The effects of co-morbidities on blood transcriptomes in tuberculosis patients before and during treatment

Clare Eckold

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy

May 2018

Department of Immunology of Infectious Disease

Faculty of Infectious and Tropical Diseases

London School of Hygiene & Tropical Medicine

Funded by EU-FP7 (TANDEM: grant 305279)

Abstract

The blood transcriptome in tuberculosis (TB) has been well described, it is distinct from healthy controls and other diseases, and is being developed as a biomarker for risk of disease progression, disease severity and treatment response. But TB patients, especially those from high TB-burden countries, often have comorbidities, so the described transcriptomic signature may not be an accurate representation of a typical population. Whether this signature remains constant in different patient phenotypes is an important question.

The effect of HIV-1 and also type 2 diabetes (T2DM) on the transcriptomic profile of TB was investigated, using microarray and RNA-seq technology, respectively. Both comorbidities increase the risk of developing active TB, but the underlying mechanism in T2DM is unknown. The potentially beneficial effects of the anti-diabetes drug metformin, on the transcriptome of healthy donors, were also investigated, as existing reports indicate it could behave as an adjuvant for TB therapy.

The effect on the TB blood transcriptome signature of HIV-1 coinfection was studied in collaboration with the PanACEA consortium, and T2DM (HbA1c > 6.5) and pre-diabetes (HbA1c > 5.7 & < 6.5) in the TANDEM consortium. It was found that the TB treatment response transcriptomic signatures could still be observed with HIV-1 coinfection. Both T2DM and pre-diabetes affected the blood transcriptome: increased inflammatory profile with down-regulated type I interferon response genes. This could be indicative of an enhanced immunopathological response in TB/DM and of the important role of type I interferons in susceptibility to TB. Because patients with pre-diabetes

had similar transcriptomes to TB/DM, albeit of lower magnitude, it shows that even at intermediate levels of hyperglycaemia there is an immune dysfunction. Metformin had an anti-inflammatory effect in healthy donors in the context of *M. tuberculosis* stimulation, indicating its potentially beneficial role in TB/DM treatment.

Acknowledgements

I want to express my thanks to my supervisor Jackie for her encouragement, judgement and her, albeit sometimes bizarre, propensity to always be able look on the bright side of things. I feel like I have gained another parent; a relationship I am very lucky to have.

Hazel and her ever wise consult and the rest of the Dockrell group for their support, notably Shampa for our coffee breaks spent gossiping and complaining making it all that bit easier.

Vinod at the UMCG and the rest of the Genetics group for being so welcoming and letting me use their equipment. Without whom, I would not have anything noteworthy to write about.

Of course, my friends and family for enduring my many crises. Thanks to Brother #2 for being a LaTeX helpdesk. Brother #1 not so much. My partner for being my personal chef and therapist, and my dad for reminding me that maybe I have come quite far from the days I used to eat soil.

-*Stay Savage*

Ibi Atemie

Contents

1 Introduction **1**

1.1 Tuberculosis 1

1.1.1 A natural history 1

1.1.2 A global healthy issue. Again 3

1.1.3 Current actions in TB control 5

1.1.4 Immunology of Tuberculosis 11

1.2 HIV-1 15

1.2.1 The AIDS epidemic 15

1.2.2 HIV treatment and diagnosis 16

1.2.3 Immunology of HIV 18

1.2.4 TB and HIV coinfection 19

1.3 Diabetes Mellitus . . . 19
1.3.1 A history . . . 19
1.3.2 A disease of the West . . . 21
1.3.3 Diagnosis and management of diabetes . . . 23
1.3.4 Diabetes as an inflammatory syndrome . . . 27
1.3.5 TB and diabetes . . . 31
1.4 Transcriptomics . . . 31
1.4.1 Transcriptomics in TB . . . 33
1.4.2 Transcriptomics in diabetes . . . 36
1.5 Aims and Objectives . . . 38

2 Transcriptomic changes through TB treatment in HIV patients 39

2.1 Introduction . . . 39
2.1.1 TB and HIV coinfection . . . 39
2.1.2 TB and HIV Immunology . . . 41
2.1.3 Diagnosis and treatment of TB/HIV . . . 43
2.1.4 Microarrays in TB . . . 44
2.2 Aims and Objectives . . . 47
2.3 Methods . . . 48
2.3.1 Sample processing . . . 48
2.3.2 Data processing and analysis . . . 49
2.4 Results . . . 51
2.4.1 Quality control of RNA extraction . . . 51
2.4.2 Microarray two-colour direct analysis . . . 52

2.4.3 Microarray separate channel analysis . . . 60
2.5 Discussion . . . 65
2.5.1 Findings . . . 65
2.5.2 Limitations of study . . . 67

3 Transcriptional effect of Metformin **68**
3.1 Introduction . . . 68
3.1.1 Type 2 diabetes mellitus (T2D) . . . 68
3.1.2 Metformin . . . 70
3.1.3 Unconventional uses for an old drug . . . 71
3.1.4 Background of study . . . 75
3.2 Aims and Objectives . . . 77
3.3 Methods . . . 78
3.3.1 Sample Processing . . . 78
3.3.2 Data processing . . . 79
3.3.3 Differential expression analysis . . . 81
3.3.4 Gene-set analysis . . . 82
3.4 Results . . . 84
3.4.1 Assessment of RNA quality . . . 84
3.4.2 The effect of metformin whole blood gene expression . . . 86
3.4.3 The effect of metformin on stimulated and unstimulated PBMC samples . . . 94
3.5 Discussion . . . 105
3.5.1 Findings . . . 105

3.5.2 Interferons and metformin 106
3.5.3 Limitations and future work 109

4 Tuberculosis and diabetes in TANDEM: cross-sectional study 111

4.1 Introduction 111
4.1.1 Tuberculosis and Diabetes: the collision of two epidemics 111
4.1.2 Clinical presentation of TB/DM 113
4.1.3 Pre-diabetes 114
4.1.4 TB/DM immunology: a dysfunction 115
4.1.5 The TANDEM consortium 117
4.1.6 Current transcriptomic findings 120
4.2 Aims and Objectives 122
4.3 Methods 123
4.3.1 Sample collection and processing 123
4.3.2 Data processing and analysis 126
4.3.3 Modular analysis 129
4.3.4 Previously published signature analysis 129
4.4 Results 131
4.4.1 Patient characterisation 131
4.4.2 Assessment of RNA quality 139
4.4.3 Gene expression in TB/DM patients in South Africa 144
4.4.4 Gene expression in TB/DM patients in Romania 150
4.4.5 Gene expression in TB/DM patients in Indonesia 157
4.4.6 Gene expression in TB/DM patients in Peru 158

4.4.7 Gene expression in TB/DM in field sites combined 159
4.5 Discussion 170
4.5.1 Findings 170
4.5.2 The role of interferons 172
4.5.3 Limitations and future work 178
4.5.4 Implications in TB control 178

5 Tuberculosis and diabetes in TANDEM: longitudinal study 180

5.1 Introduction 180
5.1.1 Diabetes effects TB treatment response 180
5.1.2 The TANDEM consortium 182
5.2 Aims and Objectives 185
5.3 Methods 186
5.3.1 Sample processing and data analysis 187
5.4 Results 190
5.4.1 Diabetes and tuberculosis treatment outcomes 190
5.4.2 Global analysis of TB/DM patients through TB treatment 193
5.4.3 Directed analysis in closely defined population 204
5.5 Discussion 210
5.5.1 Findings 210
5.5.2 Implications on TB control 213
5.5.3 Comments on the study and future work 214

6 Final Discussion and Summary 217

List of Figures

1.1 Global incidence of TB in 2015 4

1.2 The immune response to *M. tb* 13

1.3 Global HIV prevalence in 2008 15

1.4 HIV prevention advert 16

1.5 Estimated total number of adults living with diabetes in 2015 22

1.6 Glucose regulation of insulin secretion by pancreatic β cells 26

1.7 Inflammation development in type 2 diabetes 29

1.8 Inflammation induction in islets of patients with type 2 diabetes 30

2.1 Prevalence of HIV in new TB cases. 40

2.2 The effect of HIV on *M. tb* granuloma formation. 42

2.3 Comparison of RNA quality 51

2.4 Two colour: before and after normalisation 53

2.5 Volcano plots of early and late changes in TB treatment 55

2.6 Heatmaps of top 1000 genes differentially expressed in intensive phase . . 57

2.7 Panel plot of modular expression at diagnosis compared to week 2 59

2.8 Separate channel: before and after normalisation 61

2.9 Line plot of genes that change at every time-point 63

2.10 Heatmap of Zak signature 64

3.1 Schematic of dosing of metformin during the pilot trial 78

3.2 RNA quality assessment from two different kits 85

3.3 RNA quality of all samples 86

3.4 Principal component analysis of ex vivo samples 87

3.5 P-value histogram for effect of metformin in ex vivo samples 88

3.6 Volcano plot of effect of metformin in ex vivo samples 88

3.7 Bar graph of pathway enrichment in whole blood 90

3.8 Effect of metformin on oxidative phosphorylation 91

3.9 Principal component analysis of cultured PBMCs 95

3.10 P-value histograms in cultured PBMCs in response to metformin 96

3.11 Volcano plot of global effect of metformin in PBMCs 97

3.12 Heatmap of significant gene-sets in PBMCs 98

3.13 Volcano plots of effect of metformin in different stimulation conditions in PBMCs 100

3.14 Network plot of effect of metformin in unstimulated samples 101

3.15 Network plot of effect of metformin in stimulated samples 102

3.16 Normalised count of genes in interferon module in PBMC samples 103

3.17 Gene ontology analysis of stimulated PBMC samples 104

4.1 Projected prevalent diabetes cases and current worldwide tuberculosis incidence 112

4.2 TANDEM consortium map 118

4.3 The TANDEM consortium 120

4.4 TANDEM RNA-seq cross-sectional study design 127

4.5 Age of patients recruited from the four field sites 132

4.6 Gender frequencies of patients recruited from the four field sites 133

4.7 Body Mass Index of the patients recruited from the four field sites 134

4.8 HbA1c values in patients in each field site 136

4.9 Proportion of TB/DM patients on anti diabetic medication 139

4.10 RNA quality electropherograms 140

4.11 Relationship between RNA quality (RIN) and field site 141

4.12 Relationship between RNA quality (RIN) and read alignment 142

4.13 Principal component analysis of diagnosis samples in South Africa 144

4.14 Volcano plots of each disease group in South Africa 146

4.15 Three-way Venn diagram of South Africa 147

4.16 Four-way Venn diagram of South Africa 148

4.17 Panel plot of modular differentially expressed in South Africa 149

4.18 Principal component analysis of diagnosis samples in Romania 150

4.19 Volcano plots of each disease group in Romania 152

4.20 Three-way Venn diagram of Romania 153

4.21 Four-way Venn diagram of Romania 154
4.22 Panel plot of modular differentially expressed in Romania 156
4.23 Volcano plots of each disease group in Indonesia 158
4.24 Volcano plots of each disease group in Peru 159
4.25 Principal component analysis of combined populations 160
4.26 Volcano plots of disease groups in combined population 161
4.27 Panel plot of modular differential expression in combined populations . . 163
4.28 Heatmap of top modules 165
4.29 Box plot of normalised count in interferon module 166
4.30 Heatmap of Zak signature 167
4.31 Absolute sum of Zak signature 168
4.32 Zak signature stratified by HbA1c 169

5.1 HbA1c values before and after TB treatment 191
5.2 Treatment outcomes 192
5.3 Normalised count of genes within B-cell module 195
5.4 Average normalised count of genes within B-cell module 196
5.5 Normalised count of genes within T-cell module 197
5.6 Average normalised count of genes within T-cell module 198
5.7 Normalised count of genes within interferon module 199
5.8 Average normalised count of genes within interferon module 200
5.9 Interferon module sum correlated with HbA1c 202
5.10 Interferon module sum correlated with HbA1c combined populations . . 203
5.11 Heatmap of Zak signature 204

5.12 Panel plot of modular differential expression in Indonesia 206

5.13 Module activity through TB treatment in Indonesia 207

5.14 Neutrophil gene heatmap of TB treatment in Indonesia 208

5.15 Interferon gene heatmap of TB treatment in Indonesia 209

List of Tables

1.1 The first-line anti-tuberculosis drugs 8

1.2 TB transcriptomic studies and their findings 35

3.1 Significant gene-sets in whole blood in response to metformin 92

3.2 Significant KEGG pathways in whole blood in response to metformin using second analysis tool 93

3.3 Global effect of metformin in gene-sets in PBMCs 99

4.1 Total number of RNA samples thats underwent RNA-seq analysis 124

4.2 Significant difference in Body Mass Index between patient groups in South Africa 135

4.3 Significant difference in Body Mass Index between patient groups in Romania135

4.4 Significant difference of HbA1c values between each patient group in Indonesia 137
4.5 Significant difference of HbA1c values between each patient group in South Africa 137
4.6 Significant difference of HbA1c values between each patient group in Peru 137
4.7 Significant difference of HbA1c values between each patient group in Romania 138
4.8 Sample numbers in cross-sectional study for Peru and Indonesia 143
4.9 Sample numbers in cross-sectional study in Romania and South Africa . 143
4.10 Total sample numbers in the three patient groups 143

5.1 Sample number for longitudinal analysis 187
5.2 Percentage of poor treatment outcomes 193
5.3 Most differentially expressed modules through TB treatment 194
5.4 Significance results at the gene-level of interferon genes 201
5.5 Sample numbers in Indonesia longitudinal analysis 205

1

Introduction

1.1 Tuberculosis

1.1.1 A natural history

Tuberculosis (TB) has long anguished humankind, reports in mummies indicate the relationship has been around for thousands of years, with paleopathological evidence dating back 8000BCE.[1] The bacterium responsible for causing tuberculosis disease, typically in the lungs, is *Mycobacterium tuberculosis*, which was discovered by the German bacteriologist Robert Koch in 1882.[2] Prior to this, in Ancient Greece, it was known as phthisis (phthiein = to waste away), first described by Hippocrates, the Greek physician known as 'the father of medicine'. He described a disease of 'weakness of the lung', which came with a fever and a cough and was often fatal. It was hypothesised that it was a hereditary disease, rather than an infectious one, due to it occurring throughout families.[3] It was first recognised as infectious by Galen of Pergamon, who described symptoms of a 'hectic fever', bloody sputum and small hard swelling in the lungs he named phûma.

In the 19th century, tuberculosis or consumption as it was known, was responsible for 25% of deaths in Europe. One of the most famous was George Orwell, who succumbed to TB just after completing *1984*, of which a lot of his suffering was reflected in the protagonist; Winston. In the 20th century however, the death toll started to decline as living standards and nutrition began to improve. Mortality continued to decrease with the discovery of the antibiotic streptomycin in the 1940s, and again with the introduction of the BCG vaccine in the 1980s. Tuberculosis was almost eradicated in the West in the late 20th century, although continued to decimate the developing world. But now, tuberculosis has made a comeback worldwide.[4]

M. tb is primarily a pulmonary pathogen, although it can assert itself all over the body, causing skeletal, pleural and lymph node tuberculosis amongst others. *M. tb* spreads through bacilli-infected air droplets, which are then inhaled into the lungs. The bacilli can either be eliminated by the host's immune response, or remain persistent in a latent state. This latency is where the bacilli have been contained so do not cause disease; it is asymptomatic and non-infectious. Approximately 10% of individuals latently infected go on to progress into active, symptomatic tuberculosis disease at some point in their lives. Of course, the *M. tb* infection to disease is a dynamic spectrum, but for clinical simplicity, individuals are characterised as either latent or active. Such clinical symptoms of active tuberculosis disease include coughing, fatigue, loss of appetite, haemoptysis, and more classically: weight-loss. Symptoms vary from individual and overlap with other respiratory diseases like pneumonia, or even lung cancer.[5]

Comparatively, *M. tb* is not as infectious as other pathogens, like smallpox or measles. This is measured by a 'reproductive number' or R_0. It equates to the average number

of secondary infectious cases generated when one infectious case is introduced into an uninfected population. In the case of *M. tb*, its R_0=2. For perspective, measles virus is R_0=12. That being said, the average infectious period of *M. tb* can be over a year and it is also frequently fatal. Without proper treatment, 45% of active TB patients die.

1.1.2 A global healthy issue. Again

Due to a multitude of reasons, TB has now become a global health issue. It began its resurgence in the 1990s, strongly correlating with the acquired immunodeficiency syndrome (AIDS) epidemic and human immunodeficiency virus (HIV) prevalence. Immigration, an ageing population and overcrowding have contributed to its revival. This revival is evermore dangerous than before because of its acquired drug resistance to the current antituberculosis therapy, caused by poor health care services and infrastructure, and poor treatment adherence.

TB is now the leading cause of death from an infectious disease in 2015, overtaking HIV. The World Health Organisation (WHO) reported that globally in 2015, there were 10.4 million new cases (56% were men, 34% women, 10% children). Only six countries accounted for 60% of these new cases; India, Indonesia, China, Nigeria, Pakistan and South Africa, with China and India contributing 45% (Figure 1.1). Within this, almost half a million cases were multi-drug resistant TB (MDR). Together with this, there were 1.4 million deaths worldwide caused by TB. Although the number of deaths has fallen since 2000, TB continues to be one of the top ten causes of death worldwide. These new cases are defined as symptomatic active disease, but the majority of people infected with *M. tb* remain latent. Current estimates of global latent infection are at 2-3 billion people.

As approximately 10% of infected people go on to develop active TB, this means that the many people with latent TB are acting as a reservoir of potential TB disease. A number of factors contribute to the activation of disease, including smoking, HIV-1 coinfection, malnutrition, diabetes mellitus, male gender and age.[6]

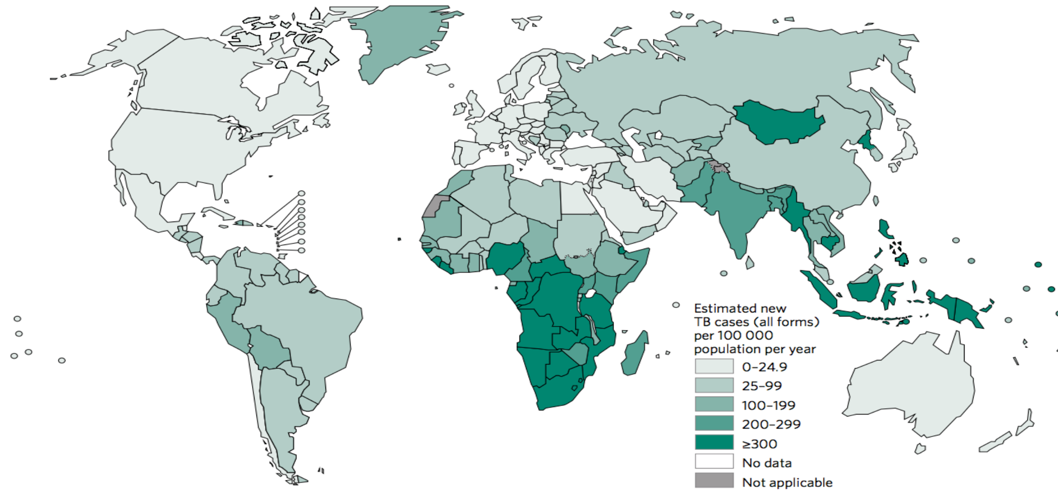


Figure 1.1: **Global incidence of TB in 2015.** New cases per 100,000 population per year. Data from WHO.

Progression to active TB can occur from months to years after initial infection, and can vary considerably depending on the individual. Progression can be effected by a myriad of factors, with most associated with an immunodeficiency. Immunocompromised individuals, like transplant patients or cancer patients undergoing chemotherapy are more susceptible to active TB disease progression. This is also the case with intravenous drug users and smokers. On the other hand, individuals in prison due to overcrowding, the homeless, and healthcare workers are at increased risk of infection because of higher exposure.

Malnutrition is a big risk factor, as it is also coupled with confounding risks, like

poverty, inequality, poor housing, and inadequate access to health services. This is of course a problem in low income countries that have high incidence of TB. Another large risk factor in the progression to active disease are comorbidities, especially HIV-1 coinfection. This is one of the most detrimental associated risk factors on the individual level. An HIV-positive individual is 10 fold more likely to progress to active TB compared to an HIV-negative individual. As well as HIV-1, other comorbidities increase susceptibility, for instance, diabetes, kidney disease, and some autoimmune disorders like lupus and rheumatoid arthritis. These comorbidities, especially diabetes, can end up being a larger risk factor on the population level. This is because a much higher number of people are exposed.[5]

1.1.3 Current actions in TB control

Vaccines

Currently there is only one licensed vaccine available against *M. tb*; the Bacillus Calmette-Guérin (BCG) vaccine. The BCG vaccine was introduced in 1921, and was heavily promoted by the WHO which lead to worldwide use during the 1950s. Now, more people have been vaccinated with BCG than any other vaccine,[7] although it has not come without its controversy, in that its impact on TB control is unclear.

The efficacy of the vaccine against adult pulmonary TB has been questionable. A meta-analysis revealed that the efficacy of BCG was highly variable, ranging 0-80%, with lower efficacy the closer one gets to the equator. The reason behind this variability is unknown, but it has been hypothesised that people in low efficacious regions have been exposed to environmental non-tuberculosis mycobacteria.[8] To further this, a meta-

analysis of paediatric BCG showed the duration of protection was up to ten years, which waned over time.[9]

Another vaccine candidate, the MVA85A vaccine recently underwent its phase II trial, which was designed to enhance the efficacy of the BCG vaccine in a prime-boosting strategy. Unfortunately however, the vaccine did not induce protection superior to BCG on its own.[10] There are several other vaccine candidates at various stages of clinical development, but none are likely to be licensed in the short-term.[11]

Diagnostics

When a patient enters a clinic with an unexplained persistent cough for two weeks, they are investigated for TB. Currently, the cornerstone of diagnosis is sputum culture, which is highly specific; able to differentiate *M. tb* from other non-tuberculosis mycobacteria (NTM). The liquid culture method is even more sensitive, which uses the Mycobacterial Growth Indicator Tube (MGIT) from BACTEC^{TM}. The main caveat of culture methods however is time; it can take a median of four weeks before a result is obtained as mycobacteria are notoriously slow growing. Another classical method is sputum smear microscopy using Ziehl-Neelson staining. This technique also has high specificity, but lower sensitivity than sputum culture (∼50%) because of the need for relatively high bacilli load to be detected. This presents a problem in HIV-positive patients who tend to have reduced numbers of bacilli because of the fewer cavitations.[12]

From microbiological methods, there are also molecular methods to diagnose TB based on polymerase chain reaction (PCR), including the Gene Xpert MTB/RIF. This technique is very fast and has high sensitivity and specificity. One of its outstanding

features is its ability to test for drug resistance to one of the principal first-line antibiotics; rifampicin. The locus that is responsible for rifampicin resistance; *rpoB* remains fairly conserved which means it can be successfully amplified and therefore detected. This is invaluable in the early detection of hard to treat cases, but comes at a high monetary cost which is not always feasible in low resource countries.

Contrasting to diagnostics that detect the presence of bacteria, methods exist to detect changes in the host, the most traditional being chest X-rays. Radiography is sensitive and has the advantage of detecting latent or asymptomatic TB. Again though, sensitivity is lower in HIV-positive patients and equipment is expensive. It is also of lower specificity, and radiologists require substantial training to accurately diagnose TB.

Another class: immunodiagnostics, detect the immune response to the pathogen rather than the pathogen itself. This includes the tuberculin skin test (TST) and the inteferon-gamma release assay (IGRA). They are both indicators of infection, not disease. The IGRA is superior in specificity to the TST in that IGRA measures the presence of *M. tb* specific antigens like early secreted antigen-6 (ESAT-6), so there are fewer false positives in for example BCG-vaccinated individuals.[5,13]

Treatment

Before the age of antibiotics in the 19th century, TB patients were prescribed rest and 'appropriate' nutrition, as TB caused significant weight-loss. Sometimes, patients would be sent to a sanatorium in the countryside for the fresh air. Times quickly changed with the discovery of the first antibiotic, penicillin by Alexander Fleming in 1928. However, this was not effective in *M. tb* because of its non-porous cell wall. Then, Selman Waksman

discovered streptomycin, which had a different spectrum of activity and was effective against *M. tb*,[14] resulting in him winning the Nobel Prize in 1952. Drugs currently used in TB treatment were discovered shortly afterwards; isoniazid in 1952, and rifampicin in 1965.[4]

The current TB regimen is six months of first-line chemotherapy drugs (Table 1.1). This is split into two phases; the intensive phase of two months, and then the continuation phase of four months. Bacteria exist in different populations within a tuberculosis lesion, and therefore have different characteristics like multiplying speed.[15] This means that a combination of drugs which also possess different characteristics is needed for successful treatment. For instance, isoniazid is particularly effective against rapidly growing bacilli. The first two weeks of chemotherapy reduces the bacilli by 90%, and sustained chemotherapy is required in order to kill the remaining 10%. Failure to do so can lead to recurrence or drug resistance. The intensive phase comprises four drugs; isoniazid, rifampicin, ethambutol and pyrazinamide, and kills most of the bacilli. Typically, patients are culture negative after the two month intensive phase, and this phase often determines the outcome of the regimen. The continuation phase is then the remaining four months with isoniazid and rifampicin, where the remaining persistent bacilli are killed.[5]

Antibiotic	Mechanism of action
Rifampicin	Inhibits bacterial RNA polymerase
Isoniazid	Inhibits mycolic acid synthesis
Pyrazinamide	Disrupts synthesis of ATP
Ethanbutol	Inhibits lipid and cell wall metabolism

Table 1.1: **The first-line anti-tuberculosis drugs and their mechanism of action.**

This standard six month regimen has a good success rate, with a recurrence rate of around 5%. But, problems arise with the regimen, centering around the treatment duration. With prolonged treatment, patients can develop drug toxicity particularly in the liver. This can undermine adherence, which in turn leads to treatment failure, relapse, development of drug resistance and ongoing transmission. Strategies have been put in place to aid drug treatment and adherence. Directly Observed Therapy-Short course (DOTS) was put in place by the WHO and involves standardised treatment with health worker visits to witness administration and provide patient support. It is part of a wider response including political commitment and sustained financing, case detection through quality-assured bacteriology, ensuring effective drug supply and management and implementing monitoring and evaluation systems.

Antibiotics have been, in a way, victims of their own success. As TB incidence fell in high income countries, the need for new drugs was neglected. There has been a lack of public and private investment in research and development since the 1970s, reflected in the lack of any new antibiotics. So, as a solution, other already available drugs are being optimised. These include traditionally second-line drugs; fluoroquinolones like moxifloxicin and gatifloxicin. These types of drugs have shown excellent in vitro activity, but in trials using fluoroquinolone-based treatment for four months have not achieved similar cure rates as the standard six month regimen.

One of these recent trials; ReMox involved shortening the regimen to four months and administering moxifloxicin instead of either isoniazid or ethanbutol. The result was there was an increase in time to culture negativity, but there was no difference in relapse rate compared to the standard six month regimen.[16] This showed that the culture

results at month two were insufficient in predicting TB treatment outcome. Another fluoroquinolone based trial, OFLOTUB also shortened TB treatment to four months and administered gatifloxacin instead of ethambutol during the intensive phase of treatment. They found that there was high variability across different populations and ultimately were unable to show non-inferiority, with more treatment failures, compared to the standard six month regimen.[17] These results show that being culture negative at two months does not always lead to sterile cure at the end of treatment, and shows why we need better ways to predict treatment outcome and relapse-free cure.

Treatment monitoring

Treatment monitoring is important because it provides information that would be used to make inferences regarding treatment success. Also, if monitoring showed that treatment was unsuccessful, it would allow for the appropriate alteration of regimen, ultimately reducing reactivation rates and drug resistance. Treatment monitoring is incredibly valuable in clinical trials involving new drugs and regimens, and also vaccine trials. In trials, drug efficacy is measured by the rate of relapse after two years post treatment. The two year relapse rate is 5%, so any new compounds need to be non-inferior. Currently the WHO recommends the two month sputum smear status as an indicator of poor treatment outcomes, however there is loss of sensitivity and risk of contamination.

Sputum conversion rate lacks precision in reliably predicting the long term clinical outcome of a regimen. The early bactericidal assay (EBA) is able to evaluate the patient's response to anti-tuberculosis medication within the first few days. The EBA is when anti-TB drugs cause significant reductions in sputum colony forming unit counts (CFUs) in

the sputum of a TB patient, measured during the first two days of treatment.[18] But, even though CFU is often used to detect bacterial load, it is not always representative of the total population of *M. tb* present. Barer et al noted that resuscitation-promoting factors were able to detect previously undetected viable *M. tb* cells, which had been invisible in standard culture demonstrating that there is a presence of non-replicating cells in sputum.[19]

A new method of measuring viable *M. tb* in sputum is the molecular bacterial load (MBL) assay which detects *M. tb* 16S rRNA. The culture-free method is incredibly rapid and accurate in quantifying the bacillary load.[20,21] A study comparing three different treatment monitoring methods; CFU, time to positivity (TTP) and MBL showed that the MBL assay was able to detect subpopulations of *M. tb* which were not detected by CFU or TTP methods.[22] It demonstrates that a molecular signatures have many advantageous. A molecular surrogate end point for TB treatment is much needed; a rapid and reliable predictor of potential poor treatment outcomes. Such a signature would be used across human populations and would therefore need to include information from different ethnic groups, geographical locations and also different comorbidities. Such a test that would reflect successful and continued cure would be beneficial in the clinical management of TB patients

1.1.4 Immunology of Tuberculosis

The immune response to *M. tb* can be complex and involves a medley of immune cells and responses, summarised in Figure 1.2. Studies have also shown that there is a fine balance between protection and pathogenesis and *M. tb* has adapted to persist in its host

through the construction of a granuloma. The fundamental control of *M. tb*, as shown in human and animal studies, involves $CD4^+$ T-cells, IFN-γ and TNF-α. The immune response to *M. tb* starts when the bacterium enters the respiratory tract via inhalation, where it encounters the first-line of defence; alveolar macrophages. The macrophages are able to internalise the bacterium via receptor mediated phagocytosis. This is done through Toll-Like Receptors (TLRs), manose receptors (MR), Fcγ receptors (FcγR) and complement receptors. These receptors behave as Pattern Recognition Receptors (PRRs) to the mannosylated lipoarabinomannan (ManLAM) present on the bacterial cell surface. Dendritic cells (DCs) also interact with *M. tb* in the early stages as an antigen presenting cell (APC) through Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin (DC-SIGN) receptors.

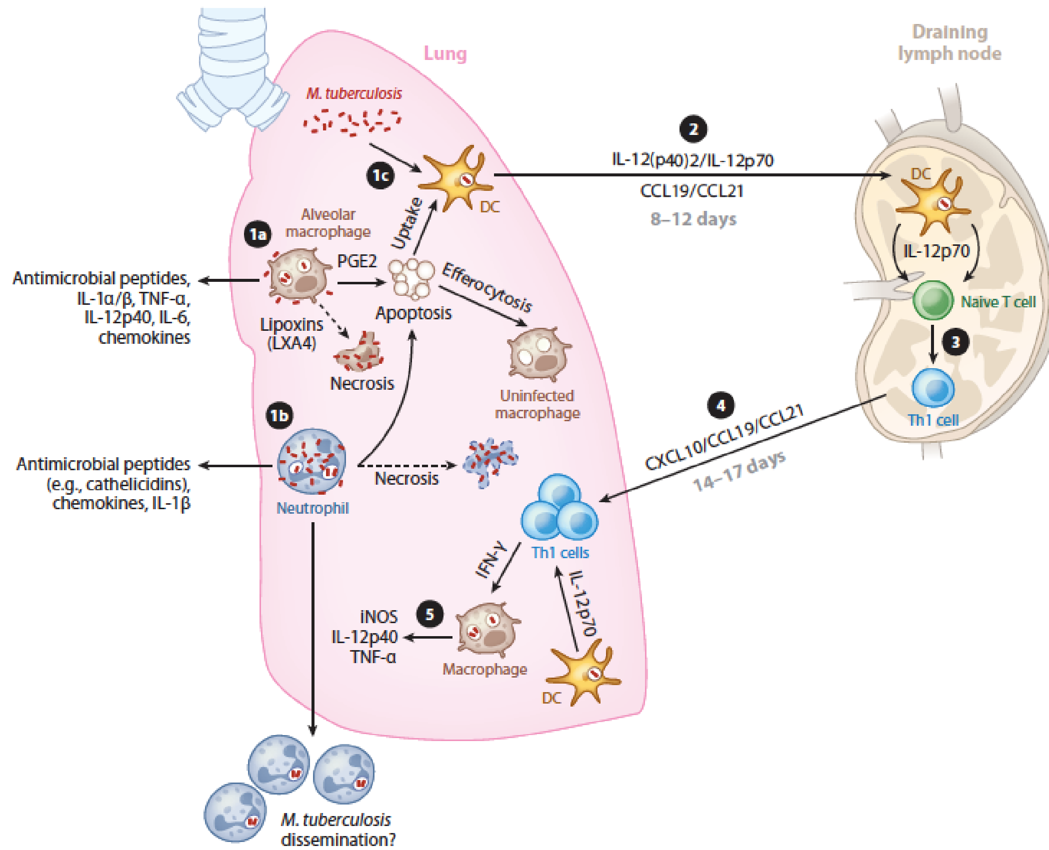


Figure 1.2: **The immune response to *M. tb*.** Following infection with *M. tb*, lung alveolar macrophages, neutrophils and DCs become infected, leading to the production of cytokines and chemokines.*M. tb* infected DCs migrate to the lung-draining lymph nodes 8-12 days post infection under the influence of IL-12, which then drive naive T-cells towards Th1 phenotype. Antigen specific Th1 cells migrate back to the lungs and produce IFN-γ, leading to macrophage activation and further cytokine production.[23]

The phagocytosed bacillus is then transported to the lysosome for degradation, but, *M. tb* can inhibit the phagosome-lysosome fusion and therefore prevent acidification and its killing. The infected macrophage can then induce necrosis, allowing the spread of bacilli, or apoptosis. Virulent strains of *M. tb* are able to manipulate macrophages and evade apoptosis. Infected DCs then transport *M. tb* to the lymph node for T-cell priming.

The naive T-cells are activated, start to proliferate and become effector cells able to migrate towards the lung tissue. They are driven towards a Th1 response by IL-12, so are able to secrete IFN-γ which is central in the killing of *M. tb.* Primed CD4$^+$ T-cells migrate to the lung, secreting IFN-γ, IL-2 and TNF-α, leading to macrophage activation. There becomes an influx of phagocytic cells causing the formation of a granuloma to contain the bacilli, like a prison. This physical barrier consists of macrophages, neutrophils and later CD4, CD8 T-cells and B-cells. But, as bacteria replicate, an imbalance can occur, causing necrosis, caseation and rupture of the granuloma, causing *M. tb* to leak into the airway.[24]

For the maintenance of the granuloma, TNF-α is crucial, the importance of which is seen in humans and animals. TNF$^{-/-}$ mice and patients on anti-TNF therapy for rheumatoid arthritis are highly susceptible to *M. tb.* Those on anti-TNF therapy are at 5 fold higher risk of developing TB.[25] Formation of the granuloma does occur in absence of TNF-α but at a slower rate and with higher bacillary numbers.

The adaptive immune response is also important in the maintenance of the granuloma. For example, CD4$^+$ T-cell deficient mice are unable to control bacterial growth because of the absence of IFN-γ. The significance of CD4$^+$ T-cells was witnessed during the HIV/AIDS epidemic, where patients were severely immunocompromised because of a depletion of their CD4$^+$ T-cell population.

1.2 HIV-1

1.2.1 The AIDS epidemic

AIDS was first recognised in 1981, and its cause; HIV was discovered not long afterwards. The HIV/AIDS epidemic was devastating, and so far 35 million people have died globally from HIV/AIDS. The world has committed to end the AIDS epidemic by 2030, but still in 2015, there were 2.1 million new HIV infections worldwide, and 1 million people died from HIV related diseases in 2016, leaving 36.7 million people living with HIV globally. There are large disparities across regions and within countries (Figure 1.3), with the majority of cases in eastern and southern Africa (19.1 million living with HIV).[26]

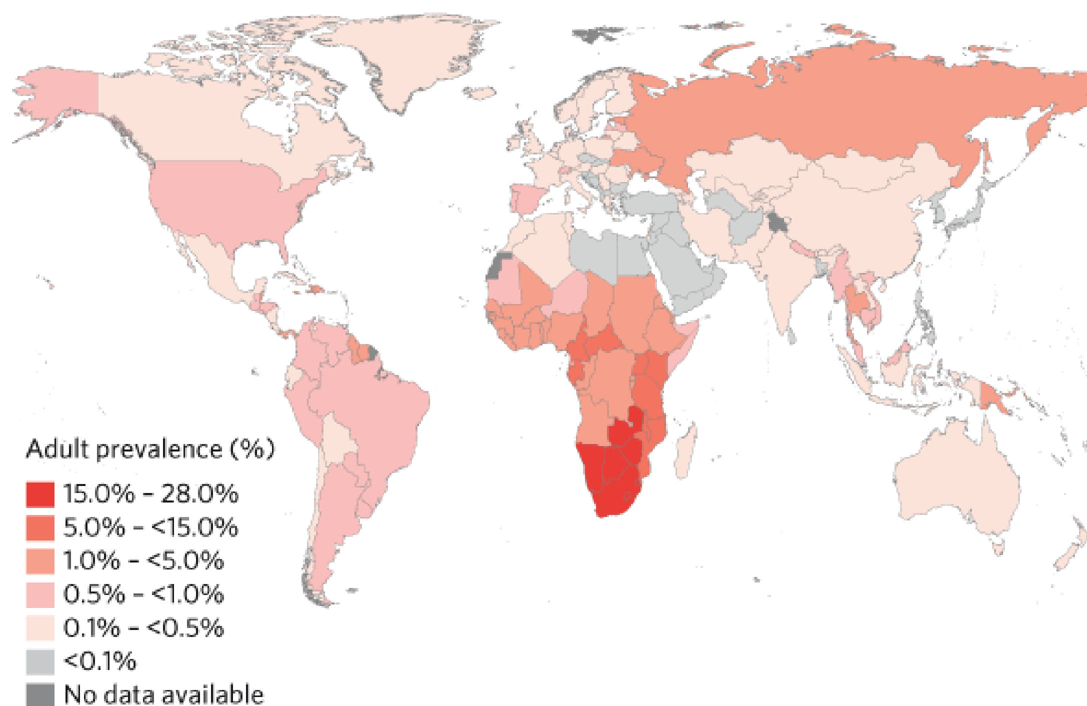


Figure 1.3: **Global HIV prevalence in 2008.** Data from WHO.

There was huge stigma surrounding HIV/AIDS because it is sexually transmitted and was associated as a 'gay disease'. This prudish mentality only added to the epidemic,

as there was such little awareness and reluctance in being diagnosed. Attitudes slowly started to change, and HIV prevention adverts were being shown on the TV and on posters, an example of which is shown in Figure 1.4. Now, fewer and fewer people are progressing to AIDS due to the development and rollout of highly active anti-retroviral therapy (HAART), also known as anti-retroviral therapy (ART).

In the absence of treatment, the progression of AIDS from HIV can take 10 years, but of course varies considerably. The hallmark of HIV infection is the progressive depletion of $CD4^+$ T-cells, and AIDS is defined as a $CD4^+$ count of below 200 cells per µl. AIDS patients succumb to opportunistic infections, like *M. tb*, pneumonia and certain cancers. Upon infection with the virus, 50-90% of individuals develop influenza-like symptoms, and due to the nonspecific nature, it is often not recognised as HIV. So, it has been recommended that people in high-risk groups that present with an unexplained fever should be tested for HIV.

Figure 1.4: **HIV prevention advert in 1991.**

1.2.2 HIV treatment and diagnosis

Diagnosis of HIV has come along way and now there is a point of care test that takes minutes, which detects HIV antigens. However, this has comparatively low specificity and sensitivity, so a PCR is used to confirm results. This detects genomic material; HIV

RNA and DNA, and also measures viral load. Serological assays are also used, namely enzyme-linked immunosorbent assays (ELISA) which detect antibodies to the virus.

Currently, there is no cure or vaccine, but ART has enabled HIV-infected individuals to have a normal quality of life. HIV can remain latent in subpopulation of cells or in reservoirs which current therapy cannot yet eliminate, so it is possible the HIV infection will reactivate. At the moment, therapy is based on a functional cure; the suppression of the virus to sub-clinical levels. ART is becoming a global success, and has reached 46% global coverage.

There are a few classes of ART drugs; reverse transcriptase inhibitor (RTI) like zidovudine, protease inhibitors (PI) like saquinavir, non-nucleoside reverse transcriptase inhibitor (NNRTI) like nevirapine and nucleotide reverse transcriptase inhibitors (NtRTs) like tenofovir. Within each class, there is variability in potency, toxicity and drug-drug interactions. For example, saquinavir interacts with rifampicin, so its use is limited in TB patients. Treatment regimens tend to follow triple therapy; two nucleoside RTIs and one NNRTI. Ongoing work involves optimising combination regimens to minimise side effects and improve adherence.[27]

Pre-exposure Prophylaxis (PrEP), commonly known as Truvada, and post exposure (PEP) prophylaxis are also available. PrEP is taken by those at substantial risk of HIV-infection and can stop the virus from establishing an infection. PEP involves taking ART after potentially being exposed to HIV in order to prevent becoming permanently infected, and is administered daily for 28 days. Although neither of these are 100% effective, both offer valuable options in the prevention of HIV.[28,29]

1.2.3 Immunology of HIV

HIV is transmitted across the genital epithelium, commonly through damage to the said epithelium, and transferred to mucosal DCs, macrophages and T-cells. HIV enters these cells through the CCR5 and also the CXCR4 receptors. The DCs then migrate to the lymph nodes to present the virus to T-cells, initiating the adaptive immune response.

A curious case; the 'Berlin patient' has been 'cured' of HIV. He was HIV-infected and underwent a bone marrow transplant for leukaemia in 2008. The donor had a CCR5 mutation so the receptor was defective, preventing HIV entry into cells. His viral load has remained below detection limits without ART.[30] This case demonstrated the central role for the CCR5 receptor and the potential for the development of CCR5-targeted treatment.

During infection, the HIV ssRNA activates the TLR7/8 receptor in DCs, increasing the production of type I IFNs which impair viral replication and recruit immune cells. The type I IFN: IFN-α recruits natural killer (NK) cells which are able to recognise and kill infected cells, however, HIV has evolved mechanisms to dysregulate NK cell function. As mentioned, HIV-1 depletes $CD4^{+}$ cells, causing loss of the cellular arm of the immune response. HIV-1 selectively infects $CD4^{+}$ cells and causes the cells to lyse and apoptose. These $CD4^{+}$ T-cells provide cytokines for proliferation and differentiation so progressive depletion of $CD4^{+}$ causes great immune function impairment. The other arm of the adaptive immune system also generates a response. Antibodies from B-cells can be detected and used in diagnostics. In vivo, neutralising antibodies are able to limit viral replication, but HIV quickly mutates to overcome it.[31] To add, the conserved regions of the virus are shielded or hidden from the immune system, so an antibody response is

not capable of killing the virus.

1.2.4 TB and HIV coinfection

M. tb is the most common opportunistic infection in HIV-positive individuals, and HIV was partly responsible for the resurgence of TB in the 1990s. TB is the leading killer of HIV-positive people; in 2015, 35% of HIV deaths were due to TB. In 2015 there were 1.2 million global new TB cases in HIV-positive individuals and 390,000 people died of HIV-associated TB. HIV is one of the largest risk factors of developing active TB with HIV-positive patients are 10 times more likely to progress to active TB disease. Considering an estimated quarter of the worlds population is latently infected with *M. tb* and 36.7 million people globally live with HIV, it is a huge amount of potential disease and also death.[6]

HIV and *M. tb* share a biological synergy; cell-mediated immunity, specifically $CD4^+$ T-cells, which are essential in the control of *M. tb*, and a depletion of this arm of immunity is the hallmark of HIV infection. This is discussed in more detail in Chapter 2.

1.3 Diabetes Mellitus

1.3.1 A history

Diabetes is a progressive chronic disorder characterised by abnormalities of metabolism. Its overriding feature is the inability to control glucose appropriately, resulting in a myriad of problems throughout the body. There are a number of type of diabetes, mainly type 1 (T1DM) and type 2 (T2DM) which differ in causation and pathogenesis.

Reports of diabetes date back centuries, it was first described in Egypt 1500BC in reference to 'a medicine to drive away the passing of too much urine'. After which the hindu physician Sushruta noted a disease of honey urine, echoed in Persia by Avicenna who detailed an account of diabetes including 'a melting down of flesh and limbs into urine'.[32] Ancient diagnosis for such a thing involved tasting the urine, or seeing if ants congregated around it. It was not until 81-133AD when a Greek physician Aretaeus of Cappadocia coined the term diabetes, from the Greek 'to pass through', describing the excessive urination in which the condition is associated (the Latin word for sweet; mellitus was later added). It was in this time that diabetes was thought of as a disease of the kidney, 'diarrhoea of the urine', a theory put forward by Galen (AD 129-210), which remained in Europe throughout the Renaissance. Later discoveries proved the diabetic urine was caused by sugar[33,34] and unravelled the role of the liver and then the pancreas. The role of the pancreas was notably discovered by Mering and Minkowski in 1889 when they performed a pancreatectomy in dogs which resulted in a condition similar to diabetes.

Diabetes had been a relatively uncommon disease before the 20th century and was classically associated with the wealthy, but changes in lifestyle meant prevalence increased, as Robert Saunby put it, 'there is no doubt that diabetes must be regarded as one of the penalties of advanced civilisation'. The first reference of diabetes as an epidemic was made in the 1920s by Eliot Joslin, one of the most prominent diabetes specialists of the 20th century. He espoused that six people in a row of houses had died from diabetes, and had the cause of death been cholera, the health authorities would have been round instantaneously.[35]

In the search of a treatment for diabetes, Banting and Best won the Nobel Prize in Physiology or Medicine for the isolation of insulin in 1921. The results of the use of insulin in humans were published the next year, with patients experiencing a reduction in blood glucose and no glycosuria. Finally, after centuries, a treatment for diabetes could be administered and mass-produced.[35] In the decades that followed, hyperbolic headlines in the mainstream media alluded to a 'cure for diabetes', and it has become an almost mundane, manageable condition, but the reality is it remains a pernicious disease.

1.3.2 A disease of the West

T2DM accounts for 90% of the diabetic population, the prevalence of which has been rising exponentially. The number of people with diabetes had risen from 108 million in 1980 to 422 million in 2014. Current projections estimate 552 million cases by the year 2030.[36] The top burden countries are China, India and the United States of America (Figure 1.5). In addition to this it has been estimated that diabetes directly caused 1.5 million deaths in 2012.[37]

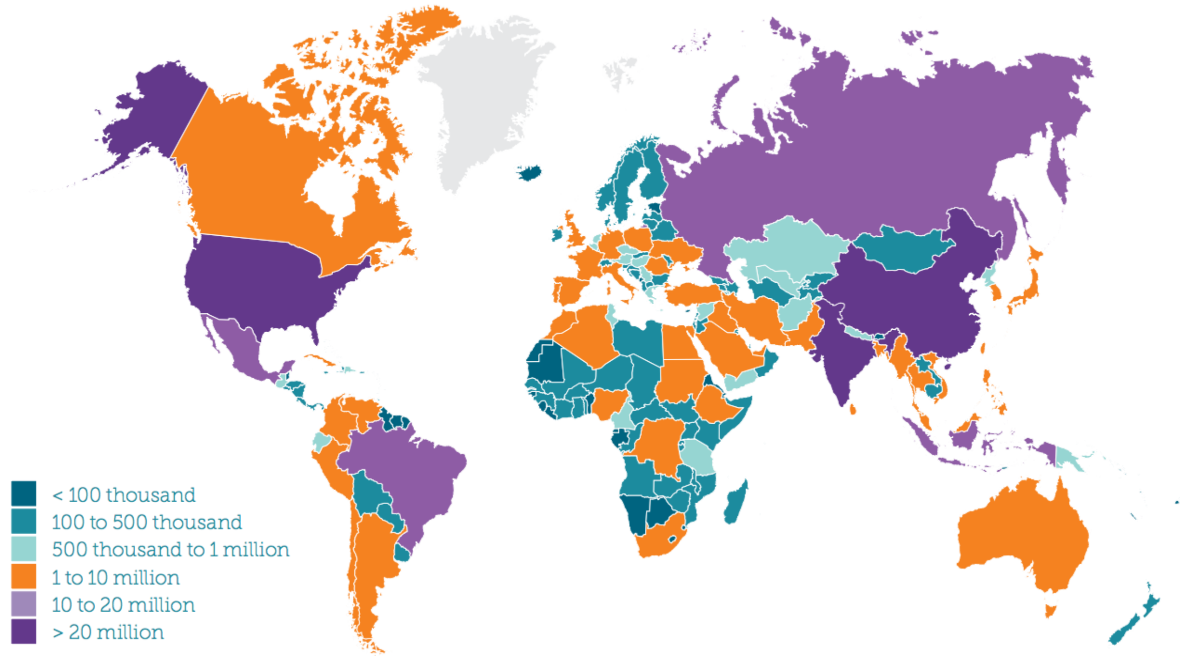


Figure 1.5: **Estimated total number of adults living with diabetes in 2015.** IDF Diabetes Atlas, 2015.[36]

This current and projected increase is a result of urbanisation, dietary changes and a rise in life expectancy. There is a strong genetic predisposition, but over time, humankind's lifestyle has become almost unrecognisable, now risk factors for T2DM like physical inactivity and obesity have become commonplace. The majority of T2DM patients are overweight (80-90%) and obesity, defined as a body mass index (BMI) > 30, is strongly correlated with T2DM. Obesity is one of the largest risk factors in developing T2DM, and in itself has become an epidemic. Cases of obesity have more than doubled since 1980 and in 2014, 600 million adults globally were classed as obese.[38]

1.3.3 Diagnosis and management of diabetes

Although often diabetes patients can be asymptomatic, they present the classical symptoms of frequent urination, excessive thirst, weight loss and blurred vision. In clinics, patients are screened for T2DM if they are over the age of 45, or have two or more risk factors; physical inactivity, a first-degree relative with diabetes, hypertension, high cholesterol, obesity or history of cardiovascular disease (CVD).

Current diagnostic tests are random plasma glucose (RPG), glycated haemoglobin A1c (HbA1c) and fasting plasma glucose (FPG). A RPG reading of $\geq 200mg/dl$ is positive for diabetes, and is confirmed using HbA1c ($\geq$ 6.5%), or FPG ($\geq 126mg/dl$). RPG is highly variable depending on when the subject last ate, so it not always reliable which is why subsequent testing is needed. HbA1c is more expensive and slightly less sensitive than FBG, but is much more convenient as the patient does not need to fast.[39]

In more detail, HbA1c is a form of a haemoglobin; the oxygen carrying protein in erythrocytes, that is bound to glucose. It is also referred to as glycated haemoglobin. Haemoglobin can assess glucose concentration because intracellular erythrocyte glucose concentrations are in equilibrium with glucose plasma levels. A reaction occurs between glucose and the amino groups in haemoglobin, forming glycated haemoglobin. This glycated haemoglobin builds up in erythrocytes, and is able to reflect the average level of glucose to which the cell has been exposed to over its life time of 120 days. HbA1c is therefore used as a longer term gauge of blood glucose levels.

Before the pharmaceutical era, management of diabetes was principally diet and exercise, and even starvation therapy. This was effective, but the rising prevalence demanded more advanced medical treatments. The implementation of pharmacotherapy was in full

swing in the 1990s. Different therapeutics with different effects are available including injectable insulin. Also available are biguanides, like metformin, which act on the liver and enhance the effects of already present insulin, and sulfonylureas, like glibenclamide which act directly on the pancreas to augment the supply of insulin. Each have pros and cons, but metformin is the most favoured as it does not induce severe hypoglycaemia. Commonly, anti-diabetes drugs are administered in combination; insulin and metformin together lower glucose more effectively, and adverse effects such as weight gain are avoided.[39] It should be noted that pharmaceutical intervention is an adjuvant, rather than a substitute for dietary modification and exercise.

Still, exercise therapy has remained an effective treatment for the prevention of T2DM as well as having other benefits. In a clinical trial in the USA, non-diabetic patients who had elevated fasting plasma glucose were administered either metformin twice daily, a placebo, or lifestyle modifications (weight-loss goals and physical activity). The follow-up was over two years, and they found that lifestyle intervention reduced the incidence of T2DM by 58%, compared to metformin by 31%.[40]

Complications resulting from uncontrolled T2DM is a major cause of disability, and even early death. These disease manifestations include kidney disease, blindness and CVD. T2DM is the leading cause of new cases of blindness in adults, and also increases the risk of CVD by 3-6 times. Therefore it is imperative diabetes patients receive that appropriate support and treatment to control their glucose levels.

Mechanism of glucose control

Glucose control is tightly regulated by two hormones, insulin and glycogen which work in antagonism of one another. Dysregulation causes severe negative effects on the body, including death, so regulation is paramount. Diabetes is a chronic disease that is characterised by pancreas either not producing enough insulin or the body not using the insulin it produces effectively. T1DM has a strong genetic component and is where the insulin-secreting cells in the pancreas have been destroyed by the immune system, leading to an absolute deficiency of insulin. On the other hand, T2DM is where inadequate amounts of insulin are secreted relative to the blood glucose concentration, while the insulin-secreting cells remain intact.

The insulin-secreting cells are located in the pancreas; an endocrine secreting tissue, comprised of groups of cells called the Islets of Langerhans. The Islets of Langerhans contain a number of cell types; primarily α cells which secrete glucagon, and β cells which secrete insulin. The regulator of these two hormones is the amount of glucose in the blood. Insulin is a small peptide protein: it is anabolic and stimulates glucose uptake through plasma membrane receptors into the tissues allowing for glucose to be metabolised and stored as glycogen and fat.[41]

When blood glucose levels rise, the glucose transporter GLUT2 transports the glucose into the β-cells, where it is converted to glucose-6-phosphate by hexokinase IV and enters glycolysis. Because of the increase of glucose catabolism, ATP increases, causing the closure of ATP-gated K^+ channels in the plasma membrane. This causes a reduced efflux of K^+, which in turn depolarises the membrane. The closing of these channels increases the intracellular positive charge and decreases the membrane potential. Membrane depo-

larisation opens voltage gated Ca^{2+} channels, resulting in the increase in cytosolic Ca^{2+} concentration. This increase triggers the release of insulin via exocytosis, summarised in Figure 1.6. The subsequent reduction in blood glucose levels are detected by the β-cells, which slows or halts the release of insulin, initiating the feedback regulation.[42]

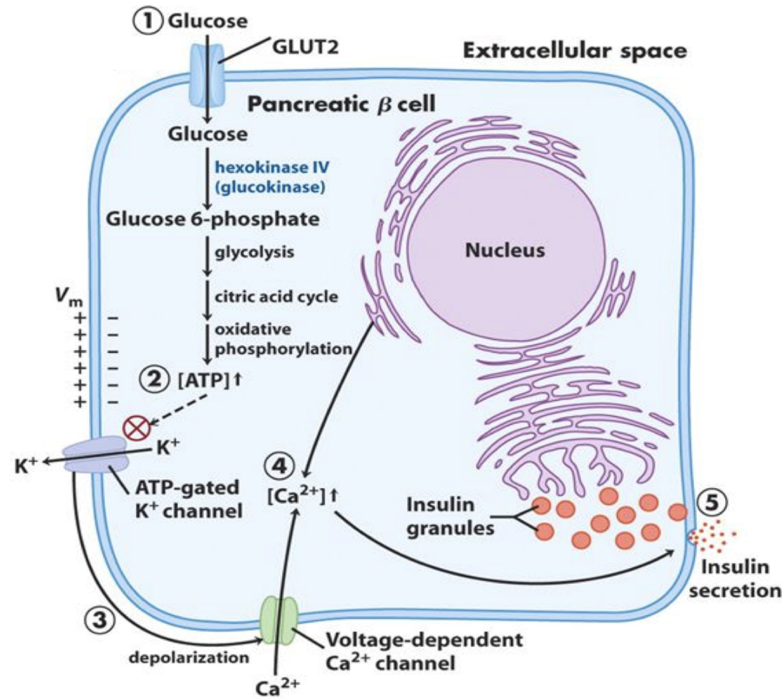


Figure 1.6: **Glucose regulation of insulin secretion by pancreatic β cells.** When blood glucose levels are high, metabolism in the β cells raises ATP, closing K^{+} channels in the plasma membrane, depolarising the membrane. In response to the change in membrane potential, voltage-gated Ca^{2+} channels open, causing an influx of Ca^{2+} into the cytoplasm, which triggers the release of insulin.[42]

Diabetes pathogenesis

T2DM is characterised by the failure of the body to respond appropriately to insulin, coupled with impaired secretion and insensitivity of target cells to insulin. This is known as insulin resistance and occurs to some extent in all obese individuals and T2DM pa-

tients. It is where there is a decrease in the ability of insulin to alter the metabolism in target tissues. This occurs when there is an oversupply of nutrients relative to energy expenditure, so more insulin is secreted to compensate. The pancreas has to continuously secrete more insulin. In time, the β-cells become exhausted from the continual extra insulin secretion needed, and start failing to secrete sufficient insulin to obtain normal blood glucose levels leading to impaired glucose tolerance and uptake in skeletal muscle.[39]

Chronic exposure to high glucose concentration is toxic and raises reactive oxygen species (ROS) levels. The pancreatic β-cells have low levels of anti-oxidative enzymes so are vulnerable to oxidative stress, therefore constantly elevated glucose levels leads to β-cell failure and loss via apoptosis.[43] The progressive β-cell death contributes further to a decrease in adequate insulin secretion.

Overabundance of free fatty acids (FFAs) also majorly contribute to adverse metabolic effects such as insulin resistance. An increase in FFAs is tightly linked with obesity, where there is an excessive growth of adipose tissue. FFAs add to insulin resistance by impairing glucose uptake in muscle because of competitive inhibition and also reduces the inhibition of glycogenolysis. With elevated adipose tissue mass, higher levels of FFAs can leave the adipose cell and enter the circulation to be taken up by other organs like the liver and skeletal muscle.

1.3.4 Diabetes as an inflammatory syndrome

T2DM has been linked to inflammation for some time; back in the late 1800s it was found certain non steroidal anti-inflammatory drugs (NSAIDS), salicylates and aspirin improved hyperglycaemia in diabetes patients.[44] This was again noted in the 1950s, when

a diabetic patient was administered high dose aspirin to treat their arthritis, and they no longer required daily insulin infections.[45] Fifty years later, it was found that salicylates' molecular target was to inhibit the phosphorylation action of the IκB kinase (IKK) complex, a central regulator of NFκB activation.[46] However, even though anti-inflammatories exhibited desirable effects like increased insulin sensitivity and reduced HbA1c values, the focus of medications was targeted on insulin secretion rather than any anti-inflammatory properties, so did not gain momentum.

T2DM has now been described as an immune dysfunction disorder, with patients displaying persistent low grade inflammation. This is down to the relationship between T2DM and obesity. In obesity, there is increased adiposity, and adipocytes are an endocrine organ which secrete fatty acids, hormones, cytokines and chemokines (Figure 1.7). Several cytokines, namely TNF-α, IL-6 and IL-1β are secreted by adipocytes. This was first noted in a study by Hotamisligil et al, where TNF-α mRNA was found in adipose tissue from rodent models of obesity and diabetes. Then, when TNF-α was neutralised in the obese rodents, there was an increase in the peripheral uptake of glucose in response to insulin.[47] In a separate study, TNF-$\alpha^{-/-}$ mice were resistant to the development of insulin-resistance in diet-induced obesity.[48] TNF-α seems to play a role in inflammation-induced insulin resistance and demonstrates how inflammation and insulin resistance are connected.

Dysfunction of some components of the immune system in T2DM has been illustrated, including altered concentrations of cytokines and chemokines, and changes in absolute number of immune cell subsets. For instance, not only TNF-α, but other pro-inflammatory cytokines are reported to be elevated in obese subjects; serum IL-6 has been

reported to be higher in T2DM patients than in healthy controls, and also positively correlated with HbA1c levels.[49] Also, IL-6 expression is elevated 10-fold in the adipose tissue of obese individuals than in lean individuals.[50–52] These pro-inflammatory cytokines in turn recruit macrophages, which are present in higher numbers in the adipose tissue of obese individuals.[53] Together, this causes a feedforward process; activation of NFκB and the chronic activation of pro-inflammatory pathways within insulin target cells, leading to further insulin resistance. With the association of obesity and T2DM, inflammation and insulin resistance, this shows the tie between inflammation and metabolism.

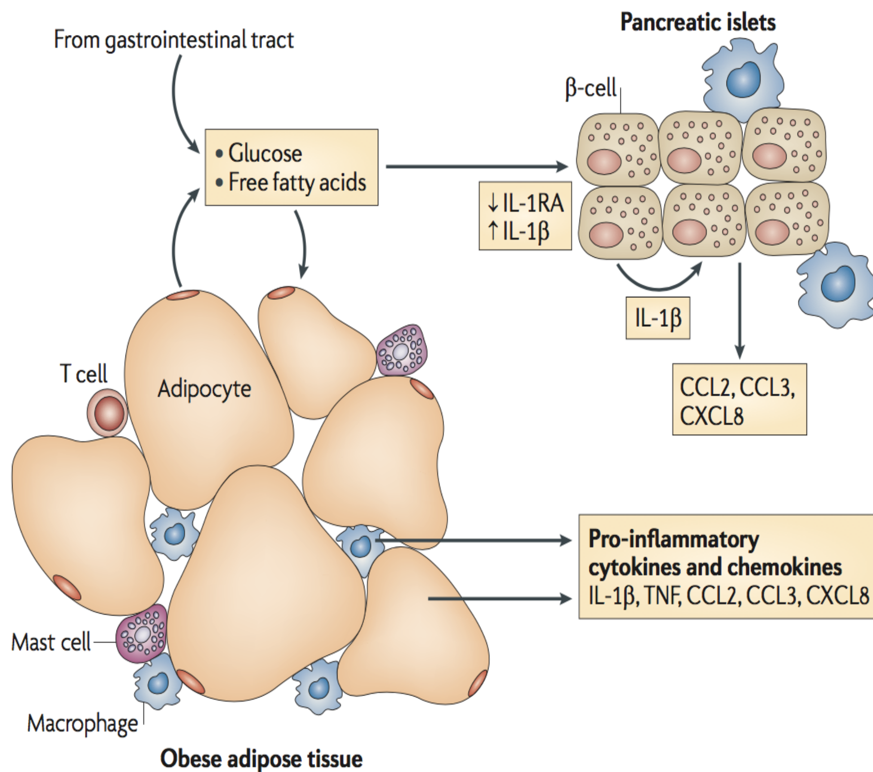


Figure 1.7: **Inflammation development in type 2 diabetes.** Excessive levels of glucose and FFAs stress the pancreatic islets and insulin-sensitive tissues like adipose tissue. This leads to the production and release of pro-inflammatory cytokines and chemokines resulting in the recruitment of immune system cells and inflammation.[54]

It is also possible for FFAs to trigger inflammation, through the stimulation of TLR signalling, Figure 1.8. A hypothesis for this is that FFAs behave as a ligand for TLRs due to their structural similarity with bacterial cell wall lipopolysaccharides (LPS).[55] Stimulation of TLR4 by FFAs for example leads to the activation of NFκB and downstream production of pro-inflammatory cytokines.[56,57] This mechanism further contributes to the subclinical inflammation displayed in T2DM patients.

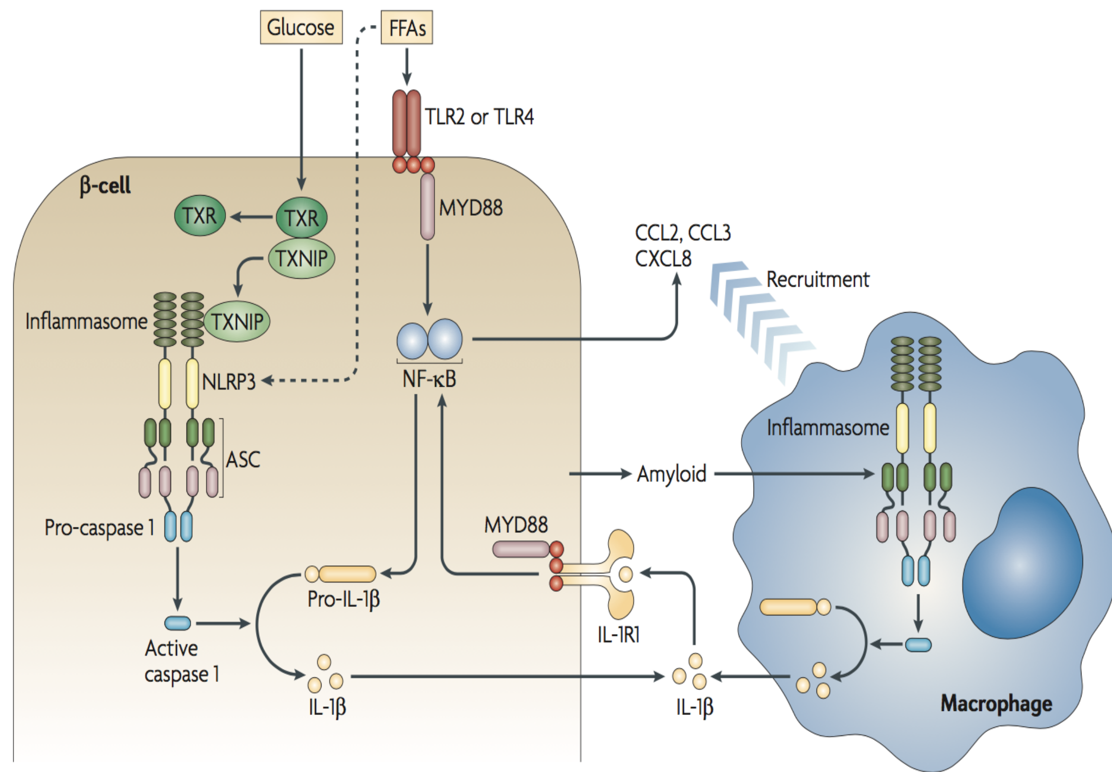


Figure 1.8: **Inflammation induction by glucose and FFAs in islets of patients with type 2 diabetes.** High concentrations of glucose promote the secretion of pro-inflammatory cytokines through NF-κB. FFAs can also directly induce activation of TLR2 or TLR4, also activating NF-κB and enhancing inflammation.[54]

Bearing in mind that T2DM is an inflammatory syndrome, it is possible that future therapies for diabetes would target inflammation. A proof-of-concept trial entitled

Targeting INflammation Using SALsalate in T2DM (TINSAL-T2DM) recently investigated the effect of administering anti-inflammatories on glycaemia in T2DM patients. They found that the salsalate group had reduced HbA1c compared to the placebo group demonstrating that modulating inflammatory pathways is a possible therapy for T2DM patients.[58]

1.3.5 TB and diabetes

A link between TB and diabetes was first noted centuries ago, but the comorbidity did not come into the spotlight. Diabetes represents a risk factor for active TB progression, and also for poor treatment outcomes. T2DM patients are 3.1 times more likely to develop active TB than non-diabetic individuals.[59] Considering the vast growing number of cases of T2DM globally, it serves as a potential reservoir for active disease. T2DM has been described as a chronic inflammatory metabolic syndrome, but the exact mechanism of how T2DM causes elevated TB susceptibility is unclear. TB/DM is discussed in more detail in Chapters 4 and 5.

1.4 Transcriptomics

The transcriptome is the complete set and quantity of gene RNA transcripts present in a cell. Analysing the transcriptome can determine and quantify the amount of gene activity in for example a certain cell type, developmental stage or physiological condition. Analysis of the transcriptome has become a decisive tool in the characterisation and understanding of the molecular basis in biological variation. What is key however, is that it shows the *relative* abundance of the transcripts between two physiological states.

It is used heavily in investigating the gene expression changes in diseases, for example, it is able to quantify the immune response to a disease.

Large scale gene expression studies first started in the 1990s when the microarray was invented, prior to which, gene expression analysis was investigated using RT-PCR on an individual gene basis. Microarrays are able to measure over 40,000 mRNA transcripts at the same time, in an unbiased approach. Over the years it has become a very popular, robust and informative method, which has seen the biological advancement of many investigations between disease and health, or pharmaceutical effects. Microarray technology is based on hybridisation of a sample onto a chip with pre-existing probes, which is discussed in more detail in the following chapter.

Microarrays have been widely used, but recently RNA-sequencing (RNA-seq) has become an attractive alternative and is slowly replacing them. In RNA-seq, an RNA sample is enriched for mRNA (or depleted for rRNA) and fragmented. The fragments are converted to a library of cDNA, with adapters attached to either end. The inserts are also tagged with a barcode and sequenced in a high throughput manner. This produces millions of short reads (30-400bp) which can then be mapped to a reference genome, or, with a method called *de novo* assembly, where reads can be overlapped into sections of sequence without the aid of a reference. Once the alignment is complete, the number of reads mapped within a genomic feature are counted. The number of reads mapped to a gene is used as the measure of its abundance in that sample.[60,61] Transcript abundance can then be compared in samples from different conditions in order to show the differences in expression levels.

There are several advantages of RNA-seq over microarray. Firstly, RNA-seq has

a larger dynamic range with less background noise, so it can detect lower abundance transcripts.[62] Because there is more background noise in microarrays, which is then subtracted, genes expressed at low levels are lost. Studies have compared the two technologies, and it has been found that RNA-seq is more accurate at calculating absolute transcript abundance.[63,64]

Limitations in RNA-seq do exist. The analysis and normalisation is more complex as there tends to be more variation within and between samples, specifically in library size and transcript length. Length of transcripts differ greatly, so longer genes have a higher total number of mapped reads than a short transcript which needs to be taken into account. Complex algorithms have been and continue to be develop in order to tackle these issues, but there is still no gold standard.

1.4.1 Transcriptomics in TB

RNA expression analysis by microarray, and now by RNA-seq has become a powerful tool for understanding disease biology, namely in cancer, but now also in infectious disease including TB. Now there is a plethora of literature on transcriptomics in TB. The first studies showing TB was detectable in blood were in 2007, where it was found that patients with TB were distinct from healthy controls and latently infected individuals,[65,66] paving the way to a possible blood biomarker. A principal study followed in 2010 by Berry et al used a much larger cohort with more detailed characterisation. Using microarray technology, a 393-gene active tuberculosis signature was found, relative to latent tuberculosis infected individuals and healthy controls. This signature was strongly associated with interferons, of which the magnitude correlated with radiographs and disease

severity. The active TB signature cleared throughout anti-tuberculosis treatment, diminishing after six months which allowed the construction of a 'molecular distance to health', quantifying the response to treatment. To add, it was also reproducible in two different populations; South Africa and the UK, and further analysis showed the TB signature was distinct from other inflammatory diseases (sarcoidosis).[67] This TB-specific distinct signature has been shown in several other studies, Table 1.2. Not only does it divulge important information regarding biological mechanisms, but a biosignature could behave as a much-needed diagnostic.

Further on from this, Cliff et al conducted a longitudinal study to investigate the changes in the transcriptome throughout anti-tuberculosis treatment. Within the first week of therapy, there was large-scale down-regulation in the transcriptomic profile, and moderate changes later in treatment.[68] Another independent study confirmed these findings with strikingly similar results, despite samples being collected from different populations and the use of different technology platforms. It does indeed seem that the signal is treatment related, hinting at a possible method for monitoring treatment and clinical management and potentially the efficacy of new anti-TB drugs.

Previous RNA expression studies have excluded HIV co-infected patients, which is not ideal as it does not accurately represent a typical TB endemic population. In order for a biosignature to be useful, it must be applicable across populations, other ethnic groups and different geographical regions. To address this, Kaforou et al performed a study involving a large cohort of TB patients with other diseases, including HIV coinfection. They also mixed different populations as they wanted to create a signature that was as inclusive as possible. They found that the signature developed was significantly

more powerful at discriminating active TB from healthy controls if all the TB patients (with and without other diseases) from different populations were combined than if they were analysed separately.[69]

Reference	Year	Population	Study design	Sample	Findings
Mistry et al[66]	2007	South Africa	TB versus LTBI	Whole blood	22 gene pattern
Jacobsen et al[65]	2007	Germany	TB versus LTBI	PBMC	Fcγ-receptor signalling
Berry et al[67]	2010	UK and South Africa	TB versus LTBI, HC, OD	Whole blood	IFN signalling
Maertzdorf et al[70]	2011	The Gambia	TB versus LTBI, HC	Whole blood	JAK/STAT, IFN signalling
Maertzdorf et al[71]	2011	South Africa	TB versus LTBI, HC	Whole blood	-
Bloom et al[72]	2012	South Africa	TB treatment	Whole blood	-
Ottenhoff et al[73]	2012	Indonesia	TB versus HC, TB treatment	PBMC	IFN signalling
Kafarou et al[69]	2013	South Africa, Malawi	TB versus LTBI, OD (HIV+/-)	Whole blood	-
Bloom et al[74]	2013	UK	TB versus OD, TB treatment	Whole blood	IFN signalling
Cliff et al[68]	2013	South Africa	TB treatment	Whole blood	Complement
Anderson et al[75]	2014	South Africa, Malawi, Kenya	TB versus OD, LTBI	Whole blood	-

Table 1.2: **TB transcriptomic studies and their findings.** Eleven different studies investigated TB blood transcriptomes, varying in population, study design and sample type. Latent TB (LTBI), other diseases (OD), or healthy controls (HC).

A more recent study demonstrated that a correlate of risk signature, likely an early indicator of active tuberculosis, could be predicted 18 months prior to clinical diagnosis, by blood transcriptomics. This was pivotal because patients at risk of progressing into active disease potentially can be predicted, permitting the possibility of intervention. Healthy latently-infected adolescents were recruited in South Africa. During the two year follow up, patients who developed active tuberculosis disease were classified as progressors. Using RNA-seq technology, a 16-gene signature was developed, differentiating progressors from matched controls. The predictive signature achieved 71.2% sensitivity in detecting tuberculosis 0-180 days before diagnosis. The expression of the genes in progressor samples increased as tuberculosis diagnosis approached, relative to matched controls. The signature was validated in household contacts of smear positive TB patients recruited in the Gambia and South Africa. Potentially, once validated in additional propective clinical trials, such as the Cortis trial, clinicians will be able to target preventative therapies to at-risk individuals identified by a blood test.[76]

So far, transcriptomic studies have involved quite simple populations; South Africa has been heavily studied. As mentioned, for a signature to be effective, it needs to be more inclusive. So, more complex situations need to be investigated to include different comorbidities and different populations.

1.4.2 Transcriptomics in diabetes

In comparison, blood transcriptomic studies in T2DM are few whereas more studies have been conducted with tissues such as muscle. A blood-based study in Brazil by Manoel-caetano et al measured the gene expression in PBMCs of T2DM patients, and found that

T2DM had a distinct profile from non-diabetic subjects. In more detail, T2DM exhibited up-regulation of genes involved in inflammation, responses to hypoxia, oxidative stress and fatty acid processing compared to non-diabetic controls. Gene ontology analysis confirmed that the most significantly differentially expressed genes were involved in the immune response and fatty acid metabolic processing.[77]

To add, Grayson et al compared the peripheral blood gene expression profiles of patients with metabolic syndrome, coronary artery disease and T2DM. They also found that the gene expression profile of T2DM was distinct from healthy controls, with many significantly differentially expressed genes involved in the activation of T-cell signalling.[78] These results complement the hypothesis that T2DM is indeed a chronic pro-inflammatory disorder and start to uncover the molecular mechanisms of T2DM pathogenesis and potential susceptibility. Transcriptomics in TB/DM comorbidity patients had not yet been performed until very recently by Prada-Medina et al[Prada˙sytems˙2017] where they found that the blood transcriptome did not distinguish TB/DM from non-diabetic TB patients and noted more quantitative differences rather than qualitative.

1.5 Aims and Objectives

Aims

I **To investigate the effect of comorbidities on the TB blood transcriptome and its resolution through TB treatment**

II **Investigate the potential for metformin to be used as host-directed therapy**

Objectives

I **To investigate the effect of HIV coinfection on TB gene expression before and during treatment, using microarrays**

II **To examine the effect of diabetes on TB gene expression before and during treatment, using RNA-seq technology**

III **To investigate the potential mechanisms of action of metformin in response to *Mycobacterium tuberculosis* using RNA-seq**

2

Transcriptomic changes through TB treatment in HIV patients

2.1 Introduction

2.1.1 TB and HIV coinfection

TB has become such a global health issue, it was listed as one of the top ten causes of death worldwide, and caused more deaths than HIV/AIDS in 2015. It was estimated there were 1.4 million deaths from TB across the globe in 2015. Amongst them, 0.39 million deaths were among HIV-positive people. This coinfection has gained track from both diseases sharing an overlap of high incidence countries, particularly sub-saharan Africa, Figure 2.1. TB/HIV together has presented such a problem, so much so that 12% (1.2 million) of new TB cases (9.6 million) and 25% of TB deaths in 2014 were among HIV-infected patients. TB is the most common opportunistic infection associated with HIV, and remains a leading cause of death among people with AIDS.[6]

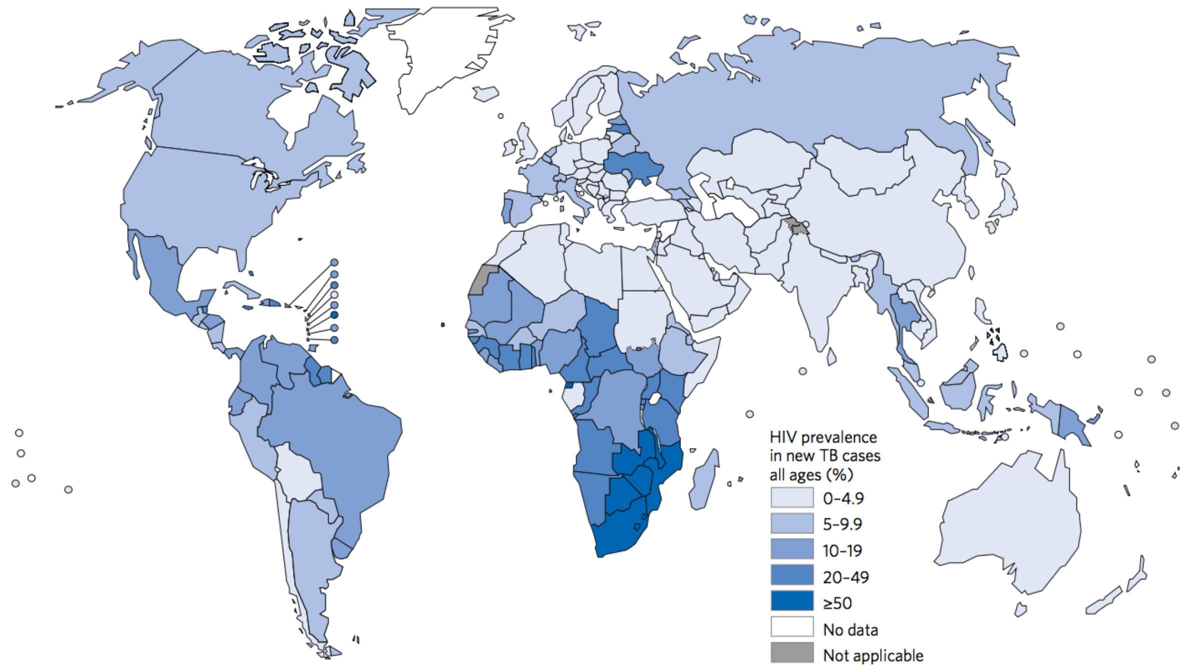


Figure 2.1: **Prevalence of HIV in new TB cases.** WHO Global TB report (2015).

Ever since the beginning of the AIDS epidemic in the 1980 and 1990s, the link between TB and HIV has played a substantial role in the resurgence of TB, and has contributed to the spread and pathogenesis of both diseases. Not only are HIV-infected patients more vulnerable in progressing to active TB disease with an increased relapse rate, but TB also accelerates the progression to AIDS.

HIV-1 is a potent risk factor of active TB disease; in general, people latently infected with *M. tb* have a 10% life-long chance of progressing to active TB disease, but this increases dramatically when patients are co-infected with HIV-1 to 10% per year. This increased risk is down to the biological synergy between HIV-1 and *M. tb*, in which both infections potentiate each other to deteriorate the immune function. This synergy can be attributed to $CD4^+$ T-cells, which were proven to play an integral role. HIV-1 infection causes $CD4^+$ lymphocytes to progressively decline in number and function, which are

crucial in the containment of *M. tb*.[79,80] So, HIV-1 infection can dramatically increase the rate of active TB disease. ART works at dampening HIV replication and restores $CD4^+$ T-cell counts.[81] However, even HIV patients who retain normal $CD4^+$ T-cell levels are still five times more likely to develop active TB than HIV-uninfected individuals.[82] All in all, the coinfection is a huge problem in the control of both diseases.

2.1.2 TB and HIV Immunology

HIV first comes into contact with macrophages at the mucosal surface. It enters the cell using the CCR5 or CXCR4 receptor. Recent studies have shown that HIV is able to disrupt aspects of the innate response. For example, HIV causes a dysfunction in the macrophage responses; studies have shown that it reduces the viability of macrophages and causes impaired levels of TNF-α.[83]

There is evidence that HIV and TB potentiate each other. HIV infection of macrophages increases *M. tb* growth. In turn, the TB induced pro-inflammatory response, including TNFα, IL-1β and IL-6 and decrease in IL-10, then drives HIV replication.[84]

HIV also causes the loss of the cellular arm of the adaptive response. As discussed, HIV primarily infects $CD4^+$ T-cells, and replicates, causing accelerating apoptosis and viral-mediated cell lysis. This progressive $CD4^+$ depletion restricts Th1 responses, which produce IFN-γ, IL-2 and TNF-α; key cytokines in the control of *M. tb*. This has been shown in *in vitro* studies where T-cell responses to *M. tb* in HIV-infected cells are significantly lower as well as a reduction in IFN-γ production.[85]

M. tb granulomas formed in HIV-positive individuals differ in their architecture compared to in HIV-negative individuals. There are fewer immune cells and defective for-

mation of the granuloma, Figure 2.2. This is particularly the case for CD4 T-cells; the lower the CD4 count, the fewer pulmonary cavities and poorer granuloma formation and containment. This inevitably results in an increase in dissemination of *M. tb*. This is why HIV causes reactivation of latent *M. tb* and faster progression to active TB disease. Conversely, HIV-positive patients are less likely to transmit *M. tb* because fewer bacilli are able to replicate extracellularly inside granulomas prior to rupture and release via aerosol through coughing.

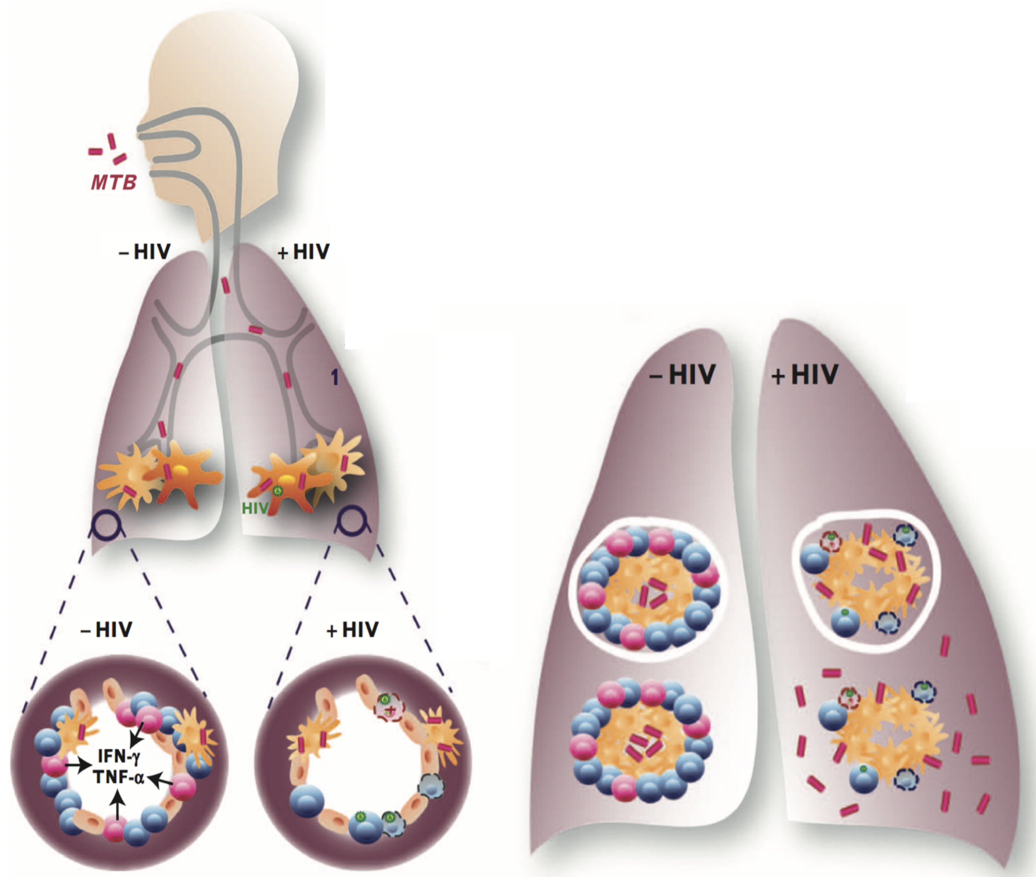


Figure 2.2: **The effect of HIV on *M. tb* granuloma formation.** How HIV-induced CD4 T-cell depletion impairs the immune control of *M. tb* infection. In HIV-positive individuals there are reduced numbers of CD4 T-cells in the lungs at the air-tissue interface. Reduced numbers result in defective granuloma formation ultimately leading to dissemination of the bacilli.[85]

2.1.3 Diagnosis and treatment of TB/HIV

Diagnosis of active TB in HIV-1 co-infected patients can be a challenge. This is because in HIV-infected patients, there tends to be a lower concentration of *M. tb* in sputum due to the fewer cavity formations. The GeneXpert is a valuable diagnostic tool as it demonstrates high sensitivity whilst also can detect rifampicin resistance. In some countries, like South Africa, it is replacing sputum smear microscopy as the first assay in diagnosing TB. The Xpert MTB/RIF was endorsed by the WHO in 2010, although it is still expensive. However, in HIV co-infected patients, the GeneXpert MTB/RIF had significantly lower sensitivity than HIV-negative TB patients.[5] Diagnosis of TB in HIV-infected individuals is difficult as the HIV immunosuppression reduces T-cell responses, which causes false negative TST and IGRA results.

A relatively new test for *M .tb* that uses urine detects lipopolysaccharide liboarabinomannan (LAM) on the cell wall of *M. tb* has proven effective (66.7%) in HIV co-infected patients with advanced immunosuppression i.e. $CD4^+$ count of $< 50/\mu L$. The TB-LAM is able to act as a rapid-diagnostic for those most at risk of mortality for TB disease so it is very much useful tool. However, new and better diagnostics are needed, which is where a biomarker could have much potential.[86]

Treatment of TB/HIV coinfection can be complex, due to the potential drug interactions between ART and anti-tuberculosis medication. The current treatment regimen for TB/HIV patients is the same as for HIV-uninfected patients: the standard six months of combination antibiotics. ART restores the immune depletion caused by the virus, reducing the progression into AIDS, and it also drastically reduces the incidence of active TB in HIV-positive patients.[87] However, there has been little research on treatment out-

comes, and rates of TB among HIV-positive patients remain very high even when they are on ART.

Treatment with ART in TB/HIV patients can develop into a serious condition; immune reconstitution inflammatory syndrome (IRIS). This is when the immune system starts to recover, and there is a rapidly normalising $CD4^+$ T-cell poopulation, which can lead to an overwhelming inflammatory response which is ultimately detrimental to the host. It occurs in 10-40% of HIV-positive patients and can be life threatening.

A number of randomised control trials have investigated different therapy regimens for TB/HIV patients. One involved administering 12 months of standard anti-tuberculosis therapy to HIV-1 co-infected individuals, and found patients had lower relapse rates than patients who were administered the standard six months of treatment.[88] Another involved administering HIV-1 patients isoniazid prophylaxis for one year after the standard TB treatment, which also found a slight reduction in relapse or reinfection.[89] A meta-analysis summarising several studies found similar results; an increase in treatment length slightly reduced relapse rate, and also patients were three times more likely to fail treatment if they were not on ART. They recommended at least 8 months of therapy, daily dosing and concurrent ART, as these were associated with better outcomes, but a randomised trial would need to confirm this.[90]

2.1.4 Microarrays in TB

Microarray analysis can be used to monitor genome wide expression levels of many genes under a particular condition. Microarrays work using a probe based method. Thousands of probes are bound to a chip, and are used to measure the relative concentration of

specific DNA sequences in a solution. These probes represent the genes in a genome, so one can analyse the 'genome wide' expression, of known genes, however. In gene expression studies, the RNA is extracted from a sample of interest, and fluorescently labelled. The probes hybridise to the labelled nucleic acids in the mixture, and can therefore quantify the relative nucleic acid sequences. The fluorescence at each probe spot is detected by a laser and the fluorescence intensity directly corresponds to the measure of expression of the probe associated gene.[5]

Microarrays have been used extensively in cancer research, but now also in infectious diseases research. So far, a number of microarray experiments have been done on patients with uncomplicated TB, and have shown their expression profile is distinct compared to the profiles of latently infected patients, other inflammatory diseases and also healthy controls.[67] However, TB patients co-infected with HIV-1 have mostly been excluded from such studies.

Kaforou et al characterised blood transcriptome signatures in active TB disease in both HIV-positive and HIV-negative individuals. This was a large microarray study with mixed populations and coinfection status. When HIV-positive patients were included in the analysis, the biosignature generated was more powerful than without them. Including coinfection patients in such studies is important as it is a more realistic representation of a typical high TB burden population, like in sub-saharan Africa, where the TB/HIV burden is the greatest.[69]

Longitudinal studies have shown that in non-complicated TB patients, rapid large-scale changes are seen in blood gene expression in the initial stages of treatment, specifically within one week.[68] Such changes could act as a marker for drug efficacy as it may

correlate with rate of relapse.[91] Again though, these longitudinal studies have excluded HIV-1 co-infected patients. So, if these early changes also occur in patients with HIV-1 coinfection, biomarkers could be developed for populations with a high prevalence of HIV-associated TB, facilitating drug development and clinical management.

This study was conducted in collaboration with the Pan-Africa Consortium for the Evaluation of Antituberculosis Antibiotics (PanACEA consortium). To add context, recent publications using the same cohort were investigating the molecular bacterial load assay as an early bactericidal.[21] Much work, including previously in our group has been conducted in a specific location or ethnic group in Cape Town, South Africa, so this project gave us the opportunity to not only determine whether our results were reproducible in HIV-positive patients, but also in a different geographical location, ethnicity and different prevalence of *M. tb* strains.

I hypothesised that coinfection does not confound TB induced blood transcriptomes before and during treatment, enabling coinfected patients to be combined with TB-only patients in treatment-response based biomarkers. A transcriptomic signature generated from TB patients would be effective when analysing HIV/TB patient data.

2.2 Aims and Objectives

Aims

I **To investigate whether peripheral blood transcriptomic changes are detectable during TB treatment in HIV coinfected patients.**

Objectives

I **Perform microarray analysis on whole blood samples from active TB patients with and without HIV-1 coinfection.**

II **Compare the early and late transcriptomic changes in HIV-negative and HIV-positive patients.**

III **Compare the transcriptomic changes in the early and late phases of TB treatment.**

2.3 Methods

2.3.1 Sample processing

Twenty five pulmonary tuberculosis patients were recruited from primary care clinics in Mbeya (NIMR-MMRP in Mbeya Referral Hospital), Tanzania in affiliation with the PanACEA consortium. The study design was originally 20 HIV-positive TB patients and 20 HIV-negative TB patients, but problems arose with the shipment of the samples.

Patients were sputum smear positive for acid fast bacilli pulmonary TB, 18-65 years old, and weighed 31-90kg. Patient exclusions were: diabetics, pregnant women and other serious lung conditions. All patients received standard TB treatment from day one onwards, with daily doses adjusted to weight range, following the guidelines of the Tanzanian National TB and Leprosy Programme (NTLP). Of these 25 patients, 15 were HIV-negative and 10 were HIV-positive. HIV-positive patients were all on HAART.

Three millilitres of venous blood were collected at TB diagnosis, and at three time points after the initiation of TB treatment; week two, week four and week eight. The blood was collected into Tempus Blood RNA Tubes (Applied Biosystems) and frozen at $-80°C$ and shipped to the London School of Hygiene & Tropical Medicine.

Total RNA was extracted using the Tempus^{TM} mRNA kit (Ambion). Samples were DNase treated and the globin transcripts were depleted using Ambion^{TM} GlobinClear kit (Life technologies), and RNA quantification wad performed using the Nanodrop. Fifty nanograms of RNA was fluorescently labelled with Cy3 (Cyanine 3-CTP) or Cy5 (Cyanine 5-CTP) using the Two-Colour Low input Quick Amp Labelling Kit (Agilent). Sample labelling followed a direct design, and were alternatively labelled between diagnosis and

week 2, and week 4 and week 8. Often, microarray experiments involve a common reference design, however, a direct design is one in which there is no single RNA source which is hybridised to every array, but rather dye swaps between two conditions. This design was chosen to detect the early and late responses in TB treatment, and it as seen that the very early changes in anti-tuberculosis therapy are the most profound. cRNA was hybridisated onto the SurePrint G3 Human Gene Expression 60K, Two-colour Microarray Based Gene expression analysis Genechip (Aglient), and Genechips were scanned using the SureScan Microarray Scanner at University College London.

2.3.2 Data processing and analysis

The Agilent Scanner creates a TIFF image. To extract the data, Agilent provides a software for feature extraction, where it subtracts each feature from the background. Data were then processed and analysed using the R (R v3.4.1 and RStudio v1.0.143) package `limma` (linear models for microarray and RNA-seq data) (v3.28.21) and normalised using the 'loess' method. `limma` is a widely used package in microarray analysis; it fits a linear model to the expression data for each transcript, and is capable of analysing multifactorial experiments.[92]

Firstly, a two colour direct analysis was done, and then a separate channel analysis was performed. The direct analysis was done because of the way the samples were labelled in two pairs, and is the traditional and most commonly used method. The analysis directly compared diagnosis with week 2 (early) and week 4 with week 8 (late). This was to investigate the early and late transcriptomic changes until two months of TB treatment. A contrast matrix was then constructed which enables one to extract results for the

second variable; HIV status. Fold change estimates were then estimated for each level in the factorial design; early changes in HIV-negative, early changes in HIV-positive, late changes in HIV-negative and finally late changes in HIV-positive. Extracting these showed the effect of HIV-1 coinfection on the early and late transcriptomic changes seen in the first two months of TB treatment.

The second analysis was a separate channel analysis. This method analyses the individual channel intensities as separate observations, instead of the log ratio. This method allows one to compare conditions that are not connected in a two-colour experimental design. For instance, loop designs, and designs that use a reference are connected, but direct designs are not always. In this particular experimental design, it was unconnected, with samples from early and late time points unlinked by hybridisation to the same array, creating 'islands' of arrays. Therefore it means that one can maximise the use of the information generated, and in this case estimate the difference between early and late samples.[93]

Modular analysis was performed using R package `tmod` (v0.31), using the modules outlined by Chaussabel et al[94] and by Li et al.[95] An adjusted p-value of < 0.05 was deemed to be significant.

A previously published TB signature, was also analysed to test whether its resolution at week 8 was confounded by HIV coinfection. The Zak signature[76] was chosen for reasons outlined in Chapter 4.The differential expression at week 8 compared to diagnosis was extracted for the 16 genes in both HIV-negative and HIV-positive patients.

2.4 Results

2.4.1 Quality control of RNA extraction

Two different RNA extraction kits were available for Tempus^{TM} blood collection tubes. The first was column based and collected total RNA including mRNA, and the second was magnetic bead bases and collected total RNA including mRNA and miRNA. For optimisation, whole blood was collected in Tempus^{TM} tubes from healthy LSHTM donors by Carolyn Stanley, and RNA was extracted using either of the two kits. Figure 2.3 shows the RNA quality obtained from the two kits. The Spin Column kit consistently extracted good quality RNA, with all RINs over 7.5. On the other hand, RNA quality yielded from the MagMAX extraction kit was highly variable and often of poor quality, with a median RIN of 4. Because of the MagMAX's inconsistency and sub-par quality yield, it was decided the samples from Mbeya should be extracted with the Spin Column kit.

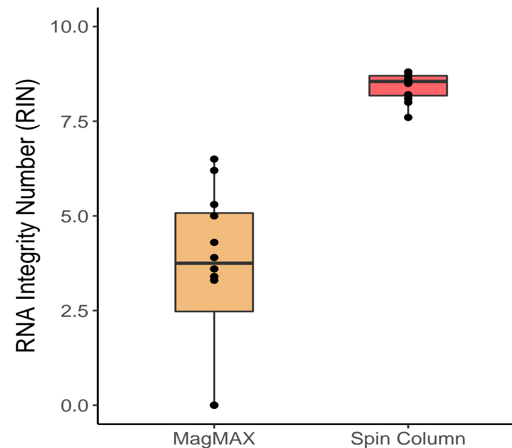
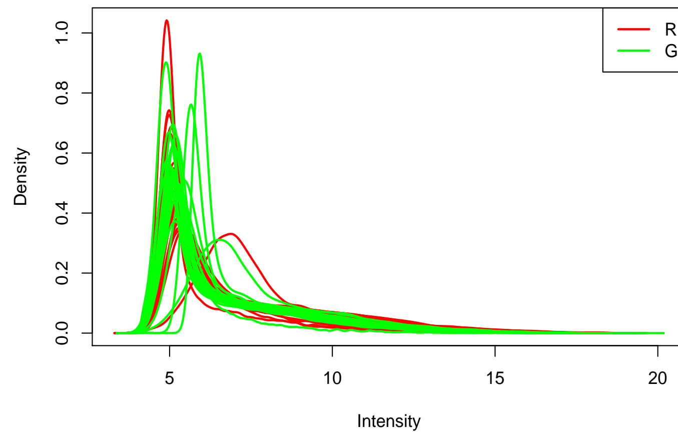


Figure 2.3: **RNA extraction from whole blood collected in Tempus^{TM} tubes, using either MagMAX or Spin Column method.** Whole blood was collected from healthy donors from LSHTM (n=9). RNA quality(RIN) was obtained from the BioAnalyser (Agilent).

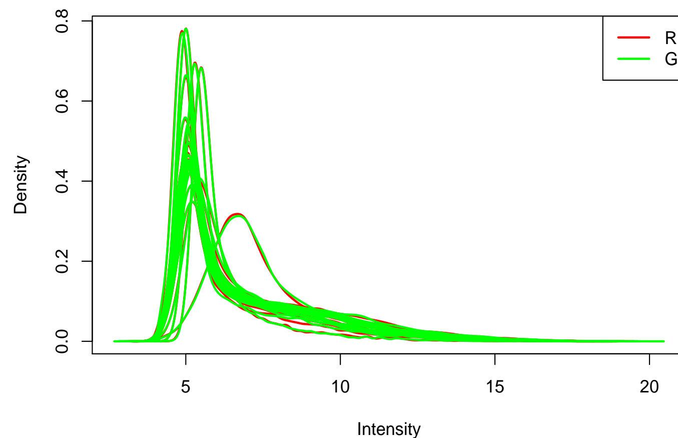
2.4.2 Microarray two-colour direct analysis

Firstly, in microarray data, normalisation needs to be performed, so that the different intensities of sample labelling are comparable. Systemic variation occurs in gene expression microarray experiments due to technical factors, like differential labelling efficiency and mRNA concentration. So, the purpose of normalisation is to eliminate such variation, leading to appropriate comparison of the biological signal.

Figure 2.4 shows before and after normalisation within the arrays. In Figure 2.4B, the green and red channels overlap, which means the loess normalisation achieved in marrying the two channels within an array together. The normalisation has aligned the two channels so the distributions become the same. Whereas prior to normalisation there is more variability between arrays and also variability between the red and green channels corresponding to Cy3 and Cy5 labelling.



(A)

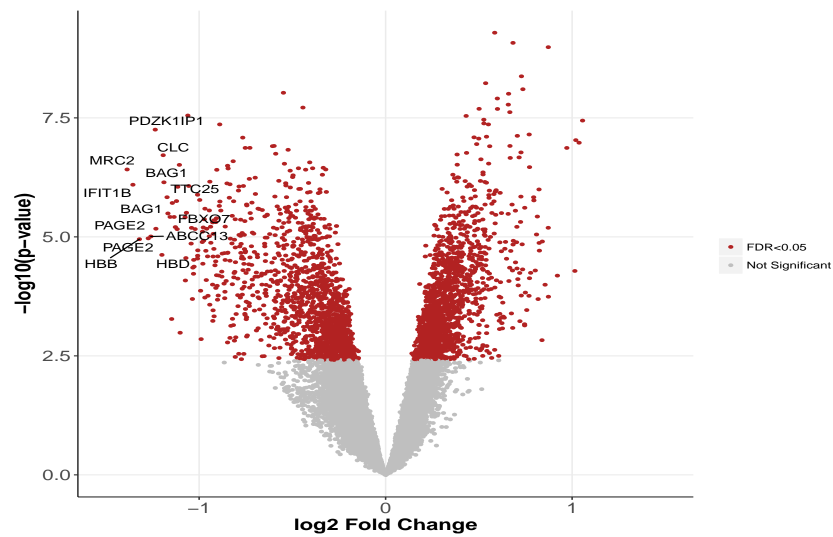


(B)

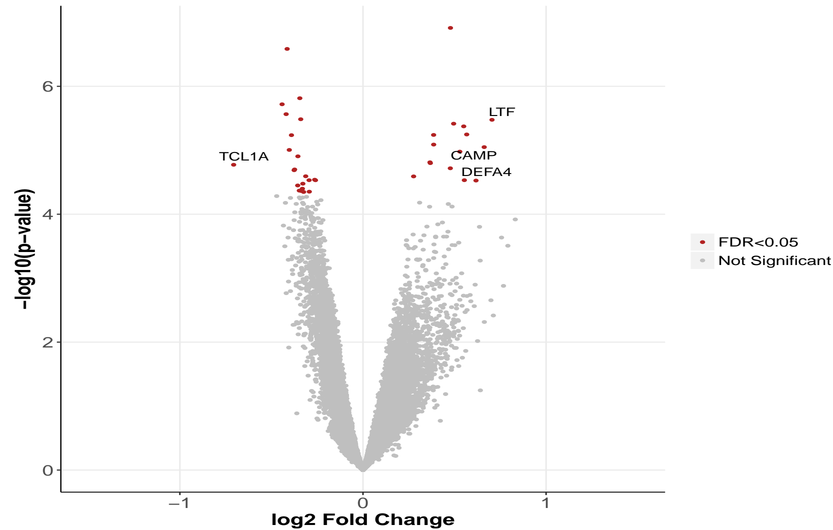
Figure 2.4: **All arrays before (A) and after normalisation (B) in direct-design analysis.** Agilent arrays were normalised using method 'loess'. Red (Cy5) and green (Cy3) are the different channels corresponding to the sample labelling.

Initially, the overall early and late transcriptomic changes in HIV-negative and HIV-positive patients combined were investigated, shown in Figure 2.5. In the early stages of TB treatment, there is a change in the expression of a large number of genes with

3113 statistically significantly expressed. Genes were mainly down-regulated at week two relative to diagnosis, coupled with some up-regulation as well. At week eight compared to week four (late stages), there was very little differential expression, with the expression of only 82 genes significantly changing over this period.



(A)



(B)

Figure 2.5: **Early (A) and late (B) transcriptomic changes in TB treatment in both HIV-negative and HIV-positive patients combined.** Early changes were the up-regulated and down-regulated genes at diagnosis compared to week 2, and late changes were the up- and down-regulated genes at week four relative to week eight. Red are statistically significantly differentially expressed genes after multiple testing correction, grey are not significant. An adjusted p-value of < 0.05 was deemed significant. The most differentially expressed genes are labelled.

Early and late changes in HIV-negative and HIV-positive samples were extracted. In HIV-negative patients, 3297 genes were significantly differentially expressed between diagnosis and week two. In HIV-positive patients, 2581 genes were significantly differentially expressed between diagnosis and week two. Figure 2.6 shows the top significantly differentially expressed genes at diagnosis compared to week two. Both HIV-positive and HIV-negative patients exhibit a similar expression profile in the early phase of TB treatment, with HIV-positive patients having slightly less down-regulation. In the late phase of TB treatment, between week four and week eight, there were no significantly differentially expressed genes. In Figure 2.5B it appears there are significant genes, but once the data are separated by HIV-status, these disappear.

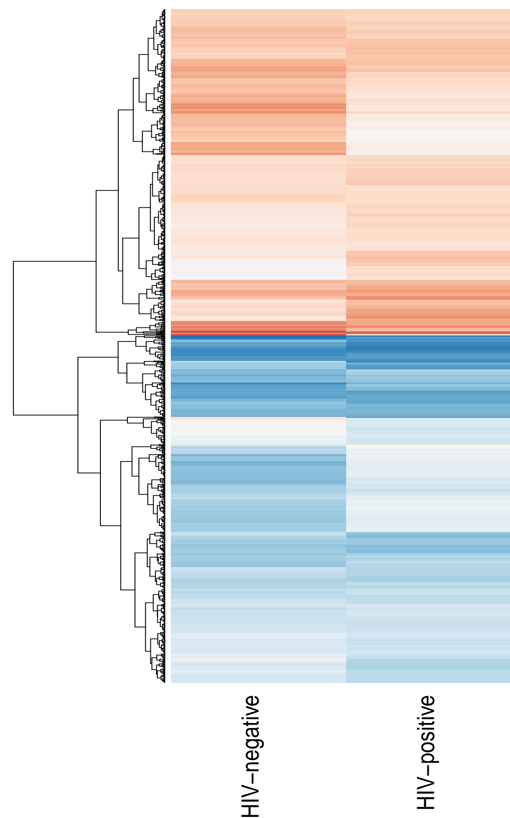
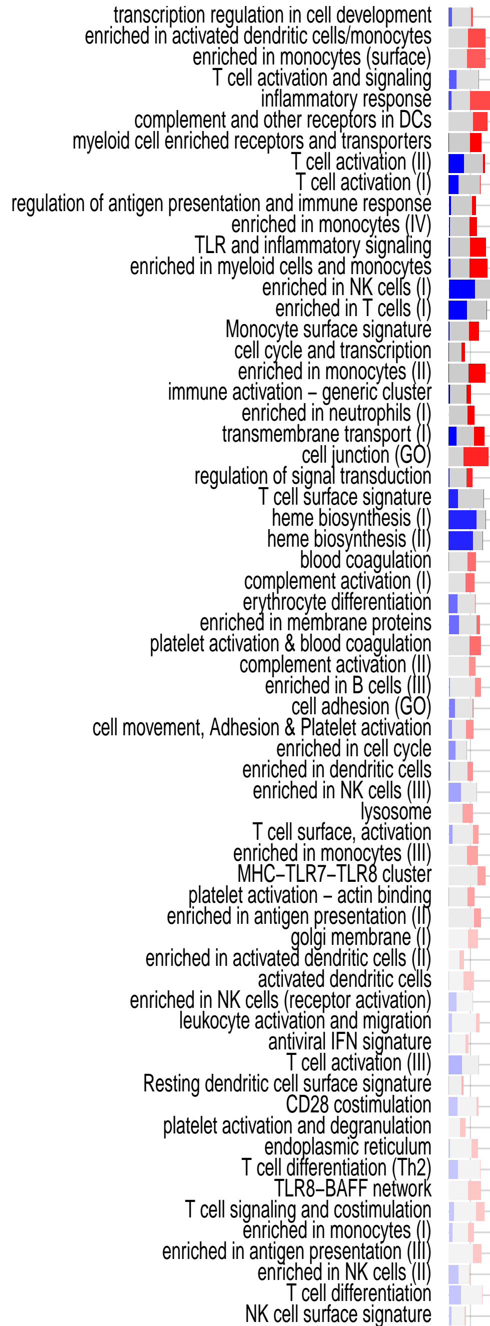


Figure 2.6: **The 1000 most significantly differentially expressed genes in the early stages of treatment in HIV-negative and HIV-positive patients.** Log_2 fold change of genes at diagnosis compared to week two. Red are up-regulated genes at diagnosis compared to week 2 and blue are down-regulated at diagnosis compared to week 2. An adjusted p-value of < 0.05 was deemed significant.

These results show that the largest transcriptomic changes in TB treatment occur at the very beginning; between diagnosis and week two, and there is subsequently relatively little changes happening later on. In addition, it appears HIV-1 coinfection has little effect on these changes. To further this, modular analysis was performed on the early changes in both HIV-negative and HIV-positive combined. So, Figure 2.7 shows the modular expression of HIV-infected and HIV-uninfected patients together at diagnosis compared

to two weeks after the initiation of TB treatment. The modular analysis looked markedly similar to what we would have expected; an up-regulation of inflammation, dendritic cells and monocytes at diagnosis, and a down-regulation of T-cell modules. It shows that the TB biosignature is still valid even when including HIV-positive patients.



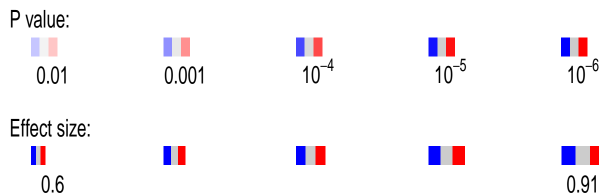
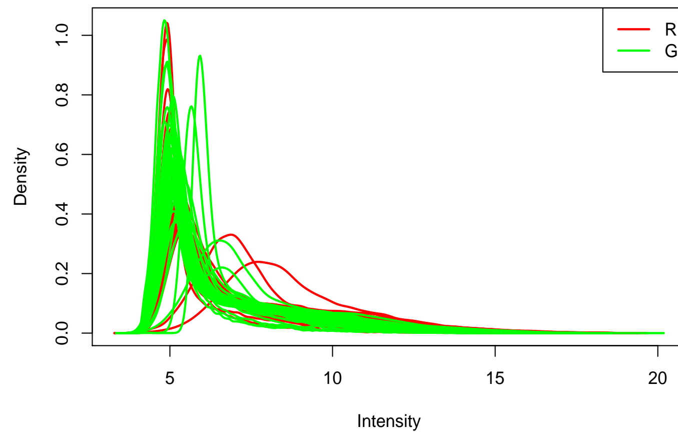


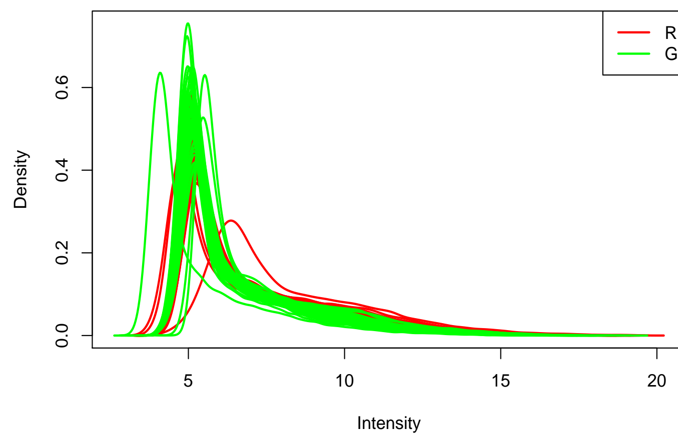
Figure 2.7: **Significantly differentially expressed transcriptional modules at diagnosis compared to two weeks after the initiation of TB treatment in HIV-positive and HIV-negative patients.** Up-regulated (red) and down-regulated (blue). The colour saturation is relative to adjusted p-value, and length of bars is relative to effect size. The amount of colour corresponds to the proportion of genes differentially expressed within that module. Modules were deemed significant with an adjusted p-value of < 0.05.

2.4.3 Microarray separate channel analysis

A second analysis was done on the same data; using the separate channel method, which looked at the individual log intensities of each sample. This allowed the comparison of all four time-points for each patient. Figure 2.8 shows the result of loess normalisation between the arrays. This does not necessarily marry the two channels together, but scales the intensities so they have the same distribution across the arrays.



(A)



(B)

Figure 2.8: **All arrays before (A) and after normalisation (B) in separate channel analysis.** Agilent arrays were normalised using method 'loess'. Red (Cy5) and green (Cy3) are the different channels corresponding to the sample labelling.

The fold changes between each time point were calculated; diagnosis−week two, week two−week four, and week four−week eight in HIV-negative and HIV-positive patients. From this differential expression analysis, 215 genes were found to be significantly dif-

ferentially expressed at all of the time-point comparisons, Figure 2.9. Most of the genes were up-regulated at diagnosis compared to week two, which were then strongly down-regulated compared to the gene expression between week two and week four. The gene expression ended up just over $0 log_2$ fold change at eight weeks after initiation of TB treatment. There does seem to be a slight difference between HIV-negative and HIV-positive at the later time-points; HIV-positive patients were more up-regulated at week 4 compared to week 8 than in HIV-negative patients.

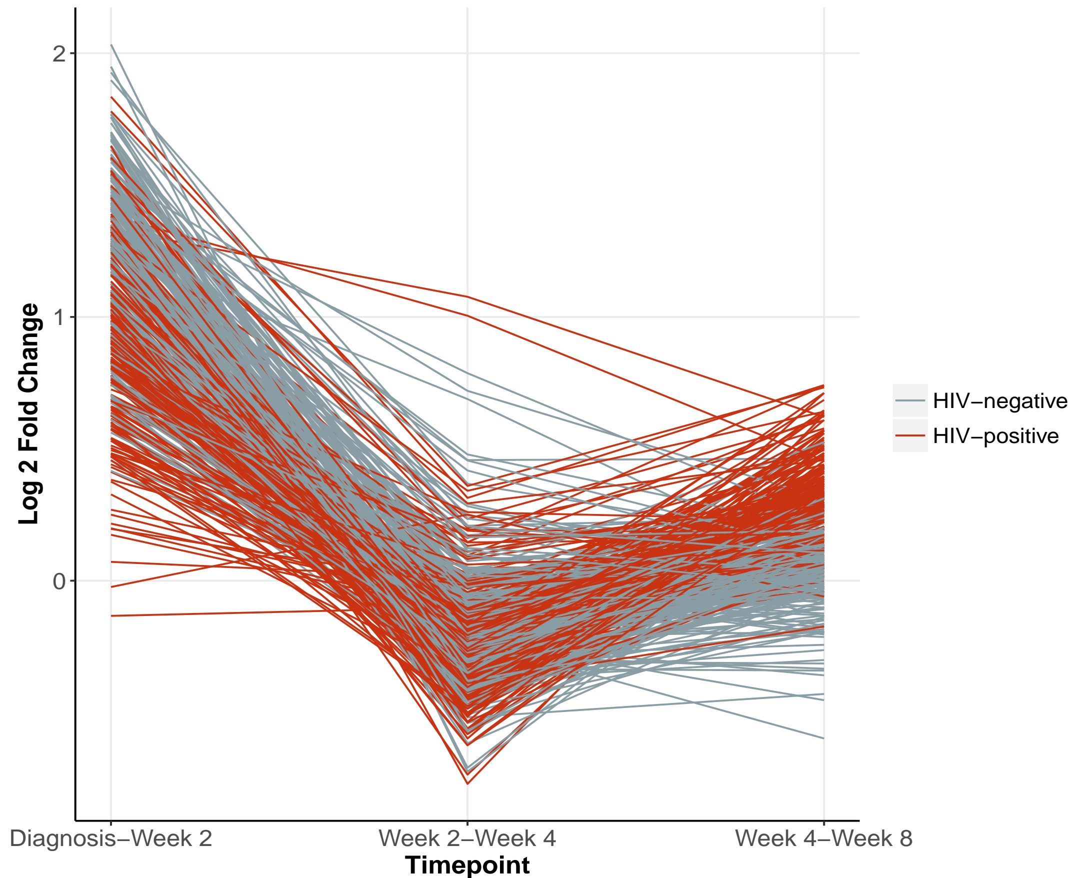


Figure 2.9: **Dynamics of gene expression changes through TB treatment in HIV-negative (grey) and HIV-positive (red) patients.** The differential expression changes of the 215 statistically significant genes between; diagnosis and week two, week two and week four, and between week four and week eight in the separate channel analysis. A p-value of < 0.05 was deemed significant.

A previously published TB signature was analysed at week 8 relative to diagnosis in HIV-negative and HIV-positive patients. Only 14 out of the 16 genes could be analysed (Figure 2.10). In HIV-negative patients, there was mainly down-regulation of the TB signature, which was to be expected. In HIV-positive patients there was a larger down-regulation of these genes. However, none of the genes were statistically significant in either group.

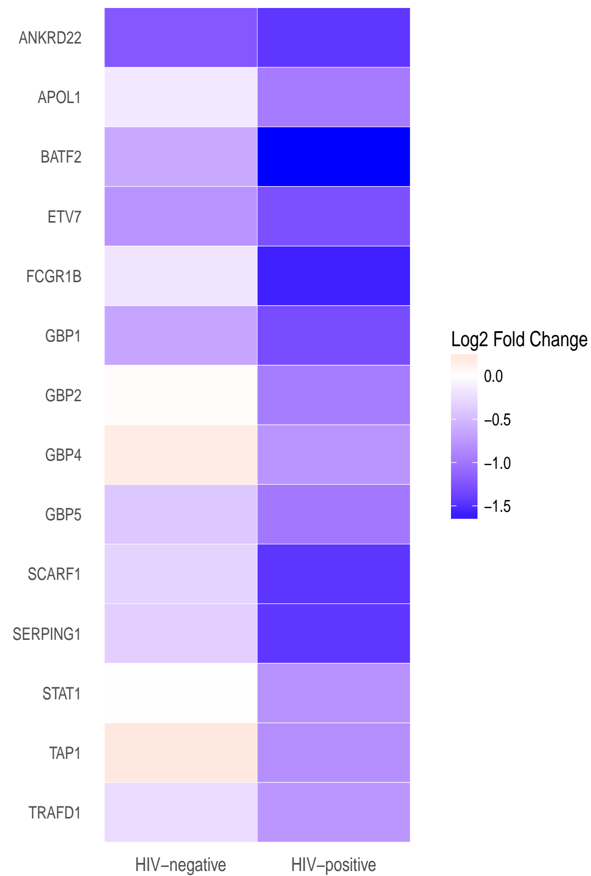


Figure 2.10: **Fold changes of the Zak signature at week 8 compared to diagnosis in HIV-positive and HIV-negative patients.** The differential expression of the 14 genes were extracted (from the separate channel analysis) at week 8 compared to diagnosis. Red is up-regulated, blue is down-regulated.

2.5 Discussion

2.5.1 Findings

Most of the transcriptomic changes in TB treatment occurred at the beginning of TB treatment; within two weeks of treatment initiation. HIV-positive co-infected patients exhibited similar changes. However, there does seem to be a difference in HIV-negative and HIV-positive patients between week four and week eight post initiation of TB treatment, which needs to be further investigated.

The biological interpretation of the microarray data showed that there was an up-regulation of inflammation at TB diagnosis in both HIV-infected and HIV-uninfected TB patients. By week 2, these was an increase in T-cell transcripts. These data show the pattern of magnitude of innate and then adaptive immune responses through treatment.

When analysing a previously published blood TB signature at week 8 compared to diagnosis, one would expect general down-regulation (as before treatment is associated with heightened inflammation), which is seen in both HIV-negative and HIV-positive patients. However, this was more so in HIV-positive patients. It would be interesting to see if this remained the case at the end of treatment.

Overall, the TB transcriptomic signature and the changes during the treatment response could still be observed with HIV coinfection. These data demonstrate that it is possible to derive transcriptomic signatures which will work in a mixture of different disease statuses, which is highly valuable for biomarker and diagnostic development.

So far, not many studies have included HIV co-infected patients which is not ideal considering there are so many HIV/TB cases. The key study that did however, was by

Kaforou et el;[69] they included adult TB patients from different populations and that were HIV-infected and HIV-uninfected. The signature that was developed had high sensitivity and specificity (over 90%) when differentiating TB from LTBI and also TB from OD. When the analysis was split by HIV status, the same signature had reduced specificity in detecting TB from LTBI and OD. The 393- (TB versus LTBI) and 86- (TB versus OD) gene signatures developed by Berry et al[67] were applied to their data set where it performed poorly; the sensitivity and specificity in differentiating TB from OD in HIV-positive patients was below 70%.

A study by Anderson et al investigated the possibility of a diagnostic for TB in HIV-positive and HIV-negative children using transcriptomics. This study also included patients from multiple countries and they developed a disease risk score which was able to distinguish TB from LTBI (sensitivity 94%, specificity 100%) and TB from OD (sensitivity 82.9%, specificity 83.6%).[75] These two studies show that HIV status can be mixed in gene expression studies and it actually produces a more powerful gene signature.

There have been other studies that have investigated the transcriptome between HIV-negative and HIV-positive TB patients using the multiplex ligation probe assay (MLPA). This technique measures the gene expression of approximately 80 genes in many samples at once, the gene probes are typically based on previous literature. Two separate studies, one which included patients from four sub-saharan countries found expression of the Fc-γ Receptor I A1 (FCGR1A) genes were the most consistent classifier for TB.[96,97] Gene expression data are able to distinguish TB regardless of HIV status or geo-genetic differences.

Data do show early changes in gene expression that are typically seen in TB patients

are also clearly seen in HIV-infected individuals. It shows that biomarkers for early treatment response could be developed to use in this population. This is probably because these changes relate to innate, inflammatory immune responses, rather than to adaptive responses.

2.5.2 Limitations of study

Ultimately, there were too few samples in this study, as we had aimed for over 20 samples in each group. Also, we did not have later time-points through the continuation phase until the end of treatment. End of treatment time-points would have been favourable to see whether the gene expression signature had completely resolved in HIV-positive patients as it would have in HIV-negative patients. However, both of these were outside of our control, but should certainly be addressed in future studies.

This study was performed in part as preparation for the subsequent analysis in TB/DM patients, discussed in Chapters 4 and 5. However, a decision was taken early on to perform the latter work using RNA-seq rather than microarray as first intended. Nonetheless, this preliminary study showed that `limma` was a good statistical approach for measurement of TB treatment response via the blood transcriptome and that there is still a place for microarrays.

To conclude, HIV and ART do not prevent transcriptomics from being a useful tool for measuring TB drug-induced cure, especially in the early phase of treatment.

3

Transcriptional effect of Metformin

3.1 Introduction

3.1.1 Type 2 diabetes mellitus (T2D)

It is estimated in global reports there are 382 million people currently living with diabetes. By cause of urbanisation and the ever hiking obesity problem, projections expect this to rise to 592 million by 2035.[36] Type 2 diabetes, or non-insulin dependent diabetes mellitus (NIDDM) accounts for 90% of diabetes cases in adults. It manifests as hyperglycaemia, in turn caused by insulin resistance, and/or impaired insulin secretion. Type 2 diabetes has a complex aetiology, combining genetic and environmental factors.

Glucose is the primary fuel for the brain and is used to produce ATP through aerobic or anaerobic respiration. It is used for a number of processes, including glycogen synthesis for storage or also oxidisation to pyruvate via glycolysis to provide ATP. Plasma glucose concentrations are tightly regulated. A drop in blood glucose levels manifests

as discomfort and mental confusion. Further reductions lead to coma, convulsion and sometimes brain damage and death. Elevated glucose also has serious physiological consequences. High blood glucose levels can cause kidney failure, cardiovascular disease, nerve damage and blindness. Therefore, maintaining normal blood glucose concentration is a high priority.[42]

A variety of mechanisms have evolved to regulate glucose levels. The most important regulators are the hormones insulin and glucagon, both produced by the pancreas in clusters known as the islets of Langerhans. Insulin is produced by the β- cells, and is released into the blood stream in response to changes in blood glucose. Insulin's role in the body is primarily directed at metabolism, with it's most imperative being the use of glucose as energy. Insulin stimulates glucose uptake, where it can be converted to glucose-6-phosphate, and later glycogen. It also stimulates the storage of this excess fuel as fat in adipose tissue.[42]

When there is persistently excessive glucose in the blood stream, is can cause the body to become de-sensitised to insulin leading to a decrease in insulin stimulated glucose transport and metabolism. This is called insulin resistance, and is a precursor state of type 2 diabetes. The body then compensates by releasing even more insulin in order to bring about the appropriate biological effects. Eventually, the β-cells fail, and the lack of insulin transpires to the inability to regulate blood glucose.[42]

Medications are used to regulate blood glucose levels, where the body fails on its own. Typically, type 2 diabetes patients are given oral medication, of which metformin is the most commonly issued worldwide. Other medication includes sulphonylureas, like glibenclamide, which increase the amount of insulin produced by the pancreas. Some-

times injectable insulin is prescribed, but it can cause dangerously low blood sugar levels.

3.1.2 Metformin

Metformin first appeared in herbal guides from the 17th century, derived from the plant *Galega officinalis* (French lilac). The ancient herbal remedy was described as having abilities suitable for treating what we now know as diabetes.[98] The metformin compound itself was synthesised in 1922. Despite this, it was overshadowed by the synthesis of insulin in the same decade. It was not until the 1950s that it received notoriety after demonstrating cardiovascular benefits. Now, it is the most widely administered T2DM medication.[99] It has been recommended as the first line of defence based on the UK Prospective Diabetes Study (UKPDS) in 1998.[100] One of the advantages of metformin is that it reduces blood glucose levels without inducing hypoglycaemia. Together with its mild side effect profile and ease of administration, it has become a favourable treatment for T2DM patients.[101]

The effects of metformin are profuse. Mainly, metformin has glucose reducing properties, but also improves insulin sensitivity and reduces inflammation. It's exact mechanism of action is largely unknown, but some studies have pointed to the role of AMP-activated protein kinase (AMPK), a cellular regulator of glucose and lipid metabolism. A prominent study in rats by Zhou et al demonstrated that metformin's mechanism of action starts with the activation of AMPK. When AMPK was inhibited, metformin was unable to induce its inhibitory effect on glucose production, showing a crucial role in the metformin cascade.[102] This cascade is then thought to result in the modulation of circulating lipids and is also known to increase glucose uptake in the skeletal muscle during

exercise,[103] showing a mechanism in which metformin augments its insulin action.

In a study by Cho et al, metformin was administered to healthy subjects, in which their urine underwent metabolic profiling. They found that there was a clear discrimination between the control and the metformin-treated group. Metabolites like cortisol were reduced after metformin administration. To further their results, the group administered metformin to rats and performed in vitro and in vivo experiments showing that the protein expression of p-AMPK increased after metformin treatment, and also AMPK phosphorylation was induced in the pituitary gland,[104] therefore confirming the AMPK pathway induction by metformin.

Prior to AMPK activation, metformin is thought to inhibit complex 1 of the mitrochondrial respiratory chain. El-Mir et al isolated mitochondria from the liver of rats previously treated with metformin, or untreated. They found that electron transfer through complex I was inhibited in the liver mitochondria treated with metformin compared to the control. The change in energy state could then activate AMPK.[105]

3.1.3 Unconventional uses for an old drug

Metformin has gained popularity in other avenues of research besides diabetes. It has been reported that metformin has anti-cancer, immunoregulatory and anti-ageing effects. Diabetes patients taking metformin live longer, in the form of experiencing fewer cardiovascular events and fewer cases of cancer. Epidemiological studies have peaked the interest in the use of metformin as an adjuvant in cancer therapy, noting that metformin reduces cancer incidence and also mortality. A meta-analysis by Gandini et al established that metformin reduced the incidence of cancer by 31% and reduced the mortality by

34% in diabetes patients, compared to patients on other medication.[106]

The mechanism is thought to be two-fold. Firstly, metformin activates AMPK, which in turn reduces insulin resistance, and secondly due to the anti-inflammatory effects of metformin. High insulin levels are thought to promote tumourigenesis through modulators like insulin growth factors. Insulin resistance is known to affect breast cancer prognosis and increase breast cancer risk. This partly explains the association between obesity and breast cancer. A study by DeCensi and colleagues conducted a trial of metformin versus placebo, and found that cancer proliferation was lower in the metformin treated group.[107]

In other cancers, metformin has been shown to inhibit cancer growth of colon, lung, pancreatic and prostate cancer cells. This was by directly modulating the inflammatory processes that play an important role in tumour progression, blocking NFκB and subduing the pro-inflammatory response.[108–110]

However, in one study in pancreatic cancer patients who were administered metformin showed no improvement in mortality compared to patients without metformin.[111] There seems to be some heterogeneity in the effects of metformin, and more research needs to be done in its beneficial effects in cancer patients, perhaps in the chronic treatment with metformin. Regardless, data are promising.

In addition to the beneficial effects on cancer, the UKPDS illustrated that treatment of patients with T2DM with metformin showed decreased risk of myocardial infarction and subsequent death. The group analysed the effect of metformin and other anti-diabetes drugs in T2DM patients. The metformin treated group had improved cardiovascular benefits; a reduction in risk of CVD by 20%, together with a 42% lower risk of diabetes-

related death. The metformin treated group had better overall health outcomes and metformin had a greater effect than other anti-diabetes drugs, like glibenclamide.[100] There is an ongoing clinical trial in the preventative effect of metformin in CVD, named The Glucose Lowering In Non-diabetic hyperglycaemia Trial (GLINT).

Metformin has recently become a trendy research topic. In the mainstream media it has been reported as an 'anti-ageing drug'. Not that it is able to combat wrinkles, but rather delay the onset of age-related diseases, like cancer, CVD and Alzheimer's disease. There is currently a novel clinical trial, aptly named TAME; Targeting Ageing with Metformin, in the U.S which aims for metformin to become the world's first FDA-approved anti-ageing drug.[108]

The suggestion of metformin having longevity benefits was voiced in a study by Martin-Montalvo et al, in which middle-aged mice were administered metformin. The mice that were treated with metformin lived longer compared to the standard mice, however, at higher concentrations, it was toxic. In addition to this, the metformin treated mice exhibited better physical performance and also reduced inflammation. These in vivo experiments demonstrate that chronic exposure to metformin has beneficial effects, but why metformin manifests these pleiotropic effects is still unclear.[112]

A retrospective observational study by Bannister et al investigated diabetes patients administered metformin compared to non-diabetic controls. This study included 180,000 subjects, and found that diabetes patients that were initiated with metformin monotherapy had longer survival than the non-diabetes subjects. However, the validity of the study design was questioned, by whether the diabetic cohort had selectively retained healthier individuals. During follow-up, if a diabetes patient needed treatment alter-

ations, i.e. intensification of treatment, they were excluded from the cohort. Also, adherent patients were only chosen in the diabetes cohort, but good adherence by itself has been shown to reduce mortality.[113] Regardless, data have shown that there are beneficial effects of metformin in combating diseases other than diabetes.

More research is going on in the beneficial effects of metformin in infectious diseases, in particular in TB. Epidemiological studies have shown that T2DM confers a three fold increased risk of active tuberculosis. Given that the incidence of T2DM in TB endemic areas is set to increase in the upcoming decades, studying the effects of metformin on the immune response to tuberculosis is an integral part in understanding the link between the two diseases.[59]

A case control study in India investigated the protective effect of metformin against active TB in diabetes patients. They found that metformin had a 3.9 fold protective effect against TB. However, it is unclear whether the improved glycaemic control induced by metformin was responsible for the reduced susceptibility to TB, rather than metformin directly inducing protection against TB.[114]

A recent study by Singhal et al demonstrated that metformin may be beneficial in the context of TB. Firstly they showed in vitro that metformin reduced the survival of *M. tb* in treated cells. In vivo experiments then demonstrated the beneficial effects of metformin; mice were infected with *M.tb* and were administered isoniazid with and without metformin. They found the group administered combination therapy had lower bacillary loads in the lung than those that were administered mono-therapy. These same metformin treated mice also exhibited less pathology in the lung tissue and enhanced immune responses once the lung tissue was stimulated with *M. tb*. RNA was also isolated

from the mice lungs infected with *M. tb* that were treated and untreated with metformin, and microarray analysis was performed. Gene-set analysis of these data revealed metformin down-regulated gene-sets involved in inflammation like interferon signalling and antigen presentation, compared to samples from untreated mice. It seems that metformin is able to augment the hosts own immune response in response to *M. tb* that favours the host, which is a potentially new therapeutic approach.[115]

3.1.4 Background of study

Our collaborators Ekta Lachmandas, Reinout van Crevel, and Mihai Netea at Radboud University Nijmegen Medical College have investigated previously the effect of metformin in vitro on PBMCs from healthy volunteers in the context of *M. tb* on the cytokine response. Metformin increased glycolysis, and phagocytosis, whilst reducing the *M.tb* induced production of TNF-α, IFN-γ, and IL-1β. A complementary study was then initiated, in the form of a pilot trial. The study involved the administration of metformin in healthy human donors for six days. Whole blood was taken and collected in PAXgene tubes, in addition to a sample for harvesting peripheral blood mononuclear cells (PBMCs) before and after administration of metformin. PBMCs were stimulated with 5 μg/mL of *M. tb* lysate or unstimulated with RPMI before and after metformin administration. The frozen samples were shipped to the London School of Hygiene & Tropical Medicine for analysis.

Samples included whole blood, and also PBMCs under different stimulation conditions, which underwent transcriptomic analysis using RNA-seq technology. This differential expression analysis of healthy donors administered with metformin allowed an

unbiased approach in investigating the effect of the anti-diabetic drug on gene expression in whole blood and also in PBMCs. This would uncover the effect of metformin in absence of any stimulation on biological pathways like cellular metabolism. Half of the PBMC samples were stimulated with *M. tb* lysate, which considers the potential effect of metformin in the context of *M. tb* infection.

In vitro work has shown metformin increases ROS production, AMP activation and phagocytosis, so one would expect the gene expression data to account for these changes. Metformin administration has also shown anti-inflammatory effects in M.tb infected mice, and therefore one would expect similar changes in human ex vivo blood and PBMC transcriptomes. I hypothesised that metformin would affect the expression of genes related to ROS production and cytokine production in response to mycobacteria in healthy donors.

3.2 Aims and Objectives

Aim

I **To determine the effects of metformin administration on leukocyte function particularly in response to *M. tuberculosis*.**

Objectives

I **Perform RNA-seq analysis of ex vivo samples collected before and after metformin administration in healthy volunteers to determine global transcript changes.**

II **Perform RNA-seq on PBMC samples unstimulated or stimulated with *M. tuberculosis* lysate for 4 hours before and after metformin administration.**

III **To determine biological pathways in the immune response to *M. tuberculosis* which are affected by metformin.**

3.3 Methods

3.3.1 Sample Processing

Eleven healthy Caucasian male subjects were administered daily doses of metformin for six days. The dose rose by 500mg each day until day five, where 2000mg metformin was administered, and then on day six the subjects received a smaller dose of 1000mg. A schematic of the daily dosage is shown in Figure 3.1.

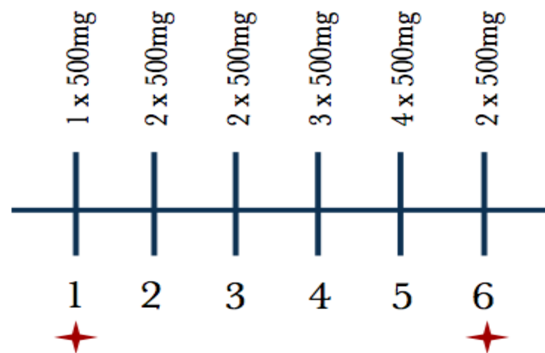


Figure 3.1: **Schematic of dosing of metformin during the pilot trial.** Stars indicate when blood samples were taken at day one, before administration, and at day six, after administration of metformin.

Whole blood (2.5ml) was taken before administration of metformin on day one, and after administration on day six and stored in PAXgene tubes. These were frozen at −80°C and shipped to LSHTM for RNA extraction (n=22). In addition, the peripheral blood mononuclear cells (PBMCs) were harvested (total number of cells per sample; 5 million) from the eleven donors before and after metformin administration. The harvested PBMC were split and subjected to two different stimulation conditions, either unstimulated with RPMI, or stimulated with *M. tb* lysate (n=44).

Of the PBMC samples taken before metformin administration (n=22), 11 were unstimulated with RPMI-1640 (Gibco) and cultured in growth medium, supplemented with 10mM L-glutamine (Life Technologies), and 11 were stimulated with *M. tb* lysate for four hours at 37°C. The 22 samples harvested after metformin administration were also unstimulated (n=11) or stimulated with *M. tb* lysate (n=11) for four hours. The PBMCs were then stored in RNA protect (Qiagen), frozen at −80°C, and shipped to LSHTM for RNA extraction. This was done at Radbound University Nijmegen by collaborators Ekta Lachmandas, Reinout van Crevel, and Mihai Netea.

After receiving the 22 frozen whole blood PAXgene samples and the 44 frozen PBMC samples, I extracted the RNA from whole blood samples using the PAXgene blood miRNA kit (Qiagen) and stored at -80C. RNA from the PBMC samples was extracted using the RNeasy Mini Kit (Qiagen) with the total RNA plus miRNA extraction method. The whole blood RNA samples were globin-reduced using GLOBINclear^{TM} (Invitrogen)

All 66 RNA samples were then processed and sequenced at University College London, Institute of Child Health (UCL). The whole blood samples were processd using the Ribo-Zero TruSeq stranded total RNA library preparation method (Illumina), whereas the cultured PBMC RNA samples were processed using the TruSeq stranded mRNA Library prep kit (Illumina). All samples were seqenced on the NextSeq500, generating ∼36-45 million 43bp paired-end reads per sample.

3.3.2 Data processing

I obtained raw RNA-seq data for each of the 66 samples in the form of FASTQ files from UCL, which I then processed and analysed. FASTQ files were aligned to the human

genome (version Human_g1k_v37) using STAR aligner (Spliced Transcripts Alignment to Reference)[116] (v2.5.2b). The STAR alignment algorithm was used because it is fast, robust, and reliable (Personal communication, Yang). Engstrom et al published an evaluation of the current RNA-seq alignment methods, in which STAR aligner was one of the top performers.[117] Alignment generates SAM files, which contain the the number of reads that overlap with 63,677 features. These reads were counted using HTSeq-count[118] (v0.6.1) using the ensembl annotation file version `Homosapiens.GRCh37.75.gtf` and with the settings described below.

`-f sam` format of the input data

`-r name` how the input data have been sorted

`-s reverse` whether the data is from a strand specific assay

`-m union` mode to handle reads overlapping more than one feature

`-i gene id` what attribute to be used as feature ID

HT-seq count produced a data frame of positive integer numbers for each feature or transcript, for each of the 66 samples. The data frame, or count data was then imported into the statistical programming language R (R v3.3.3 and RStudio v1.0.136) for differential expression analysis using the package `DESeq2`[119] (v 1.12.4). There are a number of R packages available for differential expression analysis of RNA-seq data. These packages include `DESeq2`, `edgeR`[120] and `limma`-voom.[121] `DESeq2` and `edgeR` use the negative binomial as their reference distribution, and provide their own normalisation methods. The negative binomial distribution (also known as the gamma-Poisson distribution) allows for additional variance, or overdispersion characteristic of RNA-seq data. `DESeq2` and `edgeR`

both perform similarly in gene ranking results.[122] `limma`-voom applies a normal-based microarray-like statistical method using weights to RNA-seq analysis. The packages `DESeq2`, `edgeR` and `limma` all use linear modelling as their method for statistical analysis. This method is powerful as it enables multiple variables to be analysed at the same time. In the differential expression analyses, a linear model was constructed for each transcript across all equivalent samples.

3.3.3 Differential expression analysis

The two cell sample types; PBMC and whole blood, were analysed separately. Before statistical analysis, the count data were filtered for transcript rows that were non-informative. Transcript rows that had an overall sum of less than 50 counts across all samples were removed. This can improve statistical power of the downstream differential expression analysis.[123]

In order to deduce the effect of metformin, the expression at day 6 (after metformin administration) was compared with the expression at day 1 (before metformin administration). The linear model constructed for the differential expression of the whole blood samples is summarised in Equation 3.1. This includes both before and after metformin as a variable. Next, Equations 3.2 and 3.3 show how a linear model was constructed for the effect of metformin in cultured PBMCs, and also the effect of metformin in the context of different stimulation conditions.

$$y = (\beta 1_{effect\ of\ metformin}) + \epsilon \tag{3.1}$$

$$y = (\beta 1_{effect\ of\ metformin}) + \epsilon \tag{3.2}$$

$$y = (\beta 1_{effect\ of\ metformin} + \beta 2_{stimulation\ condition}) + \epsilon \tag{3.3}$$

During differential expression analysis, a statistical test has been performed for each transcript. Thousands of transcripts equate to thousands of tests, which all have a 5% chance of being a false positive. If a straightforward approach to significance is taken, i.e. $p < 0.05$ is taken as a cut off. This would lead to the risk of many false positive results, and so correction for multiple testing must be performed. Multiple testing correction was done using the Benjamini & Hochberg false discovery rate method. An adjusted p-value of < 0.05 was determined to be significant.

3.3.4 Gene-set analysis

Gene-set analysis is an analysis method that groups genes into collections according to biological function or processes. It is advantageous as gene level analysis can leave few genes that meet the significance threshold, or, conversely, one can be left with a lengthy gene list without a unifying biological theme. Gene-set analysis eliminates such issues by finding genes with consistent but slight changes whilst introducing biological knowledge. Furthermore, gene-set analysis reduces the multiple testing load as it allows one to test collections of genes in a single step, therefore minimising the number of t-tests that are calculated.

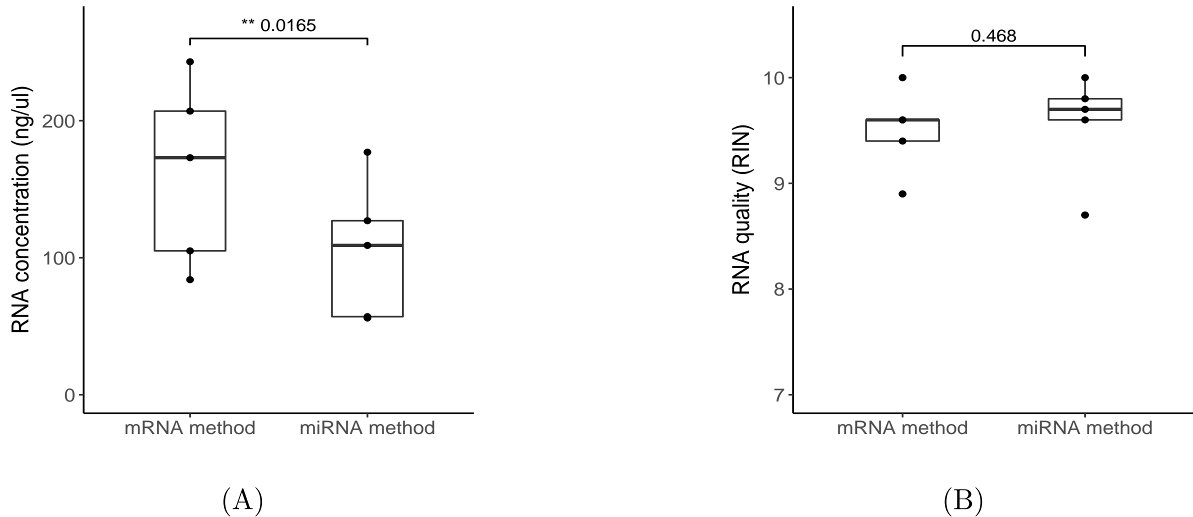
Gene-set analysis was performed using the 'R' package `piano`[124](v1.12.1). The gene-sets used were KEGG pathways (Kyoto Encyclopedia of Genes and Genomes) and Hallmark collections, both downloaded from the MSigDB (http://software.broadinstitute.org /gsea/msigdb). Hallmark sets were generated by computational methodology and summarise well-defined biological states or processes.

To consolidate the gene-set analysis, it was done again with a second package, `gage`[125] (v2.22.0). This package uses the Kolmogorov-Smirnov statistic as in Gene Set Enrichment Analysis (GSEA) from the Broad Institute (http://software.broad- institute.org/gsea/ index.jsp). It is a sensitive method, but there is evidence it has a tendency to produce false positives.[126]

3.4 Results

3.4.1 Assessment of RNA quality

The RNA extraction protocol for cultured PBMCs was optimised. This was because there were two ways of doing it. One extracted total RNA including mRNA (mRNA method), but the other extracted total RNA including mRNA and miRNA (miRNA method). Five blood samples were obtained from LSHTM healthy donors, which then underwent PBMC separation. The same sample was split into two, and underwent the two different extraction methods. Figure 3.2 shows the results of the two methods; there was no significant difference in the quality of the RNA (p-value 0.468) as determined using the BioAnalyser (Agilent), but there was in the concentration of RNA (p-value 0.0165).



(A) (B)

Figure 3.2: **RNA extraction from PBMCs using two different methods; mRNA and miRNA.** Five samples were split into two and underwent the two different extraction methods, the Qiagen mRNA and miRNA RNeasy kits. A: RNA concentration from the Nanodrop using both methods, B: RNA quality, shown as the RNA Integrity Number, from the BioAnalyser using both methods. A paired t-test was performed and a p-value of < 0.05 was deemed significant. total RNA was eluted into 30µl of RNase-free water

The miRNA method was chosen to extract the PBMC samples, as the quality was not compromised, and yield was sufficient, but it would allow us to keep the miRNA. I performed the extractions, and all 66 RNA samples had their quality assessed using the LabChip GX (PerkinElmer) prior to library preparation. Figure 3.3 shows the RIN values and the concentration of each of the whole blood and cultured PBMC samples. All the RNA from the cultured PBMCs was of very high quality. A RIN of above seven is ideal for RNA-seq, but the three whole blood samples that were below the threshold were very close so were processed anyway.

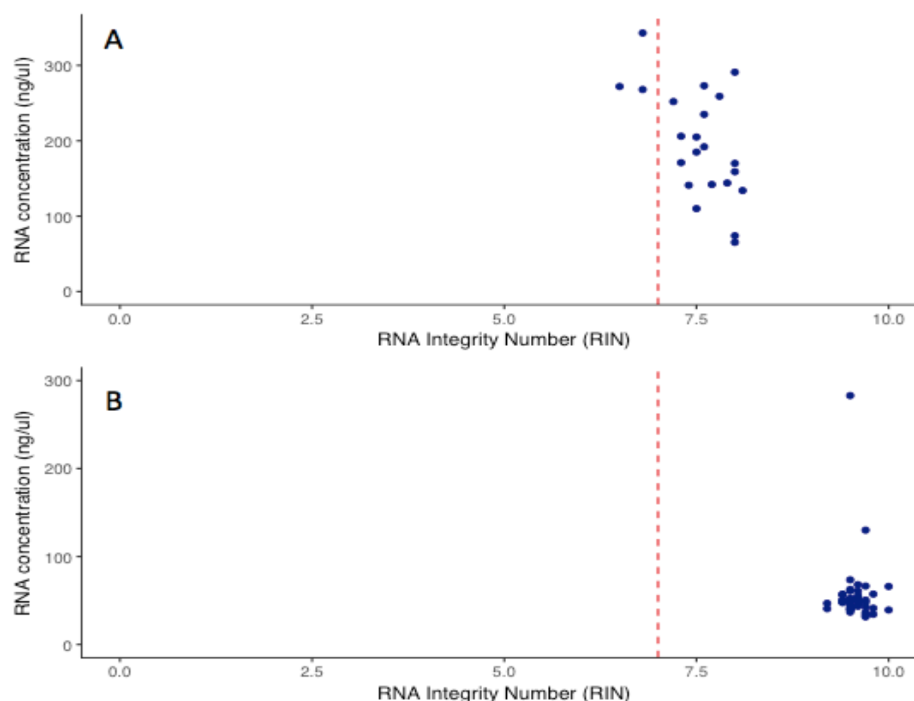


Figure 3.3: **RNA quality and concentration of all 66 samples assessed using the LabChip GX.** A: Whole blood samples, B: PBMC samples. Red dashed line at RIN = 7 indicates the threshold of ideal RIN values for downstream RNA-seq analysis.

3.4.2 The effect of metformin whole blood gene expression

Ex vivo blood was collected in PAXgene tubes from eleven healthy donors before administration of metformin (day 1) and after six days of daily metformin doses (day 6). The gene expression in the whole blood samples at day 6 was compared to day 1 in order to deduce the effect of metformin on transcription in healthy donors. A global approach was initially taken, looking at differential expression changes at the gene level. Initial principal component analysis showed that there was little variation between the two groups, before and after metformin administration (Figure 3.4). Figure 3.5 shows the global effect of metformin in the form of a p-value frequency histogram. Because

the distribution of p-values resembles a uniform distribution, it means that there is no global gene-level effect of metformin administration. As all p-values have equal chance of occurring by chance, after multiple testing correction, any seemingly significant result disappears. On the gene-level, there were was one significantly differentially expressed gene after multiple testing correction, Required For Meiotic Nuclear Division 5 Homolog A (RMND5A) (Figure 3.6). This is potentially because the relevant biological differences were relatively modest.

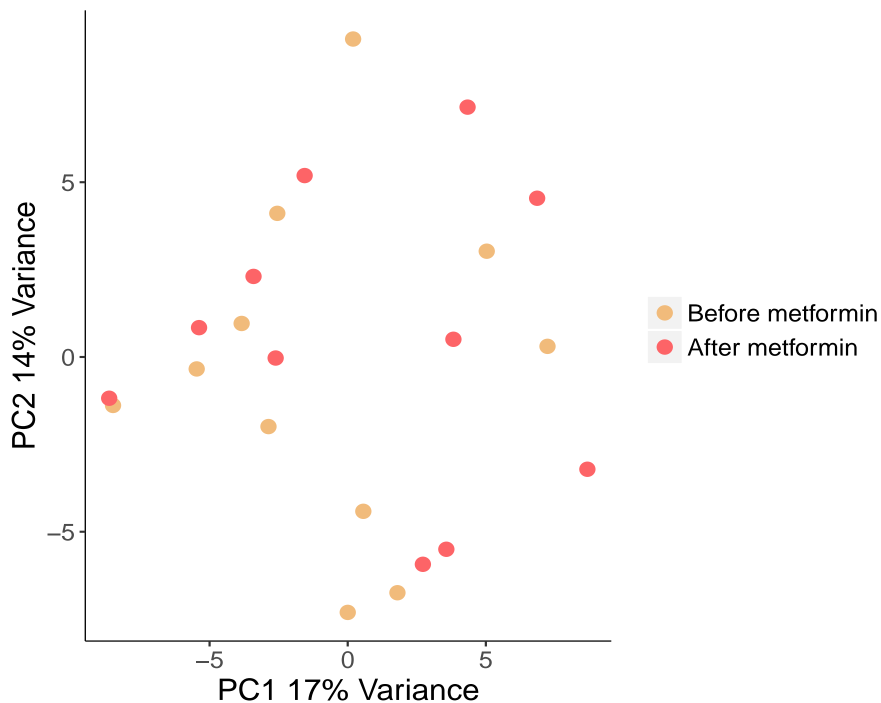


Figure 3.4: **Principal component analysis of ex vivo samples before and after six days of metformin administration.**

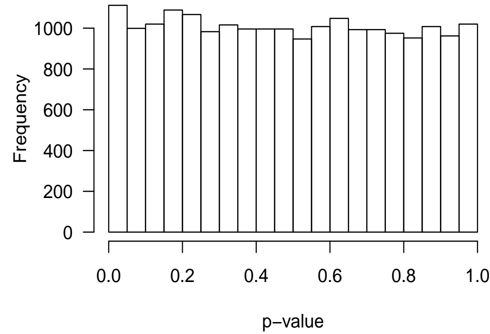


Figure 3.5: **The frequency of p-values generated from differential expression analysis in ex vivo samples before and after six days of metformin administration.** Lists of transcript counts were compared between before and after metformin using `DESeq2`.

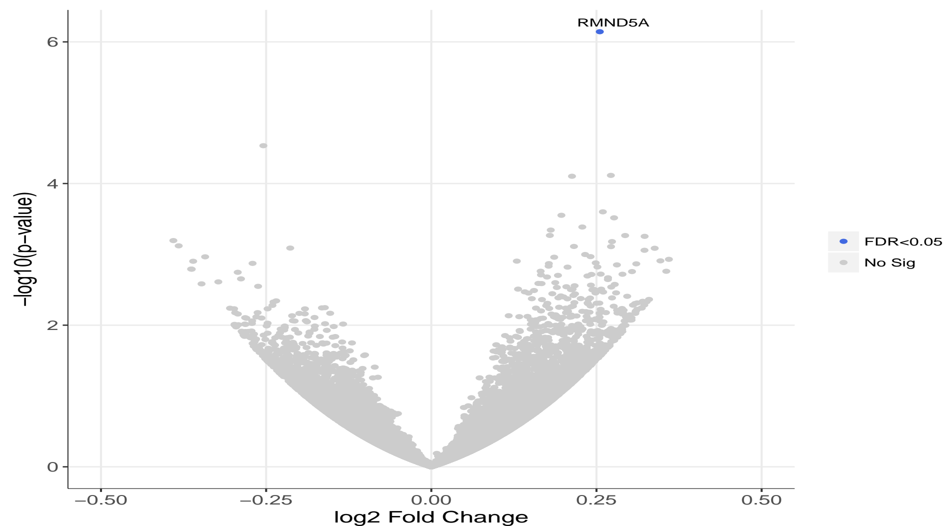


Figure 3.6: **The effect of six days of metformin administration of healthy donors on whole blood gene expression.** Grey dots show which were not statistically significantly differentially expressed, and blue dots show significant differential expression, after multiple testing correction (Adjusted p-value < 0.05).

This gene-level analysis may have missed important effects on pathways. Gene-set analysis interprets gene expression data by focusing on groups of genes that share biological functions. Studying gene-sets allows one to find consistent but small effects in collections of genes, whilst introducing biological knowledge. Gene-set analysis was performed using the R package `piano`. The KEGG pathway database from MSigDB was used for gene-set definitions. Within the whole blood samples, the differential expression of KEGG pathways after six days of metformin treatment was compared with that before administration. This demonstrated the effect of metformin intake on gene expression in whole blood. Analysis revealed several pathways that were significantly differentially expressed (Figure 3.7). The most significantly differentially expressed pathway was ribosome, which was down-regulated in response to metformin treatment. Only one other pathway, oxidative phosphorylation, was down-regulated, while a further remaining ten pathways were up-regulated. The up-regulated pathways included endocytosis, phagocytosis and insulin signalling. These changes were of a small magnitude, but nevertheless, consistent. A pathview image of the oxidative phosphorylation pathway, Figure 3.8 was generated, where the differentially expressed genes can be seen, where complex I was down-regulated in response to metformin.

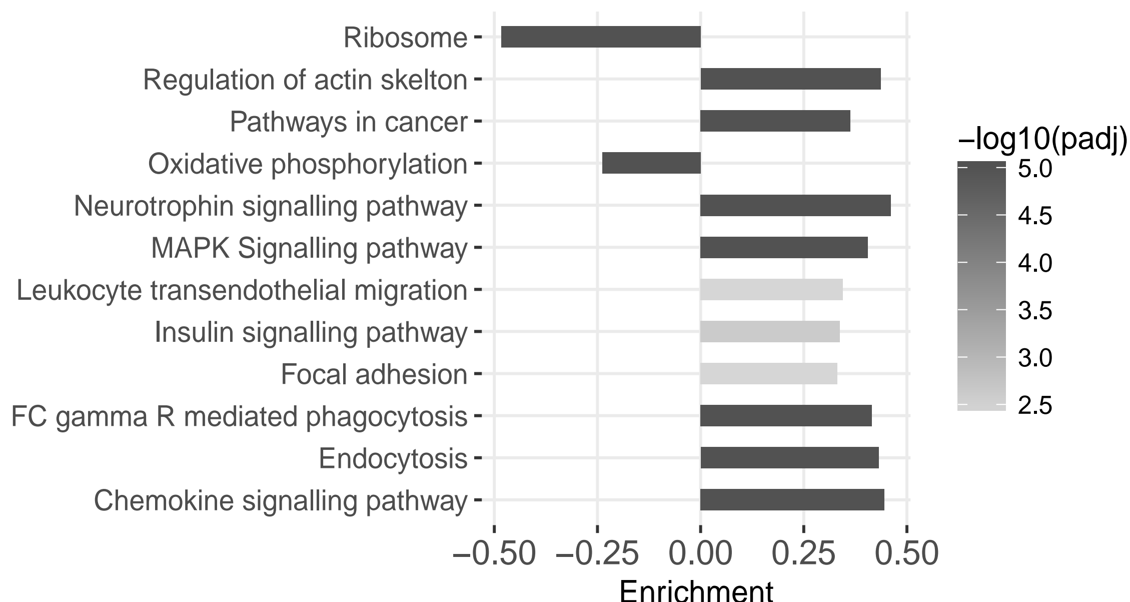


Figure 3.7: **Significantly differentially expressed KEGG pathway gene-sets in whole blood before and after metformin administration.** Data were analysed using R package `piano` and an adjusted p-value < 0.05 was deemed significant.

To extend the KEGG pathway analysis using `piano`, a second collection of gene-sets was investigated, annotated as Hallmark gene-sets. Again, gene-sets were identified which were differentially expressed in whole blood after metformin administration compared to before intake. In agreement with the previous results, of the gene collections that were significantly differentially expressed, the majority were up-regulated in response to metformin, Table 3.1. These consisted of metabolism related collections, like oxidative phosphorylation, and cell division gene-sets like mitotic spindle. In short, in whole blood, metformin had a small but significant effect on gene expression of endocytic-like pathways.

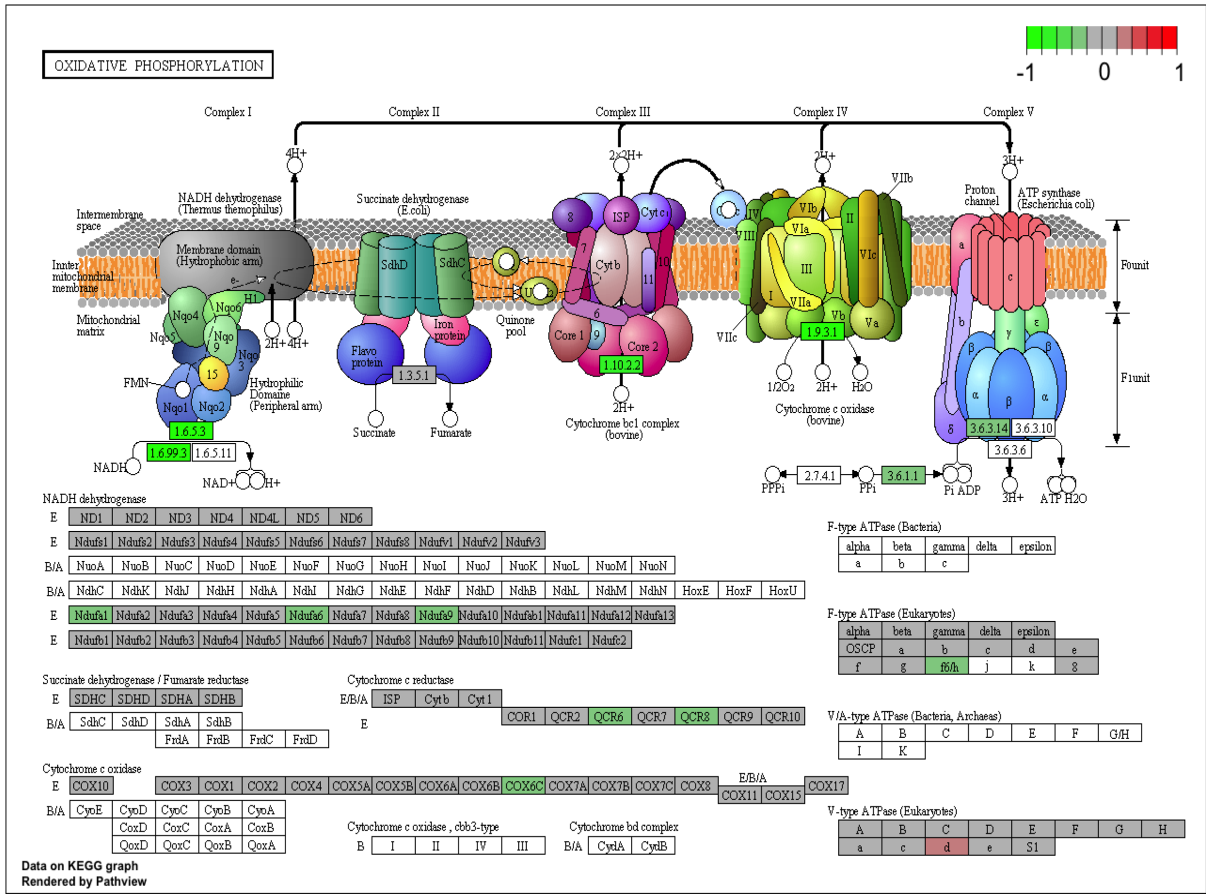


Figure 3.8: **The genes differentially expressed in the oxidative phosphorylation pathway in whole blood due to metformin administration.** Gene-set analysis found oxidative phosphorylation to be a statistically differentially expressed pathway. Green shows down-regulated and red shows up-regulated genes following metformin.

Hallmark gene-set	Adjusted p-value
TNF-α signalling via NF-κB	0
Hypoxia	0
Mitotic spindle	0
G2M Checkpoint	0
Complement	0
Inflammatory response	0
Apoptosis	0.000556
PI3K AKT mTOR signalling	0.000556
Glycolysis	0.00167
mTORC1 signalling	0.003
IL-6 JAK-STAT3 signalling	0.00523
Reactive oxygen species	0.0056
TGF-β signalling	0.00617
Oxidative phosphorylation	0.02

Table 3.1: **Hallmark gene-sets that were significantly differentially expressed in whole blood in response to metformin.** One pathway was down-regulated, seen in blue, and the rest were up-regulated, coloured in red.

To consolidate this, gene-set analysis was done with a second package, `gage`. Once more, the gene expression in whole blood after six days of metformin treatment was compared to before administration (Table 3.2). Similarly with `piano`, `gage` revealed that in whole blood, metformin up-regulates several pathways, the most significant being endocytosis, along with the only significantly down-regulated pathway; ribosome.

KEGG pathway	q-value
Ribosome	$3.15e^{-16}$
Endocytosis	$7.54e^{-10}$
Chemokine signalling pathway	$2.21e^{-08}$
Regulation of actin skeleton	$3.23e^{-08}$
Phagosome	$5.19e^{-08}$
MAPK signalling pathway	$1.86e^{-07}$
Osteoclast differentiation	$1.36e^{-06}$
Neurotrophin signalling pathway	$3.88e^{-06}$
Fc gamma R-mediated phagocytosis	$5.62e^{-06}$
Complement and coagulation cascades	$1.18e^{-05}$
Leukocyte transendothelian migration	$2.10e^{-05}$
Protein processing in endoplasmic reticulum	$1.74e^{-04}$
Cell cycle	$4.86e^{-04}$
Vascular smooth muscle contraction	$6.14e^{-04}$
Carbohydrate digestion and absorption	$6.14e^{-04}$
Insulin signalling pathway	$6.14e^{-04}$
Toll like receptor signalling pathway	$6.14e^{-04}$
Fc epsilon RI signalling pathway	$6.14e^{-04}$
Porphyrin and chlorophyll metabolism	$6.14e^{-04}$

Table 3.2: **KEGG pathways that were the most significantly differentially expressed in whole blood after metformin administration using R package gage.** Blue indicates the pathway was down-regulated in response to metformin, and red denotes up-regulated pathways. q-value is the adjusted p-value, where a value of < 0.05 was deemed significant.

3.4.3 The effect of metformin on stimulated and unstimulated PBMC samples

PBMC samples from the same eleven healthy donors before and after six days of metformin administration were incubated in vitro for 4 hours under two different stimulation conditions; stimulation with *M. tb* lysate, or left unstimulated in culture medium. Firstly, the global difference across all samples was investigated. Further from this, the data were then divided into the four variable groups for individual comparisons.

For the global analysis, the gene expression of all PBMC samples after metformin administration (day 6) was compared with all samples before metformin administration (day 1). This was also with and without a second covariate of *M. tb* lysate stimulation (Equations 3.2 and 3.3). Initial PCA analysis across all the PBMC samples revealed that the largest variance was due to the different stimulation conditions (Figure 3.9). The first model was generated (Equation 3.2) looking at all PBMC samples before and after metformin administration without regard for the different stimulation condition. The second model using Equation 3.3 was generated. This model compared after metformin with before metformin administration. This was done on the single gene-level and also gene-set level.

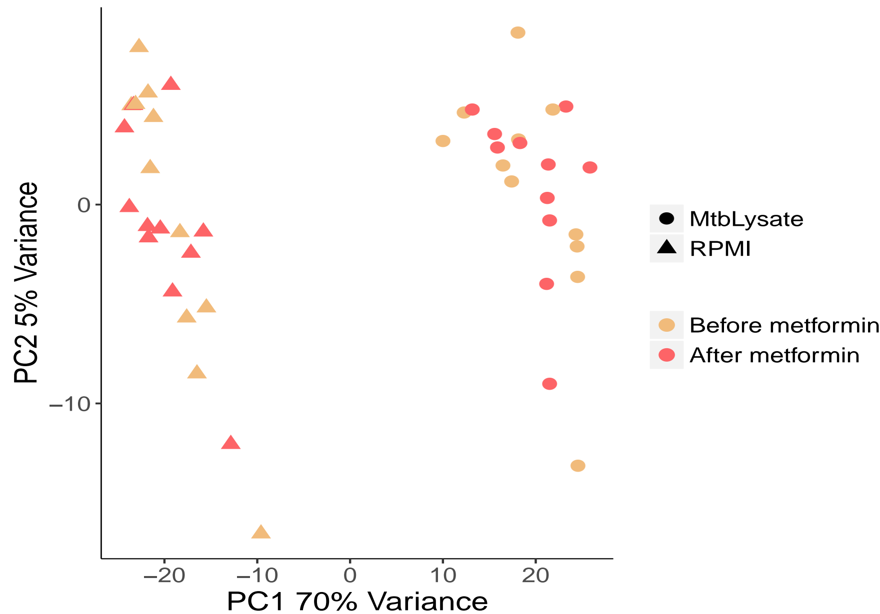
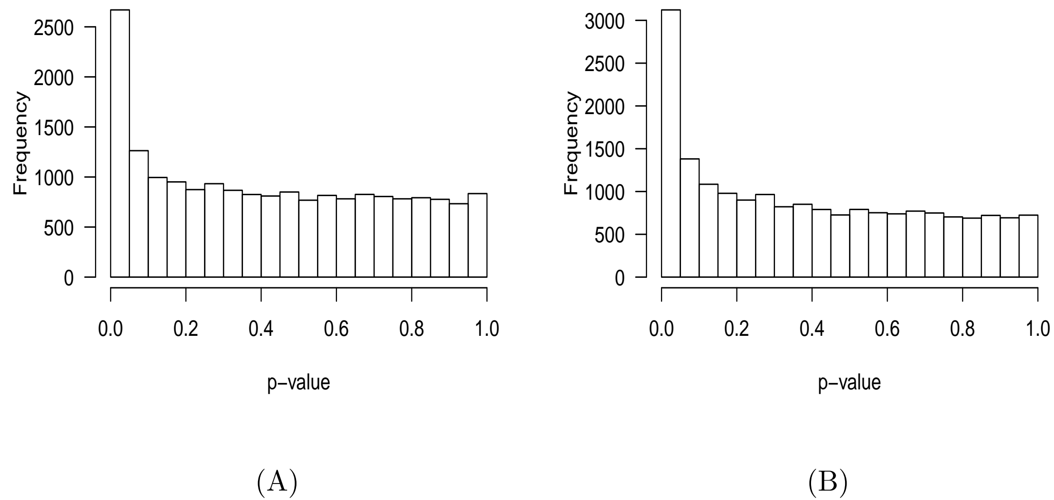


Figure 3.9: **Principal component analysis of cultured PBMCs.** PBMCs were isolated from eleven healthy donors before or after metformin administration and stimulated for four hours with *M. tb* lysate or cultured in growth medium alone. The shape of the point is equivalent to whether the sample was unstimulated with RPMI or stimulated with *M. tb* lysate. The colour represents either before administration of metformin or after six days of administration in healthy subjects.

In Figure 3.10 p-value frequency histograms revealed that in PBMC stimulated samples, there is an effect of metformin on gene expression when comparing all samples at day 6 to day 1, which slightly increases after adding a second covariate controlling for *M. tb* lysate stimulation. Volcano plots summarise the gene expression profiles for the global analysis of the effect of metformin (Figure 3.11). Both models show an effect of metformin, however, the adjusted model shows an increased gene differential expression profile. It can be summarised that metformin has a different effect depending on whether or not the PBMCs have been stimulated with *M. tb* lysate.



(A) (B)

Figure 3.10: **P-value frequencies of the global effect of six days of metformin administration in PBMCs A: before correction, B: after correcting for stimulation condition.**

If there was no difference between the stimulation conditions, the gene expression profile would remain the same after correction. After adjustment for *M. tb* lysate stimulation, 926 genes were statistically differentially expressed due to metformin (adjusted p-value < 0.05). These 926 genes were imported into the online functional annotation tool DAVID, where the gene list's biological themes were identified. The most significantly differentially expressed KEGG pathway using this tool was ribosome.

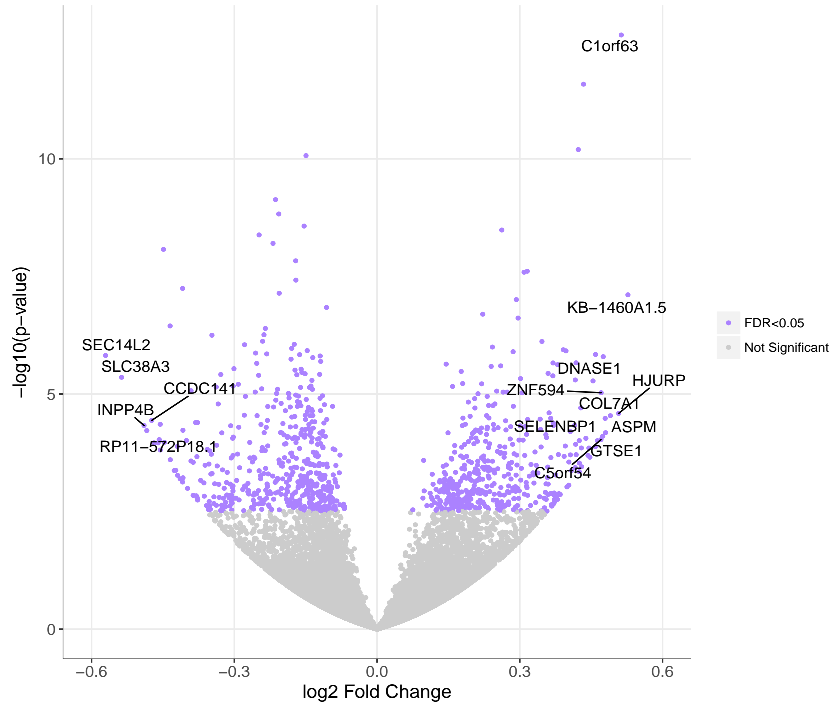


Figure 3.11: **Global effect of six days of metformin administration in healthy donors on gene expression in PBMCs, controlling for stimulation condition.** Grey dots show which genes were not statistically signifcantly differentially expressed, and purple dots show significant differential expression, after multiple testing correction (Adjusted p-value < 0.05).

Gene-set analysis was performed on the adjusted model, similarly to the whole blood samples, using two different R packages. Results from `piano` are shown in Figure 3.12. The columns of the heatmap represent the different methods of ranking the gene-sets, based on whether they were up- or down-regulated. The Hallmark gene-sets that were most statistically significantly down-regulated were interferon-α response, oxidative phosphorylation and interferon-γ response. The most significantly up-regulated gene-sets included mitotic spindle and G2M checkpoint, showing metformin promoted cell division.

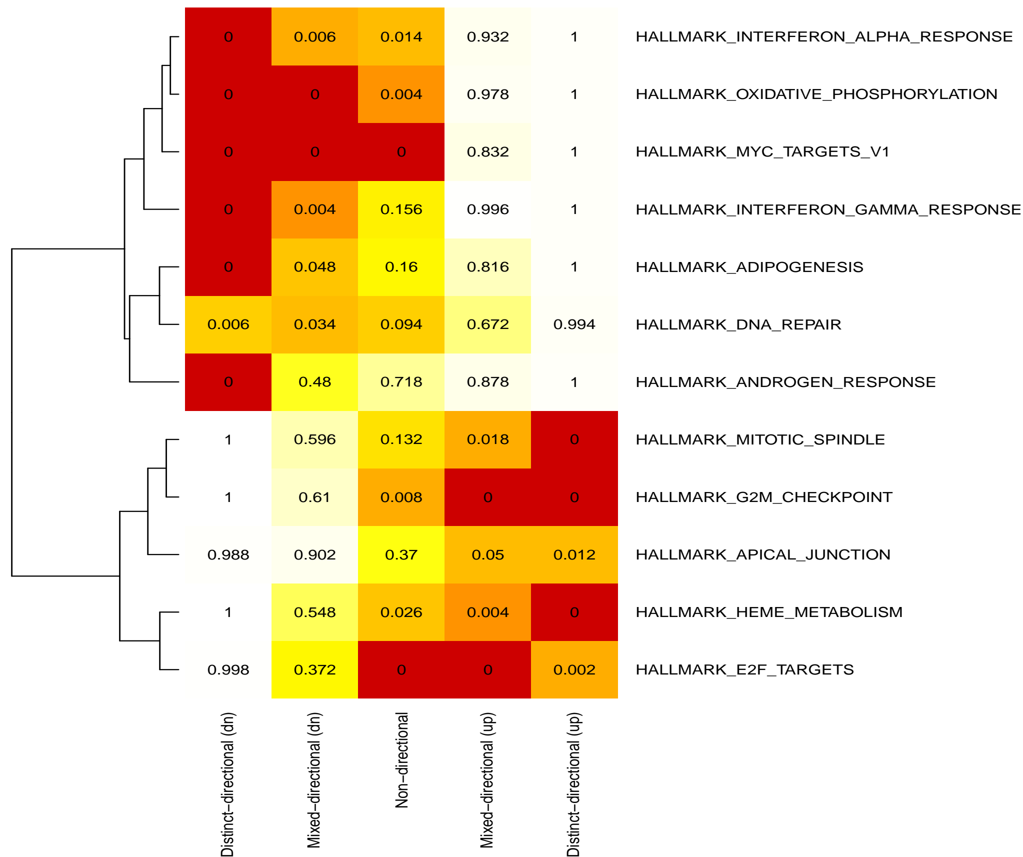


Figure 3.12: **Global effect of metformin in PBMCs from Hallmark gene-set analysis.** Significantly differentially expressed Hallmark gene-sets in three directionality classes; distinct, mixed and non. Non-directional are based on the absolute values of the gene statistics, so are independent of direction. Mixed directional uses subset up- and down- gene statistics. Distinct p-values are calculated with sign information, meaning up- and down- regulated genes cancel each other out. Numbers are the adjusted p-values, and colour saturation is equivalent to significance.

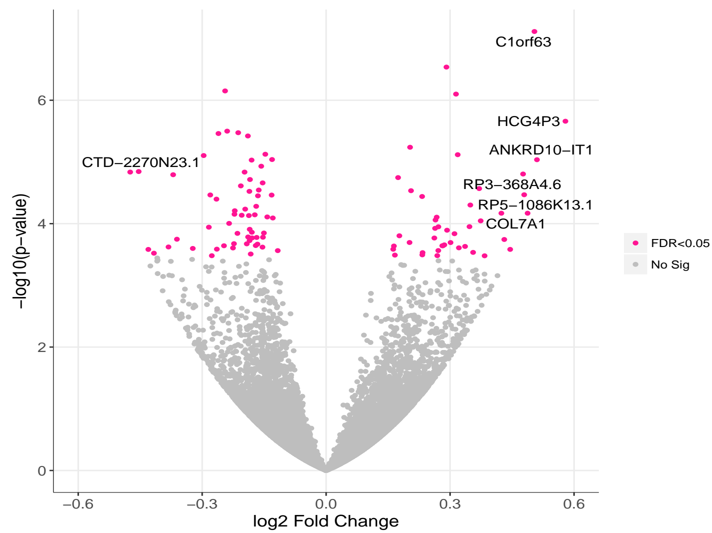
The same analysis was performed with 'R' package `gage`, with the results summarised in Table 3.3. The pathways shown were all down-regulated, as no up-regulated pathways were statistically significant. The most statistically significant pathway was ribosome.

Pathway	p-value	q-value
Ribosome	$9.03e^{-25}$	$1.42e^{-22}$
Oxidative phosphorylation	$1.28e^{-08}$	$1.01e^{-06}$
Splicesome	$2.64e^{-08}$	$1.38e^{-06}$
Proteasome	$8.56e^{-06}$	$3.37e^{-04}$
RNA transport	$7.05e^{-05}$	$2.12e^{-03}$
Ribosome biogenesis in eukaryotes	$8.09e^{-05}$	$2.12e^{-03}$
Cardiac muscle contraction	$1.18e^{-04}$	$2.66e^{-03}$
Intestinal Immune network for IgA production	$1.86e^{-03}$	$3.66e^{-02}$

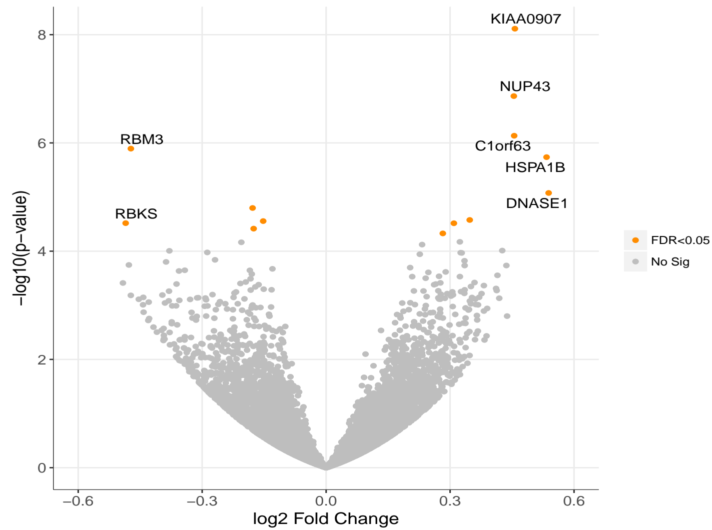
Table 3.3: **Global effect of metformin in PBMCs in a gene-set analysis using package `gage`.** KEGG pathways were all down-regulated. Samples were adjusted for *M. tb* lysate stimulation. p.val was the global p-value result, of which the q.val is the FDR adjustment using the Benjamini & Hochberg procedure. Statistics were performed using the non-parametric Komogorov-Smirnov test, with a q-value of < 0.05 deemed significant.

A second analysis was done to extract specifically what was different between the two stimulation conditions. This consisted of separating the samples by stimulation condition. Firstly, the RPMI unstimulated samples were investigated by comparing the differential expression profile after six days of metformin treatment with before administration. This infers the effect of metformin in PBMCs, in which 103 genes were differentially expressed after metformin than before administration (Figure 3.13A).

Next, the PBMC samples stimulated with *M. tb* lysate after metformin treatment were compared with the cells that were *M. tb* lysate-stimulated before metformin treatment. This reveals the effect of metformin in the context of *M. tb*, in which 13 genes were differentially expressed after metformin compared to before administration (Figure 3.13B).



(A)



(B)

Figure 3.13: **Global gene expression before and after six days of metformin administration in PBMC samples A: unstimulated with RPMI and B: stimulated with *M. tb lysate* for 4 hours.** Coloured points are statistically significantly differentially expressed genes after multiple testing correction, and grey is not significant. An adjusted p-value of < 0.05 was deemed significant.

As above, gene-set analysis was performed using `piano`, investigating the Hallmark collections, to identify the small but consistent effects in changes in a biological context. The effect of metformin in un-stimulated PBMCs was up-regulation of sets involved in cell division, like mitotic spindle and G2M checkpoint. The gene collections that were differentially down-regulated included oxidative phosphorylation. This is shown in a network plot in Figure 3.14.

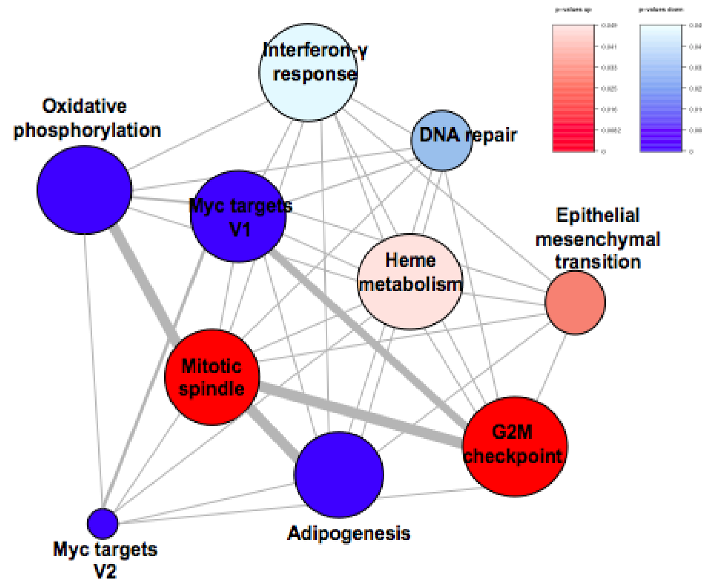


Figure 3.14: **Gene-set analysis of PBMC samples unstimulated with RPMI.** Red are up-regulated gene-sets, blue are down-regulated gene-sets. The saturation of colour is equivalent to the adjusted p-value. Stouffer method, adjusted p-value of < 0.05 was deemed significant.

The affects of metformin in PBMCs stimulated with *M. tb* lysate were down-regulation of gene collections involved in inflammation, shown in Figure 3.15. In particular, interferon-γ response genes, and TNF-α signalling via NFκB genes. This is striking as it is known IFN-γ is a critical component of the immune response to *M. tb*. It appears that in the context of *M. tb* lysate, metformin had an anti-inflammatory effect in PBMCs.

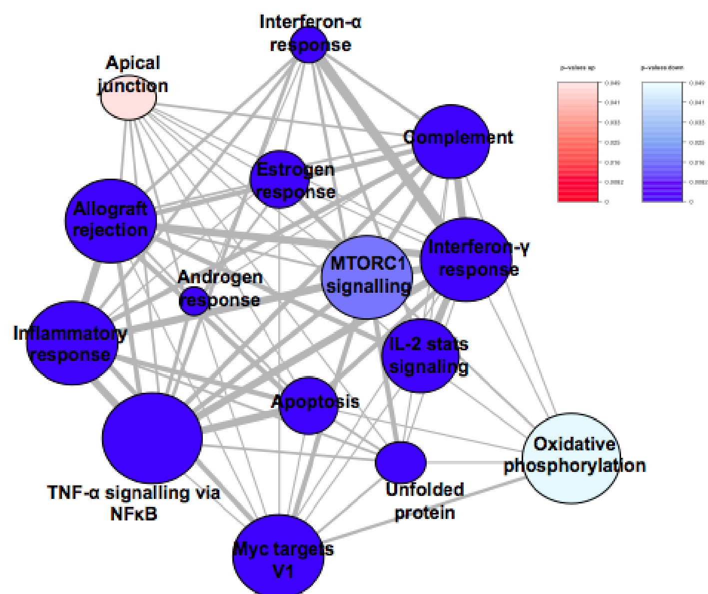


Figure 3.15: **Gene-set analysis of PBMC samples stimulated with *M. tb* lysate.** Red are up-regulated gene-sets, blue are down-regulated gene-sets. The saturation of colour is equivalent to the adjusted p-value. Stouffer method, adjusted p-value of < 0.05 was deemed significant.

Each gene-set has up to 200 genes within it. In order to see which genes were changing the most in response to metformin, genes with a 2-fold different expression were extracted. Within the IFN-γ response gene-set, 10 genes were 2-fold differentially expressed due to metformin, shown in Figure 3.16. The most differentially expressed genes were Interferon Induced Protein With Tetratricopeptide Repeats 1 and 3 (IFIT1, IFIT3), MX Dynamin Like GTPase 1 (MX1) and Radical S-Adenosyl Methionine Domain Containing 1 (RSAD1). This complements the findings using the Hallmark collections. Metformin appears to have an anti-inflammatory effect in *M. tb* lysate stimulated PBMCs, in particular on interferon response genes.

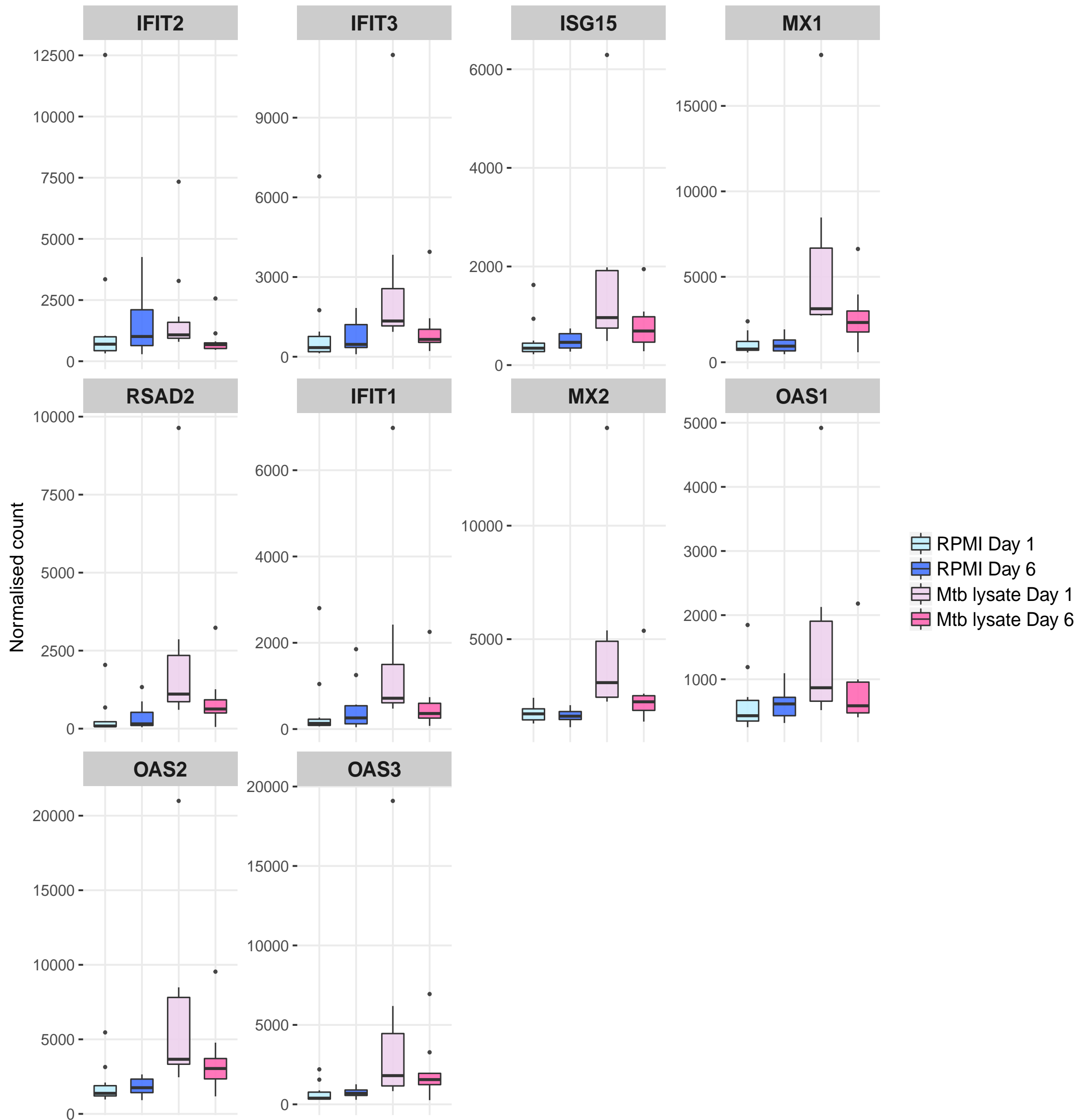


Figure 3.16: **Genes within the interferon-γ response gene-set that were $\geq$ 2 fold change differentially expressed before and after six days of metformin administration in either stimulation condition of PBMCs.** PBMCs taken from healthy donors were unstimulated in RPMI (blue) or stimulated with *M. tb* lysate (pink) before (day 1) and after six days (day 6) of metformin administration.

To confirm these gene-set findings, gene-set analysis with gene ontology terms was done. The differential expression of gene ontology terms in *M. tb* lysate stimulated PBMCs after metformin intake were compared to before intake. GO terms such as innate immune response, NFκB signalling and response to type I interferon were down-regulated in response to metformin. Type I interferons have been seen to be associated with disease progression in *M. tb* infected individuals, so the ontology term was further investigated. On the gene-level, the genes in the type I interferon set would not reach significance. But, the adjusted p-values of the genes within this GO set are significantly lower in comparison to the adjusted p-values of all the genes together (Figure 3.17).

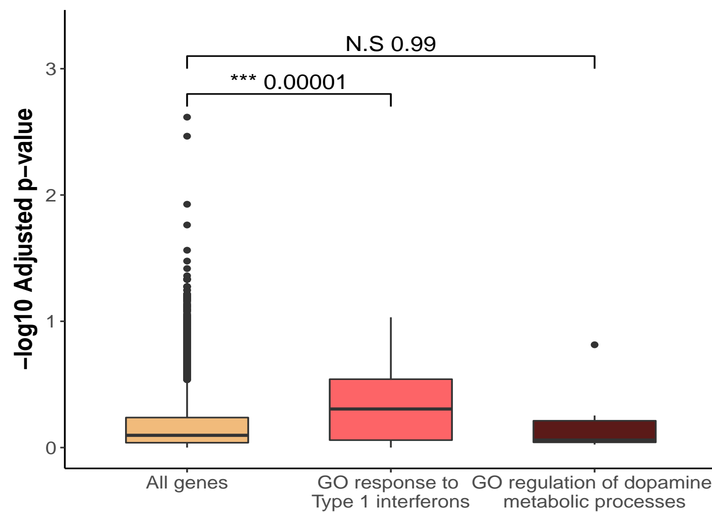


Figure 3.17: **Gene ontology analysis of the *M. tb* lysate stimulated PBMC samples.** Adjusted p-values of all the genes, the genes within the type I interferon response GO-set, and the genes within the regulation of dopamine metabolic processes, as an irrelevant control gene-set. Statistics were performed using the R package `piano`, with an adjusted p-value of < 0.05 deemed significant.

3.5 Discussion

3.5.1 Findings

In this study, a modest effect of metformin on the transcriptome in whole blood and in PBMC samples has been demonstrated. The effect of metformin seems to be primarily on cellular metabolism. In the context of *M. tb* however, metformin appears to have an anti-inflammatory effect, in the form of a reduction in the interferon response genes.

In ex vivo samples, metformin only has a significant effect on the gene-set level, not the gene-level. Such sets, in the form of pathways, were mainly up-regulated, like endocytosis and phagocytosis. If cells have an enhanced propensity for phagocytosis due to metformin, it can be envisaged that alveolar macrophages would more readily take up and kill the *M. tb* bacilli. Future work could include investigating the phagocytic capacity of antigen presenting cells, like macrophages and dendritic cells post metformin treatment to see whether metformin does improve their phagocytic ability.

In whole blood, oxidative phosphorylation was down-regulated, which fits with current literature. Studies have previously shown that complex 1 in the respiratory transport chain is directly inhibited by metformin in mice and human hepatocytes. This inhibition elevates the levels of AMP, which is then a potential cause of AMPK activation by metformin.[127] AMPK activation has been shown to enhance neutrophil chemotaxis and facilitate bacterial eradication, indicating a mechanism by which metformin exerts its beneficial immunomodulatory effects.[128] In addition, a reduction in oxidative phosphorylation means a switch towards glycolysis, also known as the Warburg effect observed in tumour cells,[129] which rapidly provides ATP as energy for phagocytosis, oxidative burst

and cytokine production in macrophages.[130]

It is worth noting that the ex vivo samples were whole blood, meaning the majority of the transcripts were likely derived from neutrophils. It is possible that the overall transcriptome signal was somewhat swamped by the overwhelming neutrophil signal, which explain why only modest effects could be detected.

In the PBMC samples, the effect of metformin was more pronounced. In the unstimulated samples as well, metformin was found to down-regulate oxidative phosphorylation. There was also with a down-regulation of interferon-γ response genes. This anti-inflammatory effect was a lot more evident in response to *M. tb* lysate stimulation, where interferon-γ response, interferon-α response, TNF-α signalling via NF-κB and inflammatory response were all strongly down-regulated. Of course, there was a large overlap in all these gene-sets. Genes that they did all have in common were the IFIT genes; IFIT1, IFIT2, IFIT3 and MX1. These genes play roles in regulating inflammatory cytokine mRNA stability, cell proliferation and apoptosis. They are mainly induced by type I interferons. The results show a novel mechanism through which metformin mediates its protective effect.

3.5.2 Interferons and metformin

The IFN signature has proven important in the response to TB, the magnitude of which correlates with disease severity.[67] A recent decisive study by Zak and colleagues found a correlate of risk transcriptomic signature for the progression to active TB disease. This signature was able to predict which latently infected individuals would progress to active disease 180 days before disease diagnosis. This correlate of risk signature comprised

of interferon genes, which was the only module to be over-represented in the signature and appeared relatively rapidly in the months leading up to diagnosis. Because of metformin's anti-inflammatory effects targeted at interferon, this could mean that it may reduce the chance of active disease progression.[76] On top of this, a study by Gideon et al in Macaques demonstrated that the degree of lung inflammation correlates with the blood transcriptional changes, which again included interferon genes. Interestingly, the interferon response modules were significantly higher in animals that would develop active disease and remained higher at later time points than animals that developed latent infection. It does appear that the interferon response modules play an important role in the progression to active disease, pointing to the option for host-directed therapy to modulate the hosts immune response.[131]

Another study by Mayer-Barber et al investigated the prospect of a host-directed therapeutic agent by targeting the type I IFN response in a mouse model. Zileuton, an arachidonic acid metabolism modulator was administered to IL1 knockout mice together with a control which were then infected with *M. tb*. They found that zileuton significantly reduced bacillary loads and the IFN-associated pathology and enhanced survival. It shows that modulating the hosts immune response with another type of host-directed therapy is beneficial in *M. tb* infection outcomes, by again targeting the type I IFNs.[132]

The unconventional use of metformin is still a relatively new research topic, but it is gaining track as data are identifying beneficial effects of metformin. Often however, the beneficial effects of metformin seem to be in chronic treatment rather than short-term administration. Regardless, it is indeed becoming an option for host-directed therapy, especially for patients with TB and diabetes, as it may improve clinical outcomes. This

was shown in the clinical arm of the Singhal study, where TB/DM patients administered metformin had better treatment outcomes, in the form of increased survival than TB/DM patients not on metformin.[115] Therefore, this study reveals a potential mechanism for the protective effect of metformin via suppression of an inappropriate IFN response. In more detail, metformin could be able to dull the pathological inflammatory response associated with TB lung lesions and disease severity, and also the progression into active disease. Metformin's potential is exciting because the WHO is very keen to drive immunomodulatory agents that could boost the efficacy of antibiotic regimens especially in the era of increasing concern of antimicrobial resistance. There are no known drug-drug interactions between anti-TB drugs and metformin, but more studies need to be formally performed that explore the relationship between them in order to advance the potential adjuvant therapy.

It is a possibility that metformin could be given as an adjunctive treatment, along with regular anti-TB therapy to TB patients. Alternatively, it could be given with prophylactic treatment to latently infected *M. tb* patients to prevent progression. However, the study by Singhal et al that presented this hypothesis has limitations. In their mouse work, where metformin was shown to improve survival in conjunction with antibiotic therapy, the difference between metformin treated and untreated animals, although significant, was of a very low magnitude. The sample numbers of this experiment were very few, so it would be interesting if this still stands with increased sample numbers. In addition to this, the gene-set analysis tool used in the transcriptional data for this study, `PAGE`, is notorious for false positives. When Luo et al compared `PAGE` with `GAGE`, they reported that `PAGE` made a large number of false positive calls.[115,125] So, whether metformin is

the new wonder drug should be taken with a pinch of salt until more research is done to validate the effect of metformin as a possible adjuvant.

3.5.3 Limitations and future work

Other diabetes drugs, like glibenclamide, also have been shown to possess anti-inflammatory effects in response to bacterial infection. Glibenclamide strongly reduced pro-inflammatory cytokine production in response to *Burkholderia pseudomallei*. Diabetes patients are more susceptible to the bacterial infection, but treatment with glibenclamide reduces the mortality of melioidosis by reducing the hyper-inflammation occurring in sepsis.[133,134] It shows that other anti-diabetes drugs besides metformin have promise outside their intended use, and should also be investigated for their beneficial effects in response for example to *M. tb* infection.

What would have been ideal, would be to have performed the PBMC stimulations with live *M. tb* than with *M. tb* lysate. Evidence has shown that more genes are significantly differentially expressed when stimulated with live *M. tb* compared to stimulation with irradiated *M. tb*.[135] Also, these current data are only for 4 hours of stimulation, and it would be valuable to investigate over further time-points. In this study, PBMCs stimulated with *M. tb* lysate for 24 hours were collected and qRT-PCR was performed on the IFN-response genes. The RNA-seq results for the 4 hour *M. tb* lysate stimulation were validated in that the IFN-response genes were down-regulated after metformin administration. At 24 hours however, the effect had dwindled, and a down-regulation of a much smaller magnitude was witnessed in response to metformin. However, this was only on a selection of genes, so it would be of value to perform genome-wide analysis on

later time-points. In addition to this, in would be helpful to have stimulated the whole blood samples with *M. tb* in order to see the difference in the immune response before and after metformin administration.

A further limitation is in the classification for the gene-sets. The Hallmark gene-sets were used, in which there are sometimes 200 genes within each set, and there are large overlaps between the sets. Because of this it is questionable as to whether the enrichment of the gene-set is truly reflective of its association. There is also variability in the gene-set analysis tools. A comparative study found that the R package `gage` was one of the worst performers, demonstrating poor specificity with a high false positive rate,[126] albeit, not as bad as `PAGE` which is why a second tool; `piano` was used. These kinds of tools inherently rely on pre-existing pathways and associations and therefore limits the scope of the analysis.

To conclude, these results do support a role of metformin in TB generally, and TB/DM specifically. Research in metformin is blossoming and it is becoming clear its beneficial properties are not just restricted to diabetes. It provides exciting avenues for research, and the positive association in TB is promising.

4

Tuberculosis and diabetes in TANDEM: cross-sectional study

4.1 Introduction

4.1.1 Tuberculosis and Diabetes: the collision of two epidemics

A link between type 2 diabetes (T2DM) and tuberculosis (TB) has been known for centuries, initially when the Persian philosopher, Avicenna described the symptoms of T2DM as including TB.[32] During the 20th century, TB was a leading cause of death among diabetic patients. Of course, the synthesis of insulin in the 1920s, and then the discovery of streptomycin and other antibiotics in the 1940s led to the comorbidity to somewhat dwindle. But, now T2DM has become a global epidemic in itself, as a consequence of rising obesity, lifestyle changes, and an ageing population. Coupled with TB, which has recently been classed the number one cause of death worldwide by an infection, they now represent a threat to global health.[6]

The TB/DM problem is exacerbated by the development of T2DM in countries where TB is also endemic, like Indonesia and India. It has been estimated that 70% of diabetics live in TB endemic countries. Indonesia has the third highest burden of TB globally, and the fourth highest burden for diabetes.[36] Countries like Romania and Peru also have a high prevalence of both diabetes and TB,[136] causing cause for concern. In a prospective cohort study, the prevalence of diabetes among newly diagnosed TB patients in Indonesia is 14.8% compared with 3.2% of the general population,[137] which is projected to rise. The increase in the prevalence of both diseases will potentially affect global TB control (Figure 4.1)

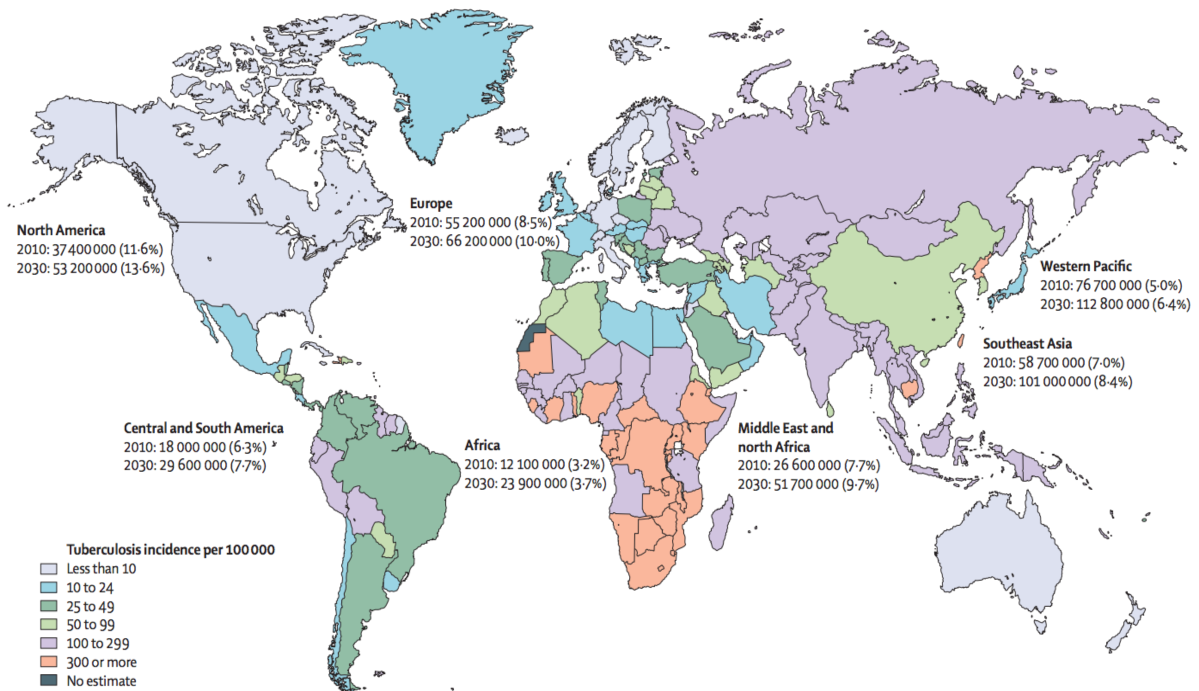


Figure 4.1: **Projected prevalent diabetes cases and current worldwide tuberculosis incidence.** Estimated number and percent of individuals with diabetes mellitus in 2010 compared with 2030 projections are shown. Tuberculosis incidence per 100 000 population data for 2007 are shown. Data from International Diabetes Foundation and WHO.

Along with HIV-1, smoking and malnutrition, T2DM has also been recorded as a risk factor for active TB disease. Even though the risk at the individual level is considerably lower than, for example, AIDS, T2DM presents a substantial effect at the population level due to the number of diabetic people.

To assess the likelihood of diabetic patients progressing to active TB disease, a systematic review of 13 observational studies and meta-analysis was performed. There was large heterogeneity between studies, potentially due to the population differences in diabetes. Overall though, diabetes does increase the risk of TB; it estimated that T2DM patients are 3.1 times more likely to develop active TB disease than non-diabetic individuals.[59] Considering an estimated quarter of the world's population is latently infected with *M. tb*, and the number of T2DM cases are projected to be 552 million by 2030,[36] it represents a huge reservoir of potential active TB disease. However, the causative link between T2DM and TB remains largely elusive.

4.1.2 Clinical presentation of TB/DM

Evidence for diabetes patients to exhibit different clinical presentation of TB is inconclusive. Alisjahbana et al reported in a prospective cohort study that diabetes patients were associated with more TB symptoms than those without diabetes. But, there was not an increase in the severity of TB disease.[137] Largely however, compared to TB patients without diabetes, diabetes patients are more likely to be sputum smear positive at diagnosis and also delayed in culture conversion.[138,139] Regarding diagnosis, screening of TB patients has shown promise in diagnosing diabetes patients by using RPG as an initial screen and then HbA1c testing (TANDEM, unpublished).

4.1.3 Pre-diabetes

Pre-diabetes, or intermediate hyperglycaemia, is on the rise as well. Current projections estimate the prevalence of pre-diabetes to be 470 million people worldwide by 2030. Pre-diabetes is defined by having glucose levels that are higher than normal, but lower than official diabetic thresholds, typically HbA1c values of $5.7\% - 6.5\%$. It can be considered a contentious category as not everyone with pre-diabetes progresses to diabetes, and the term 'intermediate hyperglycaemia' is sometimes used instead. It has been estimated that $5 - 10\%$ of pre-diabetes subjects progress to diabetes annually, although this varies considerably between populations.[140] By proxy, pre-diabetes must also be associated with insulin resistance and β-cell dysfunction, because if there were sufficient insulin sensitivity, elevated glucose levels would not be observed.[140]

The question of whether pre-diabetes leads to increased susceptibility of TB was raised when susceptibility was strongly linked to glycaemic control.[141] So, could patients have an increased susceptibility even at intermediate levels of hyperglycaemia? Such an association was shown by Viswanathan et al, who outlined that, in a cohort of TB patients in India, the prevalence of pre-diabetes was 24.5%.[142] However, it is unknown whether it is a cause or effect of TB disease. From an immunological perspective, Kumar et al wanted to find out if pre-diabetes also correlated with abnormal cytokine profiles in TB patients. They found that TB pre-diabetes patients were associated with increased circulating levels of Type I and Type 17 cytokines. Principally IFN-γ, TNF-α and IL-2 were significantly higher than TB without diabetes. This was also seen in response to stimulation with *M. tb* antigens, showing that in the context of TB, pre-diabetes causes a dysfunction in the immune response.[143]

4.1.4 TB/DM immunology: a dysfunction

The way type 2 diabetes influences the immune system could have a relevant effect on the body's defence to *M. tuberculosis*. There have been studies that investigate the effect of T2DM in the context of active TB disease. These include mouse experiments and clinical investigations, both of which demonstrate T2DM increases susceptibility to TB.

The immunological basis of TB susceptibility in T2DM was recognised some decades ago in streptozotocin (STZ)- treated mice. These mice have had their insulin producing cells depleted, causing hyperglycaemia, so they are a good mouse model for studying diabetes. When these diabetic mice were challenged with *M. tb*, more than 90% of them died, compared to 10% of the challenged non-diabetic mice.[144] Using the same model, another study found that post *M. tb* challenge, diabetic mice had a higher bacterial burden in the lung than non-diabetic controls. Martens et al demonstrated similar findings; that the *M. tb* growth phase plateaued significantly later in chronic diabetic mice than control mice, meaning a higher CFU weeks after infection.[145] In addition to when these STZ- treated mice were *M. tb* challenged, the diabetic mice produced less IFN-γ in the lung in the early stages of infection than in control mice. However, after 2 months post infection, the chronic diabetes mice had higher pro-inflammatory cytokine levels than in non-diabetic mice.[144] This could potentially indicate that diabetes delays the immune response to *M. tb*, meaning repercussions in the control and confinement of the bacteria.

Studies in cytokine profiles in TB/DM patients are mixed. Some report altered cytokine profiles in diabetic patients, whereas others report no differences. Restrepo et al reported TB patients with diabetes were associated with higher levels of pro-inflammatory cytokines like IFN-$\gamma$and IL-2 than in patients with uncomplicated TB. This was posi-

tively correlated with poor glycaemic control (HbA1c values $> 6.2\%$), in which even more cytokines like IL-6 and TNF were at higher levels.[146] To contribute to this, Kumar et al investigated the circulating levels of cytokines in TB/DM patients compared to TB patients without diabetes. They found heightened levels of Th1 and Th17 inflammatory cytokines in TB/DM patients than in uncomplicated TB. Such cytokines included IFN-γ, TNF-α and IL-17A, but also increased levels of IL-10 were detected. This was also the case in tuberculosis antigen (ESAT-6) stimulations of diabetic individuals.[147,148] In contrast, Stelenhoef et al found that IFN-γ cytokine production in TB/DM patients was not different when compared to TB patients without DM. However, diabetes patients without TB showed reduced levels of non-specific IFN-γ production,[149] and in a separate study, increased IL-6, which was positively correlated with HbA1c values.[49] They argue that it could mean defective non-specific immunity in diabetes that could cause susceptibility to TB.

The dysfunction of the immune system in diabetes patients expands to the function of immune cells. In vivo experiments have shown that diabetic mice display higher absolute numbers of CD4 and CD8 T-cells, macrophages and neutrophils.[145] This was also seen in humans; diabetic patients had elevated absolute counts of neutrophils in whole blood than in healthy controls. Not only are the number of immune cells higher, but the cell's immune function is also impaired. In general, it has been reported that T2DM patients show consistent defects in neutrophils. For example, they exhibit reduced migration,[150] reduction of phagocytic ability,[151] and also their antimicrobial activity.[152] This is especially the case in diabetic patients with poor glycaemic control. In the context of another infectious disease, this neutrophil impairment is seen in melioidosis, caused

by the bacterium, *Burkholderia pseudomallei*, where diabetes causes an impairment in migration and apoptosis, leading to a susceptibility of disease.[153]

Studies have shown that the immune cell's phagocytic capacity is also awry in response to *M. tb* infection. Restrepo et al showed that in T2DM patients, monocytes had overall lower complement receptor mediated phagocytosis, and also lower phagocytosis efficacy. Fc-γ Receptor mediated phagocytosis was also compromised in monocytes in T2DM patients, even after adjusting for age and BMI. Again, these impairments were associated with hyperglycaemia.[154] More recently, Raposo et al illustrated that phagocytosis is also inhibited in neutrophils in diabetic patients compared to non-diabetic controls.[155] Because neutrophils and monocytes are instrumental in the host defence against bacterial infection, a dysfunction would mean implications in the control of *M. tb* infection.

In short, the immune dysfunctions identified in T2DM are suggestive that these could render patients susceptible to infections, like *M. tb* where cell mediated immunity is important. It is clear that this impairment is exaggerated in diabetic patients with poor glycaemic control. However, the exact mechanism of susceptibility still remains unclear.

4.1.5 The TANDEM consortium

TANDEM (Tuberculosis and Diabetes Mellitus) is a four-year collaborative project funded by the EU. It is a multi-national consortium that aims to unravel the casual link and to improve care of concurrent tuberculosis disease and Type 2 Diabetes. The consortium partners have expertise in clinical studies, epidemiology, genetics and immunology, with leading laboratories in the United Kingdom and the Netherlands. TANDEM involves four field sites; South Africa, Indonesia, Peru and Romania (Figure 4.2).

Principally, TANDEM's objectives were:

i To identify feasible, accurate and cost effective ways of screening TB patients for diabetes and determine the prevalence of DM among TB patients and of TB in DM in different geographical locations.

ii Determine the level of DM management required during and after TB treatment, and the effect of glucose control on TB treatment outcomes.

iii To identify key pathways which may account for enhanced susceptibility to, and poorer treatment outcomes of TB/DM by comparing gene expression and biomarker profiles in TB patients with, compared to those without DM.

iv To establish the cellular and molecular basis responsible for the causal link between diabetes and TB, and in particular to determine the effect of hyperglycaemia and genetic variation on the host protective response to *M. tb*.

Figure 4.2: The TANDEM consortium is comprised of multi disciplinary partners all over the world, with field sites in South Africa, Peru, Indonesia and Romania. Their aim is to unravel the causal link between TB and type 2 diabetes, and improve care.

Currently, data are lacking to improve care of TB/DM patients, and the association is not mentioned in international and national treatment guidelines. However, screening of TB in DM and vice versa could improve case detection and ensure appropriate treatment. TANDEM involved determining the prevalence of TB/DM in the four fields; two (South Africa and Indonesia) of which are the top TB incidence countries.

Of the many facets of TANDEM, The London School of Hygiene & Tropical Medicine was the lead in the bioprofiling/gene expression study arm (Figure 4.3). The main objective was to establish gene expression profiles of TB/DM comorbidity versus uncomplicated TB. Gene expression technology is a powerful tool in looking at biological differences between two conditions. Because the biological mechanisms responsible for the increased TB incidence in T2DM are unknown, transcriptomic analysis had the potential to uncover such mechanisms. RNA-seq technology enabled us to perform an unbiased approach in finding out whether diabetes causes intrinsically different gene expression. This led onto performing a cross sectional study; of TB/DM, TB and TB preDM patients, to determine blood biomarker profiles of each.

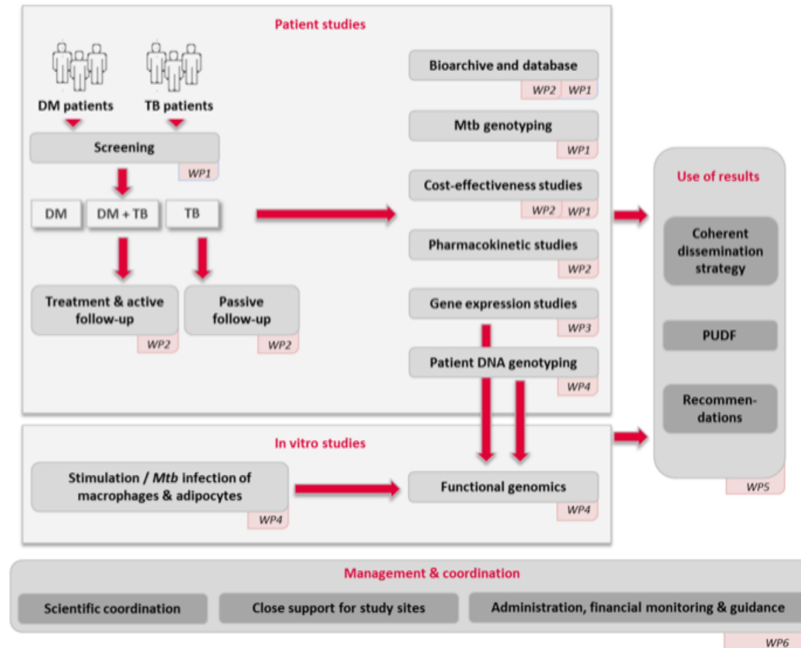


Figure 4.3: **The TANDEM consortium**

4.1.6 Current transcriptomic findings

Gene expression analysis could uncover some of the underlying biological mechanisms responsible for the increased susceptibility to TB disease in T2DM patients. The RNA expression pattern in TB patients has been well described, in which active TB patients are distinct from healthy controls, and also distinct from other diseases. The TB specific signature included a neutrophil driven Type I interferon response.[67] Since then, microarray studies have extensively investigated the TB transcriptome, in PBMCs, and whole blood in different settings.[66,74,75,156] In comparison, transcriptomic studies in T2DM are few. Manoel-caetano et al measured the gene expression in PBMCs of T2DM patients, and found there is a distinct profile of from non-diabetic subjects. The profile included

up-regulated genes involved in inflammation, and responses to oxidative stress and fatty acid processing. The TB transcriptomic studies have only included uncomplicated TB patients, which is not always reflective of the typical high TB burden population.

A study by Kaforou et al investigated the transcriptomic signature of mixed TB patients in order to generate a more inclusive result. In more detail, different populations and different disease statuses (including HIV-1 infection) were all combined. They found that when combining patients together, the signature generated was more powerful, compared to analysing the groups separately.[69] It is important to get a comprehensive analysis of the TB signature which is relevant and can be used in different settings.

So far however, T2DM patients have been excluded from such TB gene expression studies. It can be hypothesised that diabetes would modulate the effect of TB, as diabetes is itself an immune dysfunction disorder. Of course, in what way is uncertain, but because T2DM patients are more susceptible to active TB, it is presumably a detrimental effect. Therefore, investigating the gene expression profile of patients with TB and T2DM will divulge why they exhibit this increased susceptibility. We hypothesised if there were changes in common in both disease which were synergistic when combined, the expression profile of the two disease 'added together'.

Considering the pro-inflammatory profile of TB patients, and as diabetes is considered as an inflammatory condition, I hypothesised that TB/DM co-morbidity patients would have an enhanced inflammatory state to that of TB-only, which would be detectable in blood transcriptomes. Further, I hypothesied that preDM would have similar blood transcriptome to TB-only and the extent of a TB signature would correlate with extend of diabetes progression.

4.2 Aims and Objectives

Aims

I **Determine whether there were any differences of gene expression in patients with uncomplicated TB and patients with type 2 diabetes comorbidity, and to characterise any difference identified.**

II **Determine the effect of intermediate hyperglycaemia levels on the gene expression of uncomplicated TB**

Objectives

I **Perform RNA-seq analysis on whole blood samples from 4 different sites and compare the differential expression profiles of TB, TB/DM and TB preDM using an unbiased approach**

II **Compare the differential expression profile between the 4 different sites**

III **Compare the differential expression profile of TB, TB/DM, TB preDM each with healthy controls**

IV **Compare the differential expression profile of TB/DM and TB preDM each with TB**

4.3 Methods

4.3.1 Sample collection and processing

Newly diagnosed adult pulmonary TB patients with and without diabetes were recruited into the study. Untreated TB patients were bacteriologically confirmed (culture, microscopy and/or Gene Xpert depending on whether facilities were available at the field site). Informed consent was obtained and signed. Patients were excluded if they were already taking TB treatment, had MDR-TB, were HIV-positive, pregnant, taking corticosteroids, or had other serious comorbidity. TB patients previously diagnosed T2DM on anti diabetic drugs were included.

Whole blood samples were collected from four different field sites; South Africa, Indonesia, Peru and Romania. The study locations were: Stellenbosch University in Cape Town, Universidad Peruana Cayetano Heredia in Lima, University Padjadjaran in Bandung and Universitatea de Medicina si Farmacie in Craiova.

Diabetes was diagnosed by laboratory test of glycated haemoglobin (HbA1c) as $\geq 6.5\%$, according to the WHO recommendations and the American Diabetes Association. The intermediate hyperglycaemia patients were characterised with an HbA1c reading of $\geq 5.7\%$ and $< 6.5\%$. Patients without diabetes, including healthy controls, had HbA1c values of $< 5.7\%$.

Two and a half millilitres of venous blood were collected at TB diagnosis, before initiation of TB treatment, into PAXgene Blood RNA Tubes (PreAnalytiX) and frozen at $-80°C$, to then undergo RNA-seq, using the polyA tail library prep method, and single read sequencing.

Whole blood PAXgene samples (n=504) were received on dry ice by LSHTM from all four field sites; South Africa, Indonesia, Peru and Romania (Table 4.1). The PAXgene blood RNA tubes contain an additive that immediately stabilises the intracellular RNA at collection, which reduces RNA degradation. These 504 samples were from patients with TB with T2DM comorbidity, TB patients with pre-diabetes (intermediate levels of hyperglycaemia), uncomplicated TB, T2DM with no TB, and healthy controls. Sample collection was done at TB diagnosis, before the initiation of any treatment. In addition, patients with TB/DM, TB preDM and TB-only had samples collected at three other time points post initiation of TB treatment; week 2, month 2 and month 6 (discussed in Chapter 5).

Field site	Total	Cross sectional study only
Indonesia	124	38
Peru	32	32
Romania	122	70
South Africa	197	109

Table 4.1: Total number of RNA samples from each field site that underwent differential expression analysis (after any exclusions), and total number of RNA samples from each field site included in the cross sectional study.

All of the samples that underwent RNA-seq analysis were randomised using Excel, mixing all field sites, disease group and time-points. Then, each sample was assigned a new RSEQ label. This was to reduce potential batch effects, and prevent the introduction of technological bias. The 504 samples were then shipped to the Department of Genetics, University Medical Centre Groningen, University of Groningen in the Netherlands (in collaboration with Vinod Kumar and Cisca Wijmenga) where I performed the RNA-seq

lab work with Bahram Sanjabi and Desiree Brandenburg-Weening.

Total RNA (which includes mRNA, tRNA, rRNA and miRNA) was extracted using PAXgene blood miRNA kit (Qiagen) using the semi-automated QIAcube (Qiagen). RNA was quantified using the LabChip GX HiSens RNA system (PerkinElmer). The LabChip GX is a highly multiplex electrophoresis system that can analyse size and purity of RNA and DNA samples.

RNA-seq libraries were prepared using the Bioscientific NEXTflex-Rapid-Directional mRNA-seq sample preparation with the Caliper SciClone (for Illumina platform). With this kit, there were 48 different adapters available. These unique adapter barcodes were 6 bases long and used to tag the sample reads, in order to identify each sample after demultiplexing. This allows for multiple samples to be pooled and sequenced at the same time. The input RNA used for the library preparation was 1µg. The output was then measured using Caliper GX HiSens DNA kit (PerkinElmer), samples plated in triplicate into a 384 well plate. Sample pools for sequencing were made up of 24 samples, the total final concentration of which was 10mM. The samples were sequenced using the NextSeq500 High Output kit V2 (Illumina) for 75 cycles. In each flow cell, there were 24 samples, producing an average depth of 16 million reads per sample.

In total, 497 samples were sequenced. Of these, 481 were unique samples and the remaining were three 'internal normalisers' included 6 times each. These samples were included for verification of consistency in the library prep, throughout all of the plates. These 3 internal normalisers were chosen as they had high concentration and high RIN values.

4.3.2 Data processing and analysis

Patient characteristic variables were tested for normality using the Shapiro-Wilk test. If deemed normally distributed, a pairwise t-test was performed, if not, a non-parametric Wilcoxon Rank Sum Test was performed. This was done using the statistical programming language R.

RNA-seq samples were demultiplexed by Gerban van der Vries at the University of Groningen, Genetics Department. Twenty four samples were run on one flow cell, and because each flow cell had four lanes, 4 FASTQ files were generated for each sample. I obtained the 1,988 FASTQ files, which I then processed and analysed. FASTQ files were aligned to the human genome (version Human_g1k_v37) using STAR aligner (Spliced Transcripts Alignment to Reference)[116] (v2.5.1b). The 1,988 SAM files were then converted to BAM files, and the 4 BAM files for each sample were merged using SAMtools[157] (v1.2), creating 497 complete BAM files. This was done in Unix in the terminal using the GCC cluster (Genomics Coordination Centre in Groningen University). Reads were quantified using HTSeq-count[118] (v0.6.1) using the ensembl annotation file version `Homosapiens.GRCh37.75.gtf` with the settings described below.

`-f bam` format of the input data

`-r name` how the input data have been sorted

`-s no` whether the data is from a strand specific assay

`-m union` mode to handle reads overlapping more than one feature

`-i gene id` what attribute to be used as feature ID

Quality control of FASTQ data was done using the program FastQC (Babraham

Bioinformatics, v0.11.5). Count data generated by HTSeq-count, in the form of txt files were then imported into the statistical programming language R v3.3.3 and RStudio v1.0.136. Non-informative rows were removed and differential expression analysis was performed using `DESeq2` (v 1.12.4). Data were normalised using the default settings as per the metformin study in Chapter 3. Multiple testing was corrected using the Benjamini & Hochberg false discovery rate method. We had two different study designs, which are summarised in Figure 4.4.

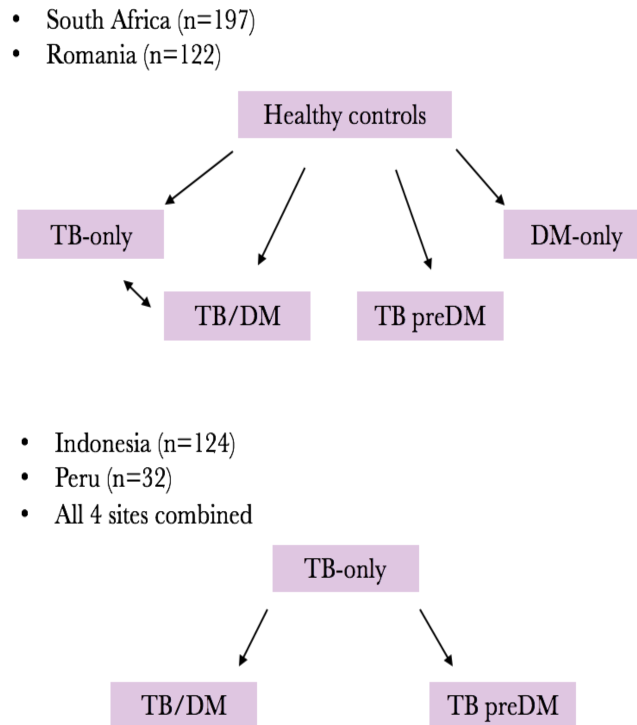


Figure 4.4: **RNA-seq cross-sectional study design.** The cross sectional analysis comprised of two study designs. The first compared all disease phenotypes with healthy controls, as in South Africa and Romania. The second used TB-only as the control, and was done for Indonesia, Peru and also when combining all 4 of the field sites.

The first study design applied to South Africa and Romania because the field sites

had collected healthy control and type 2 diabetes patients without TB. The field sites were normalised separately, and all of the disease phenotypes, TB-only, TB/DM, TB preDM and DM-only were in turn compared to their matching healthy controls.

The second study design was performed in the Indonesia and Peru populations, and later when all the field sites were combined and analysed together. This is because Indonesia and Peru only collected samples from TB/DM, TB preDM and TB-only patients. Here, TB-only was used as the control, where TB/DM and TB preDM were compared against TB-only patients. This was conducive to find out how diabetes impacts the TB gene expression profile. Another cross sectional analysis was done using this study design; combining all the TB/DM, TB preDM and TB-only samples from all 4 field sites. This substantially increased the sample numbers in the 3 disease groups, and also means any differences found would be applicable across different populations.

The two different study designs existed because in the original TANDEM consortium plan, only samples from South Africa were intended to be analysed using microarrays. The opportunity to perform RNA-seq analysis on a larger sample set became available during the course of the consortium project, and it was then decided to include samples from the other three sites in order to make any differences relevant in different populations, but we were restricted to what had been collected.

Of the combined analysis, two sub-analyses were done. One included a population correction term in the design formula (Equation 4.2), and one did not (Equation 4.1). Incorporating a second covariate can control the (unwanted) variation caused by population, and therefore further uncover the actual biological variation due to disease.

$$y = (\beta 1_{disease\ category} + \epsilon) \quad (4.1)$$

$$y = (\beta 1_{disease\ category} + \beta 2_{population} + \epsilon) \quad (4.2)$$

4.3.3 Modular analysis

Modular analysis is a technique that analyses groups of genes together as a means to identify immunological mechanisms. These gene groups are genes that tend to be expressed together, formulated from transcriptional data sets. It can provide a framework for visualisation and function interpretation and a means to identify immunological mechanisms. It was done using R package `tmod` (v0.31). This package uses the modules outlined by Chaussabel et al[94] and by Li et al.[95]

Module enrichment is calculated by ranking the genes by p-value that belong to the module, and genes that do not belong to the module using the CERNO test, described by Yamaguchi et al.[158] An adjusted p-value of < 0.05 was deemed to be significant.

4.3.4 Previously published signature analysis

To test the effect of diabetes and pre-diabetes on published TB signatures, the Zak et al[76] 16 gene signature was analysed in the differential expression results (disease category versus healthy controls). This specific signature was selected for analysis as literature is moving forward with smaller signatures and it has received the most critical acclaim. Also, it used more sophisticated and robust methods based on machine learning algorithms, rather than for example just a list of significantly differentially expressed genes.

The log fold changes for these 16 genes were extracted from the comparisons; TB/DM, TB preDM, and TB-only versus healthy controls. These fold change values were summed as per the Disease Risk Score formula from Kaforou et al.[69] The normalised count for these genes was also extracted for each patient and summed in order to stratify by HbA1c.

4.4 Results

4.4.1 Patient characterisation

The patients from the four field sites (Indonesia, South Africa, Peru and Romania) were age and sex matched. This was deliberate so there was no age or gender bias between the groups. For example, in South Africa, TB patients were selected so that they were not significantly different from the TB/DM patients. However, in the cases of the other field sites, there were no extra patients so all of them were analysed, Figure 4.5 and 4.6. Age was normally distributed, and pairwise t-test revealed no significant difference in age between any of the groups, after multiple testing correction ($p\text{-value} < 0.05$).

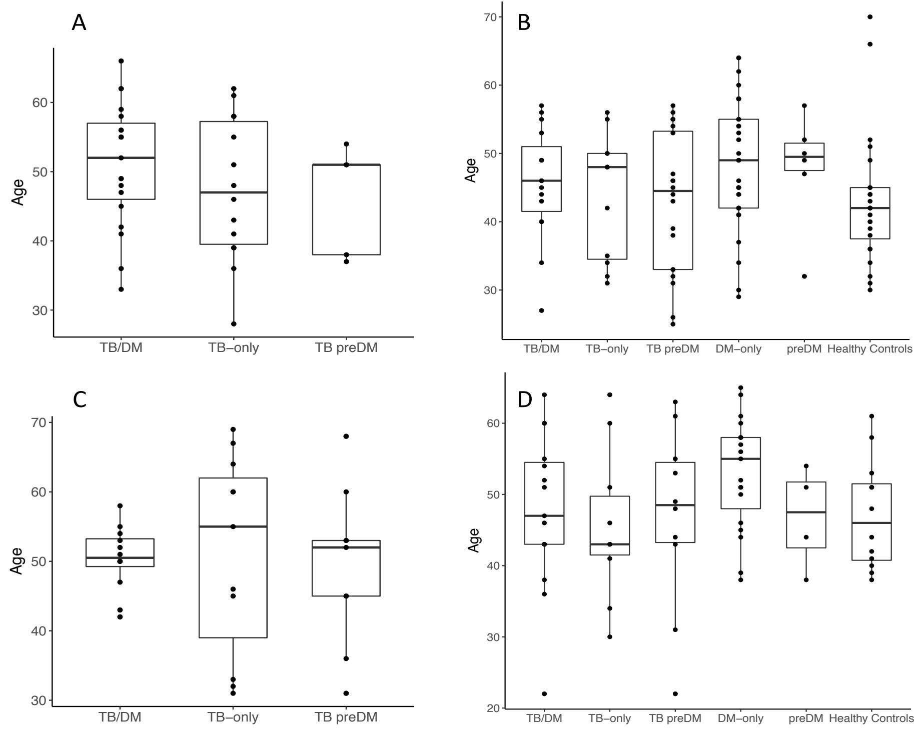


Figure 4.5: **Age of patients of all the patients groups before the initiation of TB treatment in the four field sites.** A: Indonesia, B: South Africa, C: Peru, D: Romania. Lines show the median, and the first and third quartile. Points outside the lines represent outliers, calculated by $Q1 - 1.5 * IQR$ and $Q3 - 1.5 * IQR$. Pairwise t-test was performed and a p-value < 0.05 was deemed statistically significant.

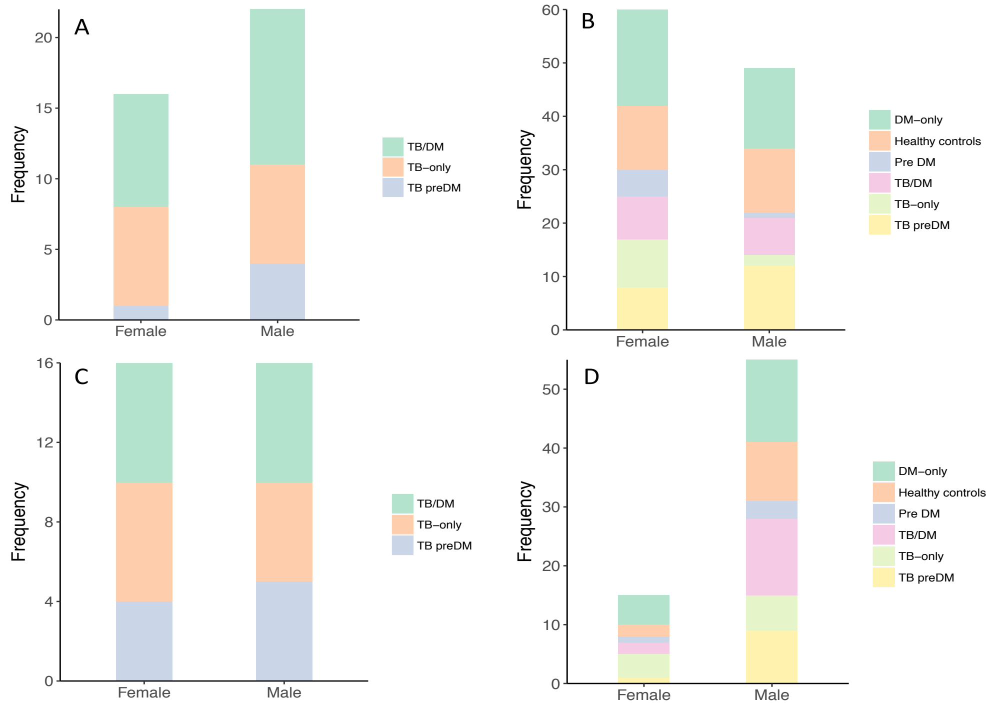


Figure 4.6: **Gender of the patients in each of the groups, in the four field sites.** A: Indonesia, B: South Africa, C: Peru, D: Romania.

In all the populations, Indonesia, South Africa, Peru and Romania, patients with TB/DM had generally higher body max index, although not statistically different from TB-only or TB preDM (Figure 4.7). In South Africa, BMI tended to be higher in diabetes patients, this was especially the case patients that were not diagnosed with TB. DM-only, preDM and Healthy Controls were all statistically significantly different from the TB disease groups; TB/DM, TB-only, and TB preDM (pairwise Wilcoxon Rank Sum Test, p-value < 0.05), results shown in Table 4.2. This is expected as TB patients have characteristic weight loss. In Romania, the BMI of the DM-only control group was statistically significant from TB/DM, TB-only and TB preDM, results shown in Table 4.3.

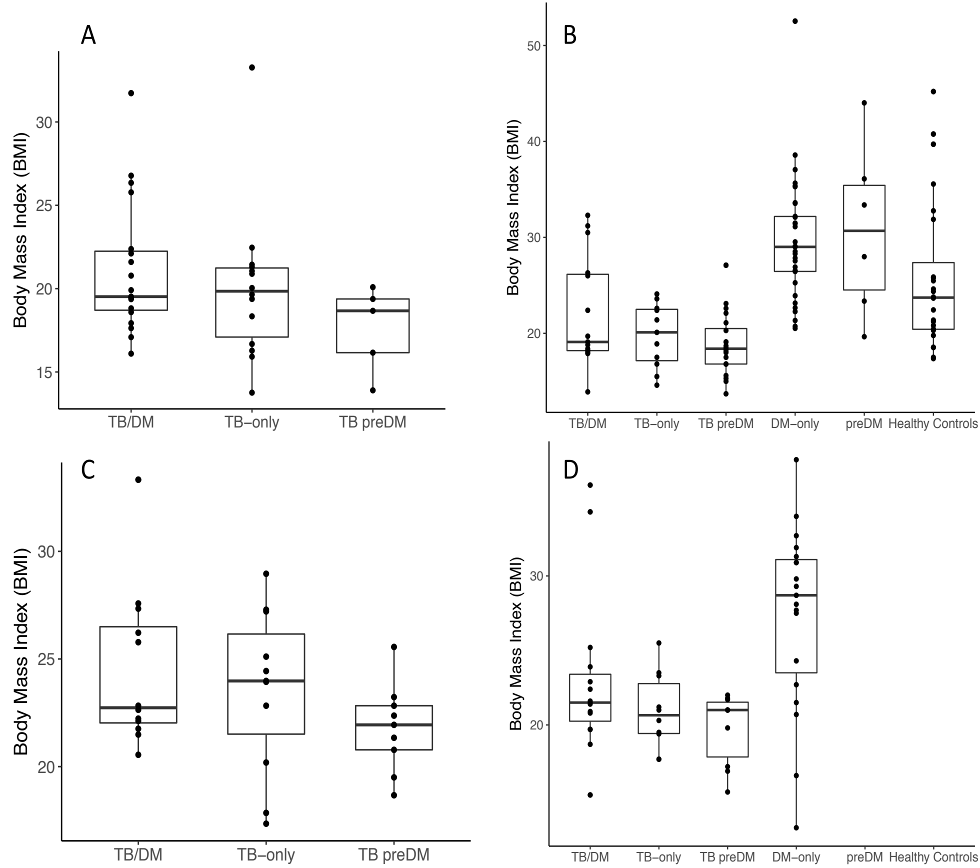


Figure 4.7: **Body mass index (BMI) of all the patient groups before the initiation of TB treatment in the 4 field sites** A: Indonesia, B: South Africa, C: Peru, D: Romania. Lines show the median, and the first and third quartile. Points outside the lines represent outliers, calculated by $Q1 - 1.5 * IQR$ and $Q3 - 1.5 * IQR$. Groups were tested for normality and a non-parametric Wilcoxon Rank Sum Test was performed.

	DM-only	Healthy controls	pre DM	TB/DM	TB-only
Healthy controls	0.083	-	-	-	-
pre DM	1	0.839	-	-	-
TB/DM	0.0041	0.617	0.152	-	-
TB-only	$2.4e^{-05}$	0.152	0.0931	1	-
TB preDM	$3.8e^{-07}$	0.0034	0.0153	0.617	1

Table 4.2: **Significance of Body Mass Index between each patient group in South Africa.** Results from a non-parametric Wilcoxon Rank Sum Test. Colour indicates a significant p-value (< 0.05)

	DM-only	TB/DM	TB-only
TB/DM	0.121	-	-
TB-only	0.021	0.663	-
TB preDM	0.014	0.635	0.663

Table 4.3: **Significance of Body Mass Index between each patient group in Romania.** Results from a non-parametric Wilcoxon Rank Sum Test. Colour indicates a significant p-value (< 0.05)

Figure 4.8 shows the HbA1c values in all the populations and groups. In all of the populations, the HbA1c values of the patient groups were all significantly different from one another, particularly in TB/DM patients than in TB-only or TB preDM, results shown in Tables 4.4, 4.5, 4.6 and 4.7 (non-parametric Wilcoxon Rank Sum Test, p-value < 0.05). In Romania and South Africa, there were also DM-only patients, which consisted of newly diagnosed, and previously diagnosed DM, some of which had well controlled diabetes. Here, DM-only was also significantly different from the other groups. The populations exhibit different patterns of glucose control. For instance, in Indonesia, the

TB/DM patients were more defined, as they had a considerably smaller IQR of HbA1c values, and also have a higher mean average than in the other populations. On the other hand, South African and Romanian diabetes patients have a much larger variability in glucose control, with lower average means.

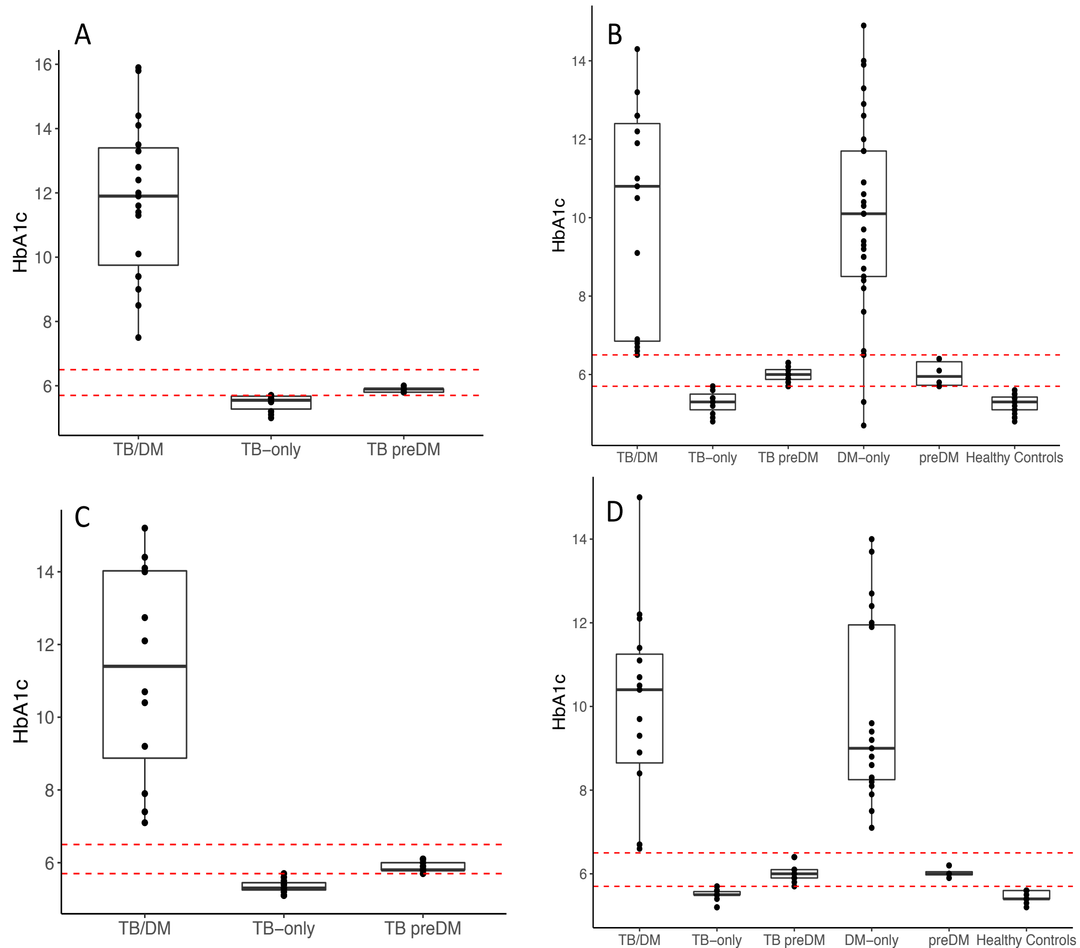


Figure 4.8: **Laboratory HbA1c values, taken before the the initiation of TB treatment of all the patients groups in the 4 field sites.** A: Indonesia, B: South Africa, C: Peru, D: Romania. Dashed red lines indicate the threshold cut offs for pre-diabetes (intermediate hyperglycaemia) (HbA1c 5.7% – 6.5%) and diabetes (HbA1c $\geq$ 6.5%). Lines show the median, and the first and third quartile. Points outside the lines represent outliers, calculated by $Q1 - 1.5 * IQR$ and $Q3 - 1.5 * IQR$.

	TB/DM	TB-only
TB-only	$4e^{-06}$	-
TB preDM	0.0017	0.0017

Table 4.4: **Significance of HbA1c values between each patient group in Indonesia.** Results from a non-parametric Wilcoxon Rank Sum Test. Colour indicates a significant p-value (< 0.05)

	DM-only	Healthy controls	pre DM	TB/DM	TB-only
Healthy controls	$9.0e^{-08}$	-	-	-	-
pre DM	0.00378	0.00136	-	-	-
TB/DM	1	$2.5e^{-06}$	0.00315	-	-
TB-only	$8.7e^{-05}$	1	0.008	0.00016	-
TB preDM	$1.4e^{-06}$	$2.1e^{-07}$	1	$6.3e^{-06}$	$8.7e^{-05}$

Table 4.5: **Significance of HbA1c values between each patient group in South Africa.** Results from a non-parametric Wilcoxon Rank Sum Test. Colour indicates a significant p-value (< 0.05)

	TB/DM	TB-only
TB-only	0.00016	-
TB preDM	0.00027	0.00027

Table 4.6: **Significance of HbA1c values between each patient group in Peru.** Results from a non-parametric Wilcoxon Rank Sum Test. Colour indicates a significant p-value (< 0.05)

	DM-only	Healthy controls	pre DM	TB/DM	TB-only
Healthy controls	$5.8e^{-05}$	-	-	-	-
pre DM	0.0163	0.0189	-	-	-
TB/DM	1	0.00016	0.0189	-	-
TB-only	0.00018	1	0.0189	0.00037	-
TB preDM	0.00018	0.00067	1	0.00037	0.00147

Table 4.7: **Significance of HbA1c values between each patient group in Romania.** Results from a non-parametric Wilcoxon Rank Sum Test. Colour indicates a significant p-value (< 0.05)

The TB/DM patients within the different field sites were receiving anti-diabetes medication (Figure 4.9). They were on either insulin, metformin, a combination of the two, in combination with other diabetes drugs, or no medication at all. The patients on no medication were mostly newly diagnosed cases. The majority of previously diagnosed diabetes patients in Romania were on insulin. But, in the other three field sites, diabetes medication was mixed, with most on metformin or metformin in combination with other oral diabetes medication (like glibenclamide).

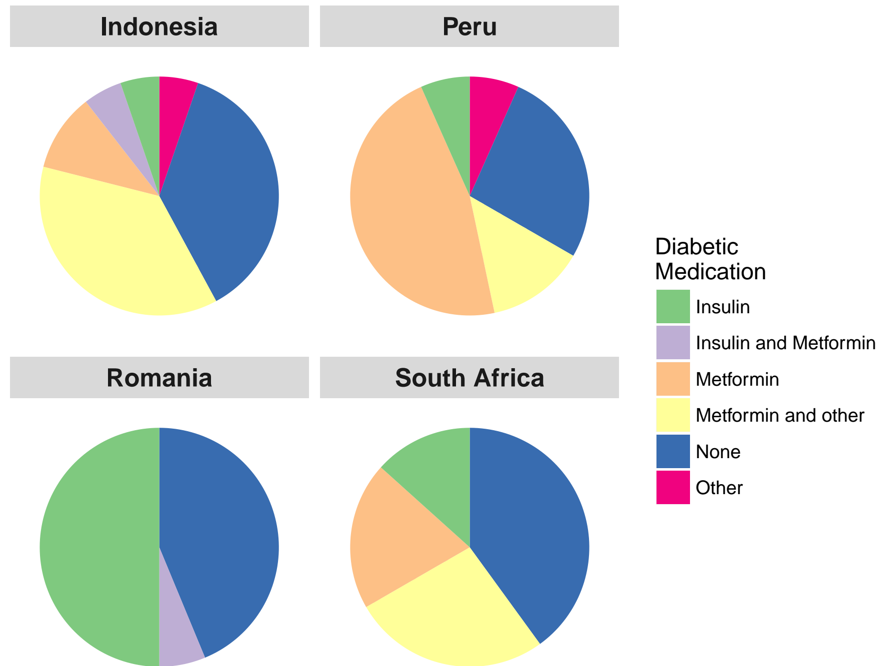


Figure 4.9: **Proportion of TB/DM patients on anti diabetic medication in the four different field sites.**

4.4.2 Assessment of RNA quality

In total (cross sectional and longitudinal study), 504 PAXgene tube samples were isolated and underwent quality assessment. Of these, 436 had a RNA integrity number (RIN) of ≥ 7, and 46 had a RIN of $\leq 7 \geq 4$. It was decided that these 482 samples would proceed to the library preparation stage. This was because although some samples that had varying degrees of quality, clear 18S and 28S RNA peaks were still seen in the LabChip GX (Figure 4.10). In 22 cases, the samples had to be excluded due to either absence of RNA or complete degradation. The field site with the poorest quality RNA was Peru, which has significantly lower RIN values than in South Africa, Romania or Indonesia (Figure 4.11).

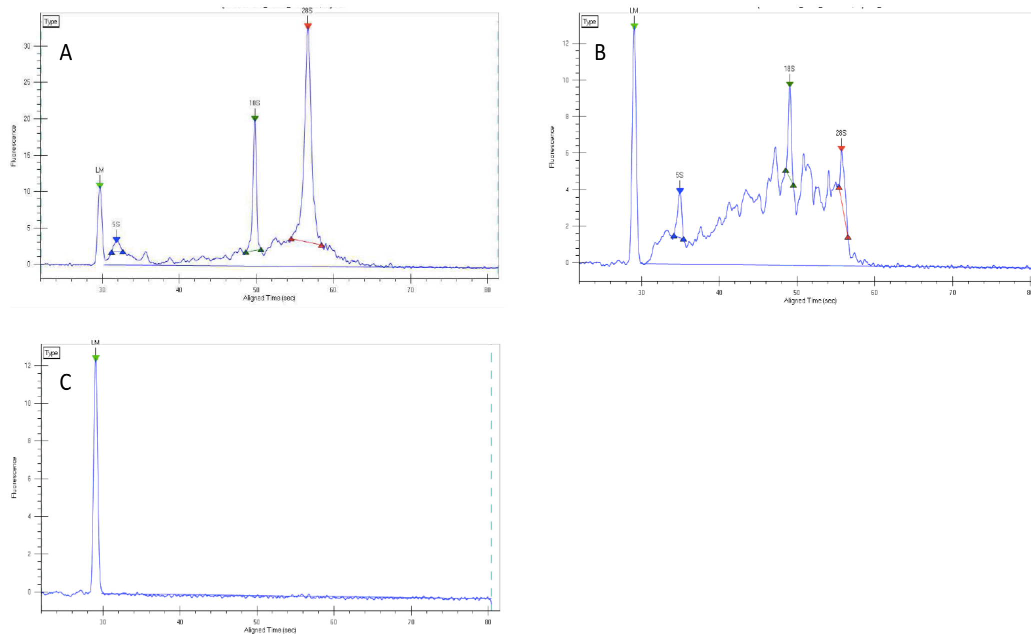


Figure 4.10: **Examples of variations of RNA quality.** RNA integrity number (RIN) was used as a measure for RNA quality. RIN is calculated from the ratio of 28S and 18S peaks, generated by the LabChip GX. The clearer the two peaks, the higher the quality of RNA. A: good quality RNA samples (RIN=8), B: partially degraded RNA sample (RIN = 4), C: sample with no RNA yield (RIN=0)

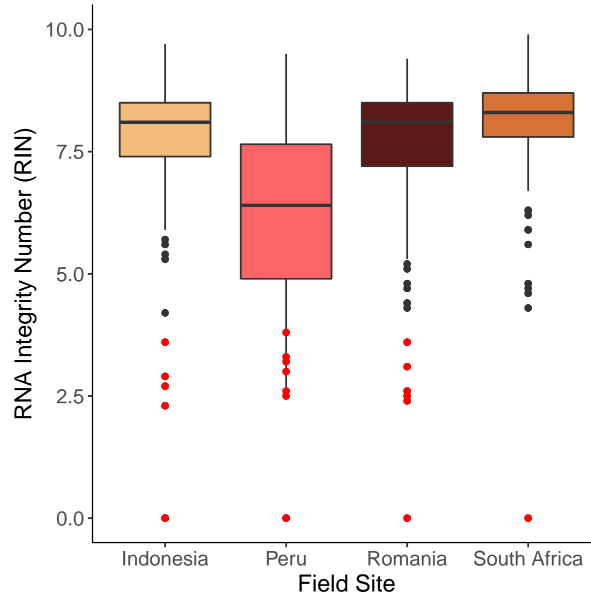


Figure 4.11: **Relationship between RNA quality (RIN) and field site.** RIN was calculated using the LabChip GX. Samples shown red, had very poor RNA quality and had to be excluded. The majority of exclusion occurred in Indonesia and Peru. Pairwise t-test revealed RIN values of samples from Peru were significantly different from the other 3 populations (p-value 1×10^{-7}).

A total of 482 unique samples were sequenced, after which, 3 samples had to be excluded due to poor number of aligned reads (Figure 4.12). The RIN values and number of aligned reads appear to have a positive association, with a higher RIN value, an increased number of reads aligning to the reference. However, 1 of the samples with poor read alignment had a RIN of 8. During cleaning of the meta data, 4 further samples had to be excluded. This brought the sum of samples for downstream analysis to 475.

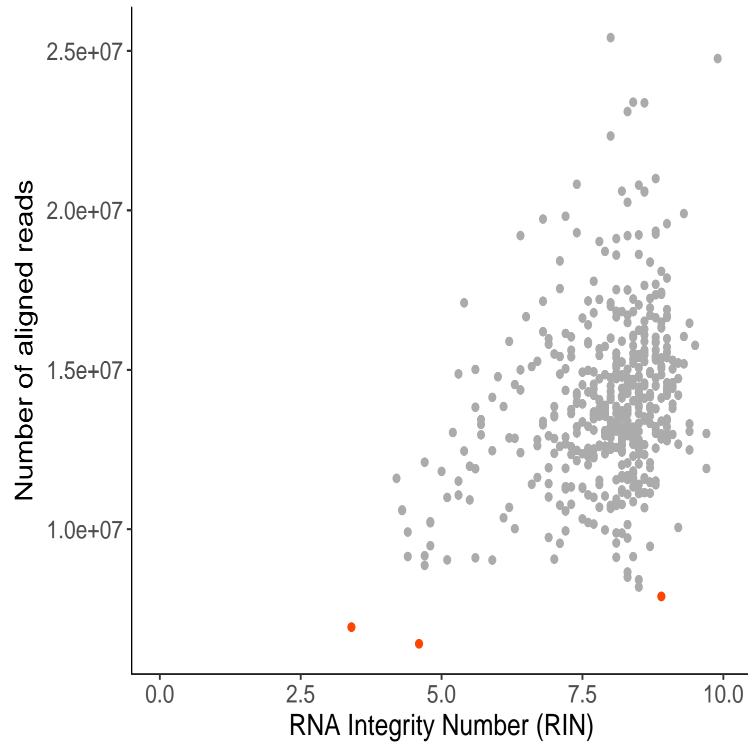


Figure 4.12: **Relationship between RNA quality (RIN) and read alignment.** Number of reads aligned to the reference human genome in each of the 481 whole blood RNA samples. RNA quality was assessed using LabChip GX. Three samples had fewer than 8,000,000 reads, and had to be excluded for down stream analysis, shown in red.

The cross-sectional study consisted of 249 samples from all 4 field sites and 6 disease categories; TB/DM comorbidity, TB preDM, TB-only, DM-only, pre-DM and healthy controls. The 3 categories; TB/DM, TB preDM and TB-only were the only groups consistent across all 4 populations, in which there were 151 samples in total. (Table 4.10)

Field site	TB/DM	TB preDM	TB-only
Peru	12	9	11
Indonesia	19	5	14

Table 4.8: **RNA sample numbers that underwent differential expression analysis in the different patient groups within the cross sectional study in Peru and Indonesia.**

Field site	TB/DM	TB preDM	TB-only	DM-only	preDM	Healthy Controls
Romania	15	10	10	19	4	12
South Africa	15	20	11	33	6	24

Table 4.9: **RNA sample numbers that underwent differential expression analysis in the different patient groups within the cross sectional study in South Africa and Romania.**

TB/DM	61
TB preDM	44
TB-only	46
total	151

Table 4.10: **RNA sample numbers of the three groups that were consistent across all the field sites within the cross sectional study.**

4.4.3 Gene expression in TB/DM patients in South Africa

The South African cohort consisted of 109 samples for 6 groups; TB/DM comorbidity, TB-preDM, TB-only, DM-only, pre-DM and healthy controls (Table 4.9). All 109 samples were normalised together, and a PCA was done to observe any clustering (Figure 4.13). The PCA plot revealed two main clusters. Healthy controls and people with type 2 diabetes (and pre diabetes) clustered together. The second main cluster comprised of TB/DM comorbidity and TB pre diabetes patients. Uncomplicated TB patients spread across these two clusters. The PCA revealed that the TB pre DM patients were more similar to TB/DM patients.

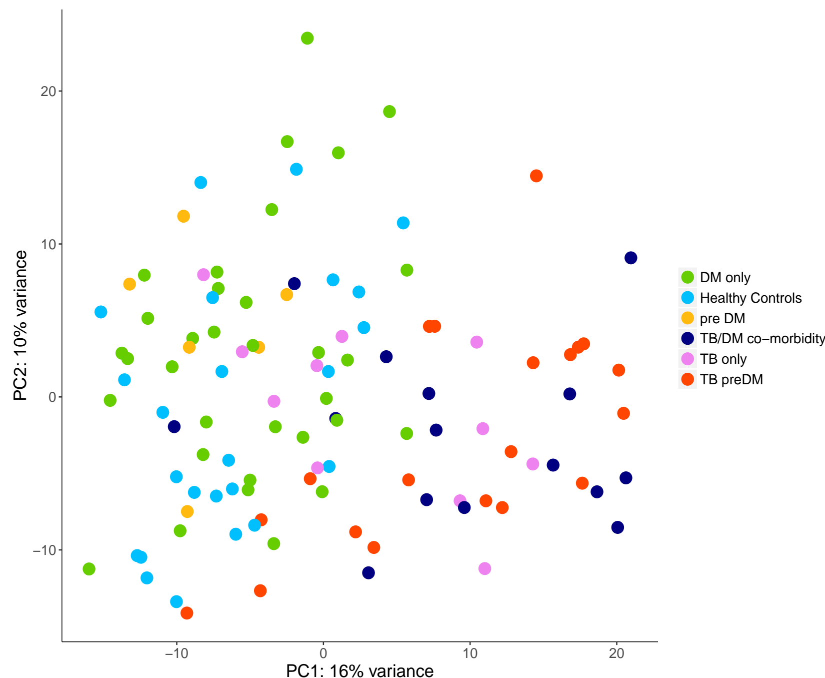


Figure 4.13: **Blood global transcriptome differences between patients groups in South Africa.** Principal component analysis of all the patient groups in South Africa before the initiation of TB treatment.

Each disease group was compared to healthy controls, to obtain a differential expression profile of each phenotype. Volcano plots in Figure 4.14 summarise the gene expression profiles for each disease category. As expected, when comparing TB-only patients to healthy controls, there is a large asymmetrical up-regulation of many genes, mirroring previous literature.[67,68] In DM patients, the expression profile was more symmetrical, with a lower magnitude of fold change. Unexpectedly however, the TB/DM and TB preDM differential expression profiles were very similar to each other, but also together very different from the expression profiles of TB-only and DM-only versus healthy controls. The two groups; TB/DM and TB preDM exhibited a comparable amplitude of gene expression, with many of the same genes up- and down-regulated.

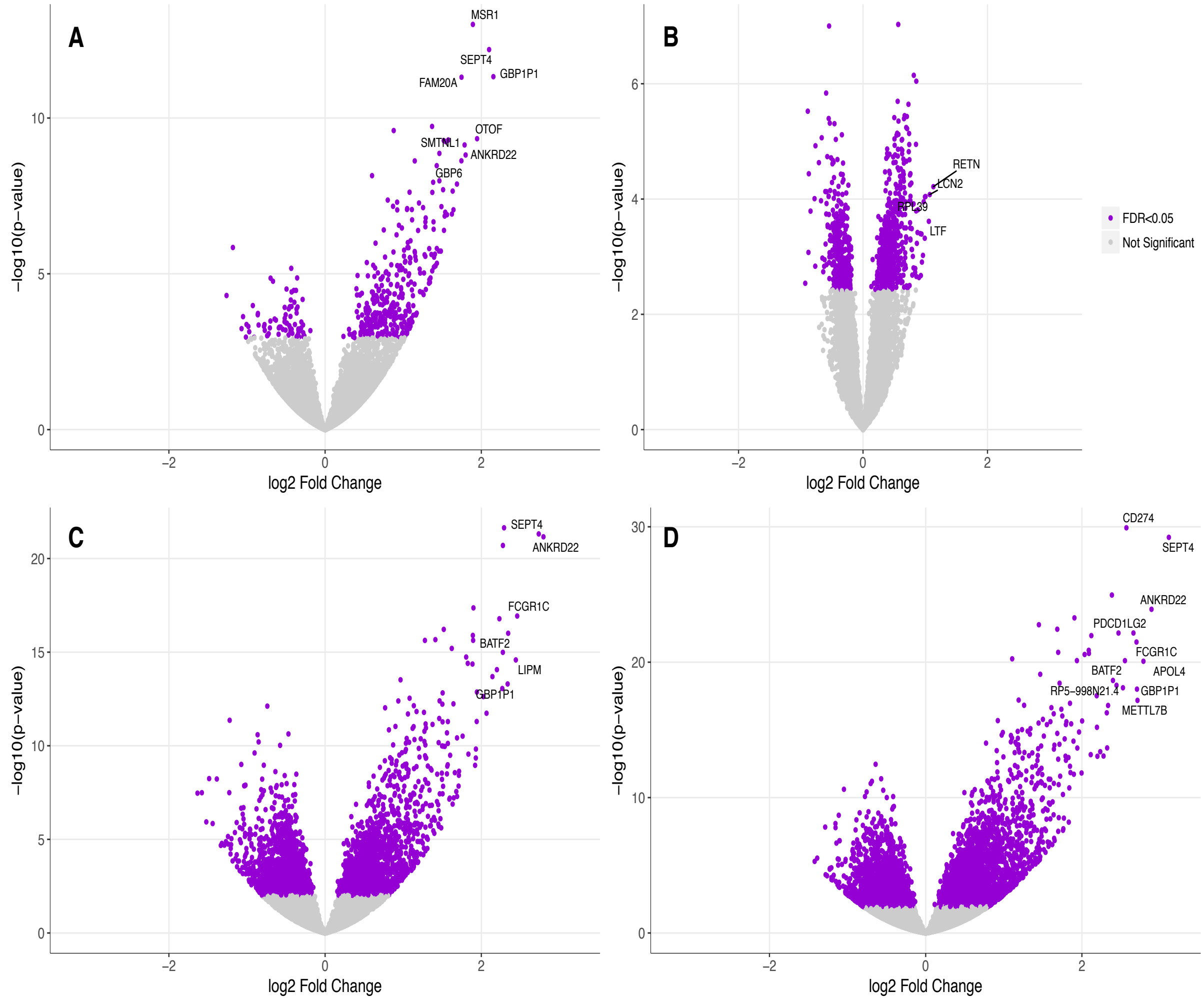


Figure 4.14: **Differential expression analysis of all the disease phenotypes in South Africa compared to healthy controls before the initiation of TB treatment.** Gene expression profiles of A: TB-only (n=11); B: DM-only (n=33), C: TB/DM (n=15), D: TB preDM (n=20), each relative to healthy controls (n=24). Genes that were deemed statistically significant had an adjusted p-value of < 0.05 after multiple testing correction (method=Benjamini & Hochberg). Purple corresponds to the genes that were significant, grey is genes that were not significant. The most differentially expressed genes were labelled.

Venn diagrams in Figures 4.15 and 4.16 encapsulate the number of genes differentially expressed when each phenotype was compared with healthy controls. Interestingly, the genes differentially expressed in diabetes are not the same as those differentially expressed in TB-only. This is curious because T2DM is seen as an inflammatory disorder, and TB stimulates a large pro inflammatory response, so one would expect there to be similarities. The genes differentially expressed in TB/DM and TB preDM overlapped considerably with DM and TB-only, but also included a large number of genes which were not differentially expressed in these two groups compared to healthy controls.

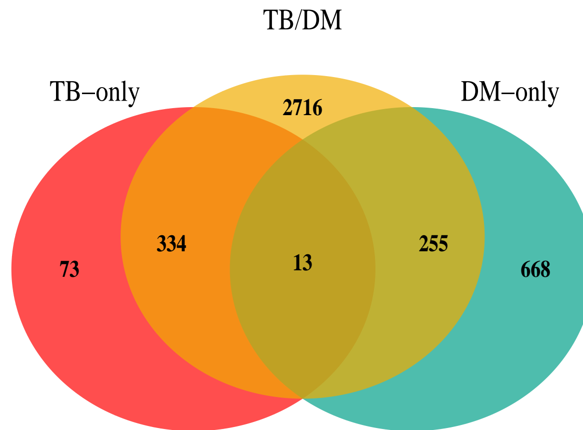


Figure 4.15: **The number of statistically significantly differentially expressed genes in TB/DM, TB-only and DM-only relative to healthy controls in South Africa.** All samples in South Africa taken before initiation of TB treatment were normalised together, and differential expression performed. An adjusted p-value of < 0.05 was deemed significant.

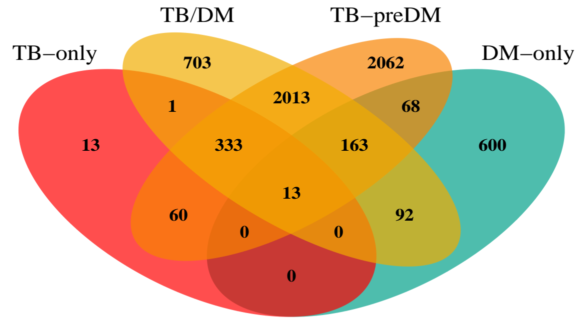


Figure 4.16: **The number of statistically significantly differentially expressed genes in TB/DM, TB preDM, TB-only and DM-only relative to healthy controls in South Africa.** All samples in South Africa taken before initiation of TB treatment were normalised together, and differential expression performed. An adjusted p-value of < 0.05 was deemed significant.

Modular analysis was performed in order to introduce biological knowledge to the groups of genes that were expressed together, shown in Figure 4.17. This was condensed into a panel plot, which shows the statistically significant modules in each of the four disease categories compared to healthy controls. In TB-only, modules that were up-regulated, included type I interferon response, complement and activated dendritic cells. In contrast, in diabetes patients, mainly respiration modules were up-regulated, that are not present in the comorbidity groups. The TB/DM and TB preDM patients exhibited similar module expression. Of the differentially expressed modules, most were up-regulated. The modules differentially expressed in TB-only were also expressed in the TB/DM and TB preDM groups. This was in addition to increased down-regulation of modules, in particular; adaptive immune response modules such as T-cell activation, differentiation and B-cells.

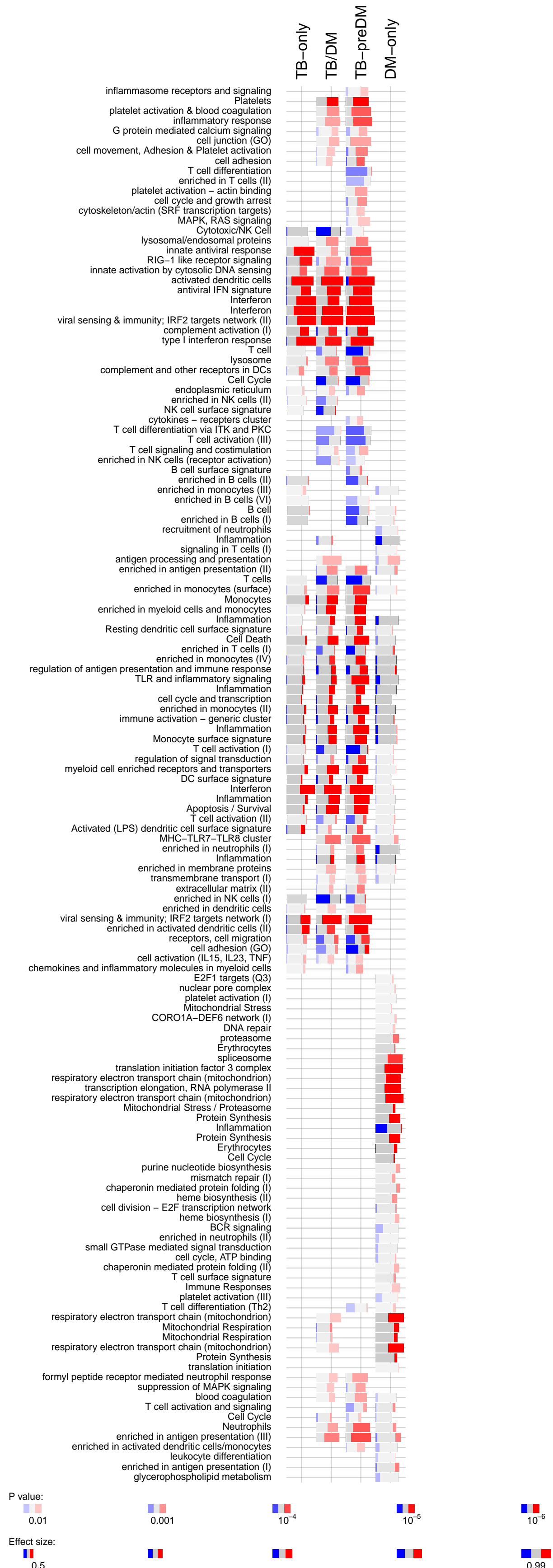


Figure 4.17: **Transcriptional modules that were significantly differentially expressed in TB-only, TB/DM, TB-preDM and DM-only compared to healthy controls in South Africa before initiation of TB treatment.** Transcripts were evaluated using a pre existing modular framework. Up-regulated (red) and down-regulated (blue) modules. The length of each bar corresponds to effect size of that module, and the colour saturation represents the adjusted p-value (< 0.05). The amount of colour represents the proportion of genes within that module that are differentially expressed.

4.4.4 Gene expression in TB/DM patients in Romania

The Romanian analysis consisted of 70 samples and was the second field site that also obtained healthy controls (Table 4.9). Again, during differential expression analysis, each disease phenotype was compared to healthy controls to get a gene expression profile of each disease group (Figure 4.4). PCA analysis in Figure 4.18 revealed some clustering of healthy controls and diabetes patients together. Also, TB/DM and TB preDM clustered primarily together.

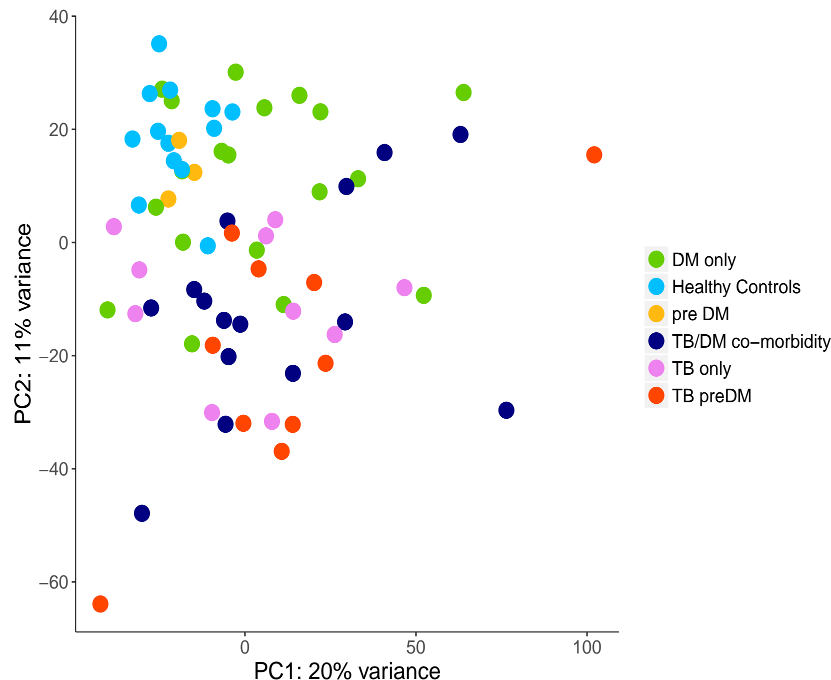


Figure 4.18: **Blood global transcriptome differences between patients groups in Romania.** Principal component analysis of all the patient groups in Romania before the initiation of TB treatment.

To demonstrate the gene expression profiles of each disease category versus healthy controls, volcano plots were generated (Figure 4.19). In the other control group; DM-only, there was a pattern of up-regulation. The gene Carbonic Anhydrase 1 (CA1) had

the largest statistically significant fold change, and has a role in respiration.

Complement C1q B Chain (C1QB) and interferon response genes, Interferon Induced Protein With Tetratricopeptide Repeats 1 and 3 (IFIT1 and IFIT3) were the top significantly differentially expressed genes in TB-only relative to healthy controls. This mirrors previous literature. Complement and interferon related genes are frequently noted as up-regulated in response to active TB disease. In this population, TB-only against healthy controls were different from in the South African population in that there was more down-regulation as well as the expected up-regulation.

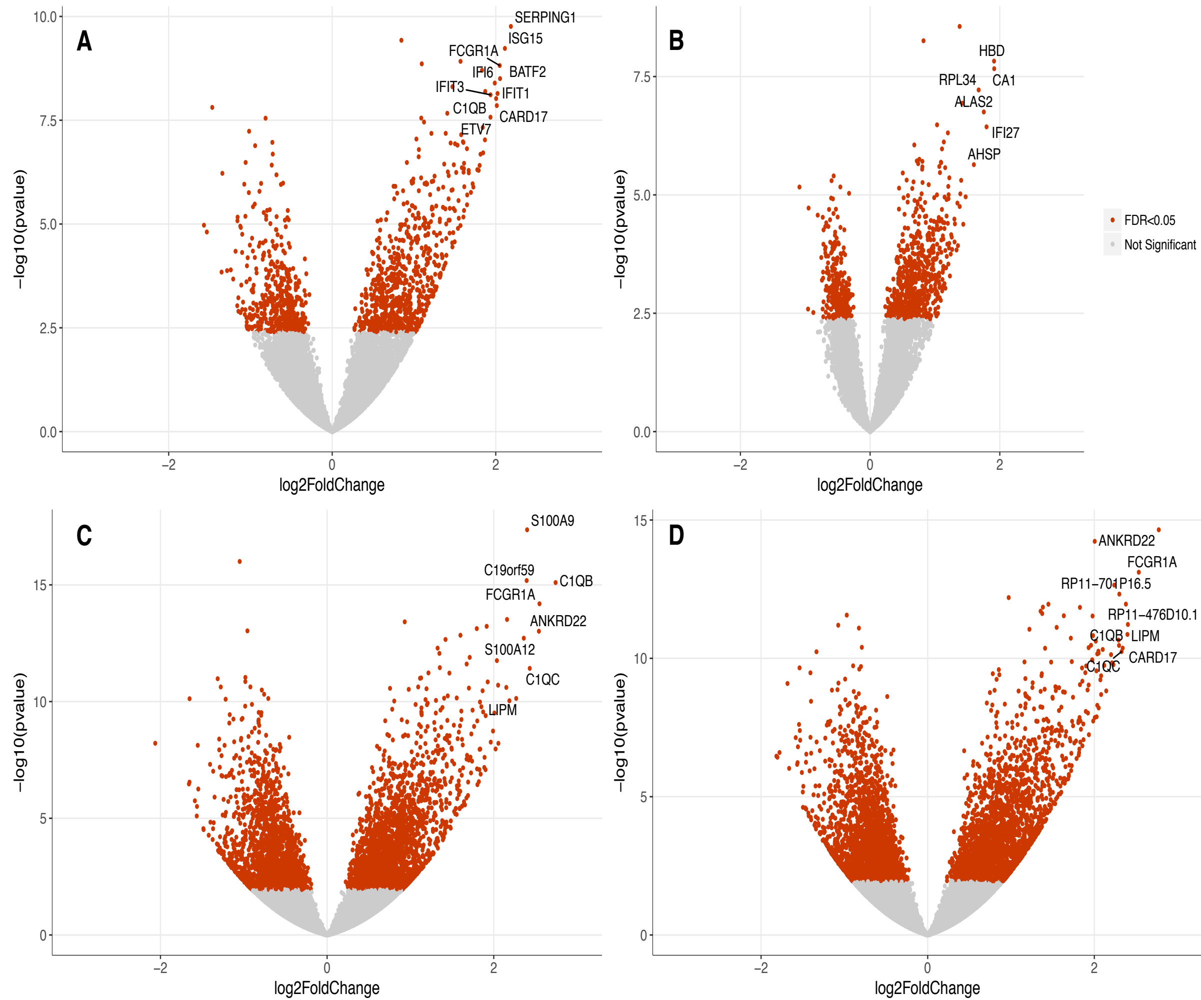


Figure 4.19: **Differential expression profiles of disease categories compared to healthy controls in Romania before initiation of TB treatment.** Blood transcriptome of; A: TB-only (n=10), B: DM-only (n=19), C: TB/DM (n=15), D: TB preDM (n=10) relative to healthy controls (n=12). Orange points represent genes that were statistically significant after multiple testing correction (adjusted p-value < 0.05), grey were genes that were not statistically significant. The most differentially expressed genes were labelled.

TB/DM and TB preDM versus healthy controls had similar expression profiles to one another, but were different to the TB-only comparison. This is alike the result from the same group comparisons in the South African population. Both disease categories had a similar magnitude of expression, with many genes over $2log_2$ fold change differentially expressed. With the large up-regulation, there was also considerable down-regulation. The genes C1QB and C1QC were highly up-regulated in both TB/DM and TB preDM than in healthy controls.

The number of statistically significant genes were plotted in Venn diagrams, Figures 4.20 and 4.21. The genes differentially expressed in TB-only or DM-only had a small overlap. The genes differentially expressed in TB/DM or TB preDM have a large overlap, but are in turn both very different from the other group comparisons.

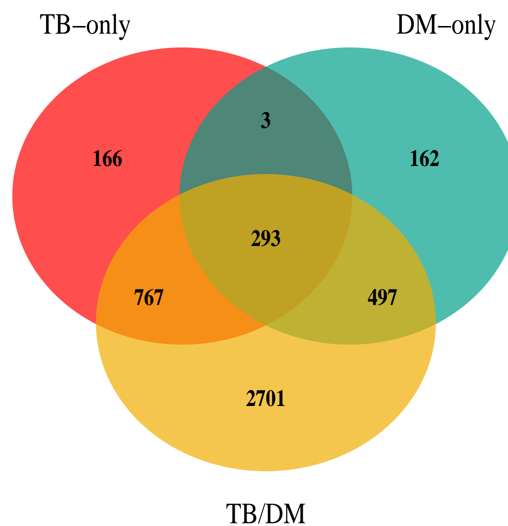


Figure 4.20: **The number of statistically significantly differentially expressed genes in TB/DM, TB-only and DM-only relative to healthy controls in Romania.** All samples in Romania taken before initiation of TB treatment were normalised together, and differential expression performed. An adjusted p-value of < 0.05 was deemed significant.

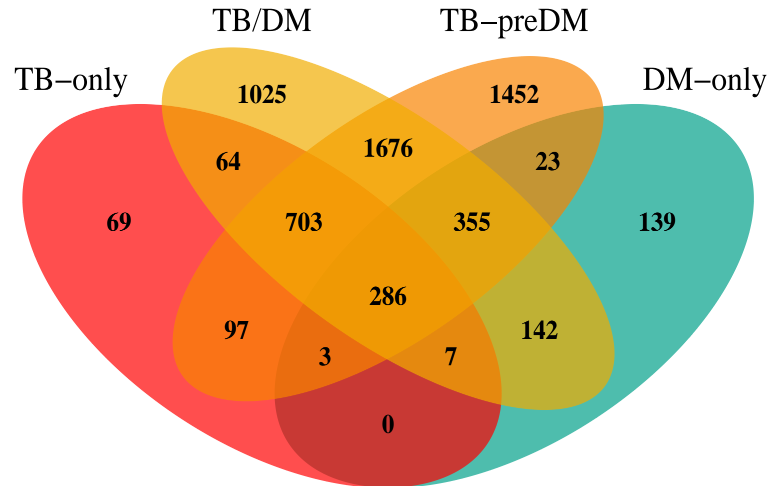


Figure 4.21: **The number of statistically significantly differentially expressed genes in TB/DM, TB preDM, TB-only and DM-only relative to healthy controls in Romania.** All samples in Romania taken before initiation of TB treatment were normalised together, and differential expression performed. An adjusted p-value of < 0.05 was deemed significant.

Figure 4.22 shows the modular analysis of the four Romanian patient groups versus healthy controls. In TB-only, as in its corresponding volcano plot, there was a large up-regulation of modules, but also reasonable down-regulation. The most statistically significant differentially expressed module was interferon. The genes within the particular interferon module are involved in IFN-γ signalling and transcriptional regulators of type I interferons. Along with interferon, monocytes, dendritic cells and inflammation were also up-regulated. Together with this, adaptive immune response modules were mostly down-regulated. These include T-cell differentiation and activation.

When comparing DM-only to healthy controls, modules were largely up-regulated. This comprised of modules involved in respiration, like the electron transport chain, but

also inflammation modules. This is contrasting to the DM-only patients in South Africa, which saw slight down-regulation of inflammation modules.

The modular expression profiles of TB/DM and TB preDM were again very similar, with considerable up-regulation. The top significantly differentially expressed modules (in both) were enriched in monocytes and inflammation. Together with this, there was also down-regulation of B-cells and T-cell activation modules. The module expression profile of these two groups in Romania was analogous to the same groups in South Africa. In both, the modules were was widely up-regulated, particularly type I interferon response and inflammation modules.

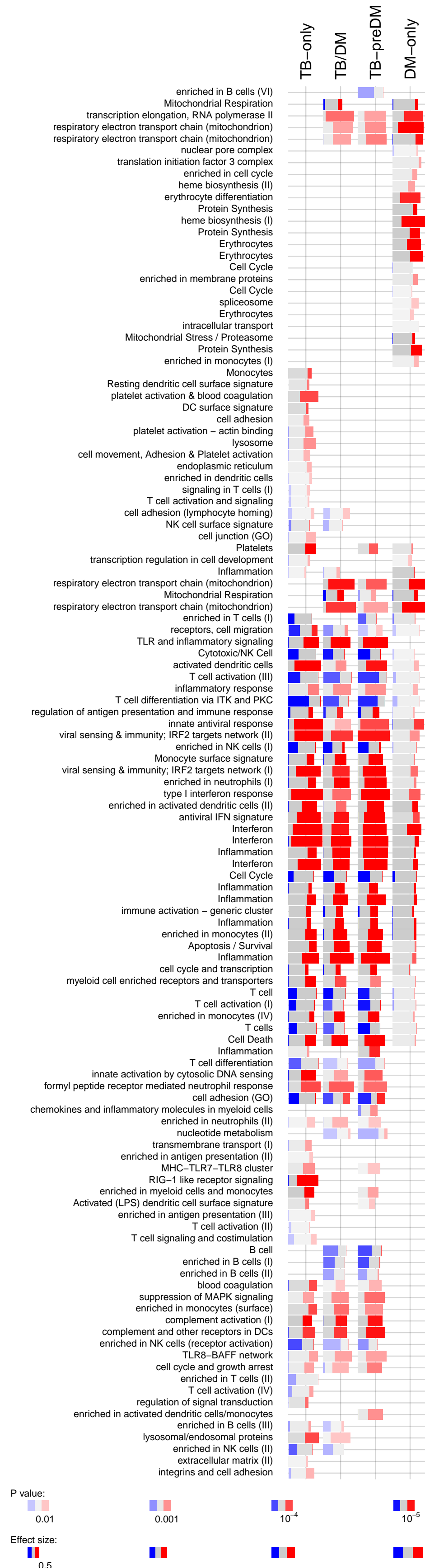


Figure 4.22: **Modular analysis of TB-only, TB/DM, TB preDM and DM-only compared to healthy controls before the initiation of TB treatment in Romania.** Up-regulated (red) and down-regulated (blue) relative to TB-only. The colour saturation is relative to adjusted p-value, and length of bars is relative to effect size. The amount of colour corresponds to the proportion of genes differentially expressed within that module. Modules were deemed significant with an adjusted p-value of < 0.05

4.4.5 Gene expression in TB/DM patients in Indonesia

The Indonesian cohort consisted of 38 samples from only three groups; TB/DM, TB preDM and TB-only (Table 4.8). As no healthy controls were collected in Indonesia, it meant that a different study design was used, in which TB-only was used as the control (Figure 4.4. TB/DM and TB preDM were each compared to TB-only, and volcano plots show the results of this. Because of the few samples in the TB preDM group, 0 genes were statistically significantly differentially expressed. In TB/DM, 24 genes were significantly differentially expressed versus TB-only. All of these except one were up-regulated. The most differentially expressed gene was a ribosomal protein (RLPL13P12) .

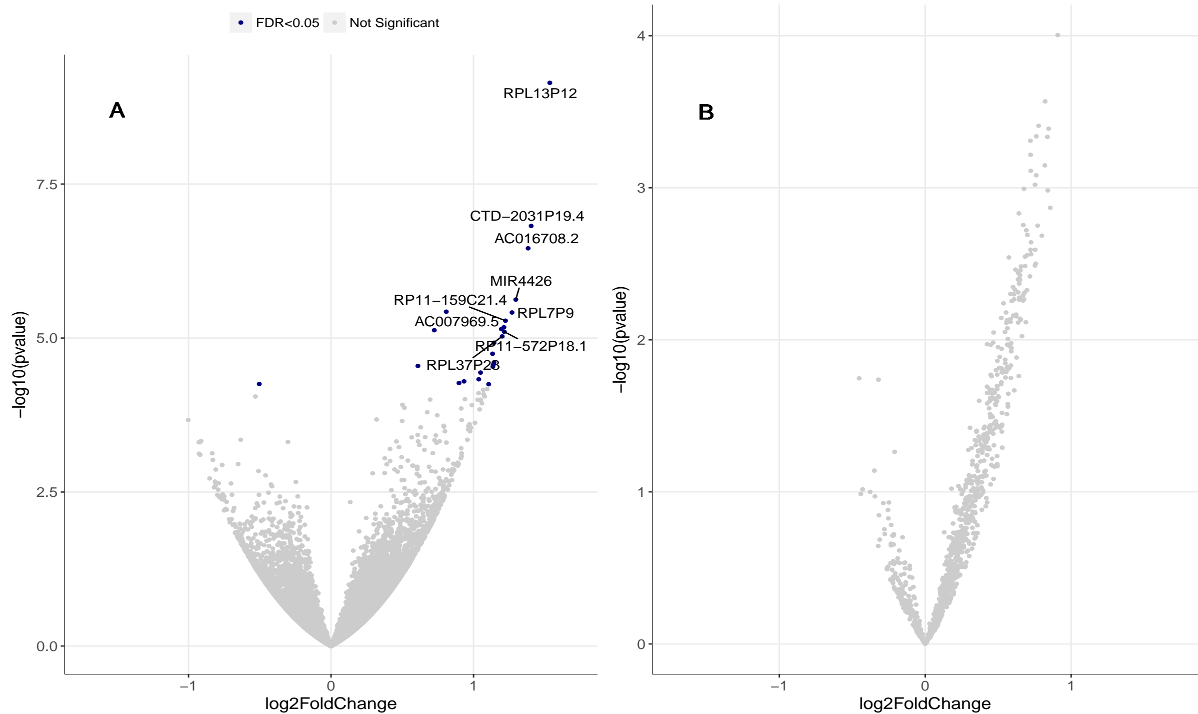


Figure 4.23: **Gene expression profile of TB/DM (A) and TB preDM (B) relative to TB-only patients from Indonesia before the initiation of TB treatment.** Navy blue points represent the genes that were significantly differentially expressed after multiple testing correction (adjusted p-value < 0.05). Grey was not significant. The genes that were most differentially expressed were labelled.

4.4.6 Gene expression in TB/DM patients in Peru

In the Peru population, 32 samples comprised of 3 groups; TB/DM, TB preDM and TB-only (Table 4.8). This cohort followed the same study design as in the Indonesian population, whereby TB-only acted as the denominator.

Volcano plots show TB/DM and TB preDM against TB-only. Statistically significant genes were mostly up-regulated in both instances. In TB/DM versus TB-only, 22 genes were statistically differentially expressed, the most up-regulated being IFI27. IFI27 is an

alpha interferon protein involved in interferon-γ signalling. When comparing TB preDM and TB-only, 39 genes were found to be significantly differentially expressed.

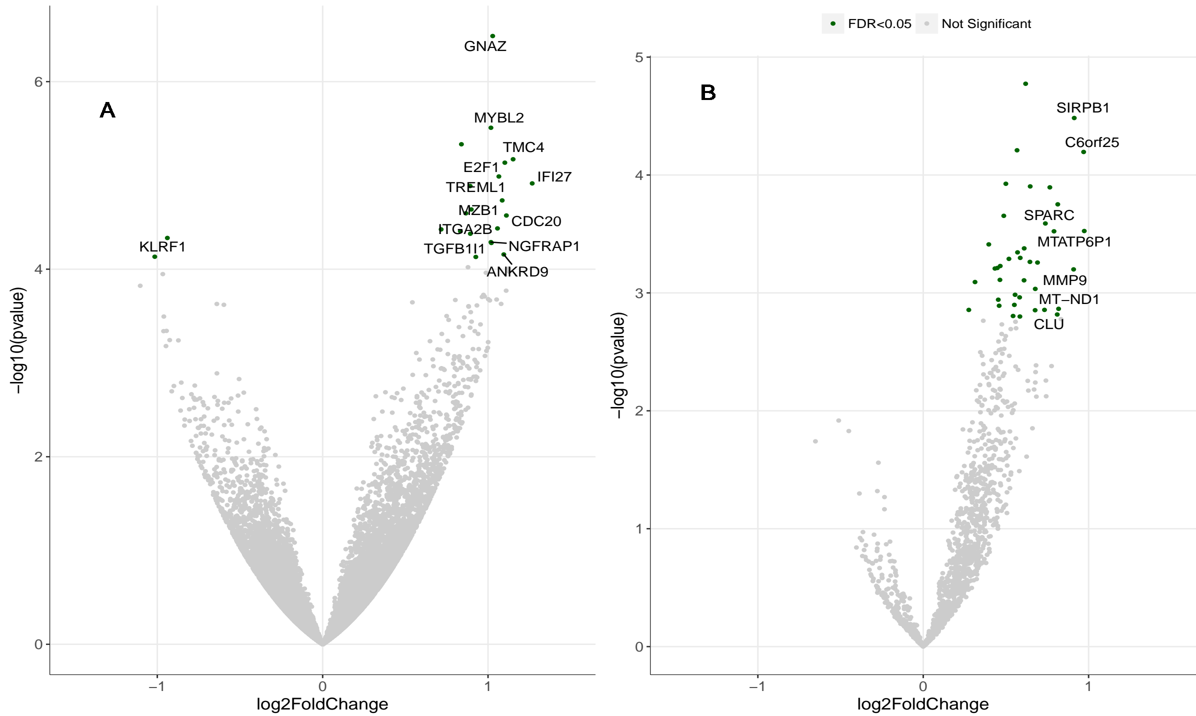


Figure 4.24: **Gene expression profiles of TB/DM (A) and TB preDM (B) compared to TB-only in Peru before the initiation of TB treatment.** Green points represent the statistically significantly differentially expressed genes after multiple testing correction (p-value < 0.05), grey was not significant. The most differentially expressed genes were labelled.

4.4.7 Gene expression in TB/DM in field sites combined

All of the TB/DM, TB preDM and TB-only groups from the four different field sites were combined. This was to increase the sample size of each of the three groups to dissect what impact diabetes has on the gene expression profile of active TB infection (Table 4.10). Because these three groups were the only ones present through the four populations, the TB-only group was used as the control.

Two analyses were run in parallel; with and without a population correction. Adding a population variable in the design can correct for some of the unwanted variation caused by the differences across the populations, thus, leaving the biological variation from the different disease conditions (Equation 4.2).

An initial PCA plot showed no clustering, Figure 4.25. This meant the disease groups from each population were not too different from one another that they could not be combined.

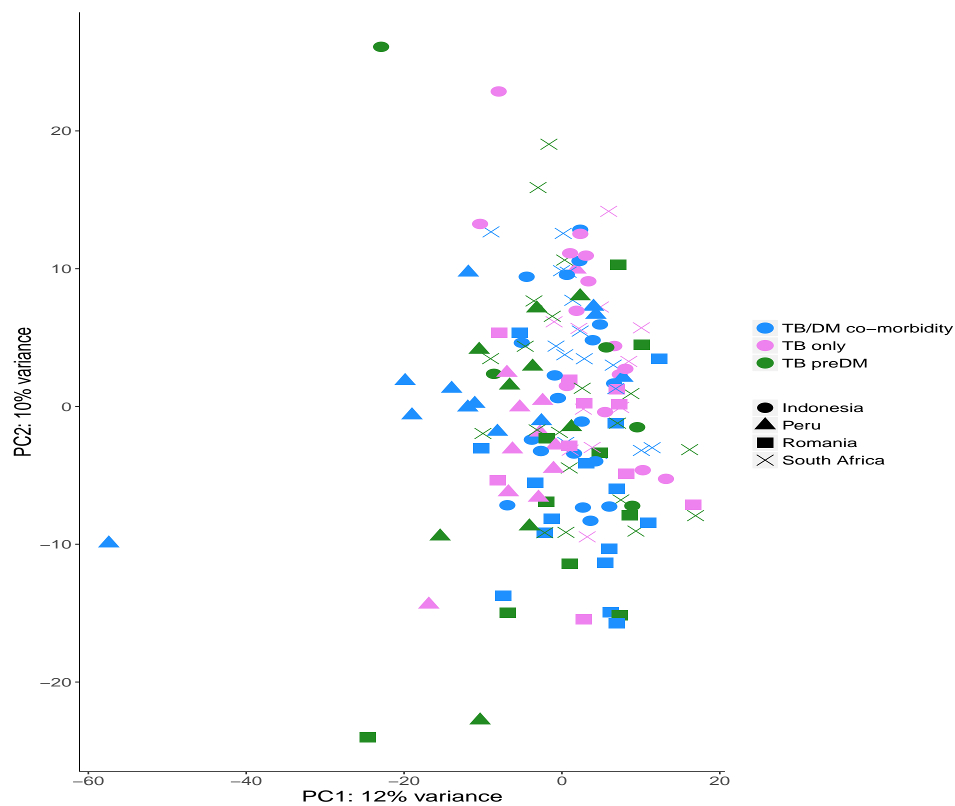


Figure 4.25: **Principal component analysis of all the samples before the initiation of treatment, in all four field sites.**

The two analyses produced slightly different results, in that without a population correction, fewer genes were found to be statistically differentially expressed. This was the case in both TB/DM and TB preDM. As the two designs produced slightly different

results, there was a small effect of population present. The population corrected model was chosen for downstream analysis.

In the population corrected design, differential expression analysis revealed 422 significantly differentially expressed genes in TB/DM against TB-only. In TB preDM, 558 genes were significantly differentially expressed. Figure 4.26 demonstrate that there was a definite effect, but of a relatively small magnitude. The pattern was quite symmetrical, but genes that were up-regulated were of a higher fold change.

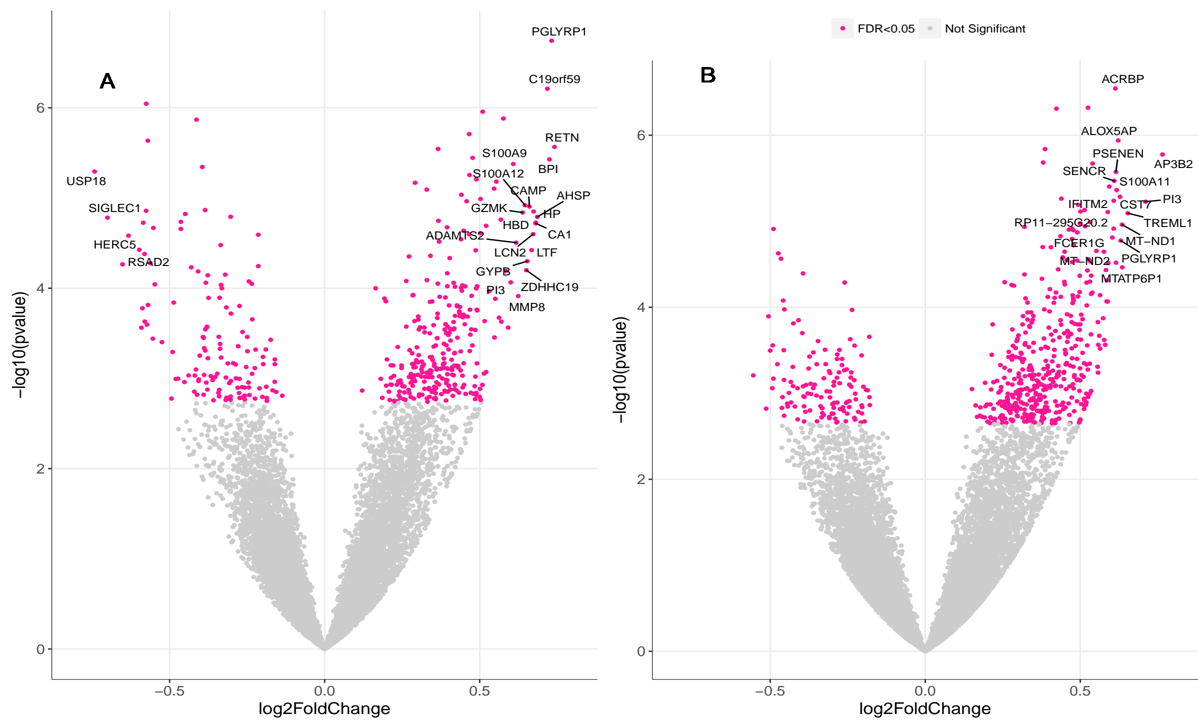


Figure 4.26: **The gene expression profile of TB/DM (A) and TB preDM (B) relative to TB-only when combining all four field sites (Indonesia, South Africa, Romania and Peru) before the initiation of TB treatment.** Genes statistically significantly differentially expressed after multiple testing correction are seen in pink (p-value$<$ 0.05). Grey are not statistically significant. The most differentially expressed genes are labelled.

To detect differences in biological mechanisms between uncomplicated TB-only and the diabetes phenotypes, modular analysis was performed on the population corrected data. A panel plot illustrates the modules that were significantly differentially expressed in the diabetes and pre-diabetes groups when comparing against TB-only.

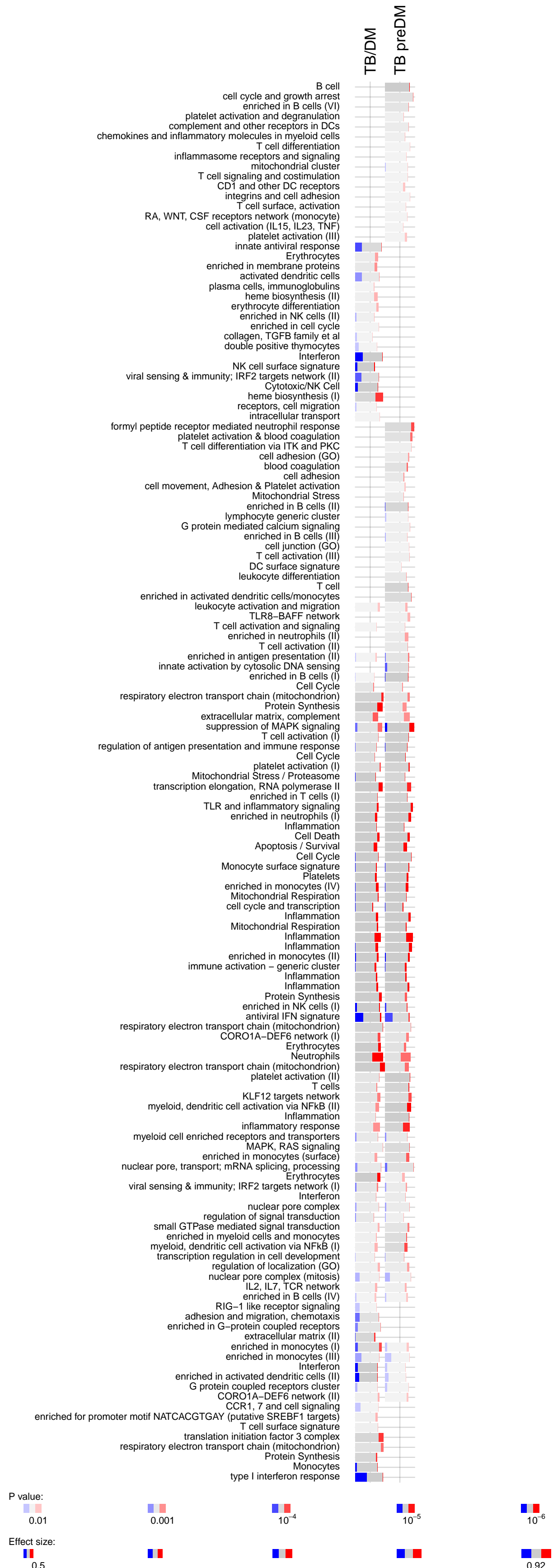


Figure 4.27: **Significantly differentially expressed transcriptional modules between; TB/DM and TB-only, and TB preDM and TB-only before initiation of TB treatment in 4 field sites (Indonesia, South Africa, Peru and Romania).** Up-regulated (red) and down-regulated (blue) relative to TB-only. The colour saturation is relative to adjusted p-value, and length of bars is relative to effect size. The amount of colour corresponds to the proportion of genes differentially expressed within that module. Modules were deemed significant with an adjusted p-value of < 0.05

TB/DM and TB preDM show similar module expression patterns when compared with TB-only. Both show a general trend of up-regulation, specifically of modules involved in inflammation. Interestingly however, patients with TB/DM also exhibit down-regulation of interferon modules compared to TB-only. The most statistically significant modules revealed from modular analysis were inflammation, inflammatory response, neutrophils, interferon, antiviral interferon and type I interferon response. The differential expression of the genes within the top modules were summarised in a heatmap, Figure 4.28. The overriding pattern is that fold change increased from TB preDM to TB/DM. In addition, whilst inflammation and neutrophils were up-regulated in both diabetes groups than in TB-only, type I interferon response was down-regulated. The greatest fold change difference was in the gene Ubiquitin Specific Peptidase 18 (USP18), which is thought to have a role in down-regulating interferon responses. In brief, it appears patients with active TB and clinically diagnosed T2DM have a reduced IFN response than uncomplicated TB. The genes within the interferon response modules that were 2 fold or greater in TB/DM or TB preDM than in TB-only were HERC5, IFIT1, LAMP3, MX1, OTOF, RSAD2 and USP18, Figure 4.29.

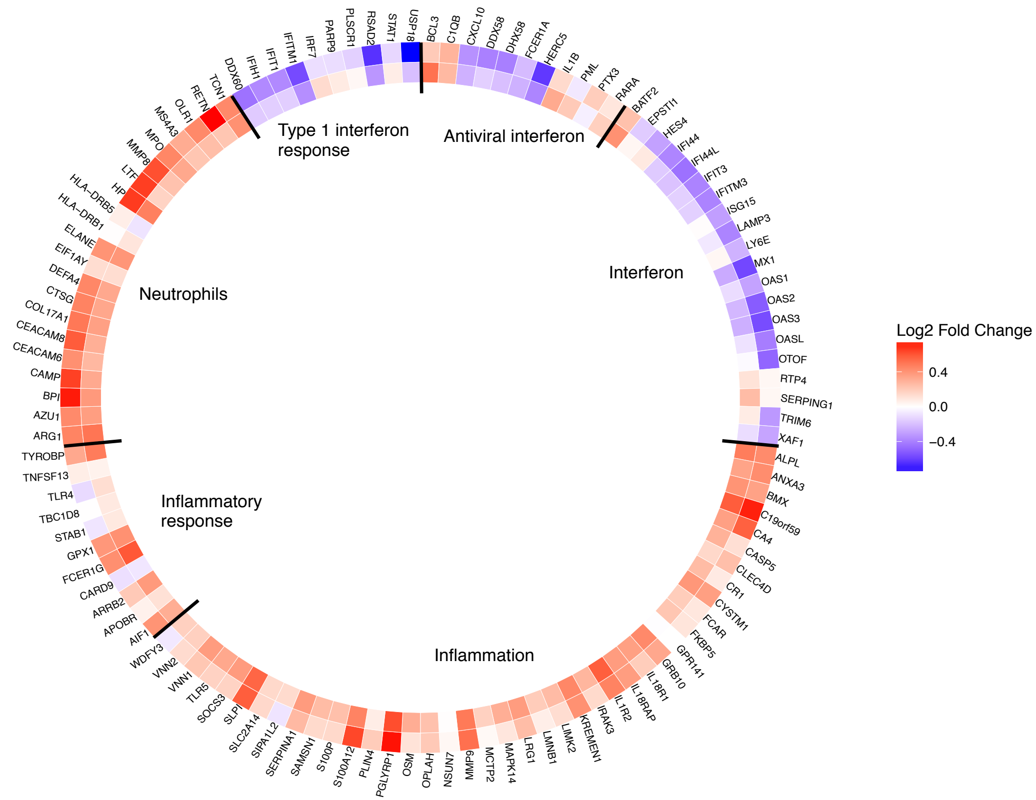


Figure 4.28: **Summary of the modular analysis results when combining all four field sites.** The fold changes of the genes within the top significantly differentially expressed modules are shown (adjusted p-value < 0.05). On the inside: TB preDM compared to TB-only. Outside: TB/DM compared to TB-only. Up-regulated genes are in red, and down-regulated genes are in blue. The saturation of colour represents the magnitude of differential expression.

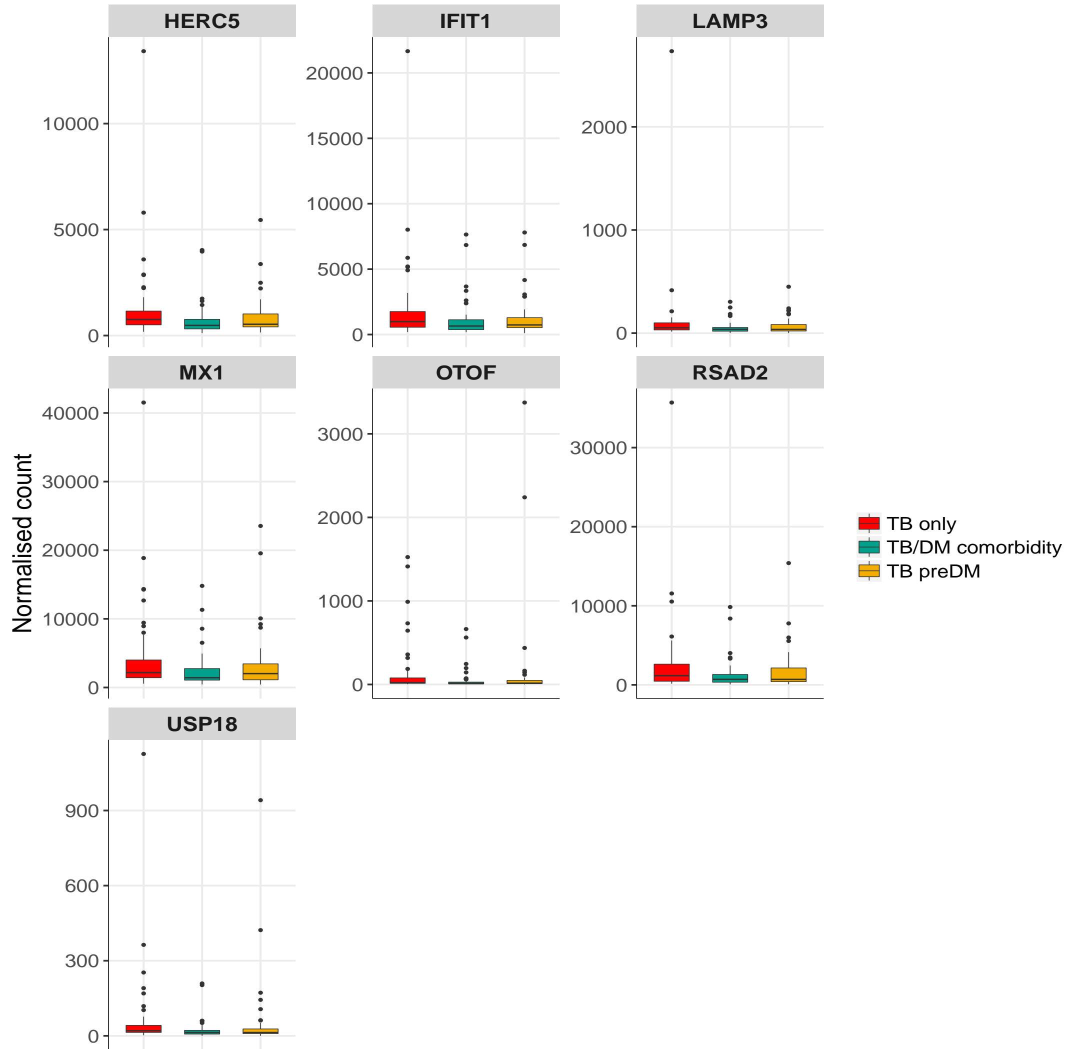


Figure 4.29: **Genes within the interferon response module that were 2 fold or greater in TB/DM or TB preDM than in TB-only.** The four populations were combined and modular analysis revealed interferon response as one of the most significantly differentially expressed in TB/DM or TB preDM and TB-only. Normalised counts for the differentially expressed genes (≥ 2) in each disease category.

An existing signature was tested in the TANDEM data to see whether diabetes comorbidity affects the performance. When plugging in the Zak 16 gene signature to the differential expression results versus healthy controls, TB/DM and TB preDM patients

exhibited a more intense expression than TB-only (Figure 4.30).This also translated to higher absolute values after summing the fold changes of all genes (Figure 4.31). All of the 16 genes were statistically significant after multiple testing correction.

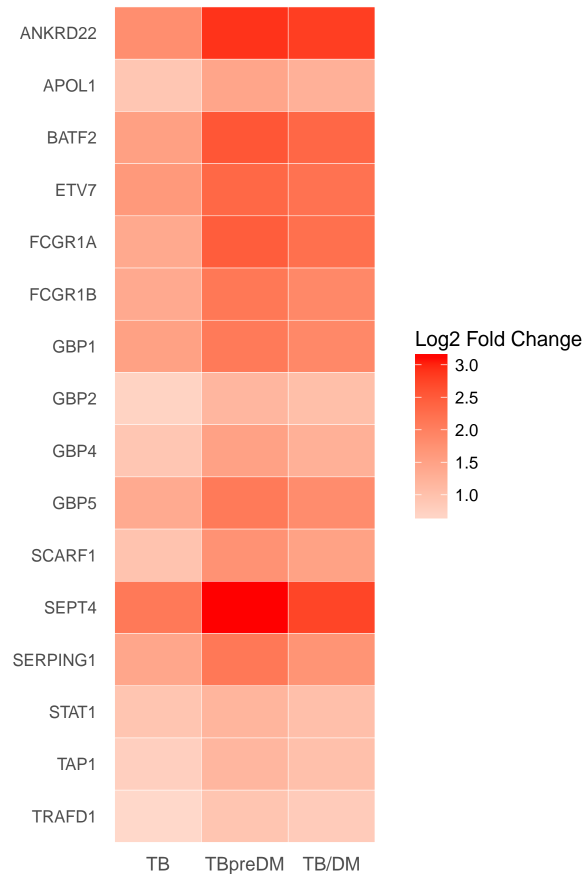


Figure 4.30: **Fold changes of the Zak signature in TB patients relative to healthy controls in South Africa.** Differential expression results of the 16 genes were extracted from TB-only, TB preDM and TB/DM versus healthy controls at diagnosis. Red indicates up-regulation.

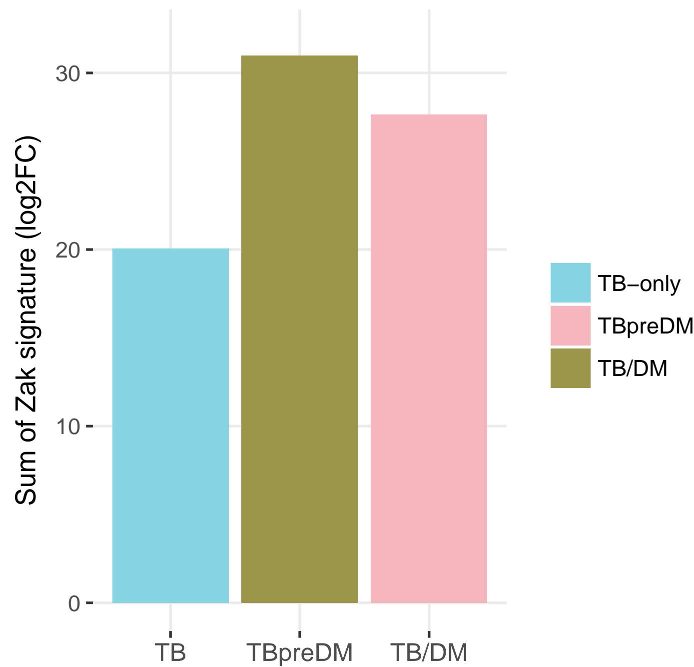


Figure 4.31: **Sum of fold changes of Zak signature.** The fold changes of the 16 genes from the Zak signature were summed in TB-only, TB preDM and TB/DM versus healthy controls.

It was also investigated whether this signature correlated with HbA1c. The sum of the normalised counts for the 16 genes for each patient was taken (Figure 4.32). Signature performance generally increased with HbA1c values in TB-only and TB preDM groups. Performance within the TB/DM group is higher than in TB-only, but with a negative correlation with HbA1c.

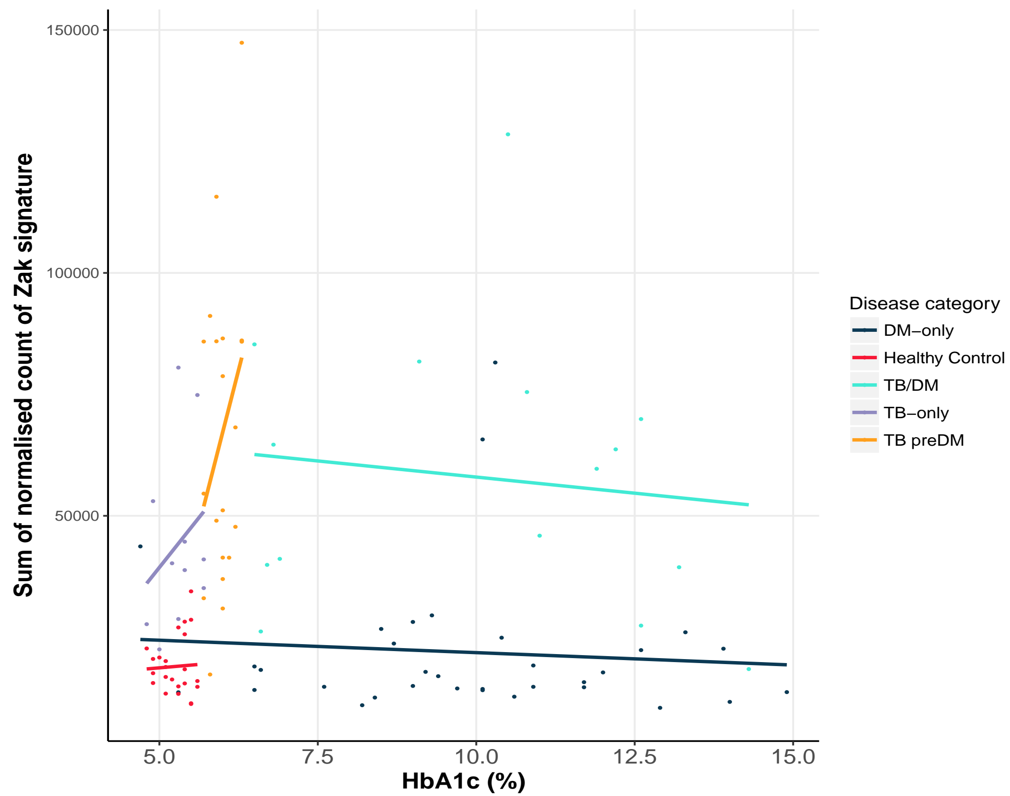


Figure 4.32: **Sum of the normalised count of the Zak signature in each patient group.** The normalised count for the 16 genes were summed for each patient category, and stratified against their HbA1c values.

4.5 Discussion

4.5.1 Findings

This cross-sectional study showed that TB/DM patients can be discriminated from uncomplicated TB patients dependent on their blood transcriptomes at baseline. Also, in contrast to expectations, the TB preDM patients were more similar to TB/DM than to TB-only. The TB-only profile reflected previously literature; a large up-regulation of inflammatory genes and modules. The TB/DM signature was of a larger magnitude than TB-only, coupled with more down-regulation as well as a large up-regulation of genes and modules. The two groups were very different from one another, indicating that diabetes does indeed affect how the immune system responds to TB.

Importantly, the gene signature of TB preDM; was remarkably similar to the TB/DM profile, than for example TB-only. Because TB/DM and TB preDM patients exhibit similar transcriptomic signatures it means that patients with even intermediate levels of hyperglycaemia, exhibit an immune dysfunction which could lead to an increase in susceptibility to active TB disease. This is contrary to current thought, as diabetologists would consider that enhanced risk of infectious disease, specifically TB, occurs at higher HbA1c values, later on in the progression of diabetes. However, these current data show there are already differences in pre-diabetes patients, which may be revealing in terms of susceptibility mechanism.

The comorbidity groups were also differential from patients with diabetes without TB. The gene expression profile of DM-only patients was of a lot smaller magnitude compared to the comorbidity groups. I had hypothesised that maybe the gene signature

of TB/DM would resemble DM-only, with further up-regulation owing to the TB. But, as the results show, this was not the case, the comorbidity causes further differential expression, beyond that of TB-only and DM-only. In addition to this, TB also seemed to modulate the diabetes, as some modules that were differentially expressed in DM-only, were then not differentially expressed in the comorbidity groups. such as the respiratory modules. It appears both diseases affect the way the host responds to each of them.

When looking at the different field sites separately, TB/DM and TB preDM versus TB-only, no genes or a small number of genes were found to be significantly differentially expressed. However, when all four populations were combined, differences were detectable, which were relatively small but consistent, between TB/DM and TB preDM with TB-only. These main differences were that in TB preDM versus TB-only, the inflammatory response was up-regulated, but interferon modules, including the type I response genes, which was one of the top significantly differentially expressed modules, were down-regulated. These differences were even greater in TB/DM versus TB-only, demonstrating that it is a spectrum. This analysis also shows that combining different populations yields more powerful and maybe more meaningful results.

The analysis of a previously published TB signature showed a larger expression in comorbidity patients than in TB-only. It was expected that an existing signature would cancel each other out, due to the down-regulated interferon, and inflammatory up-regulation, or not be as effective in DM patients. But, because the Zak signature is comprised of mainly inflammation genes, it's performance is increased, revealing that it is potentially valid for use in diabetes patients. It prompts the question that if the signature is detecting a sign of disease, is it detected more severe disease? Again, the similarities between

TB preDM and TB/DM were striking; both exhibited similar patterns of expression and absolute absolute values. There was interesting positive correlation between TB signature and HbA1c in both TB-only and TB preDM. This was lost in TB/DM, possibly due to counter-regulatory mechanisms or anti-diabetes medication.

It was speculated that the differences seen between TB/DM and TB-only might be due to the anti-diabetes medication the patients were on, but it cannot be the case for two reasons. Firstly, TB preDM patients have a similar gene and module profile to TB/DM, and the pre-diabetic patients are not on anti-diabetes treatment. Secondly, as discussed in Chapter 3 in the RNA-seq analysis of the effects of metformin in healthy donors, negligible changes were seen on the gene level in the whole blood transcriptome after the administration of metformin. Therefore, it can be concluded that these differences are due to the biological variation between TB/DM and TB-only and not the anti-diabetes medication.

4.5.2 The role of interferons

There are three groups of interferons; type I (IFN-α and IFN-β), type II (IFN-γ) and type III (IFN-ε). The type I interferons all signal through the same receptor: IFNAR1 and IFNAR2, whereas the type II interferon (IFN-γ) has the distinct receptor IFNGR. Interferons, especially the type II interferon IFN-γ are essential in the protection and containment of *M. tb*.

The importance of IFN-γ was uncovered in mouse experiments, when IFNGR$^{-/-}$ knockout mice were infected with *M. tb* and showed decreased survival. When IFN-γ was then infused into the knock out mice to reconstitute the immune response, the

mice had increased survival.[159] In another study, in which the gene for IFN-γ had been disrupted in mice and then infected with *M. tb*, there was widespread necrosis and huge bacterium levels. These data showed that when IFN-γ is lacking, the immune system struggles to contain and control *M. tb* infection.[160] This is also the case in humans, where patients with a mutation in IFNγR1 were totally susceptible to *M .tb* infection and also BCG.[161,162]

Interferons have a protective effect against infections, and are important in the containment and pathogenesis of *M. tb* and other bacteria. Type I IFNs, IFN-α and IFN-β, are particularly important in viral infections, as they activate anti-viral factors. Type I IFNs and their response can vary, from anti-viral activity and the secretion of cytokines to regulate the adaptive arm of the immune response. Type I IFNs mainly boost the immune system, by activating dendritic cells, NK cells and directly acting on $CD4^+$ and $CD8^+$ T-cells to promote differentiation into IFN-γ producing T_H1 cells.[163] Many virus have evolved mechanisms to evade or hinder type I IFNs, and type I IFN-deficient mice are more susceptible to viral infections. But, type I IFNs in bacteria are potentially detrimental to the host, and there has been research into the pro-bacterial effect of them. For example, the effect of type I IFNs has been validated in mouse experiments using $IFNAR^{-/-}$ knockout mice, which blocks the signal of type I IFNs (but not type II) through their receptor, IFNAR. The $IFNAR^{-/-}$ mouse model has been shown to be highly resistant to *Listeria monocytogenes*,[164] and also *M. tb*.

Type I interferons have been shown to be induced following *M. tb* infection, but the evidence of the role of the type I IFNs during infection is somewhat conflicting. It has been shown that type I IFNs can be harmful in the host's response to *M. tb* infection. A

study by Manca et al found that more virulent strains of *M. tb* were associated with the induction of IFN-α and IFN-β. In addition, when mice were administered type I IFN inducers, it resulted in higher bacterial loads in the lungs and exacerbated pathology.[165] To further this, Dorhoi et al investigated the effect of *M. tb* infection in the IFNAR$^{-/-}$ mouse model. They found that the knockout mice had improved mortality and harboured fewer bacteria 21 days post infection than wild type (WT) mice, demonstrating that blocking the signal of type I IFNs improves survival from *M. tb*.[166]

Type I IFNs have long been associated with active TB disease, particularly since Berry et al found a distinct interferon based transcriptional signature, which correlated with disease severity.[67] This signature was primarily neutrophil driven, and included the up-regulation of IFIT1, IFIT2, IFIT3, OAS1, OAS2, OAS3, RSAD2 and MX1 gene expression in blood.[167] Not only this, but type I IFNs have also been associated as a correlate of risk; an up-regulation of IFNs before the onset of infection was correlated with the progression to active disease in non-human primates,[76,131] indicating the IFNs are clearly key in the immune response to TB.

Another study by Mayer-Barber and colleagues investigated the effect of type I interferons in mice and the cross-talk that exists between type I IFNs and IL1. They found that IFNAR1$^{-/-}$ mice had a reduced bacterial burden compared to wild type and in the absence of type I IFN signalling, levels of IL-1β increased. Conversely, IL1R1$^{-/-}$ mice had significantly elevated bacterial load, and much lower survival, but the double knockout IL1R1, IFNAR$^{-/-}$ were significantly less susceptible to *M. tb* infection than IL1R1$^{-/-}$ single deficient mice. Finally, they tested a therapeutic agent, an arachidonic acid metabolism modulator which inhibited the type I IFN response pathway, which re-

sulted in increased survival. They concluded that the balance between type I IFNs and IL-1 is important in the disease presentation.[132]

Conversely, a study by Desvignes et al characterised the dynamics of type I and type II, and found that type I IFNs play a non-redundant protective role, particularly in the early stages of infection. They studied the double knockout mouse model; IFNGR$^{-/-}$, IFNAR$^{-/-}$, where mice did not produce type I or type II IFN responses, and compared it with the IFNGR$^{-/-}$, where only type II responses were defective. They found that the double knockout mice had significantly reduced survival, and higher lung bacillary loads than the sole IFNGR$^{-/-}$ mice. In addition to this, the double knockout had dysfunctional immune cell recruitment to the lungs and significantly different histopathology. The IFNGR$^{-/-}$, IFNAR$^{-/-}$ mice also exhibited more *M. tb* infected macrophages than in the IFNGR$^{-/-}$ mice. All these data show that the combined effects of both type I and type II IFNs contribute to a timely immune response to *M. tb* prior to the onset of adaptive immunity.[168]

Evidence for the requirement of type I IFNs in the containment of *M. tb* was also documented in Kuchtey et al. IFNR1$^{-/-}$ mice were infected with BCG, and their CFUs were measured at points throughout infection. At 14 days after infection, bacterial growth was significantly higher than in WT mice, but at 21 days post-infection, CFUs in both phenotypes were comparable. This study shows that type I IFNs may have a role in the early control of *M. tb* infection.[169]

Research is showing that IFN-α and IFN-β can exert different effects through the same receptor; IFNAR. These disparities are due to subtle differences in receptor binding, signalling cascades and feedback mechanisms.[170] The divergence of IFN-α and IFN-β

is evident in autoimmune disorders like systemic lupus erythematosus (SLE) and multiple sclerosis (MS). IN SLE, IFN-α is pathogenic, but in MS, IFN-β is used as an immunosuppressant treatment with success. MS patients administered IFN-β had reduced clinical relapses, reduced brain disease activity and slower progression to disability, showing the anti-inflammatory and tissue protective role of IFN-β.[171,172] Whereas, IFN-α serum levels are higher in SLE patients than healthy controls and is strongly associated with disease severity.[173] This is again seen in transcriptomic data, where there is a IFN-inducible gene expression signature for SLE that is distinct from healthy controls.[174] It seems that although IFN-α and IFN-β share IFNAR, they interact in a different manner with the receptor and activate the downstream pathways in a different way causing a dichotomy in their immunoregulatory roles.

Type I IFNs became available as a therapy in the 1980s for their immune stimulating and anti-viral properties, for example, in hepatitis B and C infection. Type I IFNs exhibit their antiviral capabilities by suppressing viral replication and inducing apoptosis in virally infected cells. But, type I IFN therapy in hepatitis patients increases the risk of developing T1DM, which is also closely associated with the IFN response gene OAS1.[175]

There have been some studies using type I IFNs in adjunctive therapy with antibiotic treatment in patients with active TB. When standard therapy was supplemented with IFN-α, there was a positive effect on clinical TB outcomes, ranging from improvement in clinical manifestations, and an increase in sputum smear negativity.[176–180] These studies were mainly performed on hard-to-treat-TB, like MDR-TB cases. However, when IFN-α was administered without antibiotic treatment, for example in patients with hepatitis C, there were more cases of latent TB reactivation.[181] This reactivation of TB is reported

after treatment with IFN-α therapy, but not IFN-β, however there is little known about the separate effects of IFN-α and IFN-β in *M. tb* infection. The contradictory effects in different situations reflects the heterogeneity of the type I IFN family and its downstream targets.

In the literature, there are differences between the type I IFN cytokines action, as described above, but these data and also previous TB signatures focus on the downstream inducible genes. As discussed, it was found that type I interferon response genes were down-regulated in TB/DM compared to TB-only, showing that the response of the interferons in response to *M. tb* infection is different in diabetes. Considering the nuances in the interferon responses, it is difficult to completely unravel the implications of the aberrant type I IFN response and the elevated inflammatory response.

It could be hypothesised that the TB/DM patients have an enhanced clinical phenotype, characterised by the increased inflammatory response and aberrant type I IFN response. But, in clinical studies between TB/DM and TB-only patients, there was no significant difference in the presentation or pathology of TB disease, except in one study, where TB/DM patients listed more symptoms.[137] Other work in the TANDEM consortium, yet uncompleted, involves performing unbiased computer-aided reading of digital X-rays, to further determine whether there are any chest pathology differences. So far however, there is little evidence suggesting the TB/DM have a heightened clinical TB phenotype, suggesting the altered transcriptome is a mechanism of enhanced susceptibility.

4.5.3 Limitations and future work

Of course, it would have been favourable to collect samples from healthy controls in all the field sites in order to dissect the how each disease is different in more populations. But, field sites had decided to join the RNA-seq analysis after the sample collection had been initiated, so further collection of healthy controls was not available. So, this meant the study designs were not consistent across populations which was not ideal. It would have been desirable to dissect the differences of all the disease groups between their healthy controls to see how each disease is different in each population.

Somewhat of a caveat of modular analysis is in the module classification. Sometimes the genes within a gene-set or module may not truly reflect the biological association. Module or gene-set classification can vary depending on who compiled them and by what method. A study by Pollara et al evaluated the sensitivity and specificity of particular cell-associated modules by looking at the overlapping content across different data sets. They found that different modules of the same cell type vary somewhat, and came up with a modular discrimination index (MDI) score as a predictor of module performance.[182] It would be an idea to use this tool in further analysis to validate the modular findings in these data.

4.5.4 Implications in TB control

Indeed TB/DM and TB preDM patients are an increased risk of disease development and of poor TB treatment-outcomes, including death, failure and relapse, meaning there are implications in TB control. Our data could suggest that the threshold for diabetes is lowered, as an immune dysfunction clearly exists at intermediate levels of hyperglycaemia.

Diabetes and pre-diabetes patients could be more closely followed, and undergo prophylactic TB treatment if they were diagnosed with latent *M. tb* infection. The practicalities of this may be challenging as high TB/DM burden countries are also low income, and such a control strategy is more likely to be feasible in high income, low burden countries. Diabetes clearly shapes the immune response in TB disease, so is likely to impact vaccines that induce such a response in the host. This would cause implications if diabetes patients do not respond accordingly to the vaccine, either causing an over- or under-stimulation, meaning the vaccine would not work effectively. This needs to be considered in vaccine development. This data also bring about the possibility of using host directed therapy in TB disease, where the drug targets the host, rather than the bacteria themselves. This would open new avenues of research as the immune abnormalities caused by TB/DM could be modulated by a drug which then causes the host to respond appropriately.

In the main, we have found that diabetes does effect the TB gene expression profile, and that this effect is also seen in pre-diabetes. If this transcriptomic change is indeed evidence of a mechanism of TB susceptibility, then it has real implications on efforts of TB control as pre-diabetes patients are often overlooked.

5

Tuberculosis and diabetes in TANDEM: longitudinal study

5.1 Introduction

5.1.1 Diabetes effects TB treatment response

Not only has T2DM been recorded as a risk factor for active TB disease, but also causes a greater severity of disease, and has a negative effect on treatment response. This negative effect on treatment often transpires as treatment failure, slower rates of culture conversion, higher rates of relapse and also death.[183]

There have been a few studies investigating treatment outcomes in TB patients with diabetes. In a study in Egypt, it was found diabetes conferred a 3.9 times higher risk of treatment failure. This was in patients receiving directly observed therapy so was not due to poor adherence.[184] Another study in Indonesia by Alisjahbana et al also concluded that diabetes increases the risk of TB treatment failure. In more detail, they found that

diabetes patients were at 7.65 times higher risk of being sputum culture positive after six months of TB treatment. This same study also showed that a higher proportion of diabetic patients were culture positive after two months and also six months of TB treatment than in TB patients without diabetes.[137] Results from other studies of sputum culture conversion at two months are mixed, with most reporting no difference in TB/DM and TB without diabetes. Yoon et al however reported that only uncontrolled diabetes patients showed a higher instance of sputum culture positivity, and controlled diabetes patients did not show any difference compared to non-diabetics.[185] The overarching theme however was that the mean time to culture conversion tended to be longer in TB/DM than uncomplicated TB patients.[183,186,187] Generally, it seems poor glycaemic control is strongly associated with poor TB treatment outcomes. In a prospective study of TB/DM patients, poor glycaemic control (HbA1c $\geq$ 7%) in particular was associated with higher rates of treatment failure. Of the patients with poor glycaemic control, 23% failed treatment, compared to only 4% of patients with optimal glycaemic control. These patients were also at 2.83 times higher risk of relapse.[188]

Not only are diabetes patients more likely to fail TB treatment, but they also have a higher risk of death. Retrospective cohort studies of patients with pulmonary TB have shown that diabetes patients have 6.5 increased risk of death than non-diabetics.[183,189] However, cause of death is not always recorded, so it is difficult to deduce whether it is actually down to the comorbidity. Regardless, all of the studies in TB treatment outcomes in diabetes patients point to an increased treatment failure and increased occurrence of death. It raises the question as to whether diabetes patients' immune system respond differently during treatment compared to non-diabetics.

As discussed previously, TB/DM patients exhibit altered immune system and cytokine profile compared to TB patients without diabetes. This dysfunction includes monocytes and neutrophils, and an altered expression of Th1 and Th17 cytokines, which are known to influence the pathogenesis of pulmonary TB. A couple of studies by Kumar et al have looked into the immunological changes of diabetes patients throughout TB treatment. Firstly, they found that the ex vivo percentage of dendritic cell and monocyte subsets was lower before TB treatment, and after two months of treatment. But, six months after initiation of treatment, the dendritic and monocytes subsets are increased in TB/DM patients than in TB patients without diabetes. Secondly, they found TB/DM patients had significantly increased levels of effector memory CD8 and CD4 T-cell subsets two months post treatment. At six months post initiation of TB treatment, TB/DM patients had increased levels in naive CD4 and CD8 T-cells. This was coupled with a decrease in the percentage of central memory CD4 and CD8 T-cells.[190,191] These cell types are important in the protective immunity against TB, and a dysregulation could be why diabetes patients are more likely to fail TB treatment.

5.1.2 The TANDEM consortium

The TANDEM consortium also looked at treatment response of TB/DM patients compared to uncomplicated TB, because diabetes is associated with increased TB treatment failure, death and relapse. In another work package of the consortium, a randomised control trial was performed in which TB/DM patients were randomly allocated to receive standard DM care or enhanced monitoring/adjustment of DM treatment through TB treatment, to determine whether the poor treatment outcome could be reduced.

In addition to this, we wanted to characterise TB treatment response of TB/DM patients using bioprofiling, which led to the longitudinal study. This study investigated the differences in gene expression in TB/DM, TB pre DM and TB-only patients over their treatment in different populations (Romania, Indonesia and South Africa).

The studies looking at the transcriptomic changes throughout TB treatment are comparatively few compared to cross-sectional studies. Overall, any blood transcriptomic signature present at diagnosis of TB, had diminished by two months of treatment, and disappeared by twelve months after diagnosis. This paralleled with clinical improvement, measured by chest X-ray. This signature included a type I interferon response, in which the same was also found in neutrophil-depleted samples.[73] Interestingly however, dynamic changes in whole blood are seen throughout treatment, particularly within the first one to two weeks, where there is a large down-regulation of differentially expressed genes.[68,72] This down-regulation included complement genes like C1QB and other inflammatory genes.

We wanted to see whether the same dynamic changes are seen in TB/DM patients, or are there fundamental differences. As large scale changes are seen in gene expression through treatment which correlate with TB cure, perhaps the resolution of the TB transcriptomic signature is delayed or altered by diabetes, which would correspond to poor treatment outcomes. As seen in Chapter 4, there are differences between TB and TB/DM at diagnosis, so it is likely that differences also exist at the end of treatment, if the transcriptomes are still different at six months, it may indicate why TB/DM patients are at enhanced risk of relapse, whilst also showing treatment failure. If this is indeed the case, it will help in decision making regarding treatment for TB/DM patients.

Considering diabetes comorbidity affects TB treatment outcomes in the form of increased treatment failure and relapse, this would be reflected in altered blood transcriptomes in TB/DM patients throughout and at the end of treatment compared to TB-only patients. I hypothesised that preDM patients would behave similarly to TB-only patients. I also hypothesised that when analysing a blood TB signature, TB/DM patients would have altered resolution at the end of treatment. Conversely, preDM patients would be comparable to TB-only patients, as a worse TB signature would correlate with progression of DM.

5.2 Aims and Objectives

Aims

I **To determine whether the gene expression profile through TB treatment is different in patients with TB/DM or TB preDM.**

Objectives

I **Perform RNA-seq analysis on whole blood samples from three different sites at four time points throughout TB treatment in patients with TB/DM, TB preDM and TB-only.**

II **Compare the gene expression profiles through treatment of TB/DM and TB preDM each with TB in one closely defined population, namely Indonesia.**

5.3 Methods

Whole blood samples for the longitudinal analysis were collected from uncomplicated TB patients, TB preDM and TB/DM patients from Indonesia, Romania and South Africa. Unlike in the TANDEM cross-sectional study, Peru was not included in the longitudinal study. Samples from these three field sites were collected at four time points; diagnosis (before the initiation of treatment), and at two weeks, two months and six months post treatment. These three subsequent time points were chosen because previous literature has shown dramatic changes within the first two weeks after initiation of TB treatment; the month two time point is most TB patients have become should be sputum smear or culture negative, and month six is at the end of TB treatment, when patients should have reverted to 'healthy' status. Table 5.1 shows the number of samples within each patient group, at each time point, in each field site.

		Diagnosis	Week 2	Month 2	Month 6
South Africa	TB/DM	15	15	16	12
	TB preDM	11	10	10	10
	TB-only	6	4	6	5
Indonesia	TB/DM	17	12	16	15
	TB preDM	4	4	3	5
	TB-only	12	10	10	11
Romania	TB/DM	8	10	9	8
	TB preDM	5	4	4	2
	TB-only	5	5	5	5
total	TB/DM	40	37	41	35
	TB preDM	20	18	17	17
	TB-only	23	19	21	21

Table 5.1: **Number of samples in the longitudinal study.** Samples of each patient group, at each time point, in the three field sites and in total.

5.3.1 Sample processing and data analysis

Samples and data were processed with the rest of the TANDEM samples as described in the previous chapter. Raw data were aligned and quantified with the cross-sectional samples, as a whole, with the same tools and settings as discussed in the previous chapter.

Differential expression analysis was again performed using R (R v3.4.1 and RStudio v1.0.143) package `DESeq2` (v 1.16.1). Firstly, all the samples (Table 5.1) were normalised together in a global analysis which included all three of field sites; South Africa, Romania and Indonesia and population was corrected for. In this analysis, TB-only was used as a control, with TB/DM and TB preDM compared against them. This was to determine whether TB patients with diabetes or pre-diabetes share the same transcriptomic dynamic

throughout treatment as uncomplicated TB patients. This was also because healthy controls were not collected in Indonesia.

To find the genes that were differentially expressed through time and across phenotype, the likelihood ratio test was performed. This test allows you to test the difference between two model curves, and measure 'the fit' of the two models. In other words, it creates a model curve through the time-points for each phenotype group, in which the differences are the output. It pulls out genes that are differentially expressed at any time-point but also differentially expressed across phenotype. So, the results yielded are the genes that change through TB treatment but must also be different with and without diabetes. This likelihood ratio method was done in TB/DM and TB preDM compared to TB-only. Multiple testing was corrected using the Benjamini & Hochberg false discovery rate method.

Another analysis was done; a more directed approach. This analysis involved segregating the samples by field site. In Indonesia, only TB-only and TB/DM patients were investigated because there were too few patients with TB preDM. TB-only patients were again used as the control, to compare against TB/DM. Each time-point of TB-only was compared to the same time point in TB/DM, i.e. diagnosis versus diagnosis, week 2 versus week 2 and so on. This was to dissect what was different at these crucial time points through anti TB treatment when patients have T2DM co morbidity.

Modular analysis was done to add biological knowledge, and performed using the R package `tmod` (v0.31), with the modules outlined by Chaussabel et al and Li et al.[94,95] An adjusted p-value of < 0.05 was deemed significant. From the modular analysis, module activity can be calculated for the modules that were significantly differentially expressed

in TB/DM compared to TB-only at any time-point. Equation 5.1 shows the formula to calculate the module activity, where g is the expression value of each gene within a module, and n is the number of genes within that module.

$$module\ activity = \frac{\sum g}{n} \tag{5.1}$$

A published signature was also analysed to test the hypothesis of whether DM and preDM patients resolve their transcriptome after completion of treatment as in uncomplicated TB. The Zak et al signature[76] was chosen, as in Chapter 4. TB-only patients at month 6 were compared with TB/DM and TB preDM patients as these patients would be deemed 'healthy'.

5.4 Results

5.4.1 Diabetes and tuberculosis treatment outcomes

The laboratory HbA1c values were taken in active TB patients, with and without diabetes, before initiation of TB treatment, and after completion of treatment (Figure 5.1). The general pattern in all of the three field sites and across all patient groups was that HbA1c values decreased after six months of TB treatment. This was even the case in TB patients without diabetes. The most dramatic differences in HbA1c values were seen in TB/DM patients, in all of the field sites, particularly in Indonesia. There were some exceptions however, especially in South Africa, where HbA1c values in TB/DM patients were increased at month six. This could be due to poor management of glucose control.

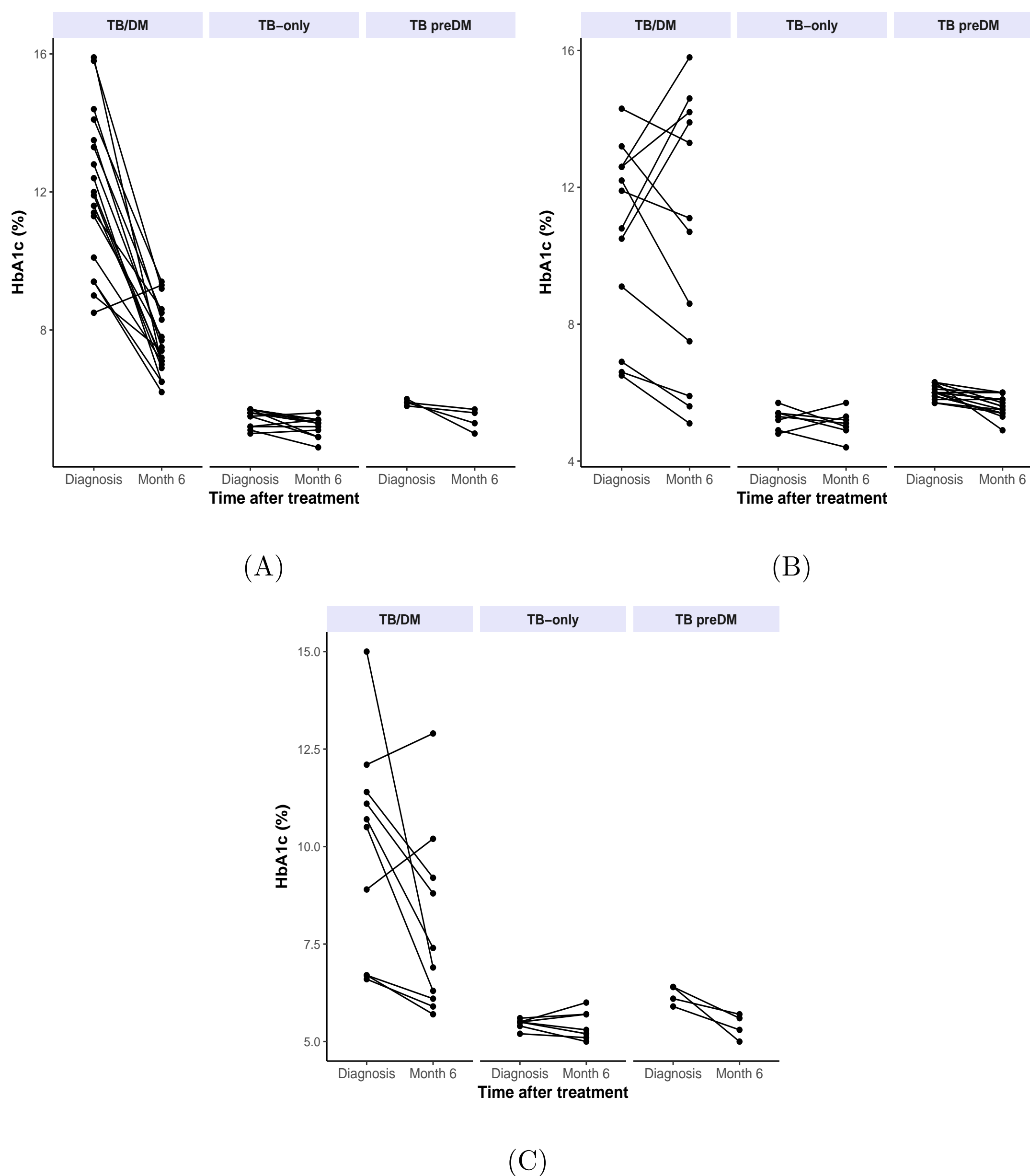


Figure 5.1: **HbA1c values in TB patients with and without diabetes or pre-diabetes before initiation of treatment, and after completion.** A: Indonesia, B: South Africa, C: Romania

TB Treatment outcomes were summarised as good or poor. Poor outcomes included treatment failure, death (due to TB) or relapse. Good outcomes included cure or treatment completion. The relapse outcomes in South Africa were assessed by whether patients

had clinically relapsed and were on treatment again alongside microbiological results. In Indonesia, relapse was defined as only culture positive after completion of treatment. Figure 5.2 shows the proportion of good and poor treatment outcomes in active TB patients with and without diabetes in the three field sites. Poor treatment outcomes in TB/DM were consistently around 27% across field sites (Table 5.2). This was higher than TB-only or TB preDM in Romania and South Africa. However, in Indonesia TB-only and TB preDM patients both had a large proportion of poor treatment outcomes, which might be due to the different definitions of poor treatment outcome.

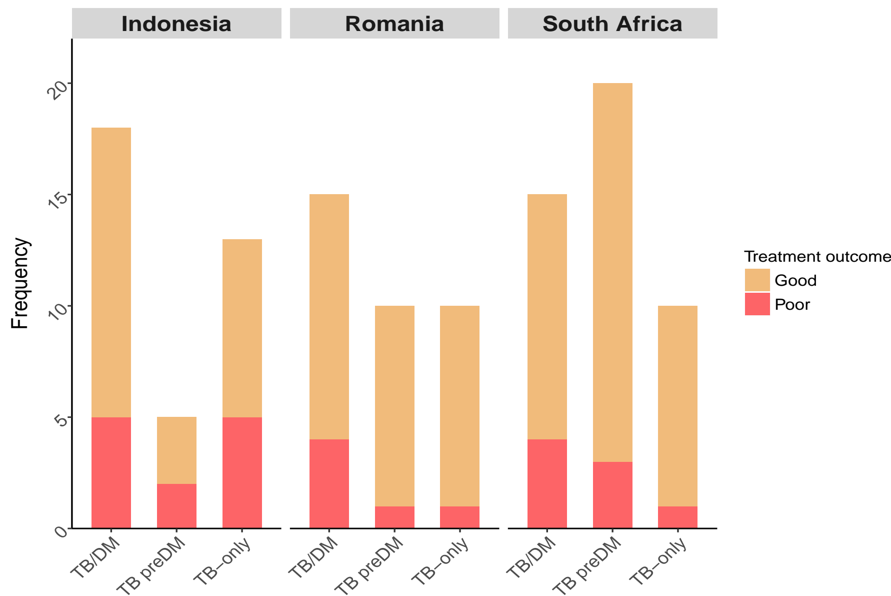


Figure 5.2: **Treatment outcomes of TB patients with and without diabetes or prediabetes from the three field sites.** Poor outcomes in South Africa were defined as patients that relapsed and had gone back onto treatment. In Indonesia, poor outcomes were based on culture results.

	South Africa	Romania	Indonesia
TB/DM	26.7%	26.7%	27.8%
TB-only	10%	10%	38.5%
TB preDM	15%	10%	40%

Table 5.2: **Percentage of poor treatment outcomes in each of the patient groups in the 3 field sites.**

5.4.2 Global analysis of TB/DM patients through TB treatment

All three field sites were combined in a global analysis, to look at the effect of TB treatment (time) in the different patient groups (phenotype). TB/DM, TB preDM and TB-only patient groups were normalised together. Using the likelihood ratio test, 8258 genes were found to be differentially expressed over time and across phenotype.

Modular analysis was performed, and the top significantly differentially expressed modules are shown in Table 5.3. These are the modules that were the most differential throughout TB treatment and across the patient groups. The most differentially expressed modules were related to B-cells and T-cells. Figures 5.3 and 5.4 show the dynamics of expression of the genes within the B-cell module in each patient group. TB/DM and TB preDM both start with a higher normalised count than in TB-only. The normalised count then decreased in all the patient groups at week two, especially in TB preDM, then increases at month six.

Module name	AUC	Adjusted p-value
Enriched in B cells (I)	0.919	$5.29e^{-28}$
Mitochondrial Respiration	0.731	$8.67e^{-19}$
B cell	0.919	$3.48e^{-18}$
Immune activation - generic cluster	0.665	$1.32e^{-16}$
Enriched in B cells (VI)	0.961	$3.57e^{-15}$
Enriched in B cells (II)	0.899	$1.04e^{-14}$
T cell	0.758	$4.09e^{-12}$
Mitochondrial Respiration	0.705	$8.14e^{-11}$
Plasma cells and B cells, immunoglobulins	0.808	$2.09e^{-09}$
B cell surface signature	0.607	$1.11e^{-08}$
T cell activation (I)	0.771	$1.69e^{-08}$
Enriched in T cells (I)	0.742	$4.20e^{-07}$
Enriched in monocytes (II)	0.647	$4.56e^{-07}$
Protein Synthesis	0.781	$7.33e^{-07}$
Inflammation	0.653	$2.53e^{-06}$
T cell surface signature	0.736	$9.11e^{-06}$
Respiratory electron transport chain (mitochondrion)	0.884	$1.17e^{-05}$
Receptors, cell migration	0.779	$3.26e^{-05}$
Enriched in B cells (III)	0.777	$4.35e^{-05}$
Enriched in naive and memory B cells	0.821	$6.32e^{-05}$
Apoptosis / Survival	0.705	$6.33e^{-05}$
Enriched in activated dendritic cells (II)	0.739	$7.71e^{-05}$
T cells	0.725	$1.09e^{-04}$
Regulation of antigen presentation and immune response	0.677	$1.09e^{-04}$
Interferon	0.856	$1.12e^{-04}$
Inflammation	0.709	$1.12e^{-04}$

Table 5.3: **Modular analysis from the results of the likelihood ratio test in TB patients with and without diabetes or pre-diabetes through TB treatment from all three field sites.** The top most significantly differentially expressed modules through time and across phenotype. AUC is the area under the curve, a measure of the effect size. P-values were adjusted using Benjamini-Hochberg correction.

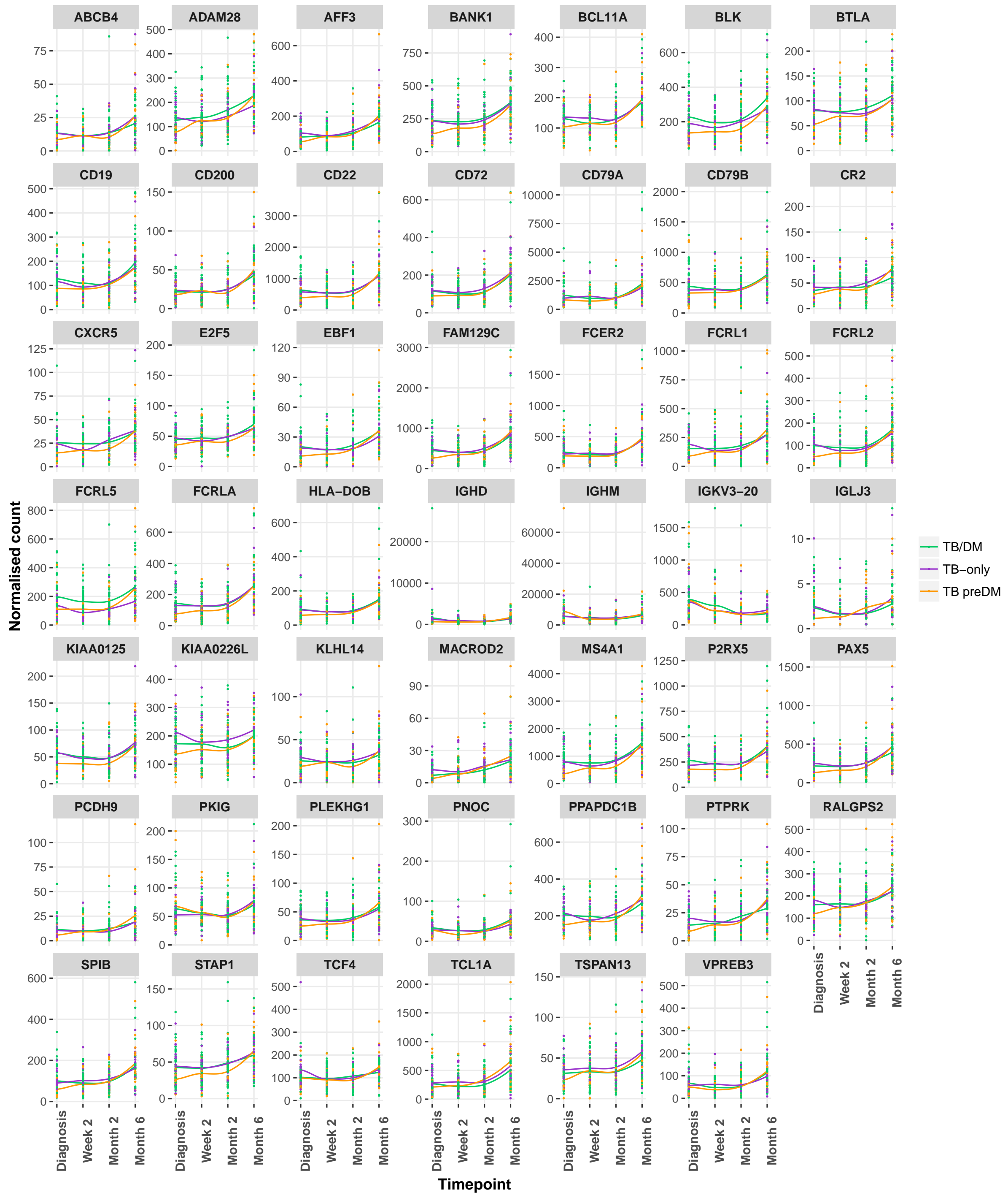


Figure 5.3: **Normalised count of genes of the most significantly differentially expressed module; B-cell, at each time point through TB treatment.** The different colours represent the three patient groups. A Loess line of best fit was plotted for each group. Data were taken from all three field sites and population corrected.

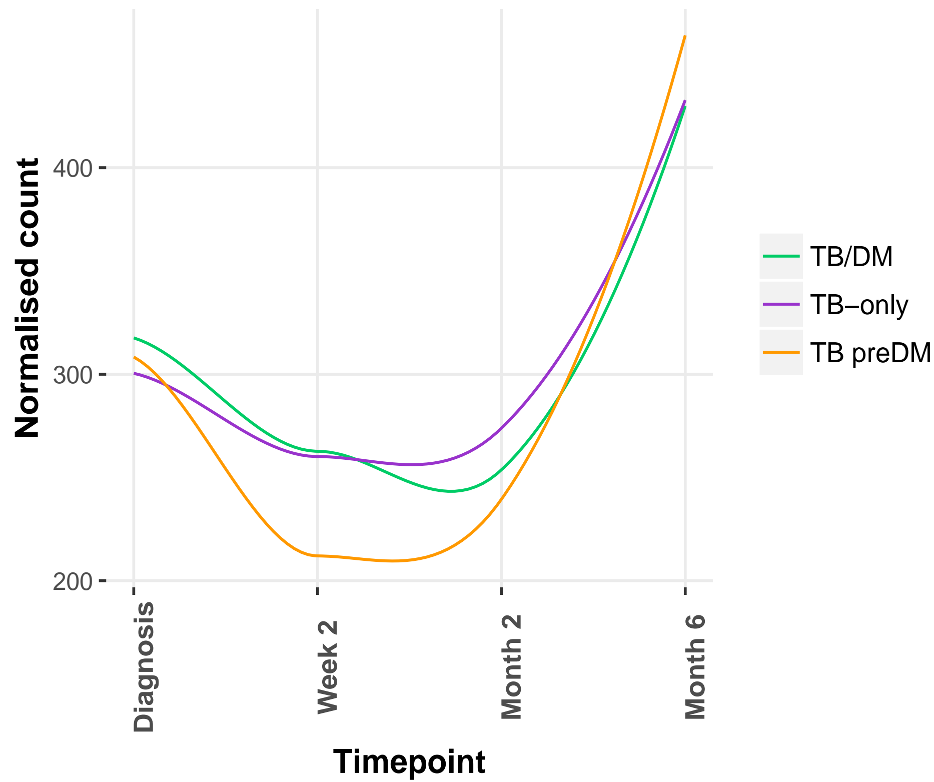


Figure 5.4: **The normalised count for all the genes from the B-cell module at each time point through TB treatment.** A loess line of best fit was fitted through all normalised counts of all the genes. Data were taken from all three field sites and population corrected

In the T-cell module, the general pattern is that the normalised count in all the patient groups increases through TB treatment. TB/DM patients have higher normalised count of each of the genes throughout TB treatment than the other two groups (Figures 5.5 and 5.6).

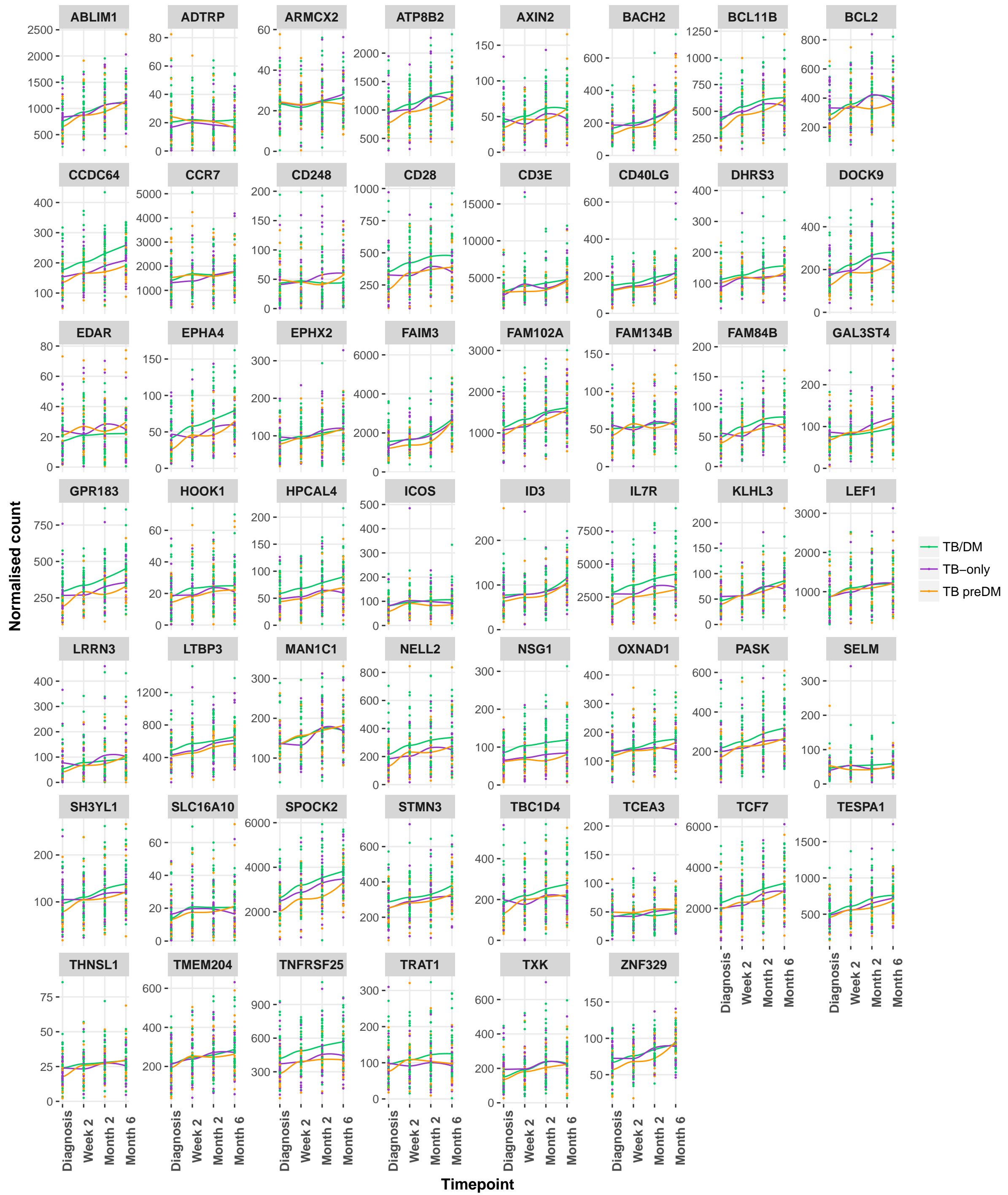


Figure 5.5: **Normalised count of genes in the significantly differentially expressed module; T-cell, at each time point through TB treatment.** The different colours represent the three patient groups. A Loess line of best fit was plotted for each group. Data were taken from all three field sites and population corrected.

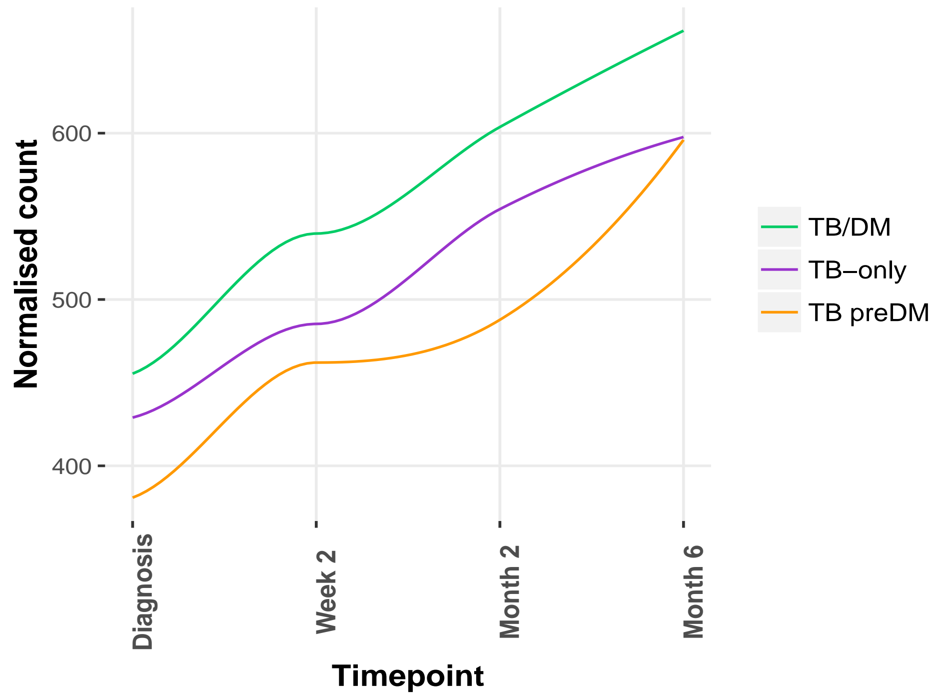


Figure 5.6: **The normalised count for all the genes from the T-cell module at each time point through TB treatment.** A loess line of best fit was fitted through all normalised counts of all the genes. Data were taken from all three field sites and population corrected

Figures 5.7 and 5.8 show the normalised counts of the genes within the interferon module. Interferon modules have consistently been significantly differentially expressed in TB/DM in the TANDEM study. The interferon genes have a similar pattern to one another; they change dramatically though time. Many are down-regulated at week two compared to diagnosis. The interferon genes in TB/DM and TB preDM behave similarly, but together behave opposite to TB-only patients at both month two and month six. The adjusted p-values from the gene-level differential analysis for each of these genes is shown in Table 5.4.

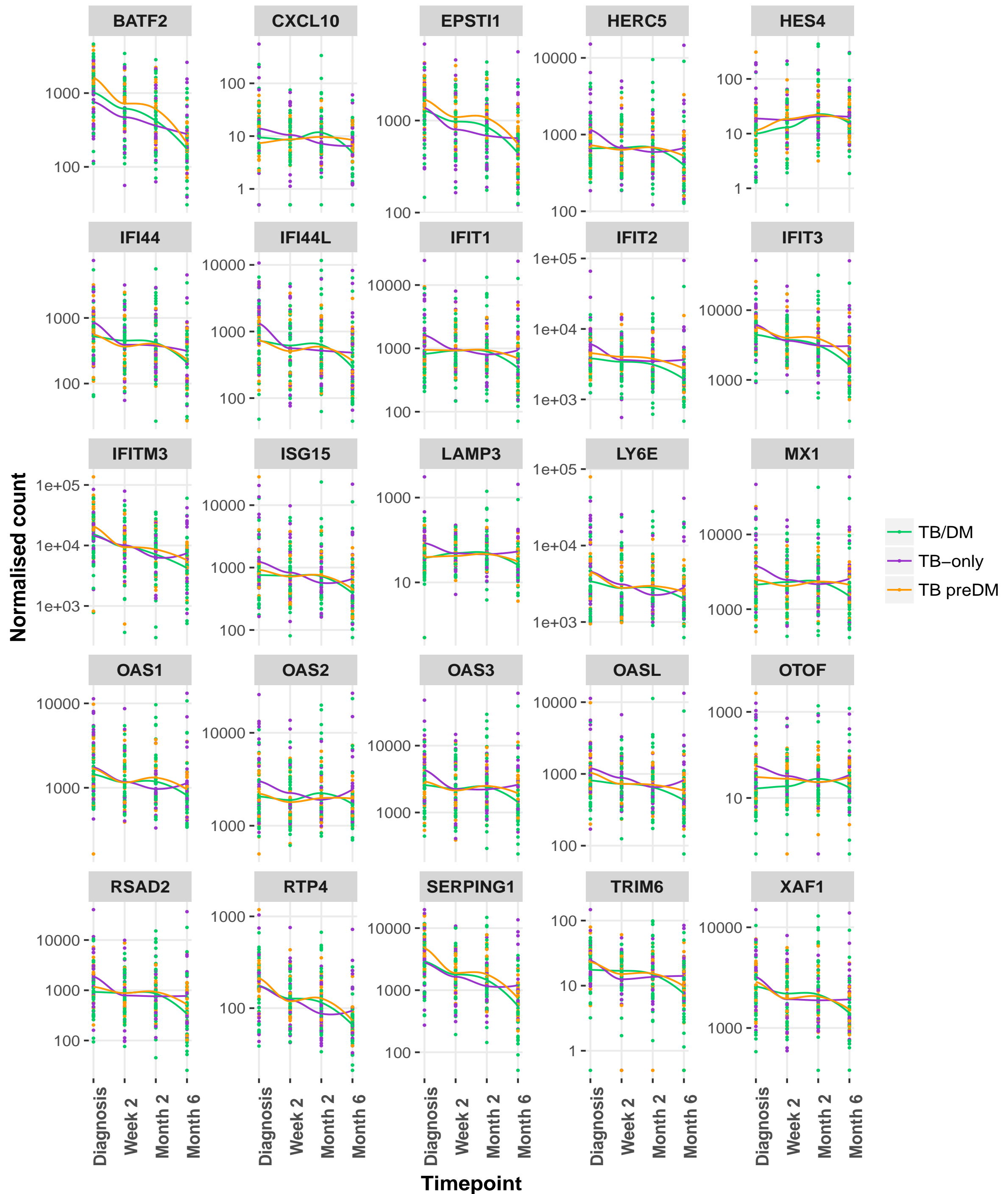


Figure 5.7: **The normalised counts of the genes within the significantly differentially expressed module; interferon.** A loess line of best was plotted through each of the time points. The colours represent the different patient groups. Data were taken from the three field sites and population corrected.

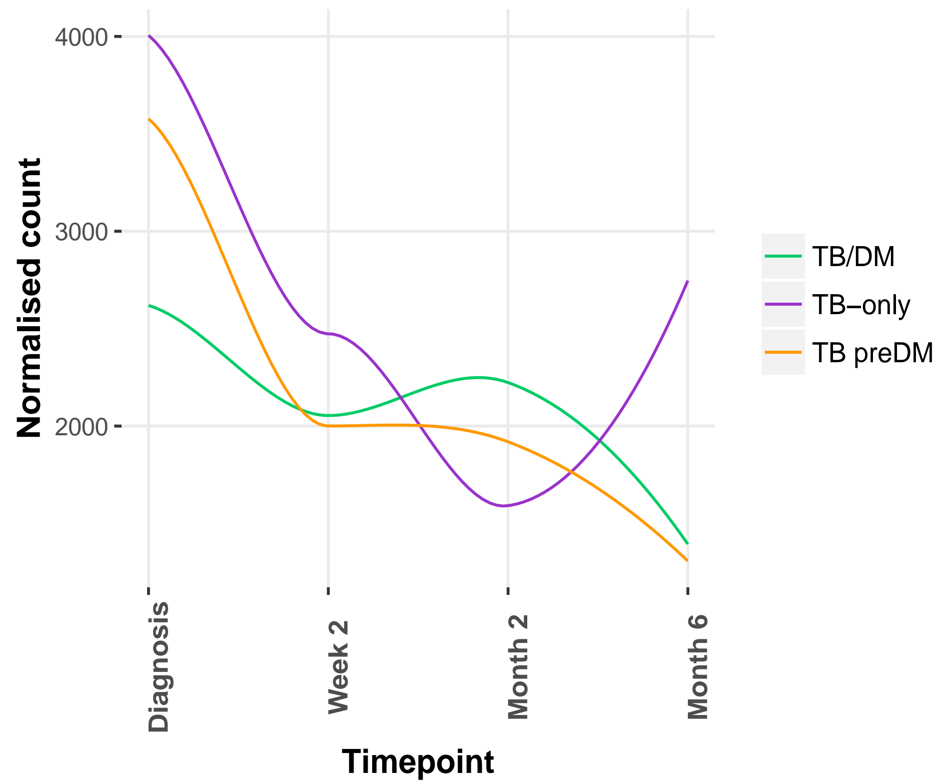


Figure 5.8: **The normalised count from all the genes in the interferon module at each time point through TB treatment.** A loess line of best fit was plotted through all of the counts of all the interferon genes. Data were taken from all three field sites and was population corrected.

Adjusted p-value	Gene Symbol
0.0713	BATF2
0.00389	CXCL10
0.00211	EPSTI1
0.000437	HERC5
0.113	HES4
0.0119	IFI44
0.00226	IFI44L
0.00116	IFIT1
$2.65e^{-05}$	IFIT2
0.00288	IFIT3
$6.42e^{-06}$	IFITM3
0.00596	ISG15
$1.12e^{-05}$	LAMP3
0.00130	LY6E
0.000231	MX1
0.00192	OAS1
0.000332	OAS2
$1.47e^{-05}$	OAS3
0.124	OASL
$5.49e^{-06}$	OTOF
$4.04e^{-06}$	RSAD2
0.0495	RTP4
0.00315	SERPING1
0.00471	TRIM6
0.00707	XAF1

Table 5.4: **Adjusted p-values for genes within interferon module produced from the likelihood ratio test.** TB/DM, TB preDM and TB-only patients groups included in the normalisation

There has been evidence in the literature that the immunological altered state in TB/DM is correlated with increasing HbA1c values. So, the sum of the normalised count of each interferon gene was plotted against the patients' HbA1c reading at diagnosis

(Figure 5.9). There appears to be positive correlation in the TB-only patients; sum of normalised count increases with HbA1c values. However, in TB/DM patients, there appears to be slight negative correlation, in that the sum of the normalised count of the interferon genes decreases with HbA1c readings. The TB preDM group is heavily skewed by the Indonesian population because of the small sample size, but also have a positive correlation, but less so than TB-only. When looking at the field sites combined, the correlation is not statistically significant, but the pattern remains true (Figure 5.10).

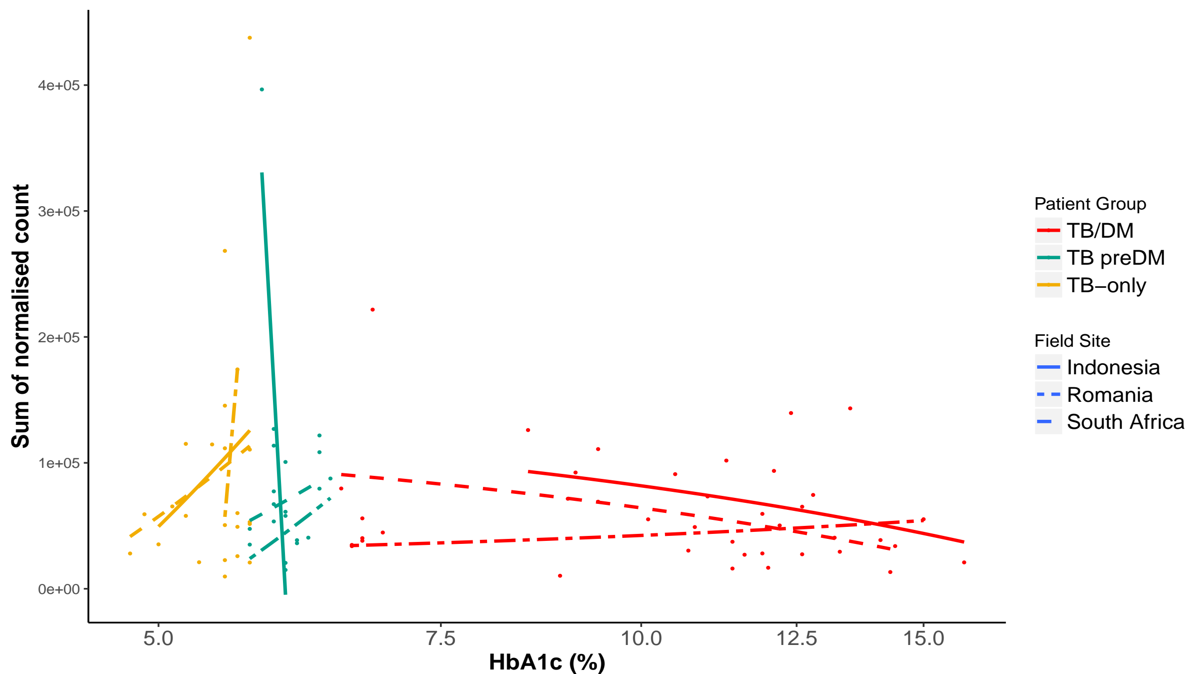


Figure 5.9: **Sum of the normalised count of each patient for each gene in the interferon module, against patient HbA1c levels taken at TB diagnosis separated by population.** Lm line of best fit was plotted for each patient group for each field site. Colour represents the patient group, and line type represents the field site.

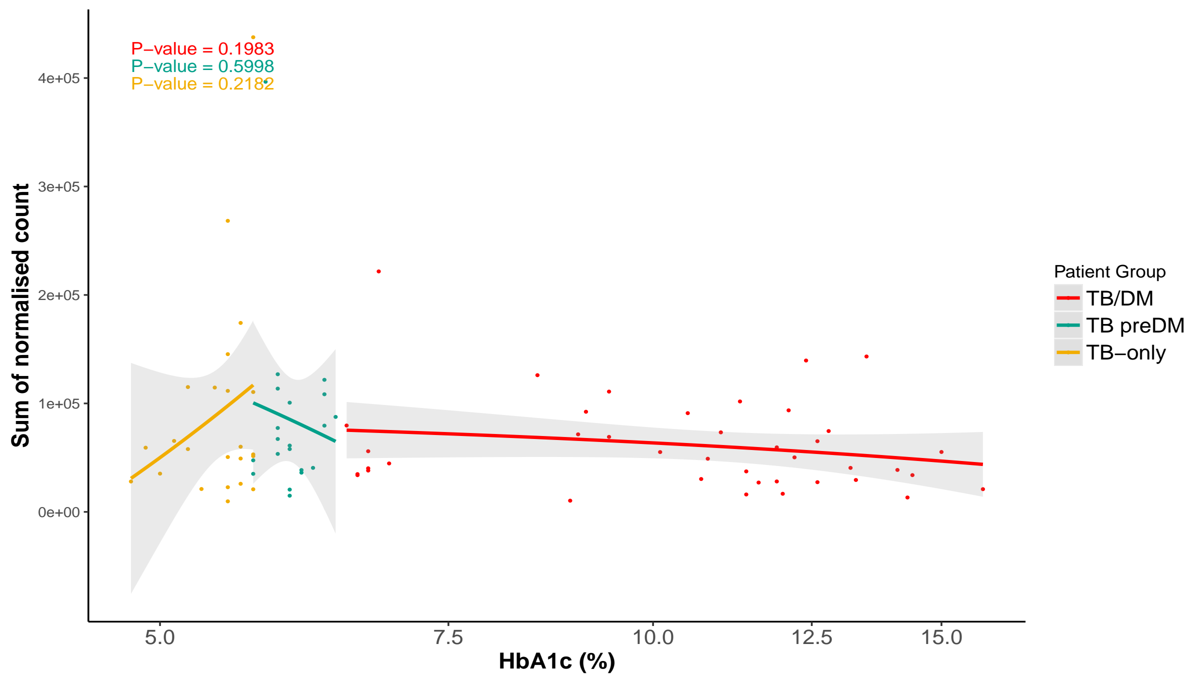


Figure 5.10: **Sum of the normalised count of each patient for each gene in the interferon module, against patient HbA1c levels taken at TB diagnosis.** Colour represents the different patient groups. Lm line of best fit plotted with confidence intervals, with the field sites combined. P-values of the correlation for each patient group are shown. A p-value of < 0.05 was deemed significant.

The Zak et al signature was analysed in TB/DM and TB preDM patients at month 6 to determine whether diabetes co-morbidity affected the resolution of the transcriptome after TB treatment. TB/DM and TB preDM at month 6 was compared to TB-only at month 6, as they would be deemed 'healthy'. Figure 5.11 shows that the 16 gene signature was down-regulated at the end of treatment in TB/DM and TB preDM patients compared to TB-only. All of these genes were statistically significant. The sum of the log fold changes of these genes in TB/DM and TB preDM were -11.38 and -10.88 respectively.

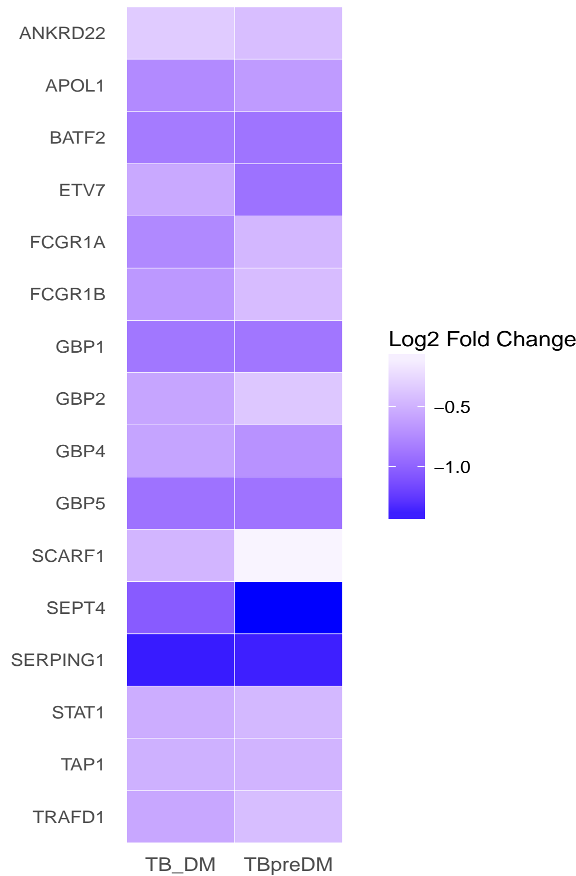


Figure 5.11: **Fold changes of Zak signature in TB/DM and TBpreDM relative to TB-only at month 6.** Differential expression results of the 16 gene signature were extracted at the end of treatment in TB/DM and TB preDM patients versus TB-only. Blue indicates down-regulation.

5.4.3 Directed analysis in closely defined population

In a directed approach, only the Indonesian samples were normalised together, and a comparison of each time point was made between TB/DM and TB-only. This was in order to look at what specifically happens at each time point in TB/DM compared to TB-only. The Indonesian population was chosen as the patients are comparatively well characterised, with two clear-cut groups; TB/DM and TB-only. The number of genes

statistically significantly differentially expressed in TB/DM and TB-only between each time-point are shown in Table 5.5. The greatest difference was at month two.

Time-point	Number of genes
Diagnosis	2
Week 2	3
Month 2	239
Month 6	8

Table 5.5: **Number of significantly differentially expressed genes in TB/DM patients compared to TB-only, between each time point through TB treatment.** A p-value of < 0.05 was deemed significant.

Modular analysis of the Indonesian samples also showed that most of the differential expression between TB/DM and TB-only was at month two (Figure 5.12). The modules that were differentially expressed in TB/DM at any time point were selected for further investigation into their module activity.

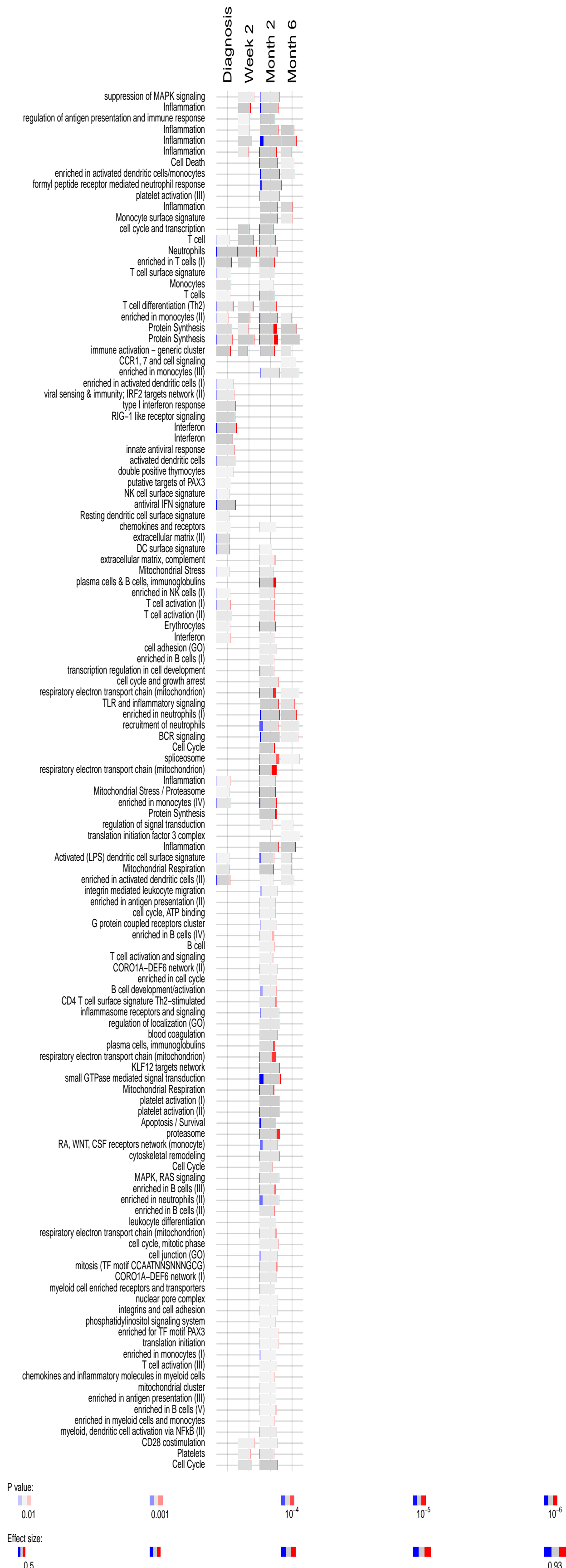


Figure 5.12: **Modules that were significantly differentially expressed in TB/DM relative to TB-only between each time point in Indonesian patients.** Time points were before, throughout, and at the end of TB treatment. Upregulated modules in red, down-regulated modules in blue. The length of the bar represents the effect size of the module, and the saturation of colour represents the adjusted p-value (< 0.05)

The module activity was calculated for the modules that were significantly differentially expressed in TB/DM compared to TB-only at any time-point, using Equation 5.1. Figure 5.13 shows the module activity dynamics in TB/DM patients throughout TB treatment. For example, the neutrophil module is up-regulated in TB/DM relative to TB-only at diagnosis, and persists throughout treatment. Another module; interferon displays a peculiar pattern. It is down-regulated in TB/DM relative to TB-only at diagnosis, but is then up-regulated at month two, oscillating through treatment.

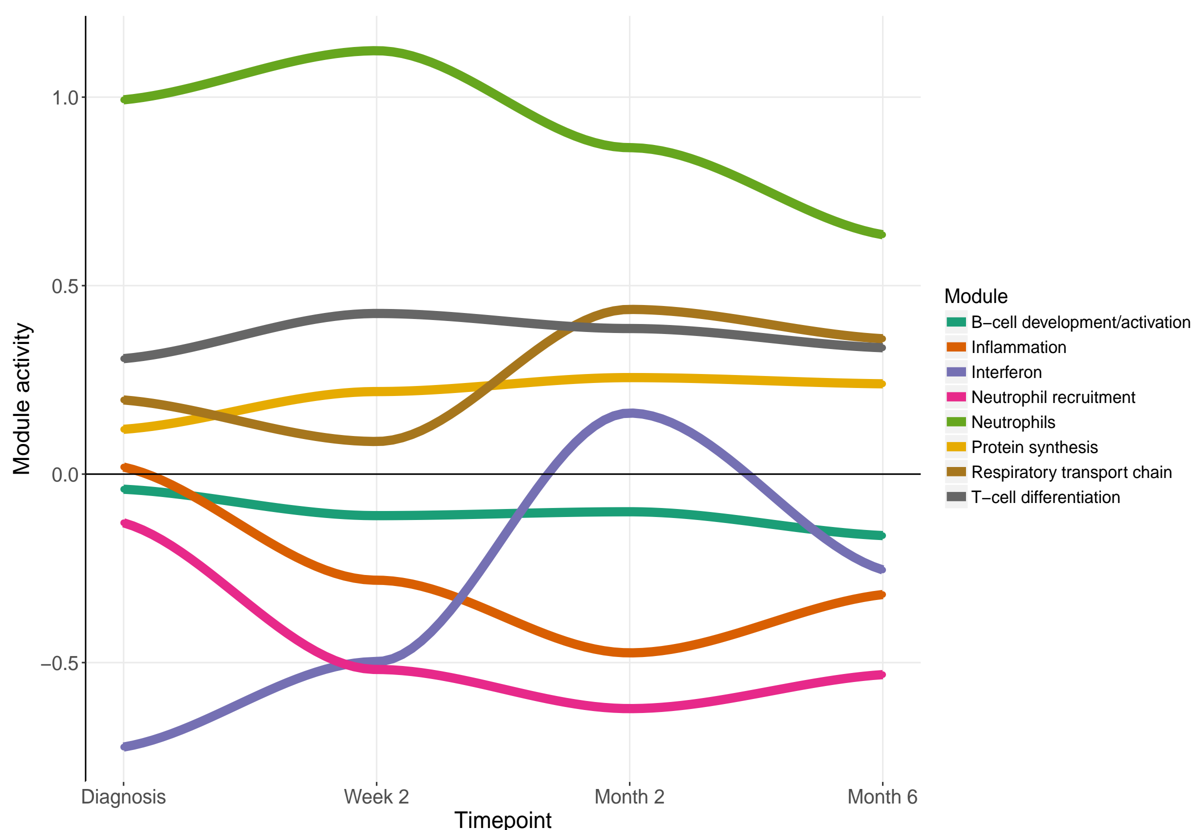


Figure 5.13: **Modules that were significantly differentially expressed between TB/DM and TB-only at any time point throughout TB treatment in Indonesian patients.** Modular activity was calculated by summing the expression of all the genes within a module at a time point and dividing by the number of genes within that module.

To look on the gene level within the significant modules, Figure 5.14 shows the gene expression levels of each gene in the neutrophil module through time in TB/DM patients compared to uncomplicated TB-only. The saturation of the colour equates to the fold change. For example, the genes within the neutrophil module are highly up-regulated; up to $2log_2$ fold change, with the most occurring at week two and month two. In the interferon module, the majority of the genes are down-regulated, as seen in the cross-sectional study. These genes are highly down-regulated; especially otoferlin (OTOF), which is over $-2log_2$ fold change. Like in Figure 5.15, there is an up-regulation at month two in a number of the genes, which then plateaus at the end of treatment at six months.

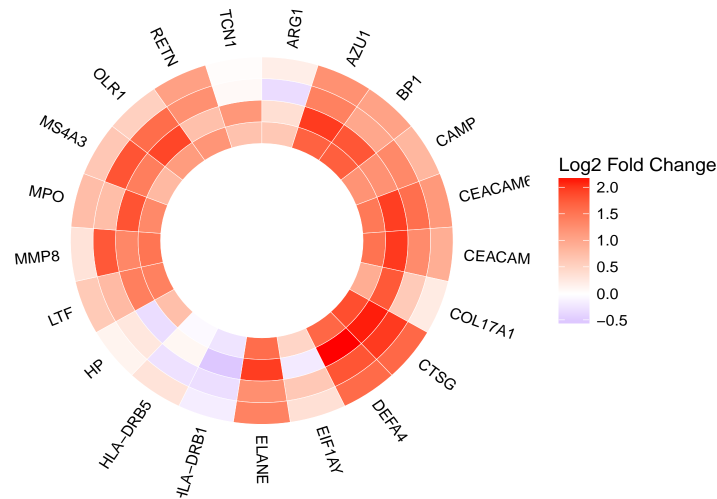


Figure 5.14: **The expression of the genes within the neutrophil module in Indonesian TB/DM patients relative to TB-only.** From the inside, outwards: diagnosis, week 2, month 2, month 6. The saturation of the colour is equivalent to the fold change.

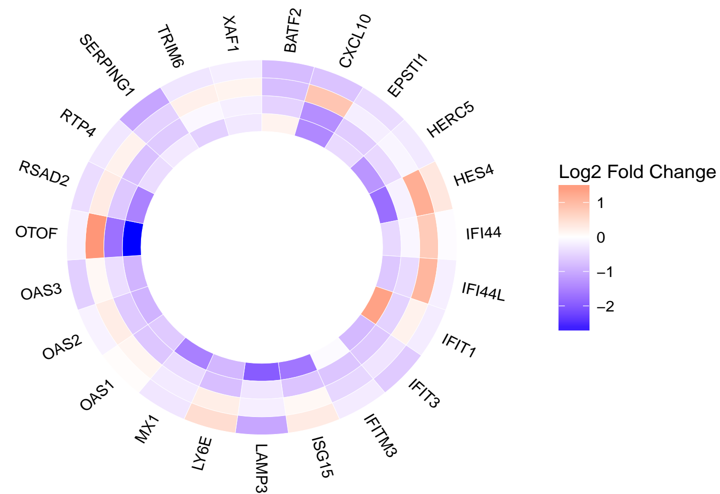


Figure 5.15: **The expression of the genes within the interferon module in Indonesian TB/DM patients relative to TB-only.** From the inside, outwards: diagnosis, week 2, month 2, month 6. The saturation of the colour is equivalent to the fold change.

5.5 Discussion

5.5.1 Findings

The longitudinal study showed that, similarly to the cross-sectional study, interferon response genes were down-regulated at diagnosis in TB/DM and TB preDM relative to TB-only. Not only this, but it does seem diabetes affects the way the host responds to TB treatment; there was something very different in the dynamic changes which occur in response to the killing of *M. tb.* This correlates phenotypically at the end of treatment, as TB/DM have increased rates of poor treatment outcomes.

Two months after the initiation of TB treatment was the critical point of contrast in the dynamic changes. The type I IFN response genes showed an unexpected dynamic; in TB/DM and TB preDM, the type I IFN response genes were up-regulated at month two compared to TB-only, after being down-regulated at diagnosis.

When analysing the Zak signature at month 6, it was found that the resolution of the signature in TB/DM and TB preDM had gone further, and dampened the inflammation at the end of treatment compared to uncomplicated TB. This is quite paradoxical as the comorbidity patients started with an increase of inflammation before the initiation of TB treatment. Again, TB/DM and TB preDM patients were very similar to one another, and together different from TB-only. It provokes interesting questions on the affect of intermediate hyperglycaemia. As DM and preDM confound the resolution of the TB blood signature, causing them to be different from 'normal', it could have implications in their capacity for relapse and also re-exposure. If they are starting from a lower level, could they be more susceptible?

This has implications for TB control, as diabetes affects the way one responds to TB treatment, resulting in increased likelihood of treatment failure or relapse. If diabetes patients are not successfully clearing *M. tb*, it can increase prevalence and peddle drug resistance. Different treatment options should be considered for comorbidity patients. Also, this same abnormal dynamic exists in TB preDM patients showing it is a continuum. This is important because currently, patients with intermediate levels of hyperglycaemia have been overlooked, as it was thought TB susceptibility was only apparent at higher levels of hyperglycaemia. However, from these data, pre-diabetes patients also displayed an altered response to TB treatment. Perhaps pre-diabetes patients should also have re-consideration when it comes to appropriate TB treatment. As discussed in the cross-sectional study in Chapter 4, it reiterates how important it is for TB patients to be screened for (intermediate) hyperglycaemia.

A recent study in macaques nicely demonstrated time-specific transcriptional effects before and through *M. tb* infection, albeit without anti-tuberculosis therapy. They compared these time-specific changes of animals that developed active disease with animals that remained latently infected, and also animals that had severe lung inflammation with animals that had less inflammation.[131] Animals that developed active TB had significantly higher interferon response module expression at infection and at later time-points post-infection. Interestingly, these transcripts were differentially expressed in the animals that went onto develop active disease before *M. tb* infection. Genes in particular were MX1 and IFIT1. The IFN response modules were also highly up-regulated in animals that had more lung inflammation, indicating type I IFNs had a negative impact on disease outcome. This supports the study by Zak et al that found a correlate of risk for

active TB disease that included a type I IFN signature.[76] There is an argument that this signature is less a reflection of the hosts' response to infection, but a predisposition of outcome. Considering that IFNs are recognised as a marker of susceptibility, the current data initially appear paradoxical as IFNs are down-regulated in TB/DM relative to uncomplicated TB patients at diagnosis, even though epidemiological data show diabetes patients are more susceptible to active TB. But, the patterns of type I IFN signatures are different during treatment, it seems that the response is delayed, so diabetes causes a prolonged persistence resulting in lack of resolution during the timeframe of conventional treatment.

Our data contradicts the dogma that there is one blood transcriptomic signature of TB which resolves uniformly during TB treatment, but rather an uncoupling of the immune response occurs in TB/DM patients. Our data show that different components within the TB signature are separately regulated; the neutrophil and inflammatory response in TB/DM are controlled by TB treatment, but a different pattern is displayed in the type I IFN response. This difference means that by the end of TB treatment, the TB/DM patients' immune status has not resolved to the same point as TB-only patients. This demonstrates that there is a more complex regulation of gene expression occurring through TB treatment than originally thought.

In this regard, a molecular distance to health single measure, based on prior TB transcriptomic data, would not be sufficient to explain the increased risk of poor TB treatment outcomes in TB/DM. It is likely that a new biosignature will be needed in order to be inclusive and measure these differences.

5.5.2 Implications on TB control

It would be beneficial to follow diabetes and pre-diabetes patients more closely in their TB treatment. TB/DM patients may still not have their HbA1c under control after completion of TB treatment, and the enhanced risk of relapse could be because the diabetes is still not managed, so intervening with glycaemic management is critical. Results from the randomised control trial (RCT) performed in TANDEM have shown that TB/DM patients had significantly improved glycaemic control when they had been intensively monitored.

Seeing as TB/DM patients have an altered response to TB, different treatment algorithms should be considered. Perhaps an extended intensive phase and/or extended continuation phase would be beneficial: these options could be tested in a clinical trial, measuring clinical outcome as the primary endpoint, with the resolution of the TB/DM transcriptomic signature measured as secondary outcome. If the cure rate and relapse rate improves, the transcriptome would be seen returning to normal. This would be indicative of disease clearing, and would mean that the altered treatment was a success. Again, the longitudinal study suggests the option of host-directed therapy, or adjunctive treatment, including something that will re-calibrate the immune system after it has been modulated by diabetes, so the host can respond effectively. It could be in the form of a non-inappropriate boost to the immune system, or modulation in the middle of treatment, like neutralising anti-type I IFN antibodies.

5.5.3 Comments on the study and future work

A large limitation from this study was the sample size. There were too few samples for each patient group at each time-point. In differential expression studies, this is crucial, so when there are too few samples, fewer genes are statistically significant. To add to this, Peru could not be included in the longitudinal study as the samples were not collected in time. There was a deadline for all samples to be received by LSHTM, as the RNA-seq experiments were on a tight time line and all the samples had to be processed together to avoid unwanted batch effects.

The grouping of diabetes patients also proved challenging, particularly in the pre-diabetes group. Some remained at intermediate HbA1c values at the end of treatment, whereas some decreased to 'normal' levels and some increase. The patients whose HbA1c values lower to normal levels were likely experiencing TB-induced hyperglycaemia, so it is difficult to understand which patients are truly pre-diabetic when they initially present in the TB clinic. But, if a patient does experience transient hyperglycaemia, they are still different from a patient who remains HbA1c $< 5.7\%$, and are at a higher risk of developing full diabetes later in life. Further work however could unravel the effect of transient hyperglycaemia and pre-diabetes. But, the addition of more and more groups runs the risk of losing resolution in the data and reducing the sample numbers in each group even further.

There were also challenges in analysing longitudinal data. Each time-point sample was from the same patient, so were highly correlated, and the linear modelling methods treated each time-point as a separate, independent variable. This is slightly erroneous. It means the relative variation between two time-points is dulled because as they were from

the same patient, are already very similar to one another. In repeated measure data, a mixed model approach is usually taken, which calculates the random effect, to calculate the added variance between individuals. However, this method is not available in current differential expression analysis tools.

Firstly, it is important to ascertain in Indonesia whether poor treatment outcomes refer to a patients clinical symptoms, and if further treatment was taken, as at the moment, poor treatment outcomes were defined on microbiological culture data. This would establish whether diabetes led to poor treatment outcomes in Indonesia. Other work in TANDEM involved investigating the pharmacokinetics of rifampicin and anti-diabetes drugs. Diabetes patients have been associated with lower concentrations of rifampicin, which could be associated with higher BMI commonly seen in T2DM patients. It could simply mean that their TB treatment needs to be weight-based dosing to achieve the desirable effects. However, these results are yet to be announced, but other drug-drug interactions need to be considered.

Of course, a study with more sample numbers would be ideal, and validation of the results of the current RNA-seq project using a larger sample set an alternative, more high throughput assay, would be logical. This option is currently being investigated by the TANDEM consortium. There was a plan to analyse a much larger set of samples (∼800) from the whole cohort using the reverse transcriptase multiplex ligation probe assay (RT-MLPA) to validate the results of the RNA-seq to dramatically expand this work, but this has not yet been completed. This would also include more time-points; month one, month four and month twelve (from South Africa). It would be of value in order to validate the results so far and unravel mechanisms of susceptibility as more

intensive points through treatment.

In conclusion we have found that TB/DM and TB preDM patient immune systems behave differently in response to TB treatment which could be indicative of why diabetes causes an increase in poor treatment outcomes. The TB treatment signature is more complex than first considered; it appears different parts of the immune response are regulated separately which has implications in biosignature development.

6

Final Discussion and Summary

Comorbidities do have an effect on the TB blood transcriptome. HIV coinfection had a comparatively small effect, and diabetes had a large effect. The distinct TB biosignature has been well described, and it is still detectable in HIV-positive patients. Diabetes causes considerable alterations to the TB transcriptome. This is contrary to what I initially expected, as HIV can be so devastating in TB patients, whereas diabetes is not notoriously as overwhelming. This is because HIV primarily affects the adaptive immune response, leaving the innate immune system relatively intact, and ART is very successful in re-stabilising the adaptive response. On the other hand, diabetes affects systemic metabolism, inflammation and the innate response, components that influence the progression and prognosis of TB. Also, the control of diabetes is highly variable within populations, so even with available treatment, sufficient management is not always obtained and therefore not curbing these harmful effects.

These differences in TB/DM and TB preDM patients span an increase inflammation and a decrease in type I IFN responses, which may reflect a mechanism of susceptibility

as diabetes affects the immune response to *M. tb.* To add, as pre-diabetes patients also show dramatic differences, this may possibly indicate enhanced susceptibility even before the clinical diagnosis of diabetes. It demonstrates that this susceptibility occurs even at intermediate levels of hyperglycaemia, which was not previously considered. Usually these harmful effects would be apparent at higher levels of hyperglycaemia, so it is debatable that the threshold for diabetes should be lowered because changes are observed, meaning the causative means (insulin resistance and chronic inflammation) are present as well.

Analysis of a TB blood signature revealed that the performance increased in DM and preDM patients, showing biosignatures are still valid in comorbidity patients. What was particularly surprising was its performance was positively correlated with HbA1c values, even in TB-only, showing an interesting relationship between inflammation and HbA1c. Resolution of the biosignature at the end of treatment was confounded by DM and also preDM, again reiterating the substantial affect even intermediate hyperglycaemia has on TB infection. Also, a divergence from 'healthy' at the end of TB treatment could cause implications in treatment outcome in the form of relapse. HIV coinfection also partially confounded biosignature resolution (at week 8), however none of these genes were statistically significant so it is difficult to make a conclusion. Further analysis at the end of treatment and their treatment outcomes would need to be done, in order to fully determine whether a treatment response biosignature would still be effective in HIV-positive patients.

In regards to TB control, it should be recommended that TB patients are screened for abnormal glycaemia levels, and diabetes and pre-diabetes patients should be more closely monitored for TB symptoms. There were also differences in diabetes and pre-diabetes

patients throughout TB treatment. TB/DM patients did not resolve to the same point as uncomplicated TB patients. This could reflect a susceptibility mechanism of why diabetes patients are more likely to suffer from poor treatment outcomes. This should be taken into account when administering TB treatment to comorbidity patients in regards to length of treatment, prophylaxis or host directed therapy. Host directed therapy is a more attractive option for TB treatment rather than more antibiotics because of looming drug resistance. It has been speculated that the anti-diabetes drug metformin has a potential role here, and these data showed that it exhibits desirable anti-inflammatory effects in response to *M. tb* infection.

The type I IFN response was highlighted as aberrant in TB/DM patients and has been reported as an important component in the control of *M. tb*. The response elements and subsequent inducible effects are pleiotropic which reflects in the literature which is somewhat contradictory; an elevated type I IFN response is pathogenic and is positively correlated with disease severity,[67] but then type I IFNs have also been used as adjuvant therapy. In these data, inflammation and the type I IFN response do not seem to correlate, suggesting that they are regulated separately, and showing the downstream IFN response is different in diabetes. It presents a challenge in unravelling IFNs' role in susceptibility.

IFNs are also a potential target of metformin which has been suggested as a candidate for adjunctive therapy; these data suggest that it has beneficial anti-inflammatory effects in the context of *M. tb*. Type I IFN response has been noted as a possible target for host directed therapy in other studies, namely Mayer-Barber et al. They found that when an arachidonic acid metabolism modulator, that works by inhibiting type I IFNs, was

administered to susceptible mice, there was reduced pathology and increased survival.[132] To add to the TANDEM data, functional studies would be required to validate the change in type I IFNs and their inducible genes. Useful work with animal models of TB/DM to validate biosignatures could be performed. A diabetes mouse model exists, and a prospective study could be performed where the mice are subsequently infected with *M. tb* and blood transcriptome analysis done through time to find a diabetes-associated correlate of risk. This could then be translated to non-human primates that would give more insight into TB human pathology.

Microarrays and RNA-seq are both valuable tools in investigating gene expression. Microarrays are still cheaper and it can be argued the analysis is more reliable as it is heavily peer-reviewed online and extensively optimised. On the other hand, with RNA-seq; it is still expensive, the lab work is more laborious and finicky, and there is no gold standard in the analysis. At each stage of data processing, assumptions need to be made, and different tools produce slightly different results. For instance, during alignment, the underlying algorithms of the software are different, in for example how the reads overlap at exon boundaries. Plus, the data processing needs to be performed in a cluster which is often impractical. There is a software that requires significantly less computer memory, but it is more inaccurate.[192] However, RNA-seq is able to detect lower abundance transcripts that would otherwise be removed as background noise, so the slight consistent changes that are seen in these data may or may not have been detected if microarrays had been used instead of RNA-seq.

In gene-set and module analysis, the large caveat is it relies on already known gene associations and known pathways. The analysis tools at the moment are good at re-

proving existing knowledge, but do not seem to be extending our knowledge about all the less well characterised genes. This is particularly the case for non-protein coding genes like long non-coding RNAs (lncRNAs) which are not included in module or gene-set groups. It means a loss of information that is a potentially crucial layer in biological regulation, as many lncRNAs and other non-coding RNAs are differentially expressed (another advantage of RNA-seq over microarray), but it is difficult to characterise and interpret them appropriately.

So what does this mean for the development of a biosignature? It is still valid as a treatment efficacy measurement; lack of resolution still correlates with poor treatment outcomes. But, a single measure of disease risk may not be appropriate as they are heavily based on up-regulated type I IFN genes. This is also the case in the correlate of risk which includes type I IFN response genes.[76]

Moving forward, combining TB treatment response blood transcriptome datasets, from TANDEM, PanACEA and other published studies, a biomarker signature of TB treatment response could be developed which could be widely applicable to measure treatment efficacy and outcome in a wide variety of settings and subjects. Such a biosignature would need to be tested in a phase III clinical trial, with changes in the signature correlated with treatment outcome and with relapse over a 2-year follow-up period. The continued presence of an aberrant biosignature at the end of TB treatment might indicate an increased risk of both relapse and reinfection, particularly in patients with diabetes comorbidity.

In conclusion, comorbidities do need special consideration in regards to global TB control in regards to screening and also treatment plans. Further work needs to be done

to fully understand the susceptibility mechanism, which could then be targeted for host directed therapy in the future.

Bibliography

1. Donoghue, H. D. Insights into ancient leprosy and tuberculosis using metagenomics. *Trends in Microbiology* **21** (2013).
2. Koch, R. *Die Aetiologie der Tuberculose* (Berliner klin. Wochenschr, 1882).
3. Hippocrates. *Of the Epidemics* (410-400BCE).
4. Bynum, H. *Spitting Blood The History of Tuberculosis* (Oxford University Press, 2012).
5. Davies, P. D. O., Gordon, S. B. & Davies, G. *Clinical Tuberculosis* (Taylor & Francis Group, 2014).
6. *Global Tuberculosis Report 2015* tech. rep. (WHO, 2015).
7. Bloom, B. R. *Tuberculosis Pathogenesis, Protection and Control* (American Society for Microbiology, 1994).

8. Mangtani, P. *et al.* Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin. Infect. Dis.* **58,** 470–480 (2014).

9. Abubakar, I. *et al.* Systematic review and meta-analysis of the current evidence on the duration of protection by bacillus Calmette-Guerin vaccination against tuberculosis. *Health Technol Assess* **17,** 1–372 (Sept. 2013).

10. Tameris, M. D. *et al.* Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. *The Lancet* **381,** 1021–1028 (2013).

11. Dockrell, H. M. Towards new TB vaccines: What are the challenges? *Pathog Dis* **74** (2016).

12. Aaron, L. *et al.* Tuberculosis in HIV-infected patients: a comprehensive review. *Clin Microbiol Infect* **10,** 388–398 (2004).

13. Kaufmann, S. H. & van Helden, P. *Handbook of Tuberculosis: Clinics, Diagnosis, Therapy and Epidemiology* (Wiley-VCH, 2008).

14. Council, M. R. Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation. *British Medical Journal* **2,** 769–782 (Oct. 1948).

15. Mitchison, D. A. Basic mechanisms of chemotherapy. *Chest* **76,** 771–781 (Dec. 1979).

16. Gillespie, S. H. *et al.* Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis. *N. Engl. J. Med.* (2014).

17. Merle, C. S. *et al.* A Four-Month Gatifloxacin-Containing Regimen for Treating Tuberculosis. *New England Journal of Medicine* **371,** 1588–1598 (2014).

18. Mitchison, D. A. Tests for streptomycin sensitivity of tubercle bacilli in tween 80 albumin liquid medium. *Lancet* **2,** 694–696 (1949).

19. Mukamolova, G. V., Turapov, O., Malkin, J., Woltmann, G. & Barer, M. R. Resuscitation-promoting Factors Reveal an Occult Population of Tubercle Bacilli in Sputum. *American Journal of Respiratory and Critical Care Medicine* **181,** 174–180 (Jan. 2010).

20. Honeyborne, I. *et al.* Molecular bacterial load assay, a culture-free biomarker for rapid and accurate quantification of sputum Mycobacterium tuberculosis bacillary load during treatment. *J. Clin. Microbiol.* **49,** 3905–3911 (2011).

21. Honeyborne, I. *et al.* The molecular bacterial load assay replaces solid culture for measuring early bactericidal response to antituberculosis treatment. *J. Clin. Microbiol.* **52,** 3064–3067 (2014).

22. De Knegt, G. J. *et al.* Assessment of treatment response by colony forming units, time to culture positivity and the molecular bacterial load assay compared in a mouse tuberculosis model. *Tuberculosis* **105,** 113–118 (2017).

23. O'Garra, A. *et al.* The Immune Response in Tuberculosis. *Annual Review of Immunology* **31,** 475–527 (2013).

24. Lyadova, I. V. & Panteleev, A. V. Th1 and Th17 Cells in Tuberculosis: Protection, Pathology, and Biomarkers. *Mediators of Inflammation* **2015** (2015).

25. Jain, A. & Singh, J. A. Harms of Tumor Necrosis Factor Inhibitors in Rheumatic diseases: A focused Systematic Review of the Literature. *Immunotherapy* **5,** 265–299 (2013).

26. WHO. *Global AIDS update* tech. rep. (2016).
27. Margolis, A. M., Heverling, H., Pham, P. A. & Stolbach, A. A review of the toxicity of HIV medications. *J Med Toxicol* **10,** 26–39 (2014).
28. Spinner, C. D. *et al.* HIV pre-exposure prophylaxis (PrEP): a review of current knowledge of oral systemic HIV PrEP in humans. *Infection* **44,** 151–158 (2016).
29. Sultan, B., Benn, P. & Waters, L. Current perspectives in HIV post-exposure prophylaxis. *HIV/AIDS* **6,** 147–158 (2014).
30. Hütter, G. *et al.* Long-Term Control of HIV by CCR5 Delta32/Delta32 Stem-Cell Transplantation. *New England Journal of Medicine* **360** (2009).
31. McCoy, L. E. & Weiss, R. A. Neutralizing antibodies to HIV-1 induced by immunization. *J Exp Med* **210** (2013).
32. Avicenna. *The Canon of Medicine* (1025).
33. Willis, T. *Pharmaceautice rationalis* (e Theatro Sheldoniano, 1674).
34. Dobson, M. Experiments and observations on the urine in diabetes. *Med. Obs. Inq.* **5,** 298–316 (1776).
35. Tattersall, R. *Diabetes: The Biography* (Oxford University Press, 2009).
36. *International Diabetes Federation* tech. rep. (IDF Diabetes Atlas, 2015).
37. *Global Report on diabetes* tech. rep. (WHO, Geneva: World Health Organisation, 2016).
38. *Global Status Report on noncommunicable diseases* tech. rep. (WHO, 2014).

39. Burant, C. F. & Young, L. A. *Medical Management of Type 2 Diabetes* (American Diabetes Association, 2012).

40. Group, D. P. P. R. Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin. *New England Journal of Medicine* **346,** 393–403 (2002).

41. Brook, C. G. & Marshall, N. J. *Essential endocrinology* 4th ed. (Blackwell Publishing Ltd, 2001).

42. Nelson, D. L. & Cox, M. M. *Lehninger Principles of Biochemistry* 5th ed. (W. H. Freeman and Company, 2008).

43. Robertson, R. P. Chronic Oxidative Stress as a Central Mechanism for Glucose Toxicity in Pancreatic Islet Beta Cells in Diabetes. *J. Biol. Chem.* **279,** 42351–42354 (2004).

44. Ebstein, W. Zur therapie des Diabetes mellitus, insbesondere über die Anwendung des salicylsauren Natron bei demselben. Berliner Klinische Wochenschrift. *Berliner Klinische Wochenschrift* **13,** 337–340 (1876).

45. Reid, J., MacDougall, A. I. & Andrews, M. M. Aspirin and diabetes mellitus. *Br Med J* **2,** 1071–1074 (Nov. 1957).

46. Yuan, M. *et al.* Reversal of obesity- and diet-induced insulin resistance with salicylates or targeted disruption of Ikkbeta. *Science* **293,** 1673–1677 (Aug. 2001).

47. Hotamisligil, G. S., Shargill, N. S. & Spiegelman, B. M. Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. *Science* **259,** 87–91 (1993).

48. De Taeye, B. M. *et al.* Macrophage TNF-alpha contributes to insulin resistance and hepatic steatosis in diet-induced obesity. *Am J Physiol Endocrinol Metab* **293** (2007).

49. Kado, S., Nagase, T. & Nagata, N. Circulating levels of interleukin-6, its soluble receptor and interleukin-6/interleukin-6 receptor complexes in patients with type 2 diabetes mellitus. *Acta Diabetologica* **36,** 67–72 (1999).

50. Fried, S. K., Bunkin, D. A. & Greenberg, A. S. Omental and Subcutaneous Adipose Tissues of Obese Subjects Release Interleukin-6: Depot Difference and Regulation by Glucocorticoid1. *The Journal of Clinical Endocrinology & Metabolism* **83,** 847 (1998).

51. Mohamed-Ali, V. *et al.* Subcutaneous Adipose Tissue Releases Interleukin-6, But Not Tumor Necrosis Factor-alpha, in Vivo. *The Journal of Clinical Endocrinology & Metabolism* **82,** 4196 (1997).

52. Eder, K., Baffy, N., Falus, A. & Fulop, A. K. The major inflammatory mediator interleukin-6 and obesity. *Inflammation Research* **58,** 727 (2009).

53. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *Journal of Clinical Investigation* **112,** 1796–1808 (2003).

54. Donath, M. Y. & Shoelson, S. E. Type 2 diabetes as an inflammatory disease. *Nature Reviews Immunology* **11,** 98–107 (2011).

55. Lee, J. Y., Sohn, K. H., Rhee, S. H. & Hwang, D. Saturated fatty acids, but not unsaturated fatty acids, induce the expression of cyclooxygenase-2 mediated through Toll-like receptor 4. *J Biol Chem* **276** (2001).

56. Simon, M. C. *et al.* Fatty acids modulate cytokine and chemokine secretion of stimulated human whole blood cultures in diabetes: Fatty acid modulated cytokine secretion. *Clinical & Experimental Immunology* **172,** 383–393 (2013).

57. Song, M. J., Kim, K. H., Yoon, J. M. & Kim, J. B. Activation of Toll-like receptor 4 is associated with insulin resistance in adipocytes. *Biochemical and Biophysical Research Communications* **346,** 739–745 (2006).

58. Goldfine, A. B. *et al.* Targeting inflammation using salsalate in patients with type 2 diabetes: effects on flow-mediated dilation (TINSAL-FMD). *Diabetes Care* **36,** 4132–4139 (Dec. 2013).

59. Jeon, C. Y. & Murray, M. B. Diabetes Mellitus Increases the Risk of Active Tuberculosis: A Systematic Review of 13 Observational Studies. *PLoS Med* **5** (2008).

60. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10,** 57–63 (2009).

61. Garber, M., Grabherr, M. G., Guttman, M. & Trapnell, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Meth* **8,** 469–477 (2011).

62. Soneson, C. & Delorenzi, M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* **14** (2013).

63. Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE* **9** (2014).

64. Nookaew, I. *et al.* A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae. *Nucleic Acids Res* **40** (2012).

65. Jacobsen, M. *et al.* Candidate biomarkers for discrimination between infection and disease caused by Mycobacterium tuberculosis. *J Mol Med (Berl)* **85,** 613–621 (2007).

66. Mistry, R. *et al.* Gene-expression patterns in whole blood identify subjects at risk for recurrent tuberculosis. *Journal of Infectious Diseases* **195,** 357–365 (2007).

67. Berry, M. P. R. *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466,** 973–977 (2010).

68. Cliff, J. M. *et al.* Distinct Phases of Blood Gene Expression Pattern Through Tuberculosis Treatment Reflect Modulation of the Humoral Immune Response. *Journal of Infectious Diseases* **207,** 18–29 (2012).

69. Kaforou, M. *et al.* Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med.* **10** (2013).

70. Maertzdorf, J. *et al.* Functional Correlations of Pathogenesis-Driven Gene Expression Signatures in Tuberculosis. *PLoS ONE* **6** (2011).

71. Maertzdorf, J. *et al.* Human gene expression profiles of susceptibility and resistance in tuberculosis. *Genes and immunity* **12,** 15–22 (2011).

72. Bloom, C. I. *et al.* Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS ONE* **7** (2012).

73. Ottenhoff, T. H. M. *et al.* Genome-Wide Expression Profiling Identifies Type 1 Interferon Response Pathways in Active Tuberculosis. *PLoS ONE* **7** (2012).

74. Bloom, C. I. *et al.* Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PLoS ONE* **8** (2013).

75. Anderson, S. T. *et al.* Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N. Engl. J. Med.* **370,** 1712–1723 (2014).

76. Zak, D. E. *et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* **387,** 2312–2322 (2016).

77. Manoel-Caetano, F. S. *et al.* Gene expression profiles displayed by peripheral blood mononuclear cells from patients with type 2 diabetes mellitus focusing on biological processes implicated on the pathogenesis of the disease. *Gene* **511,** 151–160 (2012).

78. Grayson, B. L., Wang, L. & Aune, T. M. Peripheral blood gene expression profiles in metabolic syndrome, coronary artery disease and type 2 diabetes. *Genes and immunity* **12,** 341–351 (2011).

79. Klatzmann, D. *et al.* Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. *Science* **225,** 59–63 (1984).

80. Douek, D. C. *et al.* HIV preferentially infects HIV-specific CD4+ T cells. *Nature* **417,** 95–98 (2002).

81. Ho, D. D. *et al.* Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373,** 123–126 (1995).

82. Gupta, A., Wood, R., Kaplan, R., Bekker, L.-G. & Lawn, S. D. Tuberculosis Incidence Rates during 8 Years of Follow-Up of an Antiretroviral Treatment Cohort in South Africa: Comparison with Rates in the Community. *PLOS ONE* **7,** 1–10 (2012).

83. Walker, N. F., Meintjes, G. & Wilkinson, R. J. HIV-1 and the immune response to TB. *Future virology* **8,** 57–80 (Jan. 2013).

84. Tomlinson, G. S. *et al.* HIV-1 Infection of Macrophages Dysregulates Innate Immune Responses to Mycobacterium tuberculosis by Inhibition of Interleukin-10. *J. Infect. Dis.* **209,** 1055–1065 (2014).

85. Geldmacher, C., Zumla, A. & Hoelscher, M. Interaction between HIV and Mycobacterium tuberculosis: HIV-1-induced CD4 T-cell depletion and the development of active tuberculosis. *Curr Opin HIV AIDS* **7,** 268–275 (2012).

86. Lawn, S. D., Kerkhoff, A. D., Vogt, M. & Wood, R. Diagnostic accuracy of a low-cost, urine antigen, point-of-care screening assay for HIV-associated pulmonary tuberculosis before antiretroviral therapy: a descriptive study. *Lancet Infect Dis* **12,** 201–209 (2012).

87. Passel S Jeffrey; Cohn, D. Unauthorised Immigrant Population: National and State Trends. *Pew Research Center. Pew Hispanic Center* (2010).

88. Perriëns, J. H. *et al.* Pulmonary tuberculosis in HIV-infected patients in Zaire. A controlled trial of treatment for either 6 or 12 months. *N. Engl. J. Med.* **332,** 779–784 (1995).

89. Fitzgerald, D. W. *et al.* Effect of post-treatment isoniazid on prevention of recurrent tuberculosis in HIV-1-infected individuals: a randomised trial. *Lancet* **356,** 1470–1474 (2000).

90. Khan, F. A. *et al.* Treatment of active tuberculosis in HIV-coinfected patients: a systematic review and meta-analysis. *Clin Infect Dis* **50** (2010).

91. Cliff, J. M. *et al.* Excessive Cytolytic Responses Predict Tuberculosis Relapse After Apparently Successful Treatment. *The Journal of Infectious Diseases* **213,** 485–495 (Feb. 2016).

92. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3** (2004).

93. Smyth, G. K. & Altman, N. S. Separate-channel analysis of two-channel microarrays: recovering inter-spot information. *BMC Bioinformatics* **14** (2013).

94. Chaussabel, D. *et al.* A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* **29,** 150–164 (2008).

95. Li, S. *et al.* Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat. Immunol.* **15,** 195–204 (2014).

96. Kassa, D. *et al.* Discriminative expression of whole blood genes in HIV patients with latent and active TB in Ethiopia. *Tuberculosis* **100** (2016).

97. Sutherland, J. S. *et al.* Differential gene expression of activating Fc-gamma receptor classifies active tuberculosis regardless of human immunodeficiency virus status or ethnicity. *Clin. Microbiol. Infect.* **20** (2014).

98. Bailey, C. J. & Day, C. Traditional plant medicines as treatments for diabetes. *Diabetes Care* **12,** 553–564 (1989).

99. Nathan, D. M. *et al.* Medical Management of Hyperglycemia in Type 2 Diabetes: A Consensus Algorithm for the Initiation and Adjustment of Therapy. *Diabetes Care* **32,** 193–203 (2009).

100. UKPDS. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. *Lancet* **352** (1998).

101. Thomas, I. & Gregg, B. Metformin; a review of its history and future: from lilac to longevity. *Pediatric Diabetes* **18** (2017).

102. Zhou, G. *et al.* Role of AMP-activated protein kinase in mechanism of metformin action. *The Journal of Clinical Investigation* **108,** 1167–1174 (Oct. 2001).

103. Hayashi, T., Hirshman, M. F., Kurth, E. J., Winder, W. W. & Goodyear, L. J. Evidence for 5'AMP-Activated Protein Kinase Mediation of the Effect of Muscle Contraction on Glucose Transport. *Diabetes* **47,** 1369–1373 (1998).

104. Cho, K. *et al.* Antihyperglycemic mechanism of metformin occurs via the AMPK/LXRα/POMC pathway. *Scientific Reports* **5** (2015).

105. El-Mir, M.-Y. *et al.* Dimethylbiguanide Inhibits Cell Respiration via an Indirect Effect Targeted on the Respiratory Chain Complex I. *Journal of Biological Chemistry* **275,** 223–228 (2000).

106. Gandini, S. *et al.* Metformin and Cancer Risk and Mortality: A Systematic Review and Meta-analysis Taking into Account Biases and Confounders. *Cancer Prevention Research* **7,** 867–885 (2014).

107. DeCensi, A. *et al.* Effect of Metformin on Breast Ductal Carcinoma In Situ Proliferation in a Randomized Presurgical Trial. *Cancer Prevention Research* **8,** 888–894 (2015).

108. Barzilai, N., Crandall, J. P., Kritchevsky, S. B. & Espeland, M. A. Metformin as a Tool to Target Aging. *Cell Metabolism* **23,** 1060–1065 (2017/08/29 2016).

109. Eikawa, S. *et al.* Immune-mediated antitumor effect by type 2 diabetes drug, metformin. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 1809–1814 (2015).

110. Rocha, G. Z. *et al.* Metformin amplifies chemotherapy-induced AMPK activation and antitumoral growth. *Clin Cancer Res* **17,** 3993–4005 (2011).

111. Kordes, S. *et al.* Metformin in patients with advanced pancreatic cancer: a double-blind, randomised, placebo-controlled phase 2 trial. *The Lancet Oncology* **16,** 839–847 (2017/08/29 2015).

112. Martin-Montalvo, A. *et al.* Metformin improves healthspan and lifespan in mice. *Nature communications* **4,** 2192–2192 (2013).

113. Bannister, C. A. *et al.* Can people with type 2 diabetes live longer than those without? A comparison of mortality in people initiated with metformin or sulphonylurea monotherapy and matched, non-diabetic controls. *Diabetes Obes Metab* **16,** 1165–1173 (2014).

114. Marupuru, S. *et al.* Protective effect of metformin against tuberculosis infections in diabetic patients: an observational study of south Indian tertiary healthcare facility. *The Brazilian Journal of Infectious Diseases* **21,** 312–316 (2017).

115. Singhal, A. *et al.* Metformin as adjunct antituberculosis therapy. *Sci Transl Med* **6** (2014).

116. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* (2012).

117. Engström, P. G. *et al.* Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Meth* **10,** 1185–1191 (2013).

118. Anders, S., Pyl, P. T. & Huber, W. HTSeq – A Python framework to work with high-throughput sequencing data. *Bioinformatics* (2014).

119. Love, M., Huber, W. & S, A. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15** (2014).

120. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (Jan. 2010).

121. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15** (2014).

122. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biology* **17** (2016).

123. Bourgon, R., Gentleman, R. & Huber, W. Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 9546–9551 (2010).

124. Väremo, L., Nielsen, J. & Nookaew, I. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucl. Acids Res.* (2013).

125. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10** (2009).

126. Tarca, A. L., Bhatti, G. & Romero, R. A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity. *PLoS ONE* **8** (2013).

127. Stephenne, X. *et al.* Metformin activates AMP-activated protein kinase in primary human hepatocytes by decreasing cellular energy status. *Diabetologia* **54,** 3101–3110 (Dec. 2011).

128. Park, D. W. *et al.* Activation of AMPK enhances neutrophil chemotaxis and bacterial killing. *Mol Med* **19,** 387–398 (2013).

129. Warburg, O. On the origin of cancer cells. *Science* **123,** 309–314 (1956).

130. Fritsch, S. D. & Weichhart, T. Effects of Interferons and Viruses on Metabolism. *Frontiers in Immunology* **7,** 630 (2016).

131. Gideon, H. P., Skinner, J. A., Baldwin, N., Flynn, J. L. & Lin, P. L. Early Whole Blood Transcriptional Signatures Are Associated with Severity of Lung Inflammation in Cynomolgus Macaques with Mycobacterium tuberculosis Infection. *J Immunol* **197,** 4817–4828 (2016).

132. Mayer-Barber, K. D. *et al.* Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. *Nature* **511,** 99–103 (2014).

133. Kewcharoenwong, C. *et al.* Glibenclamide reduces pro-inflammatory cytokine production by neutrophils of diabetes patients in response to bacterial infection. *Sci Rep* **3** (2013).

134. Koh, G. C. K. W. *et al.* Glyburide Is Anti-inflammatory and Associated with Reduced Mortality in Melioidosis. *Clinical Infectious Diseases* **52** (2011).

135. Andreu, N. *et al.* Primary macrophages and J774 cells respond differently to infection with Mycobacterium tuberculosis. *Sci Rep* **7,** 42225 (2017).

136. Goldhaber-Fiebert, J. D., Jeon, C. Y., Cohen, T. & Murray, M. B. Diabetes mellitus and tuberculosis in countries with high tuberculosis burdens: individual risks and social determinants. *Int J Epidemiol* **40,** 417–428 (2011).

137. Alisjahbana, B. *et al.* The Effect of Type 2 Diabetes Mellitus on the Presentation and Treatment Response of Pulmonary Tuberculosis. *Clinical Infectious Diseases* **45,** 428–435 (2007).

138. Ruslami, R., Aarnoutse, R. E., Alisjahbana, B., Van Der Ven, A. J. A. M. & Van Crevel, R. Implications of the global increase of diabetes for tuberculosis control and patient care: Implications of the global increase of diabetes. *Tropical Medicine & International Health* **15,** 1289–1299 (2010).

139. Restrepo, B. I. *et al.* Type 2 diabetes and tuberculosis in a dynamic bi-national border population. *Epidemiology and Infection* **135,** 483–491 (Apr. 2007).

140. Tabak, A. G., Herder, C., Rathmann, W., Brunner, E. J. & Kivimaki, M. Prediabetes: a high-risk state for diabetes development. *Lancet* **379,** 2279–2290 (2012).

141. Baker, M. A., Lin, H.-H., Chang, H.-Y. & Murray, M. B. The Risk of Tuberculosis Disease Among Persons With Diabetes Mellitus: A Prospective Cohort Study. *Clinical Infectious Diseases* **54,** 818–825 (2012).

142. Viswanathan, V. *et al.* Prevalence of diabetes and pre-diabetes and associated risk factors among tuberculosis patients in India. *PLoS One* **7** (2012).

143. Kumar, N. P. *et al.* Coincident Pre-Diabetes Is Associated with Dysregulated Cytokine Responses in Pulmonary Tuberculosis. *PLOS ONE* **9,** 1–8 (2014).

144. Saiki, O., Negoro, S., Tsuyuguchi, I. & Yamamura, Y. Depressed immunological defence mechanisms in mice with experimentally induced diabetes. *Infect Immun* **28,** 127–131 (1980).

145. Martens, G. W. *et al.* Tuberculosis Susceptibility of Diabetic Mice. *American Journal of Respiratory Cell and Molecular Biology* **37,** 518–524 (2007).

146. Restrepo, B. I. *et al.* Tuberculosis in Poorly Controlled Type 2 Diabetes: Altered Cytokine Expression in Peripheral White Blood Cells. *Clinical Infectious Diseases* **47,** 634–641 (2008).

147. Kumar, N. P. *et al.* Type 2 diabetes mellitus is associated with altered CD8+ T and NK cell function in pulmonary tuberculosis. *Immunology* (2014).

148. Kumar, N. P. *et al.* Expansion of Pathogen-Specific T-Helper 1 and T-Helper 17 Cells in Pulmonary Tuberculosis With Coincident Type 2 Diabetes Mellitus. *Journal of Infectious Diseases* **208,** 739–748 (2013).

149. Stalenhoef, J. E. *et al.* The role of interferon-gamma in the increased tuberculosis risk in type 2 diabetes mellitus. *European Journal of Clinical Microbiology & Infectious Diseases* **27,** 97–103 (2007).

150. Golub, L. M., Nicoll, G. A., Iacono, V. J. & Ramamurthy, N. S. In Vivo Crevicular Leukocyte Response to a Chemotactic Challenge: Inhibition by Experimental Diabetes. *Infection and Immunity* **37,** 1013–1020 (Sept. 1982).

151. Lin, J.-C. *et al.* Impaired Phagocytosis of Capsular Serotypes K1 or K2 Klebsiella pneumoniae in Type 2 Diabetes Mellitus Patients with Poor Glycemic Control. *The Journal of Clinical Endocrinology & Metabolism* **91,** 3084 (2006).

152. Mazade, M. A. & Edwards, M. S. Impairment of Type III Group B Streptococcus Stimulated Superoxide Production and Opsonophagocytosis by Neutrophils in Diabetes. *Molecular Genetics and Metabolism* **73,** 259–267 (2001).

153. Chanchamroen, S., Kewcharoenwong, C., Susaengrat, W., Ato, M. & Lertmemongkolchai, G. Human polymorphonuclear neutrophil responses to Burkholderia pseudomallei in healthy and diabetic subjects. *Infect. Immun.* **77,** 456–463 (2009).

154. Restrepo, B. I., Twahirwa, M., Rahbar, M. H. & Schlesinger, L. S. Phagocytosis via complement or Fc-gamma receptors is compromised in monocytes from type 2 diabetes patients with chronic hyperglycemia. *PLoS ONE* **9** (2014).

155. Raposo-García, S. *et al.* Immunological response to Mycobacterium tuberculosis infection in blood from type 2 diabetes patients. *Immunology Letters* **186,** 41–45 (June 2017).

156. Maertzdorf, J. *et al.* Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proceedings of the National Academy of Sciences* **109,** 7853–7858 (2012).

157. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

158. Yamaguchi, K. D. *et al.* IFN-beta-regulated genes show abnormal expression in therapy-naive relapsing-remitting MS mononuclear cells: gene expression analysis employing all reported protein-protein interactions. eng. *J Neuroimmunol* **195,** 116–120 (Mar. 2008).

159. Flynn, J. L. *et al.* An essential role for interferon gamma in resistance to Mycobacterium tuberculosis infection. *The Journal of Experimental Medicine* **178,** 2249–2254 (1993).

160. Cooper, A. M. *et al.* Disseminated tuberculosis in interferon gamma gene-disrupted mice. *The Journal of Experimental Medicine* **178,** 2243–2247 (Dec. 1993).

161. Döfflnger, R. *et al.* Partial Interferon-gamma Receptor Signaling Chain Deficiency in a Patient with Bacille Calmette-Guérin and Mycobacterium abscessus Infection. *The Journal of Infectious Diseases* **181** (Jan. 2000).

162. Dorman, S. E. & Holland, S. M. Mutation in the signal-transducing chain of the interferon-gamma receptor and susceptibility to mycobacterial infection. *The Journal of Clinical Investigation* **101,** 2364–2369 (1998).

163. McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. *Nat Rev Immunol* **15,** 87–103 (2015).

164. O'Connell, R. M. *et al.* Type I Interferon Production Enhances Susceptibility to Listeria monocytogenes Infection. *The Journal of Experimental Medicine* **200,** 437–445 (Aug. 2004).

165. Manca, C. *et al.* Virulence of a Mycobacterium tuberculosis clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha /beta. *Proc Natl Acad Sci U S A* **98** (2001).

166. Dorhoi, A. *et al.* Type I IFN signaling triggers immunopathology in tuberculosis-susceptible mice by modulating lung phagocyte dynamics. *Eur J Immunol* **44,** 2380–2393 (2014).

167. Cliff, J. M., Kaufmann, S. H. E., McShane, H., van Helden, P. & O'Garra, A. The human immune response to tuberculosis and its treatment: a view from the blood. *Immunological Reviews* **264,** 88–102 (2015).

168. Desvignes, L., Wolf, A. J. & Ernst, J. D. Dynamic roles of type I and type II interferons in early infection with Mycobacterium tuberculosis. *Journal of Immunology* **188,** 6205–6215 (2012).

169. Kuchtey, J., Fulton, S. A., Reba, S. M., Harding, C. V. & Boom, W. H. Interferon-αβ mediates partial control of early pulmonary Mycobacterium bovis bacillus Calmette–Guérin infection. *Immunology* **118,** 39–49 (2006).

170. Schreiber, G. & Piehler, J. The molecular basis for functional plasticity in type I interferon signaling. *Trends Immunol* **36,** 139–149 (2015).

171. De Jong, T. D. *et al.* Physiological evidence for diversification of IFNalpha- and IFNbeta-mediated response programs in different autoimmune diseases. *Arthritis Res Ther* **18,** 49 (2016).

172. Ng, C. T., Mendoza, J. L., Garcia, K. C. & Oldstone, M. B. Alpha and beta type 1 interferon signaling: passage for diverse biologic outcomes. *Cell* **164,** 349–352 (2016).

173. Dall'Era, M., Cardarelli, P., Preston, B., Witte, A. & Davis, J. Type I interferon correlates with serological and clinical manifestations of SLE. *Annals of the Rheumatic Diseases* **64,** 1692–1697 (2005).

174. Baechler, E. C. *et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proceedings of the National Academy of Sciences* **100,** 2610–2615 (2003).

175. Tessier, M.-C. *et al.* Type 1 diabetes and the OAS gene cluster: association with splicing polymorphism or haplotype? *Journal of Medical Genetics* **43,** 129–132 (2006).

176. Giosue, S. *et al.* Effects of aerosolized interferon-alpha in patients with pulmonary tuberculosis. *Am J Respir Crit Care Med* **158,** 1156–1162 (1998).

177. Palmero, D. *et al.* Phase II trial of recombinant interferon-alpha2b in patients with advanced intractable multidrug-resistant pulmonary tuberculosis: long-term follow-up. *Int J Tuberc Lung Dis* **3,** 214–218 (Mar. 1999).

178. Giosue, S. *et al.* Aerosolized interferon-alpha treatment in patients with multi-drug-resistant pulmonary tuberculosis. *Eur Cytokine Netw* **11,** 99–104 (2000).

179. Mansoori, D., Tavana, S., Mirsaeidi, M., Yazdanpanah, M. & Sohrabpour, H. The Efficacy of Interferon-α in the Treatment of Multidrug Resistant Tuberculosis. *Tanaffos* (Sept. 2017).

180. Zarogoulidis, P. *et al.* The effect of combination IFN-alpha-2a with usual antituberculosis chemotherapy in non-responding tuberculosis and diabetes mellitus: a case report and review of the literature. *J Chemother* **24,** 173–177 (2012).

181. Belkahla, N. *et al.* Reactivation of tuberculosis during dual therapy with pegylated interferon and ribavirin for chronic hepatitis C. *Rev Med Interne* **31** (2010).

182. Pollara, G. *et al.* Validation of Immune Cell Modules in Multicellular Transcriptomic Data. *PLOS ONE* **12,** 1–13 (Jan. 2017).

183. Dooley, K. E. & Chaisson, R. E. Tuberculosis and diabetes mellitus: convergence of two epidemics. *The Lancet Infectious Diseases* **9,** 737–746 (2009).

184. Morsy, A. M., Zaher, H. H., Hassan, M. H. & Shouman, A. Predictors of treatment failure among tuberculosis patients under DOTS strategy in Egypt. *East Mediterr Health J* **9,** 689–701 (July 2003).

185. Yoon, Y. S. *et al.* The effect of diabetes control status on treatment response in pulmonary tuberculosis: a prospective study. *Thorax* (2016).

186. Singla, R. *et al.* Influence of diabetes on manifestations and treatment outcome of pulmonary TB patients. *Int J Tuberc Lung Dis* **10,** 74–79 (2006).

187. Restrepo, B. I. *et al.* Mycobacterial clearance from sputum is delayed during the first phase of treatment in patients with diabetes. *The American journal of tropical medicine and hygiene* **79,** 541–544 (Oct. 2008).

188. Mahishale, V. *et al.* Effect of Poor Glycemic Control in Newly Diagnosed Patients with Smear-Positive Pulmonary Tuberculosis and Type-2 Diabetes Mellitus. *Iranian Journal of Medical Sciences* **42,** 144–151 (2017).

189. Dooley, K. E., Tang, T., Golub, J. E., Dorman, S. E. & Cronin, W. Impact of diabetes mellitus on treatment outcomes of patients with active tuberculosis. eng. *Am J Trop Med Hyg* **80** (2009).

190. Kumar, N. P., Moideen, K., Viswanathan, V., Kornfeld, H. & Babu, S. Effect of standard tuberculosis treatment on naive, memory and regulatory T-cell homeostasis in tuberculosis–diabetes co-morbidity. *Immunology* **149,** 87–97 (2016).

191. Kumar, N. P. *et al.* Modulation of dendritic cell and monocyte subsets in tuberculosis-diabetes co-morbidity upon standard tuberculosis treatment. *Tuberculosis (Edinb)* **101,** 191–200 (Dec. 2016).

192. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12,** 357–360 (2015).

Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
ART	Anti-Retroviral Therapy
APC	Antigen Presenting Cell
BCG	Bacillus Calmette Guérin
BMI	Body Mass Index
CFU	Colony Forming Units
CVD	Cardiovascular Disease
DC	Dendritic Cells
DC-SIGN	Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin
DM	Diabetes
DOTS	Directly Observed Therapy-Short course
EBA	Early Bactericidal Assay
ESAT-6	Early Secreted Antigen-6
FcγR	Fcγ Receptor
FFA	Free Fatty Acid
FPG	Fasting Plasma Glucose
GM-CSF	Granulocyte macrophage colony-stimulating factor
HAART	Highly Active Anti-Retroviral Therapy
HIV	Human Immunodeficiency Virus
IFN	Interferon
IGRA	Inteferon Gamma Release Assay
IKK	IκB kinase

IL	Interleukin
IRIS	Immune Reconstitution Inflammatory Syndrome
LTBI	Latent TB Infection
lncRNA	long non-coding RNA
ManLAM	Mannosylated Lipoarabinomannan
MBL	Molecular Bacterial Load
MDI	Modular Discrimination Index
MDR-TB	Multidrug-Resistant Tuberculosis
MGIT	Mycobacterial Growth Indicator Tube
miRNA	micro RNA
MLPA	Multiplex Ligation Probe Assay
MR	Mannose Receptor
mRNA	messenger RNA
MS	Multiple Sclerosis
M. tb	*Mycobacterium tuberculosis*
NSAID	Non-Steriodal Anti-inflammatory Drugs
NNRTI	Non-Nucleoside Reverse Transcriptase Inhibitor
NTM	Non-Tuberculosis Mycobacteria
NtRTs	Nucleotide Reverse Transcriptase Inhibitors
OD	Other Diseases
PBMC	Peripheral Blood Mononuclear Cell
PCR	Polymerase Chain Reaction
PEP	Post-Exposure Prophylaxis

PI	Protease inhibitor
PrEP	Pre-Exposure Prophylaxis
PRR	Pattern Recognition Receptor
RCT	Randomised Control Trial
RIN	RNA Integrity Number
RPG	Resting Plasma Glucose
rRNA	ribosomal RNA
RTI	Reverse Transcriptase Inhibitor
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SLE	Systemic lupus erythematosus
ssRNA	single stranded RNA
STZ	Streptozotocin
T1DM	Type 1 Diabetes Mellitus
T2DM	Type 2 Diabetes Mellitus
TB	Tuberculosis
TB preDM	Tuberculosis pre-diabetes co-morbidity
TB/DM	Tuberculosis diabetes co-morbidity
TLR	Toll-Like Receptor
TNF	Tumour Necrosis Factor
TST	Tuberculin Skiny Test
TTP	Time To Positivity
UKPDS	UK Prospective Diabetes Study
WHO	World Health Organisation