1.76	0.54	0.007
Pregnanolone	0.8	0.1	0.012
Succinylcarnitine	3.8	1.5	0.017
1-methylnicotinamide	2.4	1.7	0.021
Methylsuccinoylcarnitine	180.4	206.8	0.042

Table 18. Metabolites significantly increased 6-weeks post- treatment with DMF and also associated with DMF associated gastrointestinal adverse events. Fold changes, standard deviation and adjusted p-value provided (corrected using Benjamini-Hochberg).

Results Summary

- **Approximately 35% of patients had 'no evidence of disease activity (NEDA-4) 15 months post-treatment with DMF. DMF was highly effective at improving a secondary outcome; MSFC, in the cohort but did not have a significant effect on QOL measures.**

- **I observed a robust short term pharmacodynamic transcriptomic response to DMF in a subset of patients who responded to the drug in the medium-term. This effect coincided with 'pseudo-normalisation' of gene expression when comparing to healthy controls and also a medium-term stabilisation of gene expression.**

- **I found 3 disease related genes that persistently normalised in the responder but not non-responder group. 1 of these genes, KIR3DL2 has been previously implicated in MS disease pathogenesis. A further 10 disease-related genes (including RNF19B) were identified that were persistently altered by DMF in the treatment responder group but not the non-responder group.**

- **A small subset of Nrf2 and NFκB related genes were persistently altered by DMF.**

- **One disease related miRNA (mir-423) was also affected by DMF treatment in the responder but not the non-responder group. It is**

known to regulate RNF19b which I had previously identified as a responder gene. Furthermore, the levels of mir-423 were highly inversely correlated with concentrations of MMF and levels of the metabolite 1-methylnicotinamide.

- **An unbiased multi-omics analysis revealed high correlations between levels of RNF-19B and succinylcarnitine, methy-succinylcarnitine and succinate, all of which were significantly elevated 6 weeks post-treatment.**

Discussion

There is currently no reliable early treatment response prediction marker for any RRMS DMT. Here, I investigated whether gene expression changes at 6-weeks are associated with the medium-term clinical response to DMF. Using RNAseq, I observed that, in treatment responders, a robust short-term transcriptomic response to DMF in PBMCs was associated with activation of the *NRF2* and inhibition of the *NFκB* pathways. In non-responders, no early transcriptional changes were observed after starting DMF. In addition to a robust short-term pharmacodynamic response to DMF in treatment responders, I observed stabilisation of gene expression. This contrasted with medium-term expression changes in non-responders enriched for transcripts of genes involved in inflammatory pathways.

Previous studies investigating the pharmacodynamic effects of DMF on gene expression in vitro have identified genes related to anti-oxidant pathways(242, 250-252), anti-inflammatory pathways(53, 253) and NFκB(52, 254). All may be related to DMF's therapeutic effect. My results extend and confirm these in vivo; all three of these pathways may be implicated in the effect of DMF in the subset of patients for whom the drug has a medium-term (1-year) beneficial treatment effect. I identified an association between microRNA and mRNA (Mir-423/RNF-19B) that was persistently altered in the responder group post-treatment but not in the non-responder group. Mir-423 is a small RNA species known to be reduced in MS patients compared to controls(302). It has also been implicated in a number of other human inflammatory diseases(303). A separate study in PBMCs of MS patients treated with IFN-beta demonstrated a significant reduction in mir-423 post-treatment(304). I observed raised levels of mir-423 in MS patients at baseline (responders and non-responders)

152

and a reduction (in agreement with the IFN-beta) following 6-weeks treatment with DMF. Mir-423 is known to regulate RNF-19B, a gene that I found to be altered by DMF in treatment responders but not in non-responders. Little appears to be known about RNF-19B other than that it is expressed predominantly in natural killer cells(305). KIR3DL2 is also highly expressed in natural killer cells and was consistently differentially expressed in non-responders but normalised in responders post-treatment. It has been implicated in the pathogenesis of MS due to its interaction with HLA class 1 molecules(218, 306). My findings of specific RNA and small-RNA species of interest require replication in independent datasets and also validation using quantitative PCR techniques.

Alongside a robust transcriptomic pharmacodynamic response to DMF in treatment responders, I was able to demonstrate a complete stabilisation of gene expression over the medium-term which was not present in treatment non-responders. Whilst an interpretation of this finding is not straightforward, I hypothesise that conferring relative homeostasis in peripheral PBMC gene expression limits inflammatory flairs. In a subset of non-responders I saw over-expression of such pathways in the medium-term in association with evidence of breakthrough disease activity.

This study identified a large number of differentially expressed genes between RRMS patients and healthy controls. While some studies have also identified differentially expressed genes between these groups(223)(224), other results have been less clear, potentially reflecting the lower sensitivity of microarray technology, which identifies a much smaller number of RNA species(108). I was able to demonstrate that the differences between RRMS patients and healthy controls were largely attenuated

within 6-weeks of initiation of treatment with DMF. In the subgroup of responders, by 15 months, this apparent "normalisation" of the transcriptome was maintained over 15 months, although the number of differentially expressed genes largely returned to pre-treatment levels amongst the non-responders.

In a subset of the cohort for whom at the time of sampling there were therapeutic concentrations of the active metabolite of DMF, I was able to correlate my RNA and small RNA species of interest with metabolomic correlates of DMF response identified in Chapter 4. These included levels of MMF and 1-methylnicotinamide. Furthermore, I was able to demonstrate a strong correlation between RNF19B and succinylcarnitine, methylsuccinycarnitine and succinate. These findings imply that a potential downstream role for these metabolites in the mechanism of action of DMF, and specifically in the expression of genes and micro-RNAs known to be affected by the drug itself.

A recent study reported on the medium-term NEDA outcomes of the phase III trials (DEFINE/CONFIRM) of DMF in RRMS. In keeping with their findings using NEDA-3(289), I found that 33% of patients of patients achieved NEDA-4 after 15 months of treatment. I was also able to report a highly significant improvement in overall MSFC score and 9-hole peg test at 15 months. The improvement in MSFC score was considerably better than that found in those patients who participated in the initial DEFINE/CONFIRM studies which is encouraging given this is real-life data(307). Whilst others have been able to show an improvement in QOL with DMF as measured by SF-36(308), I was unable to replicate this finding.

The small sample size and using only 3 timepoints limit the confidence in my selection of response markers and estimation of responses. Further confirmatory work is needed. However, I attempted to reduce the impact of these factors by increasing the rigor of my statistical analysis (i.e. performing leave-one-out cross validation and a permutation analysis). The use of a control group at matching timepoints allowed us to compare findings in the patients with those in a healthy population over time. I also minimised the impact of artefacts arising from 'batch' effects, by sequencing all samples at the same time and in the same sequencing facility. Without an independent replication cohort, I was unable to test the predictive power of my findings formally. These could be tested in a future cohort using a regression model developed using my pilot findings as the training dataset.

It is unclear what biological differences exist between the responder and non-responder groups that may account for their differing response to DMF. Indeed, it may also be the case that they represent different subtypes or phenotypes of RRMS that have yet to be elucidated. I believe the latter explanation is less likely as there was consistently very little difference in differentially expressed genes between these two groups at all timepoints. Given the lack of a unique mechanism of action for DMF, it is difficult to predict the underlying reason why only a subset of patients seemed to respond the medication. If my findings were replicated in a much larger cohort, one could look for genetic variants that may be different between responders and non-responders potentially within transcription sites already known to be altered by DMF.

In summary, I have provided evidence that DMF can alter the short term transcriptome profiles in a sub-group of MS patients and that these changes are consistent with current hypotheses of major mechanisms of action of DMF. My results suggest that DMF enhances immune homeostasis after "normalising" gene expression in the PBMC. To my knowledge, this stabilising effect of DMF on gene expression has not been described elsewhere. If replicated, my results may have implications for the use of DMF in other inflammatory conditions or more generally(309, 310). However, my work also raises questions about the biological differences between the responder and non-responder groups that may account for their differing response to DMF. One approach to explore this could involve replicating my findings in a larger cohort, whilst also identifying genetic variants that are different between responders and non-responders in transcription sites modulated by DMF.

Chapter 6. Summary and Future Directions

The landscape for treating RRMS has evolved to such an extent over the last twenty years that it is now urgent to identify methods to stratify patients to a treatment that is likely to provide the greatest efficacy whilst avoiding unwanted side effects and adverse events. For the healthcare payer, there is also the added incentive of cost efficiency given that all of these DMTs are very expensive.

The MS community is moving towards an approach established by the rheumatology community for the treatment of rheumatoid arthritis (RA)(311). This can be described as a 'treat early, treat aggressively' mindset. However, the fundamental limitation of this approach is the lack of any validated early predictive measures of disease activity in individual patients. In the absence of these, we rely on monitoring traditional clinical and radiological measures of disease activity which take time to evolve. The 'treat-to-target' approach in RA has similar difficulties; DMARDS for RA have similar efficacies at a population level and there are no approved biomarkers to appropriately select the best medicine(312). The problem is complex not just because of patient heterogeneity, but because medications may act on common pathways but at different levels(313).

In the context of MS, the challenges posed by the range of medicines is greater. We know that certain medications, such as alemtuzumab work preferentially on CD-52 T cells, whereas ocrelizumab is a monoclonal antibody targeting CD20 on B - lymphocytes. Our limited understanding of the subtypes of MS (both on a molecular and a clinical level) is perhaps the main reason why we are unable to accurately

stratify patients to an appropriate medication. It is possible there are subtypes of MS that are predominantly T-cell or B-cell mediated. We have evidence from experience with other DMTs in MS that certain patients are indeed less likely to respond to some medications, e.g., a subset of patients on Beta-Interferon will develop neutralising antibodies that reduce the efficacy of the medication(314, 315). Stratification for risk of adverse events already is a part of routine management of RRMS patients. We closely monitor patients on natalizumab for the risk of developing PML which can, to some extent be predicted based on antibody levels to JC virus, duration of treatment and previous exposure to other DMTs(29).

With these examples in mind, the main aim of this study was to determine whether in MS patients starting DMF, gene expression and metabolite changes associated with drug action at 6 weeks could explain the clinical and radiological response at 15 months. I also aimed to use this information to better understand how the circulating immune cell responses in MS patients may differ from healthy controls and to further understand the mechanism of action of DMF.

Strengths of this Study

In this study, I was able to confirm that gene expression changes in PBMC are present in MS patients compared to controls. These genes are enriched for inflammatory pathways. Furthermore, I was able to demonstrate robust transcriptomic responses to DMF. Adding to prior literature, these support roles for NFkB and Nrf2 in the mechanism of action of DMF. I also identified a number of novel RNA and small RNA species that were differentially expressed following short-term administration of

DMF and which are also altered in MS patients compared to controls. This is important as it strengthens the hypothesis that gene expression changes are relevant to MS disease course and that these can be modulated by DMT.

With respect to the main aim of this study, I identified that in medium-term responder patients, there was a robust short-term response to DMF that was not found in non-responders. Furthermore, this response was also associated with a longer-term stabilisation of gene expression which was not observed in non-responders. The robust response to DMF in responders was associated with lack of apparent further changes suggesting a stabilisation of gene expression. This gene expression response model now needs prospective testing to assess its positive and negative predictive power.

Using a novel NMR-based lipoprotein assay I was able to identify a new subtype of lipoproteins (VLDL) elevated in MS patients and correlated with disability as measured by EDSS. I also found evidence for raised triacylglycerols in the MS patients compared to controls using mass spectrometry based metabolomics-assays. Furthermore, I was able to describe novel associations between TCA cycle intermediates succinylcarnitine and methyl succinylcarnitine and DMF. Using pairwise correlation techniques, these metabolic changes were associated with differences in the expression of genes and small RNAs of interest. My observations may extend understanding of the mechanism of action of DMF: the correlation between DMF and 1-methylnicotinamide particularly should also be further explored given the potential anti-inflammatory properties of this molecule(279, 281).

Challenges of this Study

In order to fulfil the potential of personalised medicine, it will be essential to understand the different phenotypes that exist within the MS syndrome. Using this approach, it is more likely that a treatment will target an individuals' unique pathology. The paradigm of this model is in the field of oncology. For example individualised, genetic-based diagnoses have had considerable impact in the licensing of imatinib for Philadelphia chromosome-positive chronic myeloid leukaemia and trastuzumab in HER2-positive breast cancer(79). MS is not a purely genetic disease and thus this approach poses a greater challenge.

In this study, whilst I identified a sub-group of patients who appeared to respond to the medication in the medium-term, there were no population differences on a transcriptome level between these patients and non-responders either before or after starting DMF. This was evident in both the differential expression analyses and also in the unbiased data-reduction techniques that I employed. Indeed, whilst there are a number of underlying genetic variants that increase the risk of MS, the odds ratios for individual risk variants are extremely modest and there are only a small proportion of genes with associations to MS(316). In addition, there are no known genetic discriminants for subtypes of MS. However, it is widely accepted that the transcriptome can also be affected by the environment which is known to contribute to MS risk susceptibility(61). Even within this limited study, I identified hundreds of genes that appeared to change in response to 6-weeks on a new medication. The interpretation of a lack of population heterogeneity on a transcriptome level is likely to be complex. Gene expression changes are clearly dynamic; they may be inherently

too predisposed to noise and environmental fluctuation in vivo to derive any meaningful discriminatory variables with the limited sampling employed here.

Whilst I attempted to identify possible genetic variants that may have been differentially expressed post treatment, thus potentially explaining population differences within the study cohort, this was ultimately unsuccessful. The main reason for this was that I could not infer the genotype of intronic sequences using RNA-Seq. The majority of these variants are in intronic regions which would not have been sequenced using this technology(317). In future studies, it would be advisable to also undertake genotype sequencing, potentially using the "ImmunoChip, an efficient genotyping platform for loci associated with autoimmune diseases(301).

Alongside these challenges included the choice of cell for this study. Given that this was a pilot, I chose all PBMC with the intention to validate any interesting findings in future work using a more targeted single cell type approach. This is now a possibility with a wide range of resources available that can help identify cell specific gene expression(318). Furthermore, the recent developments in single-cell RNA-Seq will eliminate further noise. Given that one of my genes of interest, KIR3DL2 is highly expressed in natural killer cells, this work has identified a specific cell subset for future analysis.

As part of my metabolomic profiling, I was able to derive concentrations of the active component of DMF in my patients. This was helpful for finding correlates of the medication that may relate to its downstream mechanism of action or to known adverse effects. Of a total of 27 patients assayed, only 12 appeared to have potentially

therapeutic concentrations of the drug in their plasma at the time of sampling. Given the short terminal half-life of monomethylfumarate (approximately 1 hour), no circulating evidence of the drug would be expected at 24 hours(319). Consequently, it is difficult to determine whether those patients with unmeasurable concentrations of MMF had been non-adherent over a longer period than just the last dose of medication.

A singular interpretation of this finding is difficult; mechanisms are likely to be multifactorial. We know that adherence to medications is generally low, and in particular for patients with chronic conditions(320) or where the medications are associated with adverse events. However, oral medications tend to be associated with better adherence(321) relative to injectables and adherence in general for MS has previously been reported in some cases as high(322).

Nevertheless, it is interesting to note that amongst those patients with high concentrations of MMF, 67% were responders as defined by the composite measure NEDA-4. In those with low concentrations, only 21% were responders. As such, perhaps the greatest predictor of a medium-term response to DMF was concentration of the drug at 6-weeks, i.e., people who adhere to the prescribed medicine dosing regime are more than 3-times more likely to benefit than those who do not.

A final challenge encountered during this study was the relative paucity of techniques available currently to integrate transcriptome and metabolomic data. A fundamental challenge is that no transcript can be directly mapped to a specific metabolite in the same way as a gene can be mapped to a protein(323). The statistical tools that are

available for data integration can be broadly defined as correlation-based, multivariate-based or pathway based. By demonstrating differences between patients and controls (Chapter 3) and also the pharmacodynamics effects of DMF (Chapter 4) using transcriptome, metabolome and cytokine datasets, I took a more "conceptual" integration approach to the data. This undoubtedly has many limitations and will miss associations that could only have been found if the datasets had been integrated together.

In order to address this issue, I also used pairwise correlation-based approaches to integrate my RNA and small RNA data with metabolomics. This is a form of statistical integration. One of the limitations was that metabolites were being measured from a different source (plasma) to the transcripts (PBMCs). Furthermore, this correlation-based approach does not take into account the timescale of changes; metabolite and transcriptome changes likely occur over different timescales. To limit any source of error due to this limitation, I could have considered obtaining a time course of samples for each omics and aligning these using techniques such as Dynamic Time Warping(324). However, despite these limitations, I was able to show high correlations between RNA and small RNA of interest and metabolites that were previously identified as altered in response to DMF. I was also able to minimise batch effects as these samples were extracted at the same time and from the same patients (i.e. a replicate-matched study).

A final integration approach uses pathway analysis where over-represented pathways from different 'omics approaches are identified. The advantage of this approach is discovery of pathways that may only become over-represented when both approaches

are combined and the fact it is a biologically informed data reduction method. The limitation of this approach is that it is based on an imperfect model. Using this approach I was able to confirm a role for IL-4 in the pharmacodynamic effect of DMF, and I was also able to identify STK11 as potentially important gene in the pathogenesis of MS.

Undoubtedly, in the future, as multi-omics studies become less expensive, tools enabling meaningful integration of omics will develop. Indeed, it may be that such tools integrate different statistical approaches and can transform these into functional networks based on pathway analysis. Such networks will be highly useful for the development of predictive tools for personalised medicine.

Limitations of this Study

One of the main limitations of this study was the small sample size; which undoubtedly reduced the statistical power of findings. This is a significant issue given the high probability of false positive results (6) in genetic based studies and also the low levels of replication that are commonly seen in biomedical studies in general (7,8). Small sample sizes are problematic because the chances of observing a true effect are less likely and the risk of a false positive rises. This may occur independent of or secondary to inherent study biases(248).

A study with low power is more likely to generate false negative results. For example, a study with a power of 30% will only identify 30% of genuine non-null events. In the context of NGS studies, power is even more problematic because it is often governed by 'non-scientific' limitations such as the cost of the technology itself. When

deciding on the sample size for a RNA-Seq differential expression analysis, a balance must be achieved between desired power and controlling for an appropriate error rate, given the thousands of genes being multiple tested simultaneously.

There are no clear recommendations for sample size in RNA-Seq studies. One study sought to examine the number of biological replicates needed and recommended at least six replicates per condition or 12 to reduce the chances of missing differentially expressed genes(325). However, this recommendation was based on a single study and may not be applicable to other study designs. A further study compared the different software tools available to assist in determining a sample size but found they produced widely differing results; and also took into account different parameters. The conclusion of this study is that pilot data should be used to determine future sample sizes(326). The findings of this study are particularly relevant to my study in that when using different alignment softwares, there was significant variation in the differentially expressed genes that were identified.

Variability can occur in the processing of RNA-Seq data. The earliest source of this may be in sample preparation where sequencing errors and bias may arise from contamination or sample degradation during isolation and preservation. In order to reduce the probability of this, I isolated PBMCs and performed RNA isolation on fresh samples on the day of collection. Furthermore, to reduce the chances of batch effects, all samples were library prepared and sequenced in the same facility at the same time. From visual inspection of the data using principal component analysis, there was no evidence of a batch effect.

I tested the variance in differentially expressed genes using technical replicates that were aligned and sequenced using different software and sequencing facilities respectively. When using different alignment approaches the consistency between the replicates was in the order of 45%. In order to determine whether the discrepancy resulted from different alignment software or from sequencing and library preparation, I repeated the bioinformatics steps identically. The result of this was a slight improvement in consistency (55%), albeit good consistency (80%) for the most differentially expressed genes. The cause of this variance is likely to be multi-factorial. Variance can occur during library preparation and sequencing. The errors in library preparation may have occurred during fragmentation(327), amplification(256) or at the barcoding stage(258). With respect to sequencing, Illumina(259) has sequence-specific error profiles and differing sequence depth can have an impact on the data integrity(255). It was also the case that those genes which were not consistently differentially expressed had lower mean counts and thus a statistically significant result for differential expression was less likely to be reproducible.

In order to have improved the reliability of my study, I could have sequenced samples on different platforms and integrated results that demonstrated concordance; termed cross-platform replication(328). This would have reduced the number of potentially false positive variants however could also have introduced different biases inherent in the respective sequencing technologies causing false negative results to occur.

Whilst this study contained a healthy control population who underwent sample collection at the same time as RRMS patients, I did not have a control population of randomised, untreated RRMS patients who were followed prospectively. The reason

for this is the ethical implications associated with not treating a RRMS population when a number of treatments are now available. For the same reason, it is no longer ethical to conduct placebo-controlled clinical trials in RRMS patients who would otherwise be eligible for treatment(329). Such a control population would have been most helpful to compare with non-responders as one would have expected similar changes both in the short-term following commencement of DMF and also in the medium-term. If consistent, this data could have helped support the concept of high levels of gene expression changes over time in MS patients compared to controls.

A further major limitation of my pilot study was the lack of a replication cohort. Whilst I was able to provide a replication for my metabolomics findings, I was not able to do so for the gene expression data due to time constraints. Replication of studies is an essential component of study validation. In the context of MS biomarker research, most biomarkers have failed replication with only 5% being successfully validated(73). This highlights the importance of study replication to increase the confidence in findings.

In this study, my main metabolomic findings were replicated in a sub-group of my patients. In addition, I was able to demonstrate successful reproducibility using technical replicates in my discovery cohort but assayed at a different facility. In the context of my gene expression study, whilst I did not have a replication cohort, I performed a number of statistical analyses to reduce the chances of results driven by outliers. These included permutation analyses and leave-one-out cross validation. I was also keen to demonstrate patterns of changes over time as well as changes in specific genes; the latter more prone to false positives. One approach I could have

used to validate my species of interest would have been quantitative PCR which is an essential component of validation of RNA-Seq findings and is the most obvious next step for future work.

In conclusion, my findings with respect to the mechanism of action of DMF on TCA cycle metabolites are novel and help to better understand the likely heterogeneous effects of this medication. It would be very interesting to determine whether the correlates of DMF that I found in the metabolome (methyl-succinylcarnitine and 1-methylnicotinamide) are therapeutic in themselves in an animal model of MS.

My novel finding of an association between VLDLs and MS sheds further light on the association between lipid dysregulation and MS. It has implications for the testing of lipid-lowering drugs in the treatment of MS, which has until now been restricted to patients with progressive MS(202).

Perhaps the most interesting and original preliminary result to arise from this study was the apparent relative stabilisation of gene expression over time, which I observed in the responder group over the medium term. I termed this 'gene expression homeostasis'. DMF is known to have anti-oxidant properties and we know from other conditions (as well as MS(330)) that oxidative stress can result in neurodegeneration and abnormal gene expression(331). Furthermore, oxidative stress is known to regulate gene expression of a number of inflammatory pathways(332) and has a role in ageing(333). Whilst there is a wealth of literature on specific gene expression changes that may arise as a result of oxidative stress and which, to some extent, may be countered by administration of DMF, this concept of gene expression homeostasis

has not been previously described. In order to provide greater credibility to this theory, it would of course be necessary to replicate the findings in a much larger cohort. Furthermore, it would be necessary to demonstrate maintenance of such an effect over a longer period of time. However, in the context of MS treatment (and in the current absence of treatments that can reverse disability), the ability to maintain stability (or remission) is in itself the definition an effective treatment. In this study, I have provided evidence of such stability at least on the transcriptome level in a subset of patients who responded to DMF as defined by the composite outcome measure NEDA-4. It is important to note that stabilising the transcriptome in such a way has implications for the repurposing of DMF for other conditions characterised by aberrant gene expression. This may include other inflammatory conditions, cancer or even ageing-related disorders. Research into the use of DMF for such conditions is already ongoing(309).

This pilot study aimed to identify short term dynamic transcriptomic and metabolomic changes in MS patients that are associated with longer term response to DMF. I successfully identified patterns of expression of genes, as well as specific genes and metabolites that were associated with such a response in a subset of patients. The true source of variability between responder and non-responder groups could not be identified on the basis of the subjects' individual transcriptomes. In order to determine the true predictive value of my findings, it would be necessary to test them prospectively on a new cohort of patients. Furthermore, it would be necessary to perform this study over a longer period of time to determine the true predictability over the long term. I identified an associated stabilisation of gene expression within my responder group. Gene expression homeostasis is a rather novel pharmacological

concept that may have implications for the use of DMF for other diseases and conditions. The single most important aspect of this study to pursue would be the development of a modelling approach using the RNA-Seq findings to determine the true predictive power of my findings. If these were confirmed and validated in a separate dataset, it would indeed be possible to predict a patients' medium term response to DMF based on gene expression changes within 6 weeks of onset of the medication. This would be a potentially practical example of personalised medicine for stratification of patients that is associated with a specific treatment in MS. Whilst the specific markers may be different with other drugs, the approach can be generalised for other medications used in MS. This could contribute meaningfully to improved outcomes for patients.

References

1.   Confavreux C, Vukusic S. The clinical course of multiple sclerosis. Handb Clin Neurol. 2014;122:343-69.

2.   Inusah S, Sormani MP, Cofield SS, Aban IB, Musani SK, Srinivasasainagendra V, et al. Assessing changes in relapse rates in multiple sclerosis. Multiple sclerosis. 2010;16(12):1414-21.

3.   Amato MP, Zipoli V, Goretti B, Portaccio E, De Caro MF, Ricchiuti L, et al. Benign multiple sclerosis: cognitive, psychological and social aspects in a clinical cohort. Journal of neurology. 2006;253(8):1054-9.

4.   Calabrese M, Favaretto A, Poretto V, Romualdi C, Rinaldi F, Mattisi I, et al. Low degree of cortical pathology is associated with benign course of multiple sclerosis. Multiple sclerosis. 2013;19(7):904-11.

5.   Correale J, Ysrraelit MC, Fiol MP. Benign multiple sclerosis: does it exist? Curr Neurol Neurosci Rep. 2012;12(5):601-9.

6.   Hawkins SA, McDonnell GV. Benign multiple sclerosis? Clinical course, long term follow up, and assessment of prognostic factors. Journal of neurology, neurosurgery, and psychiatry. 1999;67(2):148-52.

7.   Cotsapas C, Hafler DA. Immune-mediated disease genetics: the shared basis of pathogenesis. Trends Immunol. 2013;34(1):22-6.

8.   Howell OW, Reeves CA, Nicholas R, Carassiti D, Radotra B, Gentleman SM, et al. Meningeal inflammation is widespread and linked to cortical pathology in multiple sclerosis. Brain : a journal of neurology. 2011;134(Pt 9):2755-71.

9.   Frischer JM, Bramow S, Dal-Bianco A, Lucchinetti CF, Rauschka H, Schmidbauer M, et al. The relation between inflammation and neurodegeneration in multiple sclerosis brains. Brain : a journal of neurology. 2009;132(Pt 5):1175-89.

10.   Magliozzi R, Howell OW, Reeves C, Roncaroli F, Nicholas R, Serafini B, et al. A Gradient of neuronal loss and meningeal inflammation in multiple sclerosis. Annals of neurology. 2010;68(4):477-93.

11.   Serafini B, Rosicarelli B, Magliozzi R, Stigliano E, Aloisi F. Detection of ectopic B-cell follicles with germinal centers in the meninges of patients with secondary progressive multiple sclerosis. Brain Pathol. 2004;14(2):164-74.

12.   Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Annals of neurology. 2011;69(2):292-302.

13.   Uzawa A, Mori M, Hayakawa S, Masuda S, Kuwabara S. Different responses to interferon beta-1b treatment in patients with neuromyelitis optica and multiple sclerosis. European journal of neurology. 2010;17(5):672-6.

14.   Takahashi T, Fujihara K, Nakashima I, Misu T, Miyazawa I, Nakamura M, et al. Anti-aquaporin-4 antibody is involved in the pathogenesis of NMO: a study on antibody titre. Brain : a journal of neurology. 2007;130(Pt 5):1235-43.

15.   Kitley J, Leite MI, Nakashima I, Waters P, McNeillis B, Brown R, et al. Prognostic factors and disease course in aquaporin-4 antibody-positive patients with neuromyelitis optica spectrum disorder from the United Kingdom and Japan. Brain : a journal of neurology. 2012;135(Pt 6):1834-49.

16.   Nakashima I, Fujihara K, Miyazawa I, Misu T, Narikawa K, Nakamura M, et al. Clinical and MRI features of Japanese patients with multiple sclerosis positive for NMO-IgG. Journal of neurology, neurosurgery, and psychiatry. 2006;77(9):1073-5.

17.   Ramanathan S, Dale RC, Brilot F. Anti-MOG antibody: The history, clinical phenotype, and pathogenicity of a serum biomarker for demyelination. Autoimmunity reviews. 2016;15(4):307-24.

18.   McGraw CA, Lublin FD. Interferon beta and glatiramer acetate therapy. Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics. 2013;10(1):2-18.

19.   Trojano M, Pellegrini F, Paolicelli D, Fuiani A, Zimatore GB, Tortorella C, et al. Real-life impact of early interferon beta therapy in relapsing multiple sclerosis. Annals of neurology. 2009;66(4):513-20.

20.   Goodin DS, Reder AT, Ebers GC, Cutter G, Kremenchutzky M, Oger J, et al. Survival in MS: a randomized cohort study 21 years after the start of the pivotal IFNbeta-1b trial. Neurology. 2012;78(17):1315-22.

21.   Cohen JA, Barkhof F, Comi G, Hartung HP, Khatri BO, Montalban X, et al. Oral fingolimod or intramuscular interferon for relapsing multiple sclerosis. The New England journal of medicine. 2010;362(5):402-15.

22.   Ransohoff RM. Natalizumab for multiple sclerosis. The New England journal of medicine. 2007;356(25):2622-9.

23.   Kappos L, Radue EW, O'Connor P, Polman C, Hohlfeld R, Calabresi P, et al. A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. The New England journal of medicine. 2010;362(5):387-401.

24.   Yednock TA, Cannon C, Fritz LC, Sanchez-Madrid F, Steinman L, Karin N. Prevention of experimental autoimmune encephalomyelitis by antibodies against alpha 4 beta 1 integrin. Nature. 1992;356(6364):63-6.

25.   Rudick RA, Stuart WH, Calabresi PA, Confavreux C, Galetta SL, Radue EW, et al. Natalizumab plus interferon beta-1a for relapsing multiple sclerosis. The New England journal of medicine. 2006;354(9):911-23.

26.   Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. The New England journal of medicine. 2006;354(9):899-910.

27.   Langer-Gould A, Atlas SW, Green AJ, Bollen AW, Pelletier D. Progressive multifocal leukoencephalopathy in a patient treated with natalizumab. The New England journal of medicine. 2005;353(4):375-81.

28.   Ransohoff RM. PML risk and natalizumab: more questions than answers. The Lancet Neurology. 2010;9(3):231-3.

29.   Bloomgren G, Richman S, Hotermans C, Subramanyam M, Goelz S, Natarajan A, et al. Risk of natalizumab-associated progressive multifocal leukoencephalopathy. The New England journal of medicine. 2012;366(20):1870-80.

30.   Kappos L, Antel J, Comi G, Montalban X, O'Connor P, Polman CH, et al. Oral fingolimod (FTY720) for relapsing multiple sclerosis. The New England journal of medicine. 2006;355(11):1124-40.

31.   Gold R, Kappos L, Arnold DL, Bar-Or A, Giovannoni G, Selmaj K, et al. Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. The New England journal of medicine. 2012;367(12):1098-107.

32.   Fox RJ, Miller DH, Phillips JT, Hutchinson M, Havrdova E, Kita M, et al. Placebo-controlled phase 3 study of oral BG-12 or glatiramer in multiple sclerosis. The New England journal of medicine. 2012;367(12):1087-97.

33.   Kappos L, Gold R, Miller DH, MacManus DG, Havrdova E, Limmroth V, et al. Effect of BG-12 on contrast-enhanced lesions in patients with relapsing--remitting multiple sclerosis: subgroup analyses from the phase 2b study. Multiple sclerosis. 2012;18(3):314-21.

34.   Cohen JA, Coles AJ, Arnold DL, Confavreux C, Fox EJ, Hartung HP, et al. Alemtuzumab versus interferon beta 1a as first-line treatment for patients with relapsing-remitting multiple sclerosis: a randomised controlled phase 3 trial. Lancet. 2012;380(9856):1819-28.

35.   Coles AJ, Twyman CL, Arnold DL, Cohen JA, Confavreux C, Fox EJ, et al. Alemtuzumab for patients with relapsing multiple sclerosis after disease-modifying therapy: a randomised controlled phase 3 trial. Lancet. 2012;380(9856):1829-39.

36.   Investigators CT, Coles AJ, Compston DA, Selmaj KW, Lake SL, Moran S, et al. Alemtuzumab vs. interferon beta-1a in early multiple sclerosis. The New England journal of medicine. 2008;359(17):1786-801.

37.   Linker RA, Lee DH, Ryan S, van Dam AM, Conrad R, Bista P, et al. Fumaric acid esters exert neuroprotective effects in neuroinflammation via activation of the Nrf2 antioxidant pathway. Brain : a journal of neurology. 2011;134(Pt 3):678-92.

38.   Scannevin RH, Chollate S, Jung MY, Shackett M, Patel H, Bista P, et al. Fumarates promote cytoprotection of central nervous system cells against oxidative stress via the nuclear factor (erythroid-derived 2)-like 2 pathway. J Pharmacol Exp Ther. 2012;341(1):274-84.

39.   Gerdes S, Shakery K, Mrowietz U. Dimethylfumarate inhibits nuclear binding of nuclear factor kappaB but not of nuclear factor of activated T cells and CCAAT/enhancer binding protein beta in activated human T cells. Br J Dermatol. 2007;156(5):838-42.

40.   Chen H, Assmann JC, Krenz A, Rahman M, Grimm M, Karsten CM, et al. Hydroxycarboxylic acid receptor 2 mediates dimethyl fumarate's protective effect in EAE. The Journal of clinical investigation. 2014;124(5):2188-92.

41.   Huang H, Taraboletti A, Shriver LP. Dimethyl fumarate modulates antioxidant and lipid metabolism in oligodendrocytes. Redox Biol. 2015;5:169-75.

42.   Nibbering PH, Thio B, Zomerdijk TP, Bezemer AC, Beijersbergen RL, van Furth R. Effects of monomethylfumarate on human granulocytes. J Invest Dermatol. 1993;101(1):37-42.

43.   Litjens NH, Burggraaf J, van Strijen E, van Gulpen C, Mattie H, Schoemaker RC, et al. Pharmacokinetics of oral fumarates in healthy subjects. Br J Clin Pharmacol. 2004;58(4):429-32.

44.   Venci JV, Gandhi MA. Dimethyl fumarate (Tecfidera): a new oral agent for multiple sclerosis. Ann Pharmacother. 2013;47(12):1697-702.

45.   Schilling S, Goelz S, Linker R, Luehder F, Gold R. Fumaric acid esters are effective in chronic experimental autoimmune encephalomyelitis and suppress macrophage infiltration. Clin Exp Immunol. 2006;145(1):101-7.

46.   Schimrigk S, Brune N, Hellwig K, Lukas C, Bellenberg B, Rieks M, et al. Oral fumaric acid esters for the treatment of active multiple sclerosis: an open-label,

baseline-controlled pilot study. European journal of neurology. 2006;13(6):604-10.

47.   Kappos L, Gold R, Miller DH, Macmanus DG, Havrdova E, Limmroth V, et al. Efficacy and safety of oral fumarate in patients with relapsing-remitting multiple sclerosis: a multicentre, randomised, double-blind, placebo-controlled phase IIb study. Lancet. 2008;372(9648):1463-72.

48.   Duffy S, So A, Murphy TH. Activation of endogenous antioxidant defenses in neuronal cells prevents free radical-mediated damage. Journal of neurochemistry. 1998;71(1):69-77.

49.   Murphy TH, Yu J, Ng R, Johnson DA, Shen H, Honey CR, et al. Preferential expression of antioxidant response element mediated gene expression in astrocytes. Journal of neurochemistry. 2001;76(6):1670-8.

50.   Morris G, Anderson G, Dean O, Berk M, Galecki P, Martin-Subero M, et al. The glutathione system: a new drug target in neuroimmune disorders. Molecular neurobiology. 2014;50(3):1059-84.

51.   Vandermeeren M, Janssens S, Wouters H, Borghmans I, Borgers M, Beyaert R, et al. Dimethylfumarate is an inhibitor of cytokine-induced nuclear translocation of NF-kappa B1, but not RelA in normal human dermal fibroblast cells. J Invest Dermatol. 2001;116(1):124-30.

52.   Loewe R, Pillinger M, de Martin R, Mrowietz U, Groger M, Holnthoner W, et al. Dimethylfumarate inhibits tumor-necrosis-factor-induced CD62E expression in an NF-kappa B-dependent manner. J Invest Dermatol. 2001;117(6):1363-8.

53.   Wilms H, Sievers J, Rickert U, Rostami-Yazdi M, Mrowietz U, Lucius R. Dimethylfumarate inhibits microglial and astrocytic inflammation by suppressing the synthesis of nitric oxide, IL-1beta, TNF-alpha and IL-6 in an in-vitro model of brain inflammation. Journal of neuroinflammation. 2010;7:30.

54.   Ockenfels HM, Keim-Maas C, Funk R, Nussbaum G, Goos M. Ethanol enhances the IFN-gamma, TGF-alpha and IL-6 secretion in psoriatic co-cultures. Br J Dermatol. 1996;135(5):746-51.

55.   Stoof TJ, Flier J, Sampat S, Nieboer C, Tensen CP, Boorsma DM. The antipsoriatic drug dimethylfumarate strongly suppresses chemokine production in human keratinocytes and peripheral blood mononuclear cells. Br J Dermatol. 2001;144(6):1114-20.

56.   Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. Jama. 2007;298(10):1209-12.

57.   Gafson A, Craner MJ, Matthews PM. Personalised medicine for multiple sclerosis care. Multiple sclerosis. 2016.

58.   Fisniku LK, Brex PA, Altmann DR, Miszkiel KA, Benton CE, Lanyon R, et al. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. Brain : a journal of neurology. 2008;131(Pt 3):808-17.

59.   Tintore M, Rovira A, Rio J, Otero-Romero S, Arrambide G, Tur C, et al. Defining high, medium and low impact prognostic factors for developing multiple sclerosis. Brain : a journal of neurology. 2015;138(Pt 7):1863-74.

60.   Tintore M, Rovira A, Arrambide G, Mitjana R, Rio J, Auger C, et al. Brainstem lesions in clinically isolated syndromes. Neurology. 2010;75(21):1933-8.

61.   Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. Annals of neurology. 2007;61(6):504-13.

62.   Sawcer S, Franklin RJ, Ban M. Multiple sclerosis genetics. The Lancet Neurology. 2014;13(7):700-9.

63.   Kappos L, Freedman MS, Polman CH, Edan G, Hartung HP, Miller DH, et al. Effect of early versus delayed interferon beta-1b treatment on disability after a first clinical event suggestive of multiple sclerosis: a 3-year follow-up analysis of the BENEFIT study. Lancet. 2007;370(9585):389-97.

64.   Karim ME, Gustafson P, Petkau J, Zhao Y, Shirani A, Kingwell E, et al. Marginal structural Cox models for estimating the association between beta-interferon exposure and disease progression in a multiple sclerosis cohort. American journal of epidemiology. 2014;180(2):160-71.

65.   Zintzaras E, Doxani C, Mprotsis T, Schmid CH, Hadjigeorgiou GM. Network analysis of randomized controlled trials in multiple sclerosis. Clinical therapeutics. 2012;34(4):857-69 e9.

66.   Kalincik T, Horakova D, Spelman T, Jokubaitis V, Trojano M, Lugaresi A, et al. Switch to natalizumab versus fingolimod in active relapsing-remitting multiple sclerosis. Annals of neurology. 2015;77(3):425-35.

67.   Sominanda A, Hillert J, Fogdell-Hahn A. In vivo bioactivity of interferon-beta in multiple sclerosis patients with neutralising antibodies is titre-dependent. Journal of neurology, neurosurgery, and psychiatry. 2008;79(1):57-62.

68.   Calabresi PA, Giovannoni G, Confavreux C, Galetta SL, Havrdova E, Hutchinson M, et al. The incidence and significance of anti-natalizumab antibodies: results from AFFIRM and SENTINEL. Neurology. 2007;69(14):1391-403.

69.   Sormani MP, Bruzzi P. MRI lesions as a surrogate for relapses in multiple sclerosis: a meta-analysis of randomised trials. The Lancet Neurology. 2013;12(7):669-76.

70.   Disanto G, Adiutori R, Dobson R, Martinelli V, Dalla Costa G, Runia T, et al. Serum neurofilament light chain levels are increased in patients with a clinically isolated syndrome. Journal of neurology, neurosurgery, and psychiatry. 2016;87(2):126-9.

71.   Petzold A. The prognostic value of CSF neurofilaments in multiple sclerosis at 15-year follow-up. Journal of neurology, neurosurgery, and psychiatry. 2015;86(12):1388-90.

72.   Amor S, van der Star BJ, Bosca I, Raffel J, Gnanapavan S, Watchorn J, et al. Neurofilament light antibodies in serum reflect response to natalizumab treatment in multiple sclerosis. Multiple sclerosis. 2014;20(10):1355-62.

73.   Kroksveen AC, Opsahl JA, Guldbrandsen A, Myhr KM, Oveland E, Torkildsen O, et al. Cerebrospinal fluid proteomics in multiple sclerosis. Biochimica et biophysica acta. 2015;1854(7):746-56.

74.   Ioannidis JP. Indirect comparisons: the mesh and mess of clinical trials. Lancet. 2006;368(9546):1470-2.

75.   Spelman T, Kalincik T, Zhang A, Pellegrini F, Wiendl H, Kappos L, et al. Comparative efficacy of switching to natalizumab in active multiple sclerosis. Ann Clin Transl Neurol. 2015;2(4):373-87.

76.   He A, Spelman T, Jokubaitis V, Havrdova E, Horakova D, Trojano M, et al. Comparison of switch to fingolimod or interferon beta/glatiramer acetate in active multiple sclerosis. JAMA neurology. 2015;72(4):405-13.

77.   Trojano M, Tintore M, Montalban X, Hillert J, Kalincik T, Iaffaldano P, et al. Treatment decisions in multiple sclerosis - insights from real-world observational studies. Nature reviews Neurology. 2017;13(2):105-18.

78.   Willis JC, Lord GM. Immune biomarkers: the promises and pitfalls of personalized medicine. Nature reviews Immunology. 2015;15(5):323-9.

79.   Schilsky RL. Personalized medicine in oncology: the future is now. Nature reviews Drug discovery. 2010;9(5):363-6.

80.   Watanabe S, Minegishi Y, Yoshizawa H, Maemondo M, Inoue A, Sugawara S, et al. Effectiveness of gefitinib against non-small-cell lung cancer with the uncommon EGFR mutations G719X and L861Q. J Thorac Oncol. 2014;9(2):189-94.

81.   Kim S, Lee J, Oh SJ, Nam SJ, Lee JE. Differential effect of EGFR inhibitors on tamoxifen-resistant breast cancer cells. Oncol Rep. 2015;34(3):1613-9.

82.   Van Cutsem E, Kohne CH, Hitre E, Zaluski J, Chang Chien CR, Makhson A, et al. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. The New England journal of medicine. 2009;360(14):1408-17.

83.   Rodriguez-Antona C, Taron M. Pharmacogenomic biomarkers for personalized cancer treatment. Journal of internal medicine. 2015;277(2):201-17.

84.   Maronas O, Latorre A, Dopazo J, Pirmohamed M, Rodriguez-Antona C, Siest G, et al. Progress in pharmacogenetics: consortiums and new strategies. Drug Metab Pers Ther. 2016;31(1):17-23.

85.   Keefe DM, Bateman EH. Tumor control versus adverse events with targeted anticancer therapies. Nat Rev Clin Oncol. 2011;9(2):98-109.

86.   Vaidya B, Gupta V. Novel therapeutic approaches for pulmonary arterial hypertension: Unique molecular targets to site-specific drug delivery. J Control Release. 2015;211:118-33.

87.   Jones BJ, Bloom SR. The New Era of Drug Therapy for Obesity: The Evidence and the Expectations. Drugs. 2015;75(9):935-45.

88.   Lovly CM, Shaw AT. Molecular pathways: resistance to kinase inhibitors and implications for therapeutic strategies. Clin Cancer Res. 2014;20(9):2249-56.

89.   Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. Science translational medicine. 2011;3(75):75ra26.

90.   Suda K, Mizuuchi H, Maehara Y, Mitsudomi T. Acquired resistance mechanisms to tyrosine kinase inhibitors in lung cancer with activating epidermal growth factor receptor mutation--diversity, ductility, and destiny. Cancer Metastasis Rev. 2012;31(3-4):807-14.

91.   Isozaki H, Takigawa N, Kiura K. Mechanisms of Acquired Resistance to ALK Inhibitors and the Rationale for Treating ALK-positive Lung Cancer. Cancers (Basel). 2015;7(2):763-83.

92.   Nurwidya F, Takahashi F, Murakami A, Kobayashi I, Kato M, Shukuya T, et al. Acquired resistance of non-small cell lung cancer to epidermal growth factor receptor tyrosine kinase inhibitors. Respir Investig. 2014;52(2):82-91.

93.   Takezawa K, Pirazzoli V, Arcila ME, Nebhan CA, Song X, de Stanchina E, et al. HER2 amplification: a potential mechanism of acquired resistance to EGFR inhibition in EGFR-mutant lung cancers that lack the second-site EGFRT790M mutation. Cancer Discov. 2012;2(10):922-33.

94.   Aithal GP, Day CP, Kesteven PJ, Daly AK. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. Lancet. 1999;353(9154):717-9.

95.   Hegen H, Auer M, Deisenhammer F. Pharmacokinetic considerations in the treatment of multiple sclerosis with interferon-beta. Expert opinion on drug metabolism & toxicology. 2015;11(12):1803-19.

96.   Wu K, Mercier F, David OJ, Schmouder RL, Looby M. Population pharmacokinetics of fingolimod phosphate in healthy participants. Journal of clinical pharmacology. 2012;52(7):1054-68.

97.   Jersild C, Svejgaard A, Fog T. HL-A antigens and multiple sclerosis. Lancet. 1972;1(7762):1240-1.

98.   International Multiple Sclerosis Genetics C, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, et al. Risk alleles for multiple sclerosis identified by a genomewide study. The New England journal of medicine. 2007;357(9):851-62.

99.   Kantarci OH, Hebrink DD, Achenbach SJ, Atkinson EJ, Waliszewska A, Buckle G, et al. CTLA4 is associated with susceptibility to multiple sclerosis. Journal of neuroimmunology. 2003;134(1-2):133-41.

100.   Harbo HF, Celius EG, Vartdal F, Spurkland A. CTLA4 promoter and exon 1 dimorphisms in multiple sclerosis. Tissue antigens. 1999;53(1):106-10.

101.   Ligers A, Xu C, Saarinen S, Hillert J, Olerup O. The CTLA-4 gene is associated with multiple sclerosis. Journal of neuroimmunology. 1999;97(1-2):182-90.

102.   Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science (New York, NY). 2008;322(5903):881-8.

103.   Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53.

104.   Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.

105.   Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. Science (New York, NY). 2016;352(6292):1413-6.

106.   Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. Cell. 2014;157(1):77-94.

107.   Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nat Rev Genet. 2015;16(4):197-212.

108.   Ramanathan M, Weinstock-Guttman B, Nguyen LT, Badgett D, Miller C, Patrick K, et al. In vivo gene expression revealed by cDNA arrays: the pattern in relapsing-remitting multiple sclerosis patients compared with normal subjects. Journal of neuroimmunology. 2001;116(2):213-9.

109.   Lock C, Hermans G, Pedotti R, Brendolan A, Schadt E, Garren H, et al. Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. Nature medicine. 2002;8(5):500-8.

110.   Whitney LW, Becker KG, Tresser NJ, Caballero-Ramos CI, Munson PJ, Prabhu VV, et al. Analysis of gene expression in mutiple sclerosis lesions using cDNA microarrays. Annals of neurology. 1999;46(3):425-8.

111.   Corvol JC, Pelletier D, Henry RG, Caillier SJ, Wang J, Pappas D, et al. Abrogation of T cell quiescence characterizes patients at high risk for multiple sclerosis after the initial neurological event. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(33):11839-44.

112.   Satoh J, Nanri Y, Tabunoki H, Yamamura T. Microarray analysis identifies a set of CXCR3 and CCR2 ligand chemokines as early IFNbeta-responsive genes in peripheral blood lymphocytes in vitro: an implication for IFNbeta-related adverse effects in multiple sclerosis. BMC Neurol. 2006;6:18.

113.   Baranzini SE. Gene expression profiling in MS: a fulfilled promise? Multiple sclerosis. 2013;19(14):1813-4.

114.   Johnson MR, Behmoaras J, Bottolo L, Krishnan ML, Pernhorst K, Santoscoy PL, et al. Systems genetics identifies Sestrin 3 as a regulator of a proconvulsant gene network in human epileptic hippocampus. Nat Commun. 2015;6:6031.

115.   Tesson BM, Breitling R, Jansen RC. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC Bioinformatics. 2010;11:497.

116.   Marbach D, Costello JC, Kuffner R, Vega NM, Prill RJ, Camacho DM, et al. Wisdom of crowds for robust gene network inference. Nature methods. 2012;9(8):796-804.

117.   Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63.

118.   Zhang W, Yu Y, Hertwig F, Thierry-Mieg J, Zhang W, Thierry-Mieg D, et al. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. Genome biology. 2015;16:133.

119.   Waubant E, Goodkin D, Bostrom A, Bacchetti P, Hietpas J, Lindberg R, et al. IFNbeta lowers MMP-9/TIMP-1 ratio, which predicts new enhancing lesions in patients with SPMS. Neurology. 2003;60(1):52-7.

120.   Fainardi E, Castellazzi M, Bellini T, Manfrinato MC, Baldi E, Casetta I, et al. Cerebrospinal fluid and serum levels and intrathecal production of active matrix metalloproteinase-9 (MMP-9) as markers of disease activity in patients with multiple sclerosis. Multiple sclerosis. 2006;12(3):294-301.

121.   Benesova Y, Vasku A, Novotna H, Litzman J, Stourac P, Beranek M, et al. Matrix metalloproteinase-9 and matrix metalloproteinase-2 as biomarkers of various courses in multiple sclerosis. Multiple sclerosis. 2009;15(3):316-22.

122.   Bornsen L, Khademi M, Olsson T, Sorensen PS, Sellebjerg F. Osteopontin concentrations are increased in cerebrospinal fluid during attacks of multiple sclerosis. Multiple sclerosis. 2011;17(1):32-42.

123.   Comabella M, Pericot I, Goertsches R, Nos C, Castillo M, Blas Navarro J, et al. Plasma osteopontin levels in multiple sclerosis. Journal of neuroimmunology. 2005;158(1-2):231-9.

124.   Kivisakk P, Healy BC, Francois K, Gandhi R, Gholipour T, Egorova S, et al. Evaluation of circulating osteopontin levels in an unselected cohort of patients with multiple sclerosis: relevance for biomarker development. Multiple sclerosis. 2014;20(4):438-44.

125.   Tzartos JS, Friese MA, Craner MJ, Palace J, Newcombe J, Esiri MM, et al. Interleukin-17 production in central nervous system-infiltrating T cells and glial

cells is associated with active disease in multiple sclerosis. Am J Pathol. 2008;172(1):146-55.

126.   Duarte IF, Diaz SO, Gil AM. NMR metabolomics of human blood and urine in disease research. J Pharm Biomed Anal. 2014;93:17-26.

127.   Scrivo R, Casadei L, Valerio M, Priori R, Valesini G, Manetti C. Metabolomics approach in allergic and rheumatic diseases. Curr Allergy Asthma Rep. 2014;14(6):445.

128.   Manna SK, Patterson AD, Yang Q, Krausz KW, Idle JR, Fornace AJ, et al. UPLC-MS-based urine metabolomics reveals indole-3-lactic acid and phenyllactic acid as conserved biomarkers for alcohol-induced liver disease in the Ppara-null mouse model. J Proteome Res. 2011;10(9):4120-33.

129.   Blanchet L, Smolinska A, Attali A, Stoop MP, Ampt KA, van Aken H, et al. Fusion of metabolomics and proteomics data for biomarkers discovery: case study on the experimental autoimmune encephalomyelitis. BMC Bioinformatics. 2011;12:254.

130.   Kaddurah-Daouk R, Kristal BS, Weinshilboum RM. Metabolomics: a global biochemical approach to drug response and disease. Annu Rev Pharmacol Toxicol. 2008;48:653-83.

131.   Wishart DS. Applications of metabolomics in drug discovery and development. Drugs R D. 2008;9(5):307-22.

132.   Lehotay DC, Hall P, Lepage J, Eichhorst JC, Etter ML, Greenberg CR. LC-MS/MS progress in newborn screening. Clin Biochem. 2011;44(1):21-31.

133.   Chace DH, Spitzer AR. Altered metabolism and newborn screening using tandem mass spectrometry: lessons learned from the bench to bedside. Curr Pharm Biotechnol. 2011;12(7):965-75.

134.   Kaddurah-Daouk R, Weinshilboum RM, Pharmacometabolomics Research N. Pharmacometabolomics: implications for clinical pharmacology and systems pharmacology. Clin Pharmacol Ther. 2014;95(2):154-67.

135.   Sallustio BC. LC-MS/MS for immunosuppressant therapeutic drug monitoring. Bioanalysis. 2010;2(6):1141-53.

136.   Ponnayyan Sulochana S, Sharma K, Mullangi R, Sukumaran SK. Review of the validated HPLC and LC-MS/MS methods for determination of drugs used in clinical practice for Alzheimer's disease. Biomed Chromatogr. 2014;28(11):1431-90.

137.   Simone IL, Federico F, Trojano M, Tortorella C, Liguori M, Giannini P, et al. High resolution proton MR spectroscopy of cerebrospinal fluid in MS patients. Comparison with biochemical changes in demyelinating plaques. Journal of the neurological sciences. 1996;144(1-2):182-90.

138.   Aasly J, Garseth M, Sonnewald U, Zwart JA, White LR, Unsgard G. Cerebrospinal fluid lactate and glutamine are reduced in multiple sclerosis. Acta neurologica Scandinavica. 1997;95(1):9-12.

139.   Dickens AM, Larkin JR, Griffin JL, Cavey A, Matthews L, Turner MR, et al. A type 2 biomarker separates relapsing-remitting from secondary progressive multiple sclerosis. Neurology. 2014;83(17):1492-9.

140.   Mehrpour M, Kyani A, Tafazzoli M, Fathi F, Joghataie MT. A metabonomics investigation of multiple sclerosis by nuclear magnetic resonance. Magn Reson Chem. 2013;51(2):102-9.

141.   Moussallieh FM, Elbayed K, Chanson JB, Rudolf G, Piotto M, De Seze J, et al. Serum analysis by 1H nuclear magnetic resonance spectroscopy: a new tool for

distinguishing neuromyelitis optica from multiple sclerosis. Multiple sclerosis. 2014;20(5):558-65.

142.   van der Mei I, Lucas RM, Taylor BV, Valery PC, Dwyer T, Kilpatrick TJ, et al. Population attributable fractions and joint effects of key risk factors for multiple sclerosis. Multiple sclerosis. 2016;22(4):461-9.

143.   Ascherio A, Munger KL, Lunemann JD. The initiation and prevention of multiple sclerosis. Nature reviews Neurology. 2012;8(11):602-12.

144.   Lucas RM, Byrne SN, Correale J, Ilschner S, Hart PH. Ultraviolet radiation, vitamin D and multiple sclerosis. Neurodegener Dis Manag. 2015;5(5):413-24.

145.   Westerlind H, Imrell K, Ramanujam R, Myhr KM, Celius EG, Harbo HF, et al. Identity-by-descent mapping in a Scandinavian multiple sclerosis cohort. Eur J Hum Genet. 2015;23(5):688-92.

146.   Westerlind H, Ramanujam R, Uvehag D, Kuja-Halkola R, Boman M, Bottai M, et al. Modest familial risks for multiple sclerosis: a registry-based study of the population of Sweden. Brain : a journal of neurology. 2014;137(Pt 3):770-8.

147.   Koch-Henriksen N, Sorensen PS. The changing demographic pattern of multiple sclerosis epidemiology. The Lancet Neurology. 2010;9(5):520-32.

148.   Westerlind H, Bostrom I, Stawiarz L, Landtblom AM, Almqvist C, Hillert J. New data identify an increasing sex ratio of multiple sclerosis in Sweden. Multiple sclerosis. 2014;20(12):1578-83.

149.   Ponsonby AL, Lucas RM, van der Mei IA, Dear K, Valery PC, Pender MP, et al. Offspring number, pregnancy, and risk of a first clinical demyelinating event: the AusImmune Study. Neurology. 2012;78(12):867-74.

150.   Magyari M, Koch-Henriksen N, Pfleger CC, Sorensen PS. Reproduction and the risk of multiple sclerosis. Multiple sclerosis. 2013;19(12):1604-9.

151.   Villoslada P, Steinman L, Baranzini SE. Systems biology and its application to the understanding of neurological diseases. Annals of neurology. 2009;65(2):124-39.

152.   Kholodenko BN, Hancock JF, Kolch W. Signalling ballet in space and time. Nat Rev Mol Cell Biol. 2010;11(6):414-26.

153.   Morris MK, Chi A, Melas IN, Alexopoulos LG. Phosphoproteomics in drug discovery. Drug Discov Today. 2014;19(4):425-32.

154.   Saez-Rodriguez J, Alexopoulos LG, Zhang M, Morris MK, Lauffenburger DA, Sorger PK. Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. Cancer Res. 2011;71(16):5400-11.

155.   Hart ST, T. Sample Size for RNA-Seq and similar studies. 2016.

156.   Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 1998;8(3):175-85.

157.   Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2010;38(6):1767-71.

158.   F K. Trim Galore!

159.   DRAGEN.

160.   Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9.

161.   Fonseca NA, Marioni J, Brazma A. RNA-Seq gene profiling--a systematic empirical comparison. PloS one. 2014;9(9):e107026.

162. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

163. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42(Database issue):D68-73.

164. Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32(2):292-4.

165. Evans AM BB, Michell MW, Robindon RJ, Dai H, Stewart SJ, DeHaven CD, Miller LAD. High Resolution Mass Spectrometry Improves Data Quantity and Quality
as Compared to Unit Mass Resolution Mass Spectrometry in High-Throughput Profiling Metabolomics. Metabolomics. 2014;4(2):1-7.

166. Lewis MR, Pearce JT, Spagou K, Green M, Dona AC, Yuen AH, et al. Development and Application of Ultra-Performance Liquid Chromatography-TOF MS for Precision Large Scale Urinary Metabolic Phenotyping. Anal Chem. 2016;88(18):9004-13.

167. Dona AC, Jimenez B, Schafer H, Humpfer E, Spraul M, Lewis MR, et al. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Anal Chem. 2014;86(19):9887-94.

168. Bruker IVDr Lipoprotein Subclass AnalysisBruker IVDr Lipoprotein Subclass Analysis [Available from: https://www.bruker.com/products/mr/nmr-preclinical-screening/lipoprotein-subclass-analysis.html.

169. Jackson JE. A Users' Guide to Principal Components. 1991.

170. Wall ME, Dyck PA, Brettin TS. SVDMAN--singular value decomposition analysis of microarray data. Bioinformatics. 2001;17(6):566-8.

171. Reich D, Price AL, Patterson N. Principal component analysis of genetic data. Nat Genet. 2008;40(5):491-2.

172. Schwabbauer ML. Use of the latent image technique to develop and evaluate problem-solving skills. Am J Med Technol. 1975;41(12):457-62.

173. geladi PMD. Linearization and scatter-correction for near-infared reflectance spectra of meat. Appl Spectrosc. 1985;39.

174. Eriksson L, Rosen J, Johansson E, Trygg J. Orthogonal PLS (OPLS) Modeling for Improved Analysis and Interpretation in Drug Design. Mol Inform. 2012;31(6-7):414-9.

175. Bro R, Kjeldahl K, Smilde AK, Kiers HA. Cross-validation of component models: a critical look at current methods. Anal Bioanal Chem. 2008;390(5):1241-51.

176. Smit S, van Breemen MJ, Hoefsloot HC, Smilde AK, Aerts JM, de Koster CG. Assessing the statistical validity of proteomics based biomarkers. Anal Chim Acta. 2007;592(2):210-7.

177. Alter O, Golub GH. Singular value decomposition of genome-scale mRNA lengths distribution reveals asymmetry in RNA gel electrophoresis band broadening. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(32):11828-33.

178. Le Cao KA, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. BMC Bioinformatics. 2011;12:253.

179.   Le Cao KA, Rossouw D, Robert-Granie C, Besse P. A sparse PLS for variable selection when integrating omics data. Stat Appl Genet Mol Biol. 2008;7(1):Article 35.
180.   Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32.
181.   Altman N, Krzywinski M. Simple linear regression. Nature methods. 2015;12(11):999-1000.
182.   Krzywinski M, Altman N. Multiple linear regression. Nature methods. 2015;12(12):1103-4.
183.   lever. Points of Significance: Model selection and overfitting. Natue methods. 2016;13:703-4.
184.   Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1996;58(1):267-88.
185.   zou h. Regularization and variable selection via the elastic net
. Journal of the Royal Statistical Society: Series B (Statistical Methodology)
. 2005;67.
186.   Akaike. Information theory as an extension of the maximum likelihood principle. 1973.
187.   Patsopoulos NA, Bayer Pharma MSGWG, Steering Committees of Studies Evaluating I-b, a CCRA, Consortium AN, GeneMsa, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Annals of neurology. 2011;70(6):897-912.
188.   International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9.
189.   Dutta R. Gene expression changes underlying cortical pathology: clues to understanding neurological disability in multiple sclerosis. Multiple sclerosis. 2013;19(10):1249-54.
190.   Huang Q, Xiao B, Ma X, Qu M, Li Y, Nagarkatti P, et al. MicroRNAs associated with the pathogenesis of multiple sclerosis. Journal of neuroimmunology. 2016;295-296:148-61.
191.   Fawaz CN, Makki IS, Kazan JM, Gebara NY, Andary FS, Itani MM, et al. Neuroproteomics and microRNAs studies in multiple sclerosis: transforming research and clinical knowledge in biomarker research. Expert Rev Proteomics. 2015;12(6):637-50.
192.   Sanchez-Chaparro MM, Rodriguez-Sanchez IP, Barrera-Saldana HA, Martinez-Villarreal LE, Resendez-Perez D, Gamez-Escobedo IA. [MicroRNAs and their neuroimmunoregulator mechanisms in multiple sclerosis. Development of biomarkers for diagnosis]. Rev Neurol. 2015;60(12):562-71.
193.   Lim CK, Bilgin A, Lovejoy DB, Tan V, Bustamante S, Taylor BV, et al. Kynurenine pathway metabolomics predicts and provides mechanistic insight into multiple sclerosis progression. Sci Rep. 2017;7:41473.
194.   Zhornitsky S, McKay KA, Metz LM, Teunissen CE, Rangachari M. Cholesterol and markers of cholesterol turnover in multiple sclerosis: relationship with disease outcomes. Multiple sclerosis and related disorders. 2016;5:53-65.

195.   Tettey P, Simpson S, Jr., Taylor BV, van der Mei IA. Vascular comorbidities in the onset and progression of multiple sclerosis. Journal of the neurological sciences. 2014;347(1-2):23-33.

196.   Marrie RA, Rudick R, Horwitz R, Cutter G, Tyry T, Campagnolo D, et al. Vascular comorbidity is associated with more rapid disability progression in multiple sclerosis. Neurology. 2010;74(13):1041-7.

197.   Feingold KR, Grunfeld C. The Effect of Inflammation and Infection on Lipids and Lipoproteins. In: De Groot LJ, Beck-Peccoz P, Chrousos G, Dungan K, Grossman A, Hershman JM, et al., editors. Endotext. South Dartmouth (MA)2000.

198.   Malmendier CL, Lontie JF, Sculier JP, Dubois DY. Modifications of plasma lipids, lipoproteins and apolipoproteins in advanced cancer patients treated with recombinant interleukin-2 and autologous lymphokine-activated killer cells. Atherosclerosis. 1988;73(2-3):173-80.

199.   Weinstock-Guttman B, Zivadinov R, Mahfooz N, Carl E, Drake A, Schneider J, et al. Serum lipid profiles are associated with disability and MRI outcomes in multiple sclerosis. Journal of neuroinflammation. 2011;8:127.

200.   Tettey P, Simpson S, Jr., Taylor B, Blizzard L, Ponsonby AL, Dwyer T, et al. An adverse lipid profile is associated with disability and progression in disability, in people with MS. Multiple sclerosis. 2014;20(13):1737-44.

201.   Mandoj C, Renna R, Plantone D, Sperduti I, Cigliana G, Conti L, et al. Anti-annexin antibodies, cholesterol levels and disability in multiple sclerosis. Neuroscience letters. 2015;606:156-60.

202.   Chataway J, Schuerer N, Alsanousi A, Chan D, MacManus D, Hunter K, et al. Effect of high-dose simvastatin on brain atrophy and disability in secondary progressive multiple sclerosis (MS-STAT): a randomised, placebo-controlled, phase 2 trial. Lancet. 2014;383(9936):2213-21.

203.   Hafiane A, Genest J. High density lipoproteins: Measurement techniques and potential biomarkers of cardiovascular risk. BBA Clin. 2015;3:175-88.

204.   O'Connell KE, Mok T, Sweeney B, Ryan AM, Dev KK. The use of cytokine signature patterns: separating drug naive, interferon and natalizumab-treated multiple sclerosis patients. Autoimmunity. 2014;47(8):505-11.

205.   Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.

206.   benjamini Y HY. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Stat Soc Ser B Methodol. 1995;57:289-300.

207.   Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11(9):1650-67.

208.   Compston A, Coles A. Multiple sclerosis. Lancet. 2002;359(9313):1221-31.

209.   Tu YK, Kellett M, Clerehugh V, Gilthorpe MS. Problems of correlations between explanatory variables in multiple regression analyses in the dental literature. Br Dent J. 2005;199(7):457-61.

210.   Lindeman RH MP, Gold RZ. Introduction to Bivariate and Multivariate Analysis. Scott, Foresman, Glenview, IL1980.

211.   Gromping U. Relative Importance for Linear Regression in R: The Package relaimpo. Journal of Statistical Software. 2006;17(1):1-27.

212.   H A. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974;19:716-23.

213.   Ripley WNVaBD. Modern Applied Statistics with S. 2002.

214.   Jake Lever MK, Naomi Altman. Points of Significance: Regularization. Nature methods. 2016;13:803-4.

215.   Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1-22.

216.   Le Cao KA, Gonzalez I, Dejean S. integrOmics: an R package to unravel relationships between two omics datasets. Bioinformatics. 2009;25(21):2855-6.

217.   Friese MA, Jakobsen KB, Friis L, Etzensperger R, Craner MJ, McMahon RM, et al. Opposing effects of HLA class I molecules in tuning autoreactive CD8+ T cells in multiple sclerosis. Nature medicine. 2008;14(11):1227-35.

218.   Kaur G, Trowsdale J, Fugger L. Natural killer cells and their receptors in multiple sclerosis. Brain : a journal of neurology. 2013;136(Pt 9):2657-76.

219.   Baranzini SE. The role of antiproliferative gene Tob1 in the immune system. Clin Exp Neuroimmunol. 2014;5(2):132-6.

220.   Didonna A, Cekanaviciute E, Oksenberg JR, Baranzini SE. Immune cell-specific transcriptional profiling highlights distinct molecular pathways controlled by Tob1 upon experimental autoimmune encephalomyelitis. Sci Rep. 2016;6:31603.

221.   Schulze-Topphoff U, Casazza S, Varrin-Doyer M, Pekarek K, Sobel RA, Hauser SL, et al. Tob1 plays a critical role in the activation of encephalitogenic T cells in CNS autoimmunity. The Journal of experimental medicine. 2013;210(7):1301-9.

222.   Villoslada P, Alonso C, Agirrezabal I, Kotelnikova E, Zubizarreta I, Pulido-Valdeolivas I, et al. Metabolomic signatures associated with disease severity in multiple sclerosis. Neurol Neuroimmunol Neuroinflamm. 2017;4(2):e321.

223.   Ratzer R, Sondergaard HB, Christensen JR, Bornsen L, Borup R, Sorensen PS, et al. Gene expression analysis of relapsing-remitting, primary progressive and secondary progressive multiple sclerosis. Multiple sclerosis. 2013;19(14):1841-8.

224.   Achiron A, Gurevich M, Friedman N, Kaminski N, Mandel M. Blood transcriptional signatures of multiple sclerosis: unique gene expression of disease activity. Annals of neurology. 2004;55(3):410-7.

225.   Vukusic S, Confavreux C. Natural history of multiple sclerosis: risk factors and prognostic indicators. Current opinion in neurology. 2007;20(3):269-74.

226.   Newcombe J, Li H, Cuzner ML. Low density lipoprotein uptake by macrophages in multiple sclerosis plaques: implications for pathogenesis. Neuropathology and applied neurobiology. 1994;20(2):152-62.

227.   Jiang X, Guo M, Su J, Lu B, Ma D, Zhang R, et al. Simvastatin blocks blood-brain barrier disruptions induced by elevated cholesterol both in vivo and in vitro. Int J Alzheimers Dis. 2012;2012:109324.

228.   Hardardottir I, Grunfeld C, Feingold KR. Effects of endotoxin on lipid metabolism. Biochem Soc Trans. 1995;23(4):1013-8.

229.   Mannisto VT, Simonen M, Soininen P, Tiainen M, Kangas AJ, Kaminska D, et al. Lipoprotein subclass metabolism in nonalcoholic steatohepatitis. J Lipid Res. 2014;55(12):2676-84.

230.   den Hartigh LJ, Altman R, Norman JE, Rutledge JC. Postprandial VLDL lipolysis products increase monocyte adhesion and lipid droplet formation via

activation of ERK2 and NFkappaB. Am J Physiol Heart Circ Physiol. 2014;306(1):H109-20.

231.   Sampedro MC, Motran C, Gruppi A, Kivatinitz SC. VLDL modulates the cytokine secretion profile to a proinflammatory pattern. Biochem Biophys Res Commun. 2001;285(2):393-9.

232.   Li WH, Tanimura M, Luo CC, Datta S, Chan L. The apolipoprotein multigene family: biosynthesis, structure, structure-function relationships, and evolution. J Lipid Res. 1988;29(3):245-71.

233.   Watanabe J, Charles-Schoeman C, Miao Y, Elashoff D, Lee YY, Katselis G, et al. Proteomic profiling following immunoaffinity capture of high-density lipoprotein: association of acute-phase proteins and complement factors with proinflammatory high-density lipoprotein in rheumatoid arthritis. Arthritis and rheumatism. 2012;64(6):1828-37.

234.   Noorbakhsh F, Baker GB, Power C. Allopregnanolone and neuroinflammation: a focus on multiple sclerosis. Front Cell Neurosci. 2014;8:134.

235.   Lee SK, Kim HJ, Na SY, Kim TS, Choi HS, Im SY, et al. Steroid receptor coactivator-1 coactivates activating protein-1-mediated transactivations through interaction with the c-Jun and c-Fos subunits. J Biol Chem. 1998;273(27):16651-4.

236.   Bonetti B, Stegagno C, Cannella B, Rizzuto N, Moretto G, Raine CS. Activation of NF-kappaB and c-jun transcription factors in multiple sclerosis lesions. Implications for oligodendrocyte pathology. Am J Pathol. 1999;155(5):1433-8.

237.   Boullerne AI, Skias D, Hartman EM, Testai FD, Kalinin S, Polak PE, et al. A single-nucleotide polymorphism in serine-threonine kinase 11, the gene encoding liver kinase B1, is a risk factor for multiple sclerosis. ASN Neuro. 2015;7(1).

238.   Loewe R, Holnthoner W, Groger M, Pillinger M, Gruber F, Mechtcheriakova D, et al. Dimethylfumarate inhibits TNF-induced nuclear entry of NF-kappa B/p65 in human endothelial cells. Journal of immunology (Baltimore, Md : 1950). 2002;168(9):4781-7.

239.   Wallbrecht K, Drick N, Hund AC, Schon MP. Downregulation of endothelial adhesion molecules by dimethylfumarate, but not monomethylfumarate, and impairment of dynamic lymphocyte-endothelial cell interactions. Exp Dermatol. 2011;20(12):980-5.

240.   Ockenfels HM, Schultewolter T, Ockenfels G, Funk R, Goos M. The antipsoriatic agent dimethylfumarate immunomodulates T-cell cytokine secretion and inhibits cytokines of the psoriatic cytokine network. Br J Dermatol. 1998;139(3):390-5.

241.   Lee DH, Gold R, Linker RA. Mechanisms of oxidative damage in multiple sclerosis and neurodegenerative diseases: therapeutic modulation via fumaric acid esters. International journal of molecular sciences. 2012;13(9):11783-803.

242.   Zhao G, Liu Y, Fang J, Chen Y, Li H, Gao K. Dimethyl fumarate inhibits the expression and function of hypoxia-inducible factor-1alpha (HIF-1alpha). Biochem Biophys Res Commun. 2014;448(3):303-7.

243.   Brennan MS, Matos MF, Richter KE, Li B, Scannevin RH. The NRF2 transcriptional target, OSGIN1, contributes to monomethyl fumarate-mediated cytoprotection in human astrocytes. Sci Rep. 2017;7:42054.

244.   Reily MD, Tymiak AA. Metabolomics in the pharmaceutical industry. Drug Discov Today Technol. 2015;13:25-31.

245.   Bhargava P, Calabresi PA. Metabolomics in multiple sclerosis. Multiple sclerosis. 2016;22(4):451-60.

246.   Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. Science (New York, NY). 2002;297(5586):1551-5.

247.   Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105-11.

248.   Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nature reviews Neuroscience. 2013;14(5):365-76.

249.   Sun WP, Zhai MZ, Li D, Zhou Y, Chen NN, Guo M, et al. Comparison of the effects of nicotinic acid and nicotinamide degradation on plasma betaine and choline levels. Clin Nutr. 2016.

250.   Ha CM, Park S, Choi YK, Jeong JY, Oh CJ, Bae KH, et al. Activation of Nrf2 by dimethyl fumarate improves vascular calcification. Vascul Pharmacol. 2014;63(1):29-36.

251.   Lin SX, Lisi L, Dello Russo C, Polak PE, Sharp A, Weinberg G, et al. The anti-inflammatory effects of dimethyl fumarate in astrocytes involve glutathione and haem oxygenase-1. ASN Neuro. 2011;3(2).

252.   Han R, Xiao J, Zhai H, Hao J. Dimethyl fumarate attenuates experimental autoimmune neuritis through the nuclear factor erythroid-derived 2-related factor 2/hemoxygenase-1 pathway by altering the balance of M1/M2 macrophages. Journal of neuroinflammation. 2016;13(1):97.

253.   Tahvili S, Zandieh B, Amirghofran Z. The effect of dimethyl fumarate on gene expression and the level of cytokines related to different T helper cell subsets in peripheral blood mononuclear cells of patients with psoriasis. Int J Dermatol. 2015;54(7):e254-60.

254.   Hund AC, Lockmann A, Schon MP. Mutually enhancing anti-inflammatory activities of dimethyl fumarate and NF-kappaB inhibitors--implications for dose-sparing combination therapies. Exp Dermatol. 2016;25(2):124-30.

255.   Robasky K, Lewis NE, Church GM. The role of replicates for error mitigation in next-generation sequencing. Nat Rev Genet. 2014;15(1):56-62.

256.   Hutchison CA, 3rd, Smith HO, Pfannkoch C, Venter JC. Cell-free cloning using phi29 DNA polymerase. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(48):17332-6.

257.   Kircher M, Heyn P, Kelso J. Addressing challenges in the production and analysis of illumina sequencing data. BMC genomics. 2011;12:382.

258.   Bystrykh LV. Generalized DNA barcode design based on Hamming codes. PloS one. 2012;7(5):e36852.

259.   Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, et al. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res. 2011;39(13):e90.

260.   Sadhukhan S, Liu X, Ryu D, Nelson OD, Stupinski JA, Li Z, et al. Metabolomics-assisted proteomics identifies succinylation and SIRT5 as important regulators of cardiac function. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(16):4320-5.

261.   Chouchani ET, Pell VR, Gaude E, Aksentijevic D, Sundier SY, Robb EL, et al. Ischaemic accumulation of succinate controls reperfusion injury through mitochondrial ROS. Nature. 2014;515(7527):431-5.

262.   Park J, Chen Y, Tishkoff DX, Peng C, Tan M, Dai L, et al. SIRT5-mediated lysine desuccinylation impacts diverse metabolic pathways. Mol Cell. 2013;50(6):919-30.

263.   Hirschey MD, Zhao Y. Metabolic Regulation by Lysine Malonylation, Succinylation, and Glutarylation. Mol Cell Proteomics. 2015;14(9):2308-15.

264.   Rardin MJ, He W, Nishida Y, Newman JC, Carrico C, Danielson SR, et al. SIRT5 regulates the mitochondrial lysine succinylome and metabolic networks. Cell Metab. 2013;18(6):920-33.

265.   Jaberi E, Chitsazian F, Ali Shahidi G, Rohani M, Sina F, Safari I, et al. The novel mutation p.Asp251Asn in the beta-subunit of succinate-CoA ligase causes encephalomyopathy and elevated succinylcarnitine. J Hum Genet. 2013;58(8):526-30.

266.   Carrozzo R, Dionisi-Vici C, Steuerwald U, Lucioli S, Deodato F, Di Giandomenico S, et al. SUCLA2 mutations are associated with mild methylmalonic aciduria, Leigh-like encephalomyopathy, dystonia and deafness. Brain : a journal of neurology. 2007;130(Pt 3):862-74.

267.   Lamperti C, Fang M, Invernizzi F, Liu X, Wang H, Zhang Q, et al. A novel homozygous mutation in SUCLA2 gene identified by exome sequencing. Mol Genet Metab. 2012;107(3):403-8.

268.   Van Hove JL, Saenz MS, Thomas JA, Gallagher RC, Lovell MA, Fenton LZ, et al. Succinyl-CoA ligase deficiency: a mitochondrial hepatoencephalomyopathy. Pediatr Res. 2010;68(2):159-64.

269.   Yang SP, Yang XZ, Cao GP. Acetyl-l-carnitine prevents homocysteine-induced suppression of Nrf2/Keap1 mediated antioxidation in human lens epithelial cells. Mol Med Rep. 2015;12(1):1145-50.

270.   Li J, Zhang Y, Luan H, Chen X, Han Y, Wang C. l-carnitine protects human hepatocytes from oxidative stress-induced toxicity through Akt-mediated activation of Nrf2 signaling pathway. Can J Physiol Pharmacol. 2016;94(5):517-25.

271.   Hota KB, Hota SK, Chaurasia OP, Singh SB. Acetyl-L-carnitine-mediated neuroprotection during hypoxia is attributed to ERK1/2-Nrf2-regulated mitochondrial biosynthesis. Hippocampus. 2012;22(4):723-36.

272.   Zhao YY, Wang HL, Cheng XL, Wei F, Bai X, Lin RC, et al. Metabolomics analysis reveals the association between lipid abnormalities and oxidative stress, inflammation, fibrosis, and Nrf2 dysfunction in aristolochic acid-induced nephropathy. Sci Rep. 2015;5:12936.

273.   Parvin R, Pande SV. Enhancement of mitochondrial carnitine and carnitine acylcarnitine translocase-mediated transport of fatty acids into liver mitochondria under ketogenic conditions. J Biol Chem. 1979;254(12):5423-9.

274.   Jones LL, McDonald DA, Borum PR. Acylcarnitines: role in brain. Prog Lipid Res. 2010;49(1):61-75.

275.   O'Connor JE, Costell M, Miguez MP, Portoles M, Grisolia S. Effect of L-carnitine on ketone bodies, redox state and free amino acids in the liver of hyperammonemic mice. Biochem Pharmacol. 1987;36(19):3169-73.

276.	Tang H, Lu JY, Zheng X, Yang Y, Reagan JD. The psoriasis drug monomethylfumarate is a potent nicotinic acid receptor agonist. Biochem Biophys Res Commun. 2008;375(4):562-5.

277.	Rahman M, Muhammad S, Khan MA, Chen H, Ridder DA, Muller-Fielitz H, et al. The beta-hydroxybutyrate receptor HCA2 activates a neuroprotective subset of macrophages. Nat Commun. 2014;5:3944.

278.	Aksoy S, Szumlanski CL, Weinshilboum RM. Human liver nicotinamide N-methyltransferase. cDNA cloning, expression, and biochemical characterization. J Biol Chem. 1994;269(20):14835-40.

279.	Przygodzki T, Kazmierczak P, Sikora J, Watala C. 1-methylnicotinamide effects on the selected markers of endothelial function, inflammation and haemostasis in diabetic rats. Eur J Pharmacol. 2010;640(1-3):157-62.

280.	Tanaka Y, Kume S, Araki H, Nakazawa J, Chin-Kanasaki M, Araki S, et al. 1-Methylnicotinamide ameliorates lipotoxicity-induced oxidative stress and cell death in kidney proximal tubular cells. Free Radic Biol Med. 2015;89:831-41.

281.	Milani ZH, Ramsden DB, Parsons RB. Neuroprotective effects of nicotinamide N-methyltransferase and its metabolite 1-methylnicotinamide. J Biochem Mol Toxicol. 2013;27(9):451-6.

282.	Bryniarski K, Biedron R, Jakubowski A, Chlopicki S, Marcinkiewicz J. Anti-inflammatory effect of 1-methylnicotinamide in contact hypersensitivity to oxazolone in mice; involvement of prostacyclin. Eur J Pharmacol. 2008;578(2-3):332-8.

283.	Szafarz M, Kus K, Walczak M, Zakrzewska A, Niemczak M, Pernak J, et al. Pharmacokinetic Profile of 1-Methylnicotinamide Nitrate in Rats. J Pharm Sci. 2017;106(5):1412-8.

284.	Havrdova E, Galetta S, Hutchinson M, Stefoski D, Bates D, Polman CH, et al. Effect of natalizumab on clinical and radiological disease activity in multiple sclerosis: a retrospective analysis of the Natalizumab Safety and Efficacy in Relapsing-Remitting Multiple Sclerosis (AFFIRM) study. The Lancet Neurology. 2009;8(3):254-60.

285.	Arnold DL, Calabresi PA, Kieseier BC, Sheikh SI, Deykin A, Zhu Y, et al. Effect of peginterferon beta-1a on MRI measures and achieving no evidence of disease activity: results from a randomized controlled trial in relapsing-remitting multiple sclerosis. BMC Neurol. 2014;14:240.

286.	Nixon R, Bergvall N, Tomic D, Sfikas N, Cutter G, Giovannoni G. No evidence of disease activity: indirect comparisons of oral therapies for the treatment of relapsing-remitting multiple sclerosis. Advances in therapy. 2014;31(11):1134-54.

287.	Giovannoni G, Turner B, Gnanapavan S, Offiah C, Schmierer K, Marta M. Is it time to target no evident disease activity (NEDA) in multiple sclerosis? Multiple sclerosis and related disorders. 2015;4(4):329-33.

288.	Kappos L, De Stefano N, Freedman MS, Cree BA, Radue EW, Sprenger T, et al. Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity' (NEDA-4) in relapsing-remitting multiple sclerosis. Multiple sclerosis. 2016;22(10):1297-305.

289.	Havrdova E, Giovannoni G, Gold R, Fox RJ, Kappos L, Phillips JT, et al. Effect of delayed-release dimethyl fumarate on no evidence of disease activity in relapsing-remitting multiple sclerosis: integrated analysis of the phase III DEFINE and CONFIRM studies. European journal of neurology. 2017.

290.   Dubey D, Kieseier BC, Hartung HP, Hemmer B, Warnke C, Menge T, et al. Dimethyl fumarate in relapsing-remitting multiple sclerosis: rationale, mechanisms of action, pharmacokinetics, efficacy and safety. Expert review of neurotherapeutics. 2015;15(4):339-46.

291.   Jain S, Sima DM, Ribbens A, Cambron M, Maertens A, Van Hecke W, et al. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. Neuroimage Clin. 2015;8:367-75.

292.   Smeets D, Ribbens A, Sima DM, Cambron M, Horakova D, Jain S, et al. Reliable measurements of brain atrophy in individual patients with multiple sclerosis. Brain Behav. 2016;6(9):e00518.

293.   Jain S, Ribbens A, Sima DM, Cambron M, De Keyser J, Wang C, et al. Two Time Point MS Lesion Segmentation in Brain MRI: An Expectation-Maximization Framework. Front Neurosci. 2016;10:576.

294.   Fischer J, Jak AJ, Kniker JE, Rudick RA, Cutter G. Administration and Scoring Manual for the Multiple Scle- rosis Functional Composite (MSFC). National Multiple Sclerosis Society. 2001.

295.   Polman CH, Rudick RA. The multiple sclerosis functional composite: a clinically meaningful measure of disability. Neurology. 2010;74 Suppl 3:S8-15.

296.   Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473-83.

297.   Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11 0 1-33.

298.   Xia J, Wishart DS. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. Nucleic Acids Res. 2010;38(Web Server issue):W71-7.

299.   Hayes JD, Dinkova-Kostova AT. The Nrf2 regulatory network provides an interface between redox and intermediary metabolism. Trends Biochem Sci. 2014;39(4):199-218.

300.   TD G. Target Genes of NF-kB.

301.   International Multiple Sclerosis Genetics C, Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kemppinen A, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet. 2013;45(11):1353-60.

302.   Jernas M, Malmestrom C, Axelsson M, Nookaew I, Wadenvik H, Lycke J, et al. MicroRNA regulate immune pathways in T-cells in multiple sclerosis (MS). BMC Immunol. 2013;14:32.

303.   Keller A, Leidinger P, Bauer A, Elsharawy A, Haas J, Backes C, et al. Toward the blood-borne miRNome of human diseases. Nature methods. 2011;8(10):841-3.

304.   Hecker M, Thamilarasan M, Koczan D, Schroder I, Flechtner K, Freiesleben S, et al. MicroRNA expression changes during interferon-beta treatment in the peripheral blood of multiple sclerosis patients. International journal of molecular sciences. 2013;14(8):16087-110.

305.   Haabeth OA, Lorvik KB, Hammarstrom C, Donaldson IM, Haraldsen G, Bogen B, et al. Inflammation driven by tumour-specific Th1 cells protects against B-cell cancer. Nat Commun. 2011;2:240.

306.   Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: A comprehensive review. J Autoimmun. 2015;64:13-25.

307.   Giovannoni G, Gold R, Kappos L, Arnold DL, Bar-Or A, Marantz JL, et al. Delayed-release dimethyl fumarate and disability assessed by the Multiple Sclerosis Functional Composite: Integrated analysis of DEFINE and CONFIRM. 2016;0(0):1-4.

308.   Kappos L, Gold R, Arnold DL, Bar-Or A, Giovannoni G, Selmaj K, et al. Quality of life outcomes with BG-12 (dimethyl fumarate) in patients with relapsing-remitting multiple sclerosis: the DEFINE study. Multiple sclerosis. 2014;20(2):243-52.

309.   Al-Jaderi Z, Maghazachi AA. Utilization of Dimethyl Fumarate and Related Molecules for Treatment of Multiple Sclerosis, Cancer, and Other Diseases. Frontiers in immunology. 2016;7:278.

310.   Xie X, Zhao Y, Ma CY, Xu XM, Zhang YQ, Wang CG, et al. Dimethyl fumarate induces necroptosis in colon cancer cells through GSH depletion/ROS increase/MAPKs activation pathway. Br J Pharmacol. 2015;172(15):3929-43.

311.   Smolen JS, Aletaha D, Bijlsma JW, Breedveld FC, Boumpas D, Burmester G, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Annals of the rheumatic diseases. 2010;69(4):631-7.

312.   Smolen JS, Aletaha D, Koeller M, Weisman MH, Emery P. New therapies for treatment of rheumatoid arthritis. Lancet. 2007;370(9602):1861-74.

313.   Smolen JS, Aletaha D. Forget personalised medicine and focus on abating disease activity. Annals of the rheumatic diseases. 2013;72(1):3-6.

314.   Sorensen PS. Neutralizing antibodies against interferon-Beta. Therapeutic advances in neurological disorders. 2008;1(2):125-41.

315.   Sorensen PS, Ross C, Clemmesen KM, Bendtzen K, Frederiksen JL, Jensen K, et al. Clinical importance of neutralising antibodies against interferon beta in patients with relapsing-remitting multiple sclerosis. Lancet. 2003;362(9391):1184-91.

316.   Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Human molecular genetics. 2009;18(4):767-78.

317.   Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. Am J Hum Genet. 2013;93(4):641-51.

318.   Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, et al. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv Bioinformatics. 2008;2008:420747.

319.   Compendium EM. Tecfidera 120mg and 240mg gastro-resistant hard capsules. 2017.

320.   Schroeder K, Fahey T, Ebrahim S, Peters TJ. Adherence to long-term therapies: recent WHO report provides some answers but poses even more questions. J Clin Epidemiol. 2004;57(1):2-3.

321.   Lugaresi A. Addressing the need for increased adherence to multiple sclerosis therapy: can delivery technology enhance patient motivation? Expert Opin Drug Deliv. 2009;6(9):995-1002.

322.   Devonshire V, Lapierre Y, Macdonell R, Ramo-Tello C, Patti F, Fontoura P, et al. The Global Adherence Project (GAP): a multicenter observational study on adherence to disease-modifying therapies in patients with relapsing-remitting multiple sclerosis. European journal of neurology. 2011;18(1):69-77.

323.   Palsson B, Zengler K. The challenges of integrating multi-omic data sets. Nat Chem Biol. 2010;6(11):787-9.

324.   Cavill K, Briede. Dynamic time wariping for omics. PloS one. 2013.

325.   Schurch NJ, Schofield P, Gierlinski M, Cole C, Sherstnev A, Singh V, et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA. 2016;22(6):839-51.

326.   Poplawski A, Binder H. Feasibility of sample size calculation for RNA-seq studies. Brief Bioinform. 2017.

327.   Walsh PS, Erlich HA, Higuchi R. Preferential PCR amplification of alleles: mechanisms and solutions. PCR Methods Appl. 1992;1(4):241-50.

328.   Lam HY, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, et al. Performance comparison of whole-genome sequencing platforms. Nat Biotechnol. 2011;30(1):78-82.

329.   Polman CH, Reingold SC, Barkhof F, Calabresi PA, Clanet M, Cohen JA, et al. Ethics of placebo-controlled clinical trials in multiple sclerosis: a reassessment. Neurology. 2008;70(13 Pt 2):1134-40.

330.   Gilgun-Sherki Y, Melamed E, Offen D. The role of oxidative stress in the pathogenesis of multiple sclerosis: the need for effective antioxidant therapy. Journal of neurology. 2004;251(3):261-8.

331.   Moujalled D, Grubman A, Acevedo K, Yang S, Ke YD, Moujalled DM, et al. TDP-43 mutations causing amyotrophic lateral sclerosis are associated with altered expression of RNA-binding protein hnRNP K and affect the Nrf2 antioxidant pathway. Human molecular genetics. 2017;26(9):1732-46.

332.   Kunsch C, Medford RM. Oxidative stress as a regulator of gene expression in the vasculature. Circ Res. 1999;85(8):753-66.

333.   Finkel T, Holbrook NJ. Oxidants, oxidative stress and the biology of ageing. Nature. 2000;408(6809):239-47.

GENEWIZ NGS Laboratory
Sample QC Form

Name: Arie Gafson
Institute: Imperial College London
Quotation Number: SA1611232_R1

GENEWIZ ID	Sample Vol. (ul)	Bioanalyzer RIN (for RNA)
1	3	8.9
2	26	9.3
3	45	9.4
4	23	9.3
5	26	9.3
6	46	9.3
7	50	9.5
8	36	9.4
9	33	9.1
10	30	9.4
11	26	9.4
12	45	9.6
13	27	9.3
14	20	9.4
15	39	9.7
16	55	9.5
17	44	9.6
18	50	9.6
19	45	9.6
20	45	9.7
21	55	9.6
22	45	9.6
23	80	9
24	46	9.6
25	44	9.1
26	43	9.1
27	44	9.5
28	41	9.5
29	42	9.6
30	45	9.6
31	41	9.5
32	48	9.1
33	47	9.4
34	68	9.5
35	41	9.3
36	45	9.5
37	43	9.6
38	25	9.3
39	45	9.6
40	25	9.6
41	48	9.3
42	44	9.4
43	49	9.3
44	55	9.4

45	45	9.3
46	48	9.4
47	49	9.2
48	45	9.3
49	58	9.4
50	45	9.4
51	45	9.5
52	54	9.5
53	45	9.4
54	43	9.5
55	46	9.5
56	55	9.9
57	44	9
58	44	9.4
59	46	9.2
60	45	9.4
61	45	9.3
62	50	9.4
63	44	9.3
64	44	9.5
65	45	9.4
66	49	9.4
67	45	9.5
68	45	9.4
69	44	9.3
70	43	9.2
71	29	9.6
72	41	9
73	45	9.6
74	44	9.7
75	44	9.5
76	45	9.5
77	44	9.5
78	45	9.5
79	49	9.7
80	45	9.6
81	24	9.5
82	30	9.7
83	30	8.7
84	30	9.5
85	30	9.6
86	30	9.5
87	30	9.1
88	30	9.4
89	30	9.4
90	30	9.3
91	30	9.5
92	30	9.5
93	30	9.5
94	30	9.4
95	30	9.5
96	30	9.6
97	N/A	
98	25	8.9
99	30	9.3

100	30	9.4
101	14	9.5
102	30	9.6
103	30	9.4
104	30	9.5
105	30	9.2
106	30	9.4
107	30	9.1
108	30	9.5
109	30	9.4
110	30	9.6
111	30	9.3
112	30	9.5
113	30	9.5
114	30	9.6
115	20	9.7
116	30	9.7
117	30	9.4
118	30	9.7
119	30	9.5
120	30	9.6
121	30	9.6
122	30	9.5

FastQC Report

Summary

✅ [Basic Statistics](#)

✅ [Per base sequence quality](#)

✅ [Per tile sequence quality](#)

✅ [Per sequence quality scores](#)

❌ [Per base sequence content](#)
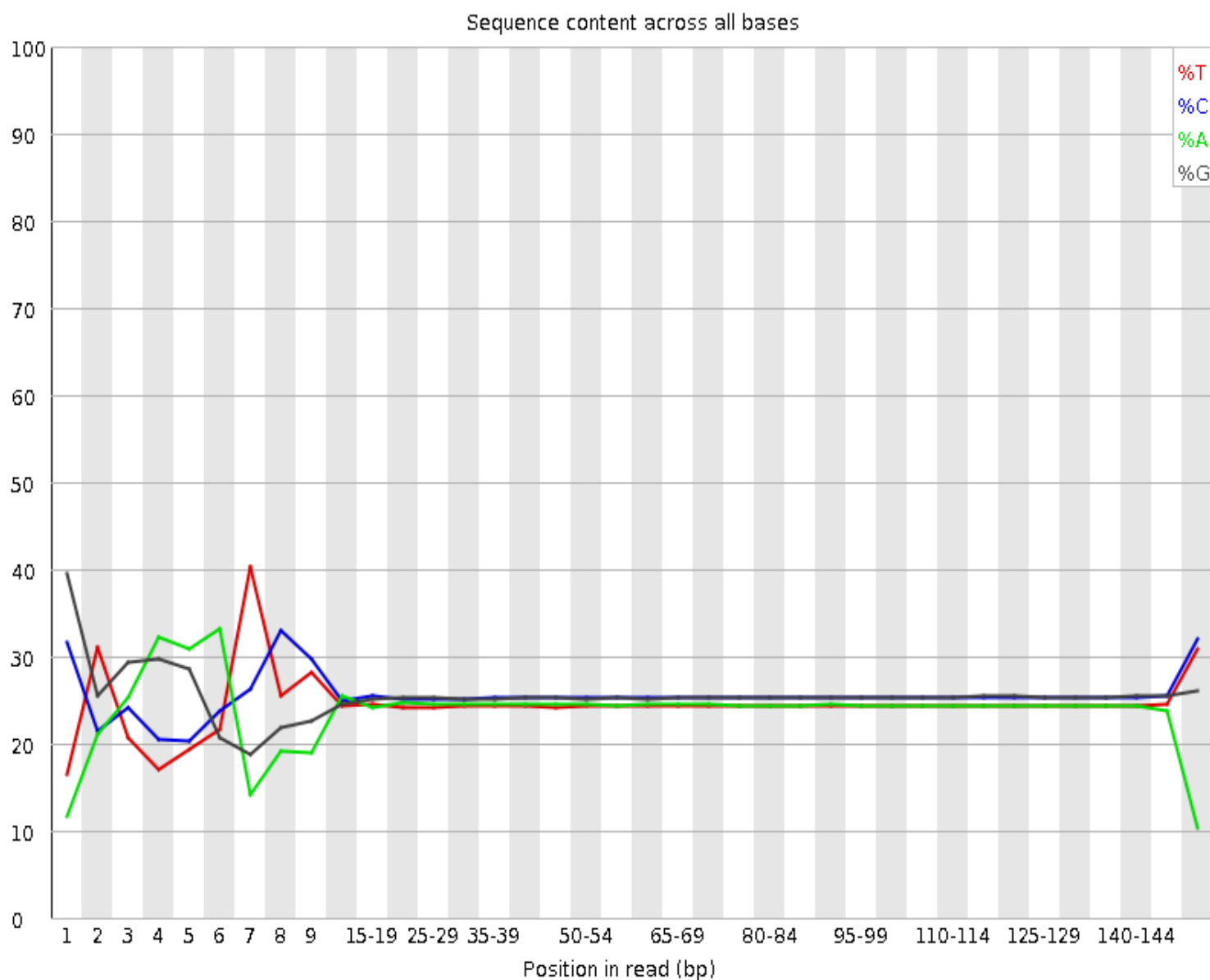
✅ [Per sequence GC content](#)
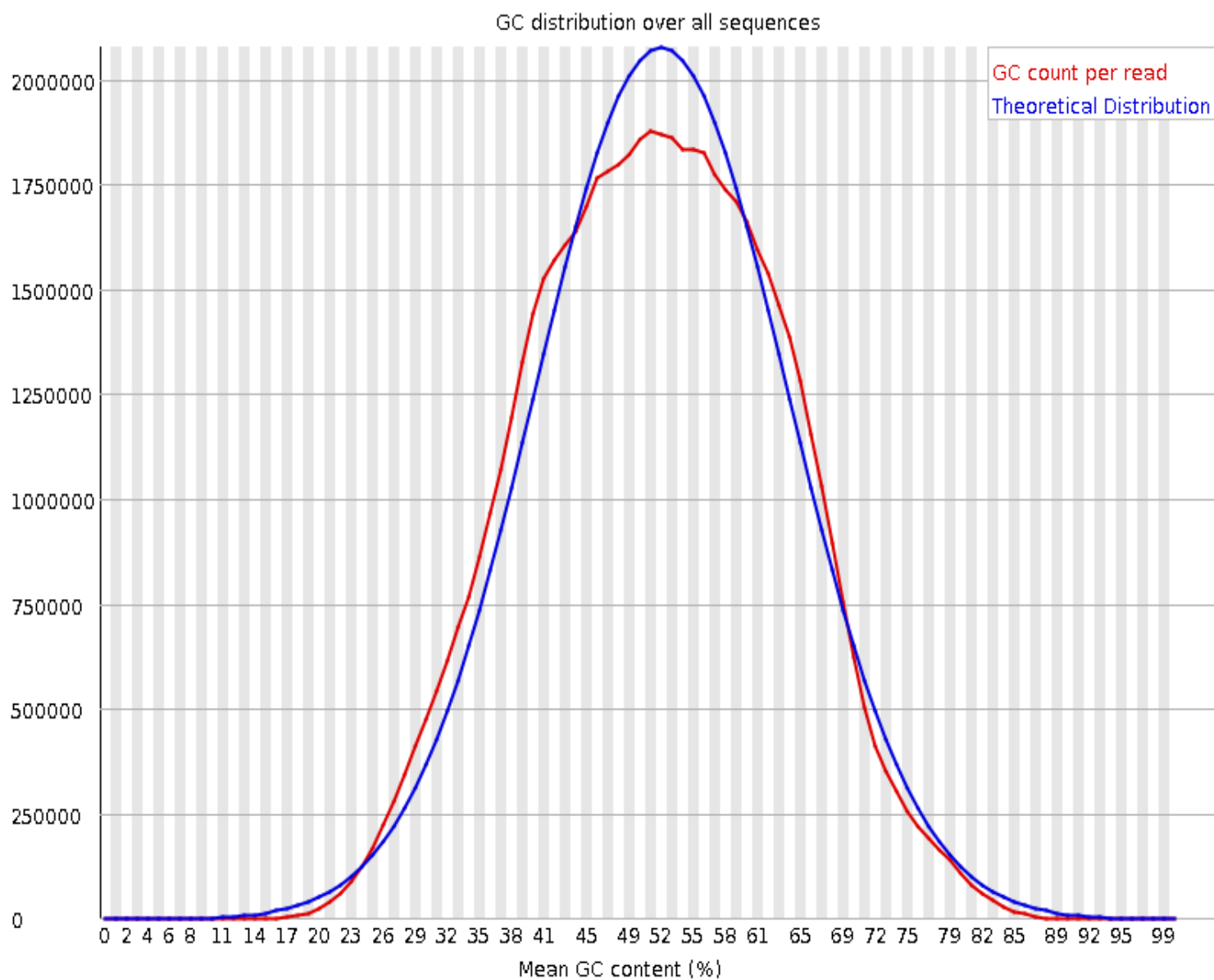
✅ [Per base N content](#)
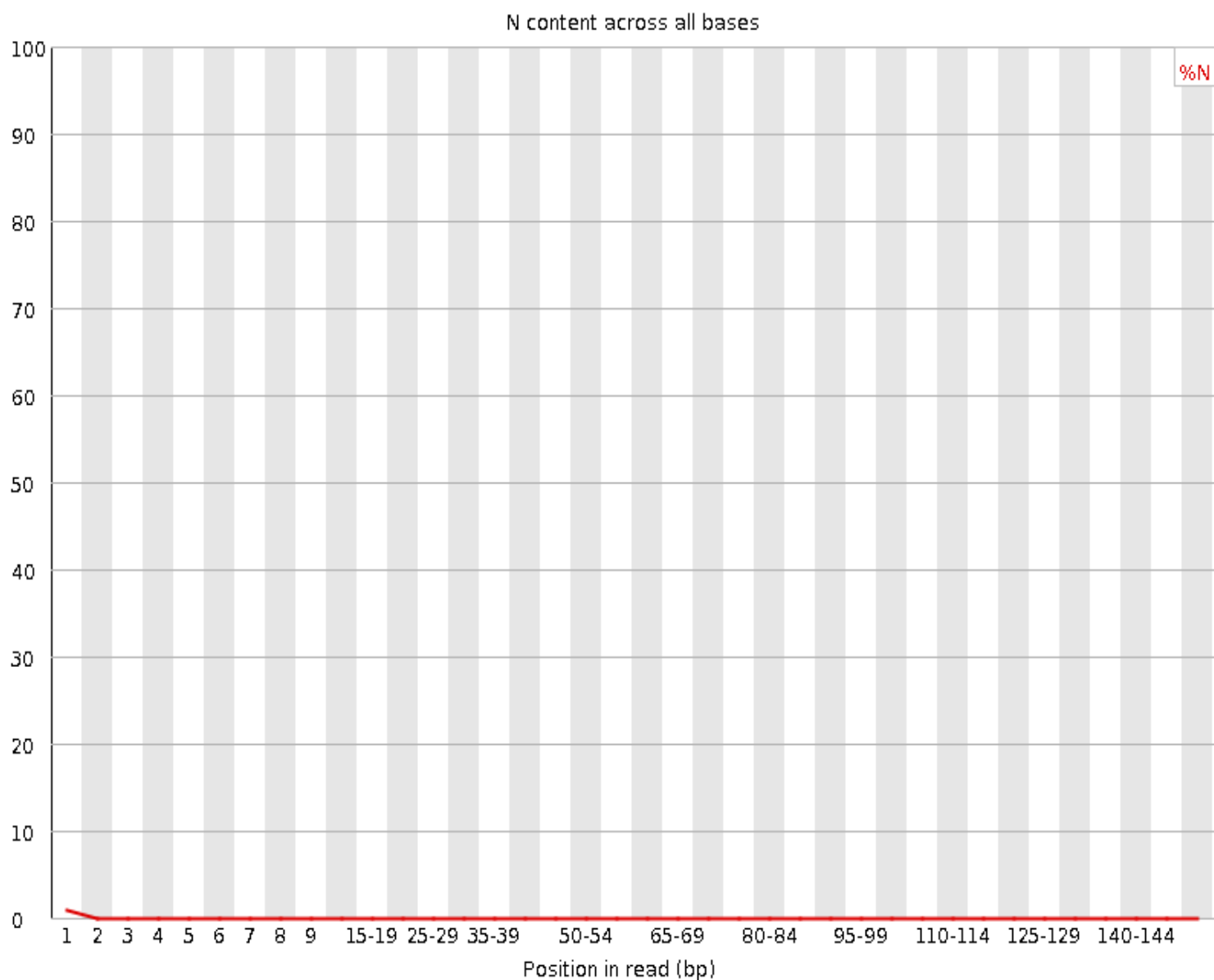
⚠️ [Sequence Length Distribution](#)
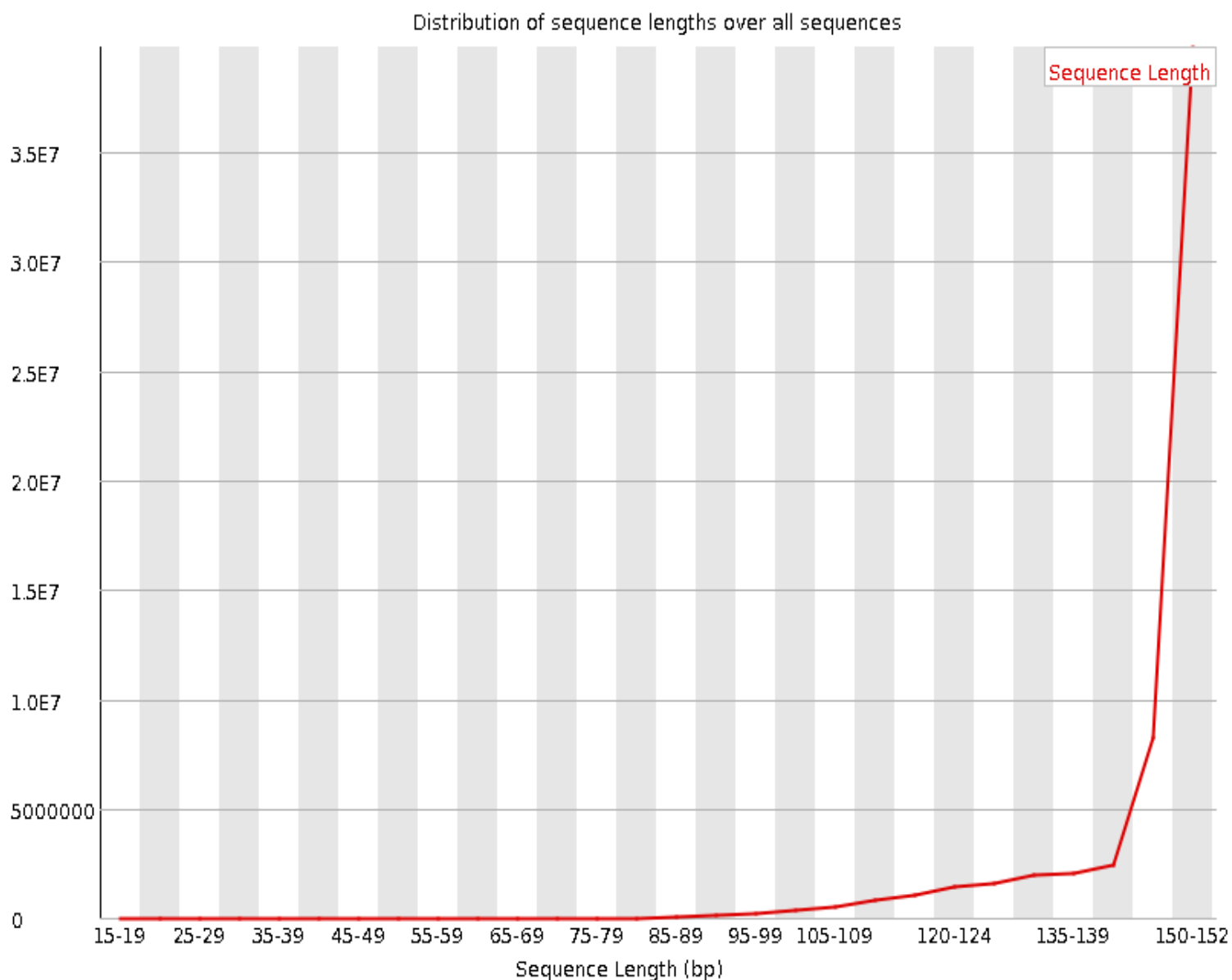
❌ [Sequence Duplication Levels](#)

✅ [Overrepresented sequences](#)

✅ [Adapter Content](#)

❌ [Kmer Content](#)

✅ Basic Statistics

| Measure	Value
Filename	AA01A-25-10-16_S14_R1_001_val_1.fq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	61407451
Sequences flagged as poor quality	0
Sequence length	20-151
%GC	51

✅ Per base sequence quality

Quality scores across all bases (Sanger / Illumina 1.9 encoding)

✅ Per tile sequence quality

Quality per tile

✅ Per sequence quality scores

Quality score distribution over all sequences

Per base sequence content

Sequence content across all bases

Per sequence GC content

GC distribution over all sequences

✅ Per base N content

N content across all bases

⚠️ Sequence Length Distribution

Distribution of sequence lengths over all sequences

❌ Sequence Duplication Levels

Percent of seqs remaining if deduplicated 39.41%

Overrepresented sequences

No overrepresented sequences

Adapter Content
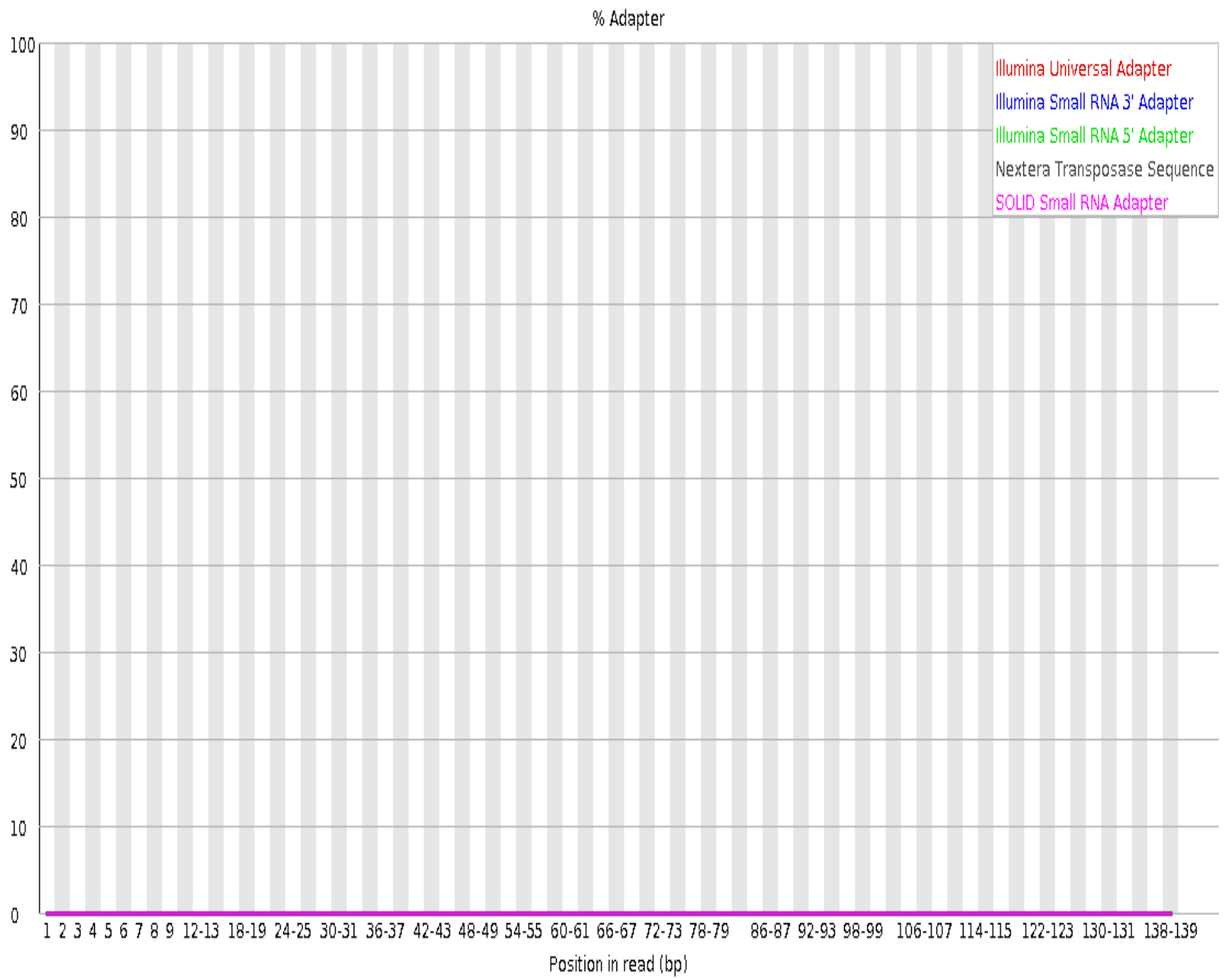
Kmer Content
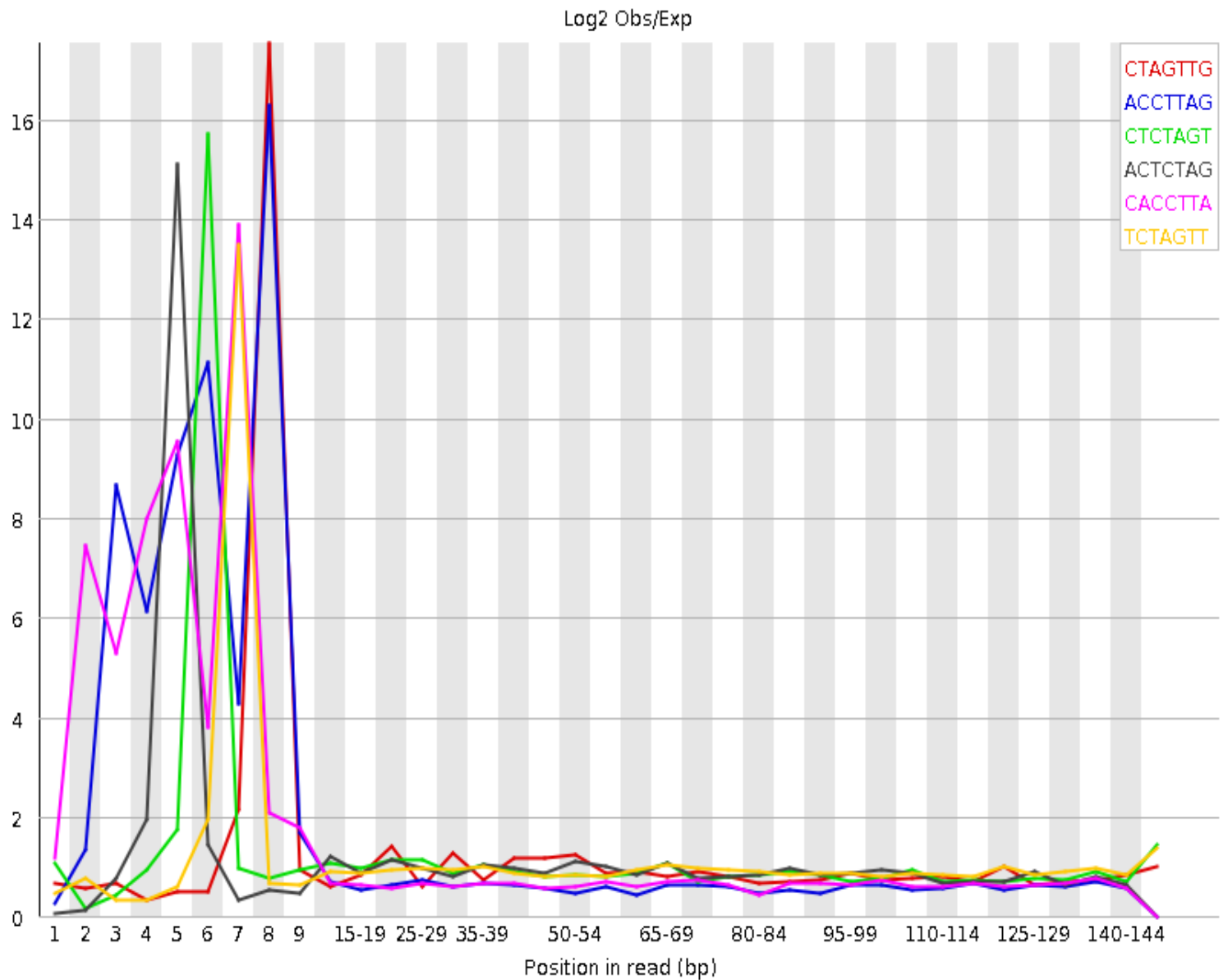
Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
CTAGTTG	22365	0.0	17.520712	8
ACCTTAG	38225	0.0	16.303713	8
CTCTAGT	29950	0.0	15.728111	6
ACTCTAG	30205	0.0	15.112215	5
CACCTTA	45425	0.0	13.903098	7
TCTAGTT	35615	0.0	13.479933	7
CGACCTA	5745	0.0	12.940136	4
CCGGTAT	15405	0.0	12.592897	1
CGCGAGT	11095	0.0	12.416726	1
GCGGAAT	14055	0.0	12.352204	1
CCTTAGA	52375	0.0	12.323477	9

```
GTCGAAC    15360    0.0    12.217952    1

CGAATCG     8720    0.0    12.192246    3

CGGCGAT    22745    0.0    11.990136    1

TCGGGTC    29065    0.0    11.977776    2

GCGGGTT    15975    0.0    11.747589    1

GCGAGAT    21235    0.0    11.419449    1

GCCGTAT     9475    0.0    11.201497    1

GGCGAAT    15890    0.0    11.191155    1

TAGTTGT    36875    0.0    10.984445    9
```

Produced by [FastQC](#) (version 0.11.5)

	Model R^2	Model P-Value	V1PL	V2TG	V2PL
LOO1	0.48	8.47907E-06	0.014	0.004	0.014
LOO2	0.48	7.69336E-06	0.009	0.003	0.014
LOO3	0.51	2.08679E-06	0.006	0.001	0.006
LOO4	0.49	5.44127E-06	0.013	0.005	0.020
LOO5	0.47	8.69054E-06	0.010	0.004	0.014
LOO6	0.51	2.20107E-06	0.011	0.003	0.010
LOO7	0.48	7.42451E-06	0.009	0.003	0.014
LOO8	0.49	4.09666E-06	0.008	0.003	0.011
LOO9	0.49	4.46746E-06	0.004	0.002	0.007
LOO10	0.48	6.5764E-06	0.012	0.004	0.015
LOO11	0.49	4.2431E-06	0.007	0.002	0.009
LOO12	0.48	6.41244E-06	0.010	0.004	0.014
LOO13	0.46	1.3048E-05	0.010	0.003	0.013
LOO14	0.48	6.2049E-06	0.009	0.003	0.014
LOO15	0.48	7.53966E-06	0.010	0.004	0.015
LOO16	0.48	6.86041E-06	0.014	0.006	0.022
LOO17	0.48	7.5189E-06	0.009	0.005	0.019
LOO18	0.51	2.31859E-06	0.005	0.001	0.005
LOO19	0.52	1.16125E-06	0.005	0.003	0.014
LOO20	0.48	6.71033E-06	0.008	0.003	0.014
LOO21	0.50	3.36502E-06	0.002	0.002	0.010
LOO22	0.48	5.8468E-06	0.010	0.003	0.011
LOO23	0.48	7.53797E-06	0.010	0.004	0.015
LOO24	0.47	9.38576E-06	0.029	0.008	0.024
LOO25	0.46	1.58951E-05	0.016	0.004	0.014
LOO26	0.48	6.45236E-06	0.010	0.004	0.015
LOO27	0.48	8.4988E-06	0.013	0.006	0.020

Supplementary Lipid Table. Results of leave-one-out cross validation for optimal regression model. Statistical measures of fit for each leave-one-out model displayed as Model R^2 value. Corresponding p-value for the model and for each VLDL predictor coefficient provided. Total of 27 leave-one-out cross validations performed.

Metabolite	Correlation with RNF19B
nervonoylcarnitine (C24:1)*	0.97
pyridoxate	0.97
LPC(18:1)	0.96
1,2,3-benzenetriol sulfate (2)	0.96
lactate	0.96
ximenoylcarnitine (C26:1)*	0.95
N-acetylvaline	0.95
palmitoyl ethanolamide	0.94
ribose	0.94
glycolithocholate	0.94
pregnanolone/allopregnanolon	0.94
N-acetylisoleucine	0.93
LPE(18:3)	0.93
FFA(16:1)	0.93
LPC(16:0)	0.92
Total LPC	0.92
17alpha-hydroxypregnenolone	0.92
pro-hydroxy-pro	0.91
3-methoxycatechol sulfate (1)	0.91
methyl indole-3-acetate	0.91
2-hydroxyphenylacetate	0.91
LPC(18:3)	0.91
LPE(20:4)	0.90
LPE(20:3)	0.90
LPC(20:4)	0.90
uridine	0.90
TAG48:5-FA18:3	0.90
2-hydroxyacetaminophen sulfa	0.89
PI(18:0/20:2)	0.89
4-vinylguaiacol sulfate	0.89
N-acetylcarnosine	0.89
tyramine O-sulfate	0.89
N-acetyl-cadaverine	0.88
LPC(20:5)	0.88
LPC(16:1)	0.88
cystathionine	0.88
4-hydroxyphenylacetatoylcarn	0.88
N-acetyltyrosine	0.87
succinylcarnitine (C4-DC)	0.87
N-methylalanine	0.87
TAG56:5-FA20:3	0.87
PC(17:0/22:5)	0.87
5-acetylamino-6-formylamino-	0.86
glucuronate	0.86

DAG(18:1/20:4)	0.86
TAG54:5-FA18:3	0.86
TAG52:5-FA18:1	0.86
LPC(20:1)	0.86
LPC(22:5)	0.86
TAG56:7-FA20:5	0.86
TAG56:7-FA18:3	0.86
TAG50:4-FA20:3	0.86
TAG52:6-FA18:1	0.85
N6-methyladenosine	0.85
arabonate/xylonate	0.85
TAG50:5-FA16:1	0.85
alpha-ketoglutarate	0.85
methylsuccinoylcarnitine (1)	0.85
sphingosine 1-phosphate	0.85
hippurate	0.85
LPC(18:0)	0.85
cyclo(pro-val)	0.85
pyruvate	0.85
creatinine	0.84
succinate	0.84
cholate	0.84
N-acetylhistidine	0.84
N-acetylphenylalanine	0.84
DAG(18:1/20:3)	0.84
PC(18:1/18:1)	0.84
TAG50:5-FA16:0	0.84
TAG52:4-FA18:0	0.84
phenylacetylglutamate	0.84
4-acetamidobutanoate	0.83
TAG52:3-FA16:1	0.83
PI(18:0/22:4)	0.83
TAG48:4-FA18:1	0.83
N-formylanthranilic acid	0.83
1-methylhistidine	0.83
N-acetylputrescine	0.83
TAG50:5-FA20:5	0.83
LPC(17:0)	0.83
TAG48:5-FA18:2	0.83
N-acetylalanine	0.82
TAG52:6-FA20:5	0.82
chenodeoxycholate	0.82
N-acetylserine	0.82
PC(18:1/22:5)	0.82
TAG54:6-FA20:5	0.82

PI(18:0/20:3)	0.82
LPC(20:3)	0.82
TAG48:4-FA18:3	0.82
PC(18:1/16:1)	0.81
5-(galactosylhydroxy)-L-lysine	0.81
TAG56:5-FA18:1	0.81
5-acetylamino-6-amino-3-meth	0.81
1-methylurate	0.81
TAG54:6-FA16:1	0.81
TAG54:8-FA20:4	0.81
LPE(22:5)	0.81
heme	0.81
TAG54:4-FA18:3	0.80
TAG52:6-FA16:0	0.80
TAG56:6-FA18:1	0.80
TAG54:4-FA20:3	0.80
5-oxoproline	0.80
TAG46:0-FA12:0	0.80
CE(22:4)	0.80
LPC(22:6)	0.80
TAG56:8-FA18:3	0.80
TAG54:5-FA16:1	0.80
tiglylcarnitine (C5:1-DC)	0.80
TAG50:5-FA18:1	0.80
TAG52:5-FA20:5	0.80
beta-hydroxyisovalerate	0.80
TAG46:3-FA18:3	0.80
TAG52:6-FA16:1	0.80
TAG54:7-FA18:1	0.80
TAG54:5-FA20:4	0.80
dihomo-linolenoylcarnitine (20	0.80
glutarate (pentanedioate)	0.79
TAG46:2-FA16:1	0.79
TAG46:3-FA12:0	0.79
phenylacetylcarnitine	0.79
TAG56:6-FA20:5	0.79
LPE(18:1)	0.79
TAG54:8-FA20:5	0.79
TAG58:8-FA20:3	0.79
oleoylcarnitine (C18:1)	0.79
TAG52:4-FA20:3	0.79
imidazole propionate	0.78
N-carbamoylaspartate	0.78
TAG54:7-FA16:1	0.78
TAG54:4-FA16:1	0.78

TAG52:5-FA20:4	0.78
TAG52:4-FA18:3	0.78
TAG54:3-FA18:3	0.78
TAG44:1-FA12:0	0.78
TAG56:4-FA20:3	0.78
TAG52:7-FA20:5	0.78
TAG52:6-FA18:3	0.78
arabitol/xylitol	0.78
PC(20:0/18:1)	0.78
TAG50:5-FA20:4	0.77
LPE(16:0)	0.77
TAG52:2-FA16:1	0.77
TAG54:6-FA18:1	0.77
PC(18:1/22:4)	0.77
ethylmalonate	0.77
TAG52:5-FA20:3	0.77
TAG44:0-FA12:0	0.77
gamma-glutamylhistidine	0.77
PC(18:0/22:5)	0.77
TAG50:4-FA16:0	0.77
TAG44:1-FA16:1	0.77
LPC(22:4)	0.77
erythronate*	0.77
PI(16:0/20:4)	0.76
N1-Methyl-4-pyridone-3-carbo	0.76
HCER(18:0)	0.76
TAG50:4-FA18:3	0.76
CE(18:4)	0.76
TAG56:4-FA20:4	0.76
TAG54:3-FA20:3	0.76
DAG(16:1/18:1)	0.76
TAG48:2-FA16:1	0.76
TAG54:5-FA20:5	0.76
glutarylcarnitine (C5-DC)	0.76
TAG56:7-FA16:1	0.76
dihomo-linolenoyl-choline	0.76
TAG58:5-FA18:1	0.76
TAG48:2-FA12:0	0.76
LCER(20:1)	0.76
TAG56:4-FA20:2	0.76
PI(16:0/16:1)	0.75
TAG50:4-FA20:4	0.75
TAG58:6-FA18:1	0.75
phosphate	0.75
TAG48:4-FA16:0	0.75

TAG48:4-FA14:0	0.75
DAG(16:0/20:4)	0.75
TAG53:7-FA18:3	0.75
TAG50:3-FA18:3	0.75
N1-Methyl-2-pyridone-5-carbo	0.75
TAG56:3-FA16:0	0.75
TAG56:6-FA18:3	0.75
TAG48:4-FA16:1	0.75
TAG52:3-FA20:3	0.75
TAG46:1-FA12:0	0.75
CE(18:1)	0.75
eicosenoylcarnitine (C20:1)*	0.75
TAG50:3-FA20:3	0.75
TAG50:5-FA18:3	0.75
cysteinylglycine	0.75
O-sulfo-L-tyrosine	0.74
TAG53:3-FA18:0	0.74
behenoylcarnitine (C22)*	0.74
TAG56:6-FA20:4	0.74
TAG48:3-FA18:3	0.74
PC(18:0/16:1)	0.74
PC(16:0/22:5)	0.74
LPC(20:2)	0.74
TAG48:1-FA12:0	0.74
TAG46:2-FA14:1	0.74
TAG48:2-FA18:1	0.74
cyclo(ala-pro)	0.74
TAG56:7-FA20:3	0.74
TAG52:5-FA14:0	0.73
hydantoin-5-propionic acid	0.73
TAG50:2-FA16:1	0.73
TAG56:7-FA18:0	0.73
oleoylcholine	0.73
LPE(16:1)	0.73
TAG48:4-FA20:4	0.73
N1-methyladenosine	0.73
TAG58:7-FA18:1	0.73
gulonate*	0.73
PI(16:0/18:1)	0.73
PC(12:0/18:1)	0.73
PC(14:0/22:5)	0.73
TAG46:3-FA16:0	0.73
PC(18:0/18:3)	0.72
mannitol/sorbitol	0.72
TAG56:5-FA20:4	0.72

Metabolite	Correlation with KIR3DL2
TAG56:5-FA22:4	0.72
DAG(16:0/18:3)	0.72
N-acetylleucine	0.72
TAG50:2-FA18:1	0.72
TAG52:2-FA18:1	0.72
PC(14:0/14:0)	0.72
TAG55:6-FA18:1	0.72
TAG52:3-FA18:3	0.72
TAG56:6-FA16:0	0.72
TAG52:4-FA20:4	0.72
p-cresol-glucuronide*	0.72
gamma-glutamyl-2-aminobuty	0.72
PC(16:0/12:0)	0.72
PC(20:0/20:3)	0.72
LPC(14:0)	0.72
PC(18:1/18:3)	0.72
TAG54:7-FA20:5	0.72
PC(16:0/18:1)	0.71
propionylcarnitine (C3)	0.71
indoleacetate	0.71
TAG54:5-FA16:0	0.71
PC(18:0/20:1)	0.71
PC(16:0/16:1)	0.71
ornithine	0.71
TAG46:1-FA16:1	0.71
TAG54:6-FA20:3	0.71
TAG56:4-FA18:0	0.71
TAG54:3-FA16:1	0.71
PC(18:0/12:0)	0.71
TAG56:4-FA16:0	0.71
gamma-glutamylvaline	0.71
TAG46:1-FA16:0	0.71
stearoylcarnitine (C18)	0.70
DAG(18:1/18:1)	0.70
TAG56:3-FA20:2	0.70
isoursodeoxycholate	0.70
carnitine	0.70
TAG58:7-FA16:0	0.70
TAG54:4-FA20:4	0.70
DAG(16:0/20:3)	0.70
bilirubin (E,E)*	0.70
O-methylcatechol sulfate	0.70
glycerophosphorylcholine (GPC	0.70
N-acetyl-1-methylhistidine*	0.70

pregnanediol-3-glucuronide	0.97
xanthine	0.95
N-acetyltaurine	0.95
androstenediol (3beta,17beta)	0.94
pregnen-diol disulfate*	0.94
5alpha-pregnan-3beta,20beta-	0.94
12-HETE	0.93
3-phosphoglycerate	0.93
5alpha-pregnan-3(alpha or bet	0.93
5alpha-pregnan-3beta,20alpha	0.93
bilirubin (E,Z or Z,E)*	0.93
solanidine	0.93
glycerophosphoserine*	0.93
PE(18:2/22:4)	0.92
dehydroisoandrosterone sulfat	0.91
phenylalanylglycine	0.90
gamma-glutamylglutamine	0.90
androstenediol (3alpha, 17alph	0.90
21-hydroxypregnenolone disul	0.90
sphingadienine*	0.89
2,3-dihydroxyisovalerate	0.88
pregn steroid monosulfate*	0.88
5alpha-pregnan-3beta,20alpha	0.87
choline phosphate	0.87
4-hydroxyphenylacetylglutami	0.87
pregnenolone sulfate	0.87
androstenediol (3alpha, 17alph	0.87
inosine 5'-monophosphate (IM	0.87
guanidinoacetate	0.86
PE(O-18:0/16:1)	0.86
S-methylcysteine sulfoxide	0.86
cysteine-glutathione disulfide	0.85
androstenediol (3beta,17beta)	0.85
glycochenodeoxycholate	0.85
glycerate	0.85
ethyl glucuronide	0.85
cytidine 5'-diphosphocholine	0.85
biliverdin	0.84
ribose	0.84
glycine	0.84
3-hydroxysebacate	0.84
acesulfame	0.83
2,3-dihydroxy-2-methylbutyrat	0.83
HCER(18:0)	0.83
phosphoethanolamine	0.83

5-hydroxyhexanoate	0.83
PE(17:0/22:5)	0.82
tartronate (hydroxymalonate)	0.82
S-methylcysteine	0.82
pyroglutamine*	0.81
lanthionine	0.81
sphingosine	0.81
hypotaurine	0.80
5alpha-androstan-3alpha,17be	0.80
androstenediol (3beta,17beta)	0.80
perfluorooctanesulfonic acid (F	0.80
C-glycosyltryptophan	0.79
etiocholanolone glucuronide	0.79
6-oxopiperidine-2-carboxylate	0.79
guanosine	0.79
oxalate (ethanedioate)	0.79
stachydrine	0.79
taurine	0.78
3beta,7alpha-dihydroxy-5-chol	0.78
N-methylpipecolate	0.78
bilirubin (Z,Z)	0.78
N6-carboxyethyllysine	0.78
N6-carboxymethyllysine	0.78
androstenediol (3beta,17beta)	0.78
methyl-4-hydroxybenzoate sul	0.77
N-methylproline	0.77
glutamine	0.77
3beta-hydroxy-5-cholestenoat	0.77
HCER(20:0)	0.76
eugenol sulfate	0.76
succinimide	0.76
gamma-glutamylglycine	0.76
phenylpyruvate	0.74
N-acetylaspartate (NAA)	0.74
Total LCER	0.74
linoleoyl ethanolamide	0.74
5alpha-androstan-3alpha,17be	0.73
uridine 5'-monophosphate (UM	0.73
LPE(20:1)	0.73
inosine	0.73
beta-cryptoxanthin	0.72
octadecanedioate	0.72
pipecolate	0.71
glycylvaline	0.71
deoxycarnitine	0.71

Metabolite	Correlation with mir-423
3,7-dimethylurate	0.71
glycocholate	0.71
andro steroid monosulfate (1)*	0.70
cytidine 5'-monophosphate (5'	0.70
Metabolite	Correlation with mir-423
1-methylnicotinamide	-0.85
methionine sulfone	-0.84
PE(18:1/20:1)	-0.83
PE(P-18:1/18:3)	-0.83
alliin	-0.83
PI(18:0/16:1)	-0.83
PC(20:0/20:4)	-0.83
LCER(22:0)	-0.82
PE(P-18:1/18:1)	-0.82
PI(16:0/16:0)	-0.82
PC(18:2/18:3)	-0.81
PE(P-18:1/16:0)	-0.81
gamma-tocopherol/beta-tocop	-0.81
SM(20:1)	-0.81
PI(18:0/18:1)	-0.80
pristanate	-0.80
PE(P-18:0/18:1)	-0.80
EDTA	-0.79
DAG(14:0/16:0)	-0.79
PE(P-16:0/20:1)	-0.78
TAG42:0-FA14:0	-0.78
glycoursodeoxycholate	-0.78
PE(P-18:1/18:0)	-0.78
linoleoyl ethanolamide	-0.78
trans-urocanate	-0.77
PI(16:0/18:2)	-0.77
LCER(20:1)	-0.77
TAG56:9-FA20:4	-0.76
TAG56:8-FA20:4	-0.76
ursodeoxycholate	-0.76
PE(P-18:2/22:6)	-0.75
PI(18:0/18:2)	-0.75
alpha-tocopherol	-0.75
2-linoleoylglycerol (18:2)	-0.75
PC(15:0/20:4)	-0.75
PE(18:1/18:3)	-0.75
PC(18:1/20:1)	-0.74
retinol (Vitamin A)	-0.74
carboxyethyl-GABA	-0.74
PI(18:0/22:6)	-0.73

3-hydroxyhippurate	-0.73
PE(16:0/20:2)	-0.73
PE(18:1/20:2)	-0.73
LCER(18:1)	-0.73
PC(18:1/18:2)	-0.72
PI(18:1/18:2)	-0.72
PE(O-18:0/18:2)	-0.72
PC(18:1/20:4)	-0.72
DAG(14:0/18:2)	-0.72
5-hydroxyindole sulfate	-0.71
PE(P-18:0/20:2)	-0.71
3-hydroxybutyrylcarnitine (2)	-0.71
PC(18:2/22:4)	-0.70
TAG56:7-FA20:4	-0.70
3-(3-hydroxyphenyl)propionate	-0.70
alpha-ketobutyrate	-0.70