Transcriptomic Approach to Understanding and Characterising Disease Pathogenesis in Sarcoidosis.

Yvonne René Kendrick
Worcester College





Submission for the degree of Doctor of Philosophy

Trinity Term 2017

Supervisor: Professor Ling-Pei Ho

Yvonne Rene Kendrick, Worcester College, University of Oxford
Thesis submitted for the degree of Doctor of Philosophy, Trinity term 2017.

Abstract

Transcriptomic Approach to Understanding and Characterising Disease Pathogenesis in Sarcoidosis.

Sarcoidosis is a multisystem immunological disorder, centred on the dysregulation of T cell activity, macrophage activation and granuloma formation. Most pulmonary sarcoidosis patients have a good prognosis, but up to 20% can go on to develop lung fibrosis. Previous research in sarcoidosis has generally focused on sarcoidosis as a single disease entity, whereas this thesis has focused on the importance of phenotypical characterisation, in particular active fibrotic pulmonary disease, and the standardisation of measuring disease activity.

I first developed a scoring system to measure disease activity based on typical findings on high resolution CT scanning (CTAS). I then identified a set of readily available blood markers and a chest radiographic score as a surrogate for the CT activity score, which enables the physician to easily assess the level of disease activity in the clinic setting (SCAS).

Using a transcriptomic approach, I used the activity score to identify the immunological pathways and individual genes correlating directly with increasing activity, both at the site of disease and in the periphery. Enrichment of immune response and defense response gene sets, particularly in BAL cell samples, were observed as disease activity increased, with CCR2, TREM2, and CCL18 featuring as prominent genes involved in the propagation of activity. CXCL9 and CXCL10 were identified as peripheral markers of active disease reflecting the activity in various compartments. In addition, after phenotyping patients with active disease into those with and without fibrosis, differential gene expression showed enrichment of immune related gene sets in the fibrotic cohort, and highlighted the roles of several significantly upregulated genes including CCL19, CCR7, CXCR3, CCR2, ADAMDEC1 and MMP12 in the pathogenesis of fibrotic sarcoidosis. Finally, enrichment of an IPF signature derived from differential expression to sarcoidosis, was observed in the fibrotic sarcoidosis cohort. This information may lend itself to further therapeutic treatments options already used in other fibrotic lung conditions in this interesting and understudied phenotype.

Acknowledgements

Firstly, I would like to thank my supervisor Professor Ling-Pei Ho, who has displayed an endless amount of energy, motivation and drive in the pursuit of scientific research, and guided me throughout my DPhil studies.

I would like to thank all those involved in this study, with particular thanks to Dr. Rachel Benamore, Dr. Emma Helm, Dr. Helen Lockstone, Dr. Karene Argoud, Dr Eleni Giannoulatou and Dr. Steve Taylor for their assistance with various aspects of this work. I would like to thank past and present members of the Ho group – Anjali, Suzanne, Hong, Sharon and Emily for their support, friendship, and laboratory advice.

Finally, I would like to thank my husband Ben, who has lived the work of my thesis as if it were his own, never complaining, and always supporting, and my two dear children, Barnaby and Georgia, without whom this thesis would have been completed long ago, but the extra challenge and constant distraction for its completion were enjoyed.

Publications arising from this thesis

Benamore R, **Kendrick Y R**, Repapi E, Helm E, Cole S L, Taylor S, Ho L P. CTAS: a CT score to quantify disease activity in pulmonary sarcoidosis. Thorax 2016;**71**(12):1161-63.

Kendrick Y, Helm E, Benamore R, Ho LP. A Combination Score Of Raised sACE, Lymphopenia And Hypergammaglobulinemia Correlates With Active Disease On Thoracic CT Scan Of Patients With Pulmonary Sarcoidosis. Am J Respir Crit Care Med 2013; 187:A1391.

Contents

ABSTRACT ..... **I**

ACKNOWLEDGEMENTS ..... **II**

PUBLICATIONS ARISING FROM THIS THESIS ..... **III**

LIST OF FIGURES ..... **VII**

LIST OF TABLES ..... **XI**

1 INTRODUCTION **1**

1.1 SARCOIDOSIS 1

1.1.1 History *1*

1.1.2 Epidemiology *2*

1.1.3 Aetiology *5*

1.1.4 Immune pathology *7*

1.1.5 Diagnosis *12*

1.1.6 Treatment *14*

1.2 DISEASE PHENOTYPE 17

1.3 DISEASE ACTIVITY 21

1.4 LUNG FIBROSIS 29

1.5 GENE EXPRESSION PROFILING 32

1.6 PROJECT AIMS 35

2 MATERIALS AND METHODS **37**

2.1 PATIENT RECRUITMENT 37

2.2 RADIOLOGY METHODS 39

2.2.1 HRCT Protocol *39*

2.3 LABORATORY METHODS 39

2.3.1 Isolation of human peripheral blood mononuclear cells *39*

2.3.2 Isolation of human bronchoalveolar lavage cells *40*

2.3.3 Isolation of RNA from transbronchial lung tissue *41*

2.3.4 Serum Collection *41*

2.3.5 Flow Cytometry *42*

2.3.6 Multiplex-cytokine detection from serum *44*

2.3.7 Enzyme-linked immunosorbent assay (ELISA) *44*

2.3.8 RNA extraction *45*

2.4 MICROARRAY ANALYSIS 46

2.5 STATISTICAL ANALYSIS 47

3 MEASURING DISEASE ACTIVITY USING CT SCANS – CTAS **48**

3.1 INTRODUCTION 48

3.2 AIM 50

3.3 METHODS 50

3.3.1 Patient selection and study design *50*

3.3.2 HRCT protocol, lung function and serum markers *51*

3.3.3 Scoring methods for HRCT abnormalities *51*

3.3.4 Statistical methods *55*

3.4 RESULTS 56

3.4.1 Patients demographics *56*

3.4.3 HRCT findings in patient cohort *57*

3.4.4 Correlation of individual CT abnormality score with markers of disease activity and burden *57*

3.4.5 CTAS predicts response to treatment *60*

3.4.6 Reproducibility of CTAS between 2 radiologists *61*
3.5 DISCUSSION 62
3.6 CONCLUSION 66

4 SARCOIDOSIS CLINICAL ACTIVITY SCORE (SCAS) – DESIGN, DEVELOPMENT AND VALIDATION OF A CLINIC BASED TOOL 67
4.1 INTRODUCTION 67
4.2 AIM 71
4.3 METHODS 71
4.3.1 Patient selection *71*
4.3.2 HRCT Scanning protocol and clinical parameter measurements. *73*
4.3.3 Scoring method for chest radiograph (RCAS) *73*
4.3.4 Scoring for serum and cellular biomarkers (CCAS) *74*
4.3.5 Sarcoidosis Clinical Activity Score (SCAS) *75*
4.3.6 CTAS score *75*
4.3.7 Laboratory methods used in validation subsets *76*
4.3.8 Statistical Analysis *77*
4.4 RESULTS 77
4.4.1 Patient demographics *77*
4.4.2 Correlation of measured parameters to CTAS *82*
4.4.3 Correlation of SCAS to cell phenotype, cytokine production, and serum proteins *85*
4.4.4 Effects of treatment on SCAS *90*
4.4.5 Inter- and Intra-observer correlation for SCAS *90*
4.5 DISCUSSION 91
4.6 CONCLUSION 96

5 INVESTIGATION OF GENES UPREGULATED IN HIGH DISEASE ACTIVITY IN PULMONARY SARCOIDOSIS. 97
5.1 INTRODUCTION 97
5.2 AIM 99
5.3 METHODS 99
5.3.1 Patient selection *99*
5.3.2 Experimental design *100*
5.3.3 Data analysis of gene expression profiles *101*
5.3.4 Data Interpretation *103*
5.4 RESULTS 105
5.4.1 Patient demographics *105*
5.4.2 Quality control of microarray samples *107*
5.4.3 Correlation of gene expression profile of BAL cells to disease activity *109*
5.4.4 Correlation of gene expression profile of PBMCs to disease activity *122*
5.4.5 Comparison of significant genes, gene sets, pathways and processes from different sampling modalities. *132*
5.5 DISCUSSION 137
5.6 CONCLUSION 145

6 INVESTIGATION OF THE PROCESSES, PATHWAYS AND GENES INVOLVED IN FIBROTIC SARCOIDOSIS. 146
6.1 INTRODUCTION 146
6.2 AIM 149
6.3 METHODS 150
6.3.1 Patient selection *150*
6.3.2 Experimental design *150*
6.3.3 Quality control of microarray samples *152*
6.3.4 Data analysis of gene expression profiles *153*

6.3.5 Data Interpretation *154*
6.4 RESULTS 156
6.4.1 Patient demographics *156*
6.4.2 Analysis of pathways, processes and genes contributing to active fibrotic sarcoidosis in lung samples *157*
6.4.3 Analysis of pathways, processes and genes contributing to active fibrotic sarcoidosis in PBMC samples *174*
6.4.4 Analysis of pathways, processes and genes contributing to active fibrotic sarcoidosis in BAL cell samples *179*
6.5 DISCUSSION 185
6.6 CONCLUSION 195
7 FINAL DISCUSSION AND FUTURE DIRECTION 196
7.1 MAIN FINDINGS 196
7.2 LIMITATIONS 207
7.3 FUTURE WORK 209
8 APPENDICES 212
8.1 DEFINITION OF TERMS FOR ABNORMALITIES ON HRCT SCORING PROFORMA. ALL FROM FLEISCHNER GLOSSARY OF TERMS APART FROM ASTERISKED. 212
8.2 DEFINITION OF ZONES ON HRCT SCORING PROFORMA 214
8.3 GRAPHICAL REPRESENTATION OF TABLE 3.6 215
8.4 GSE ACCESSION DETAILS FOR GSEA ANALYSIS 216
8.5 KEGG_CHEMOKINE_SIGNALING_PATHWAY GENE SET 217
8.6 IPF V SARCOIDOSIS GENE SIGNATURE (DOWNREGULATED) – LENG ET AL (279) 218
8.7 IPF V SARCOIDOSIS GENE SIGNATURE (UPREGULATED) – LENG ET AL (279) 219
9 REFERENCES 220

List of Figures

FIGURE 1.1 IMMUNOPATHOGENESIS OF SARCOIDOSIS. INFECTIOUS, ORGANIC AND INORGANIC PARTICLES ARE POSSIBLE ANTIGENS IN SARCOIDOSIS. AFTER PRESENTATION BY ANTIGEN PRESENTING CELLS TO CD4+ T CELLS, RELEASE OF CYTOKINES INCLUDING TNF α, IL-12, -15, -18, MCP-1, MIP AND GM-CSF OCCURS. THE EFFICIENCY OF ANTIGEN PROCESSING, PRESENTATION AND CYTOKINE RELEASE IS PROBABLY UNDER GENETIC GOVERNANCE, AND MACROPHAGE HLA AND BTNL2 ALLELES HAVE STRONG ASSOCIATIONS WITH DISEASE SUSCEPTIBILITY. THE ACTIVATED CD4+ T CELLS UNDERGO DIFFERENTIATION AND PROLIFERATION, WITH A BIAS TOWARDS TH1 CELLS AND RELEASE OF IL-2 AND IFN-γ. GRANULOMA FORMATION OCCURS, WHICH MAY PERSIST, RESOLVE OR PROGRESS TO FIBROSIS. FROM IANNUZZI MC ET AL (51) 8
FIGURE 1.2 THE NINE CLINICAL PHENOTYPES RELATING TO OUTCOME AS PROPOSED BY JUDSON ET AL (114) 20
FIGURE 1.3 PHASES OF WOUND REPAIR AND THE POTENTIAL MECHANISMS FOR FIBROGENESIS AFTER LUNG INJURY, EPITHELIAL CELLS RELEASE INFLAMMATORY MEDIATORS THAT INITIATE AN ANTIFIBRINOLYTIC COAGULATION CASCADE, WHICH TRIGGERS PLATELET ACTIVATION AND BLOOD CLOT FORMATION. THIS IS FOLLOWED BY ENTRY OF LEUKOCYTES (E.G., NEUTROPHILS, MACROPHAGES, AND T CELLS). THE RECRUITED LEUKOCYTES SECRETE PROFIBROTIC CYTOKINES SUCH AS IL-1B, TNF, IL-13, AND TGF-B. THE ACTIVATED MACROPHAGES AND NEUTROPHILS ALSO REMOVE DEAD CELLS AND ELIMINATE ANY INVADING ORGANISMS. IN THE SUBSEQUENT PHASE, FIBROCYTES FROM THE BONE MARROW AND RESIDENT FIBROBLASTS PROLIFERATE AND DIFFERENTIATE INTO MYOFIBROBLASTS, WHICH RELEASE EXTRACELLULAR MATRIC (ECM) COMPONENTS. FIBROBLASTS AND MYOFIBROBLASTS MAY ALSO BE DERIVED FROM EPITHELIAL CELLS UNDERGOING EPITHELIAL-MESENCHYMAL TRANSITION (EMT). IN THE FINAL REMODELING AND RESOLUTION PHASE, ACTIVATED MYOFIBROBLASTS CAN PROMOTE WOUND REPAIR, LEADING TO WOUND CONTRACTION AND RESTORATION OF BLOOD VESSELS. HOWEVER, FIBROSIS OFTEN DEVELOPS IF ANY STAGE IN THE TISSUE REPAIR PROGRAM IS DYSREGULATED OR WHEN THE LUNG-DAMAGING STIMULUS PERSISTS. FROM WYNN ET AL (156). 30
FIGURE 2.1 PHENOTYPING ALGORITHM USED TO RECRUIT SARCOIDOSIS PATIENTS FOR STUDY. GREEN PATHWAY HIGHLIGHTS THOSE QUALIFYING FOR RECRUITMENT. 38
FIGURE 3.1 A. DISTRIBUTION OF CTAS FOR THE PATIENT COHORT (N=97). B. CTAS IN NON-FIBROTIC AND FIBROTIC PULMONARY SARCOIDOSIS. P VALUE MEASURED BY MANN-WHITNEY TEST. 59
FIGURE 3.2 A-C CORRELATION OF CTAS WITH CONTEMPORANEOUS LUNG FUNCTION IN ALL PATIENTS. R AND P VALUE MEASURED BY SPEARMAN CORRELATION TEST 60
FIGURE 3.3 ASSOCIATION BETWEEN CTAS AND CHANGES IN FVC AT 12 MONTHS IN ALL PATIENTS (A) AND IN THE FIBROTIC SUBSET (B) WHO RECEIVED NEW OR INCREASED TREATMENT AFTER CT SCAN. FOR ALL PATIENTS R AND P VALUES SHOWN ARE FROM LINEAR REGRESSION CALCULATIONS. SPEARMAN CORRELATION VALUES ARE (A-B) R=0.40, P=0.03 AND R=0.42, P=0.02 RESPECTIVELY. 61
FIGURE 3.4 A. AGREEMENT BETWEEN TWO RADIOLOGISTS FOR CTAS. PEARSON CORRELATION ANALYSIS (N=85), AND B. BLAND ALTMAN TEST. MEAN DIFFERENCES (BIAS) OF -0.4, AND 95% LIMIT OF AGREEMENTS (PRECISION) (-8.5 TO 7.7) ARE SHOWN. 62
FIGURE 4.1 FLOWCHART DEPICTING SCORING SYSTEM FOR CHEST RADIOGRAPHS. FIRST A SCORE FOR ABNORMALITY IS GIVEN (RANGE 0-2). IF NODULES OR CONSOLIDATION IS PRESENT (ABNORMALITY SCORE 2), THEN A FURTHER SCORE IS APPLIED TO ACCOUNT FOR EXTENT. EACH LUNG (LEFT AND RIGHT) IS DIVIDED INTO THREE ZONES, AND FOR EACH ZONE THAT THE ABNORMALITY IS PRESENT IN A FURTHER ADDITIONAL POINT IS GIVEN (RANGE 1-6). A MAXIMUM SCORE OF 8 CAN BE ACHIEVED WITH THE CHEST RADIOGRAPH SCORE. 74
FIGURE 4.2 RANGE OF SERUM ACE VALUES (IU/L) SEEN IN THE 65 RECRUITED PATIENTS. EACH DOT REPRESENTS AN INDIVIDUAL PATIENT. 75
FIGURE 4.3 DISTRIBUTION OF SCAS ACROSS THE TWO COHORTS. A – WITHIN 90 DAYS OF PARAMETER MEASUREMENTS; B – WITHIN 35 DAYS OF PARAMETER MEASUREMENTS. 78
FIGURE 4.4 NUMBER OF PATIENTS FROM N=65 COHORT THAT HAVE A RAISED SERUM ACE, RAISED IMMUNOGLOBULIN G OR LYMPHOPAENIA. 80
FIGURE 4.5 CTAS VERSUS SCAS WHEN PARAMETERS ARE MEASURED WITHIN 35 DAYS OF EACH OTHER. 83

FIGURE 4.6 CTAS VERSUS SCAS WHEN PARAMETERS ARE MEASURED WITHIN 90 DAYS OF EACH OTHER. 84
FIGURE 4.7 A - GRAPH DEPICTS STRONG CORRELATION OF CD14+ MONOCYTES AS A PERCENTAGE OF LIVE PBMCS VERSUS SCAS. FACS PLOTS ABOVE SHOW GATING STRATEGY FOR CD14+ CELLS; B:- FIRST ALL PBMCS ARE GATED; C:- CD14+ CELLS ARE GATED AND PERCENTAGE OF ALL CALCULATED. 86
FIGURE 4.8. A - SIGNIFICANT POSITIVE CORRELATION BETWEEN PERCENTAGE OF TNFA EXPRESSING UNSTIMULATED CD4+ CELLS AND SCAS. FACS PLOTS ABOVE SHOW GATING STRATEGY FOR CD4+ CELLS – B:- LYMPHOCYTE POPULATION GATED, C:- CD8 +VE AND –VE POPULATIONS GATED, THE CD8-VE POPULATION IS THEN USED WHEN ASSESSING TNF ALPHA PRODUCTION WITHOUT (C) AND WITH (D) PMA/IONOMYCIN STIMULATION. 87
FIGURE 4.9 – STRONG POSITIVE CORRELATION OF SCAS WITH SERUM SIL2-R. 88
FIGURE 4.10 CORRELATION BETWEEN SCAS AND SERUM CCL18 AS MEASURED BY ELISA 89
FIGURE 4.11 CHANGE IN SCAS FOLLOWING TREATMENT BETWEEN TWO CLINIC VISITS (81 TO 279 DAYS APART)(P = 0.004 ANOVA) 90
FIGURE 4.12 A – INTER-OBSERVER AGREEMENT BETWEEN TWO CLINICIANS SCORING SCAS FOR N=30 PATIENTS; B – INTRA-OBSERVER AGREEMENT FROM THE SAME CLINICIAN SCORING THE SAME PATIENTS AT PRESENTATION 1 YEAR APART. 91
FIGURE 5.1 VENN DIAGRAM SHOWING OVERLAP IN PBMC AND BAL SAMPLING GROUPS. 106
FIGURE 5.2 THE RANGE OF SCAS VALUES IN THE BAL AND PBMC COHORT RESPECTIVELY 106
FIGURE 5.3 DENDOGRAM SHOWING CLUSTERING OF SAMPLES IN EACH SAMPLING MODALITY (A = PBMC, B= BAL CELL), AND IDENTIFICATION OF OUTLIERS PRIOR TO RMA NORMALIZATION (RH438 IN BLOOD AND TL305 IN BAL). 108
FIGURE 5.4 ENRICHMENT PLOT OF TOP GENE SET REVEALED BY GSEA ANALYSIS OF BAL GENE EXPRESSION CORRELATED TO INCREASING SCAS. THE ENRICHMENT PLOT ILLUSTRATES INCREASING SIGNAL WITH A LEFT SHIFT INDICATING ENRICHMENT AS THE ACTIVITY INCREASES. THE RED-BLUE STRIP REPRESENTS THE 15534 GENES EXPRESSED IN THE SAMPLES, RANKED ACCORDING TO CORRELATION OF EXPRESSION TO INCREASING SCAS. EACH LINE IN THE 'BAR CODE' REPRESENTS PRESENCE OF A GENE FROM THE GO-DEFINED GENE-SET (E.G. IMMUNE RESPONSE GENE SET) WITHIN THE RANKED GENES. THE ENRICHMENT SCORE (ES) IS A CUMULATIVE STATISTIC, WHICH INCREASES WHEN THE GENE IN THAT PARTICULAR GENE-SET IS FOUND IN THE RANKED GENE LIST AND DECREASES OTHERWISE. THE CHANGES ARE WEIGHTED SUCH THAT THE MAGNITUDE IS GREATER IF THERE IS A STRONG CORRELATION OF THE GENE FROM THE DEFINED GENE-SET WITH PHENOTYPE (I.E. FOUND TOWARDS THE EXTREME ENDS OF THE RANKED LIST). THE LEADING-EDGE SUBSET IS DEFINED AS GENES APPEARING AT OR BEFORE THE CUMULATIVE SUM REACHES MAXIMUM DEVIATION FROM ZERO. 111
FIGURE 5.5 LEADING EDGE ANALYSIS OF TOP TEN ENRICHED GENE SETS IN GSEA OF RANKED BAL GENE EXPRESSION CORRELATED TO SCAS. EVERY 6TH GENE, RANKED BY THE NUMBER OF GENE SETS THEY ARE FOUND IN, IS SHOWN ON THE X-AXIS. THE INDIVIDUAL GENES WHICH WERE REPRESENTED IN 4 OR MORE OF THE 10 GENE SETS ANALYSED ARE SHOWN ALONGSIDE THE HISTOGRAM. 113
FIGURE 5.6 ENRICHMENT PLOT OF TOP NEGATIVELY ENRICHED GENE SET REVEALED BY GSEA ANALYSIS OF BAL GENE EXPRESSION CORRELATED TO INCREASING SCAS. THE ENRICHMENT PLOT ILLUSTRATES ENRICHMENT WITH A RIGHT SHIFT AS THE ACTIVITY DECREASES (I.E. NEGATIVE CORRELATION TO INCREASING SCAS). THE RED-BLUE STRIP REPRESENTS THE 15534 GENES EXPRESSED IN THE SAMPLES, RANKED ACCORDING TO CORRELATION OF EXPRESSION TO INCREASING SCAS AS DESCRIBED BEFORE. THE LEADING-EDGE SUBSET IS DEFINED AS GENES APPEARING AT OR BEFORE THE CUMULATIVE SUM REACHES MAXIMUM DEVIATION FROM ZERO. 115
FIGURE 5.7 LEADING EDGE ANALYSIS OF TOP TEN NEGATIVELY ENRICHED GENE SETS IN GSEA OF RANKED BAL GENE EXPRESSION CORRELATED TO SCAS. EVERY 6TH GENE, RANKED BY THE NUMBER OF GENE SETS THEY ARE FOUND IN, ARE SHOWN ON THE X-AXIS. THE INDIVIDUAL GENES WHICH WERE REPRESENTED IN 4 OR MORE OF THE 10 GENE SETS ANALYSED ARE SHOWN ALONGSIDE THE HISTOGRAM. 116
FIGURE 5.8 A) ENRICHMENT PLOT FOR THE TOP GENE SET IN PBMC GSEA ANALYSIS – AMINO ACID CATABOLIC PROCESS. B) ENRICHMENT PLOT FOR THE DEFENSE RESPONSE GENE SET, WHICH IS ALSO PRESENT IN THE BAL CELL GSEA ANALYSIS. 124
FIGURE 5.9 LEADING EDGE ANALYSIS OF TOP TEN ENRICHED GENE SETS IN GSEA, REVEALING IDO1 AS THE TOP LEADING EDGE GENE. LEADING EDGE GENES PRESENT IN FOUR OR MORE GENE SETS ARE SHOWN ALONGSIDE THE HISTOGRAM. 126

FIGURE 5.10 OVERLAP BETWEEN GENES WITH CORRELATION COEFFICIENT >0.5 AND P VALUE <0.05 FROM THE TWO SAMPLING MODALITIES, BAL AND PBMC (I.E. POSITIVE CORRELATION AS SCAS INCREASES). 133
FIGURE 5.11 OVERLAP OF GENES FROM 'DEFENSE RESPONSE' GENE SET IN PBMC AND BAL. THE 33 GENES THAT OVERLAP ARE SHOWN NEXT TO VENN DIAGRAM. 135
FIGURE 5.12 OVERLAP OF GENES FROM GO - IMMUNE SYSTEM PROCESS (METACORE) BETWEEN BAL AND PBMC. THE 34 GENES THAT OVERLAP ARE SHOWN NEXT TO VENN DIAGRAM. 136
FIGURE 5.13 OVERLAP OF GENES FROM GO - IMMUNE RESPONSE (METACORE) BETWEEN PBMC AND BAL. THE 24 GENES THAT OVERLAP ARE SHOWN NEXT TO VENN DIAGRAM. 136
FIGURE 6.1 PRINCIPAL COMPONENT ANALYSIS OF GENE EXPRESSION IN LUNG TISSUE FOR FIBROTIC VERSUS NON-FIBROTIC SAMPLES. THIS SHOWS STRONG SIMILARITIES BETWEEN THE THREE ACTIVE FIBROTIC PATIENTS (F1-3) IN THE COHORT. 157
FIGURE 6.2 LEADING EDGE ANALYSIS OF TOP TEN ENRICHED GENE SETS IN GSEA OF LUNG TISSUE GENE EXPRESSION RANKED ACCORDING TO FOLD CHANGE OF DIFFERENTIAL EXPRESSION IN THE ACTIVE FIBROTIC GROUP COMPARED TO THE ACTIVE NON-FIBROTIC GROUP. EVERY 12TH GENE, RANKED BY THE NUMBER OF GENE SETS THEY ARE FOUND IN, IS SHOWN ON THE X-AXIS. THE TOP 50 GENES ARE SHOWN ALONGSIDE THE HISTOGRAM. 159
FIGURE 6.3 HEATMAP DISPLAYING TOP 50 UPREGULATED DIFFERENTIALLY EXPRESSED GENES IN THE ACTIVE FIBROSIS COHORT IN LUNG SAMPLES. NOTE ADAMDEC1 AND MMP12 ARE 3RD AND 4TH IN THE LIST. (F1-3 ACTIVE FIBROSIS SAMPLES, NF1-12 ACTIVE NON FIBROSIS SAMPLES)....161
FIGURE 6.4 GENE EXPRESSION DATA FROM LUNG SAMPLES PLOTTED FOR ADAMDEC1 (MANN-WHITNEY TEST P = 0.004) AND MMP12 (MANN-WHITNEY TEST P = 0.018) DISPLAYING THE DIFFERENCE BETWEEN ACTIVE FIBROSIS (F) AND ACTIVE NO FIBROSIS (NF) 161
FIGURE 6.5 HEATMAP DISPLAYING TOP 50 DOWN-REGULATED DIFFERENTIALLY EXPRESSED GENES IN THE ACTIVE FIBROSIS COHORT IN LUNG SAMPLES. (F1-3 ACTIVE FIBROSIS SAMPLES, NF1-12 ACTIVE NON FIBROSIS SAMPLES). 162
FIGURE 6.6 TOP 5 GENE SETS IN >2FC OR <-2FC DIFFERENTIALLY EXPRESSED GENE LIST IN ACTIVE FIBROTIC SARCOIDOSIS (N=3) WHEN COMPARED TO ACTIVE NON-FIBROTIC SARCOIDOSIS (N=12). YELLOW – HIGH EXPRESSION, BLUE – LOW EXPRESSION. FOR THE ENRICHMENT PLOTS THE LEFT HAND LEADING EDGE REFLECTS THOSE GENES HIGHLY EXPRESSED IN THE ACTIVE FIBROTIC COHORT (I.E. THOSE DISPLAYED IN YELLOW ON THE HEATMAP). 163
FIGURE 6.7 ENRICHMENT PLOTS SHOWING THE TWO SIGNIFICANTLY NEGATIVELY ENRICHED GENE SETS IN THE ACTIVE FIBROTIC COHORT FROM LUNG SAMPLES. 164
FIGURE 6.8 IPF SIGNATURES OBTAINED FROM GEO FROM VARIOUS SOURCES AND THEIR ENRICHMENT IN THE ACTIVE FIBROTIC SARCOIDOSIS LUNG COHORT. THE LEFT HAND SIDE OF THE ENRICHMENT PLOT REPRESENTS UPREGULATION IN THE ACTIVE FIBROTIC GROUP. 166
FIGURE 6.9 A) IPF V SARCOIDOSIS UPREGULATED GENE SIGNATURE SHOWING SIGNIFICANT ENRICHMENT IN THE ACTIVE FIBROTIC SARCOIDOSIS COHORT. B) IPF V SARCOIDOSIS DOWNREGULATED GENE SIGNATURE SHOWING SIGNIFICANT ENRICHMENT IN THE ACTIVE NON-FIBROTIC SARCOIDOSIS COHORT. 168
FIGURE 6.10 ENRICHMENT PLOTS FOR VARIOUS M1 AND M2 GENE SIGNATURES OBTAINED FROM THE GEO, SHOWING ENRICHMENT FOR BOTH MACROPHAGE PHENOTYPES, BUT MORE PROMINENCE AND SIGNIFICANCE IN THE M1 SIGNATURES. NB 'ALL M2 GENES' IS A COMPOSITE LIST OF ALL GENES FROM THE FOUR CONTRIBUTING GEO M2 SIGNATURES SHOWN ABOVE. 170
FIGURE 6.11 GENE SIGNATURE FOR COMPLICATED SARCOIDOSIS AS DESCRIBED BY ZHOU ET AL. NO ENRICHMENT IS SEEN IN THE ACTIVE FIBROTIC COHORT 171
FIGURE 6.12 ENRICHMENT PLOTS FOR TH1, TH2 AND TH17 GENE SIGNATURES, SHOWING NO SPECIFIC TH PREDOMINANCE IN THE ACTIVE FIBROTIC SARCOIDOSIS LUNG COHORT. ALONGSIDE THE TABLES SHOW THE ENRICHMENT SCORES, NORMALIZED ENRICHMENT SCORES, NOMINAL P VALUES AND FDR Q VALUES FOR EACH SIGNATURE TESTED. 172
FIGURE 6.13 ENRICHMENT OF THE CHEMOKINE SIGNALING PATHWAY (KEGG) GENE SET IS DEMONSTRATED IN THE ACTIVE FIBROTIC COHORT COMPARED TO ACTIVE NO FIBROSIS. 70/161 GENES SHOWED CORE ENRICHMENT. 173
FIGURE 6.14 PRINCIPAL COMPONENT ANALYSIS OF GENE EXPRESSION IN PBMC FOR ACTIVE FIBROTIC VERSUS ACTIVE NON-FIBROTIC SAMPLES. (F1-5 ACTIVE FIBROSIS; NF1-14 ACTIVE NO FIBROSIS) 174
FIGURE 6.15 HEATMAP DEPICTING TOP 50 MOST POSITIVELY AND NEGATIVELY DIFFERENTIALLY EXPRESSED GENES IN THE ACTIVE FIBROTIC COHORT COMPARED TO THE ACTIVE NON-FIBROTIC COHORT IN BLOOD SAMPLES FROM SARCOIDOSIS PATIENTS. YELLOW – HIGH EXPRESSION, BLUE – LOW EXPRESSION. F = FIBROTIC PATIENTS, NF = NON-FIBROTIC PATIENTS. 177

FIGURE 6.16 A) ENRICHMENT PLOT OF SIGNATURE TH2 V TH17 IN BLOOD COHORT RANKED ACCORDING TO FOLD CHANGE OF DIFFERENTIAL EXPRESSION IN ACTIVE FIBROSIS VERSUS ACTIVE NO FIBROSIS AND B) ENRICHMENT PLOT OF SIGNATURE TH1 V TH17 IN BLOOD COHORT RANKED ACCORDING TO FOLD CHANGE OF DIFFERENTIAL EXPRESSION IN ACTIVE FIBROSIS VERSUS ACTIVE NO FIBROSIS. BOTH SHOW TH17 GENES TO BE SIGNIFICANTLY UPREGULATED IN THE ACTIVE NO-FIBROSIS COHORT. 178
FIGURE 6.17 ENRICHMENT PLOT OF CHEMOKINE SIGNALING PATHWAY GENE SET OBTAINED FROM KEGG SHOWING UPREGULATION IN THE BLOOD SAMPLES OF ACTIVE FIBROTIC PATIENTS. 179
FIGURE 6.18 PRINCIPAL COMPONENT ANALYSIS OF GENE EXPRESSION IN BAL CELLS FOR ACTIVE FIBROTIC (F1-F3) VERSUS ACTIVE NON-FIBROTIC (NF1-NF5) SAMPLES. 180
FIGURE 6.19 HEATMAP DEPICTING TOP 50 MOST POSITIVELY AND NEGATIVELY DIFFERENTIALLY EXPRESSED GENES IN THE ACTIVE FIBROTIC COHORT COMPARED TO THE ACTIVE NON-FIBROTIC COHORT IN BAL CELL SAMPLES FROM SARCOIDOSIS PATIENTS. YELLOW – HIGH EXPRESSION, BLUE – LOW EXPRESSION. F = FIBROTIC PATIENTS, NF = NON-FIBROTIC PATIENTS. 183
FIGURE 6.20 ENRICHMENT PLOTS FOR FOUR M1 GENE SIGNATURES AND THEIR ASSOCIATED ES, NES, NOMINAL P VALUE AND FDRQ VALUE, SHOWING UPREGULATION IN THE BAL CELLS OF ACTIVE FIBROTIC PATIENTS. 184
FIGURE 7.1 PROGRAMMING OF MACROPHAGES INTO A FUSION COMPETENT STATUS (FROM HELMING ET AL 2009 (314)) IL4 AND IL13 BIND TO THEIR COMMON IL4 RECEPTOR AND INDUCE A FUSION-COMPETENT STATUS THROUGH ACTIVATION OF THE TRANSCRIPTION FACTOR STAT6. FUSION COMPETENCY HOWEVER ALSO REQUIRES MACROPHAGE TO MACROPHAGE CONTACT, POSSIBLY DEPENDENT ON CCL2 INDUCED CHEMOTAXIS. 202

List of Tables

TABLE 2.1 TABLE OF ANTIBODIES AND ISOTYPE CONTROLS USED IN FLOW CYTOMETRY ANALYSIS.... 43
TABLE 3.1 PROFORMA OF ALL RELEVANT CT ABNORMALITIES FOR PULMONARY SARCOIDOSIS, AND EXTENT OF THE ABNORMALITIES. FULL DEFINITIONS ARE FOUND IN APPENDIX 8.1. FULL DEFINITION OF ZONES ARE FOUND IN APPENDIX 8.2. 'CENTRAL'- PREDOMINANTLY OCCUPYING INNER 2/3 OF THE LUNG. 'PERIPHERAL' - PREDOMINANTLY OCCUPYING OUTER 1/3 OF THE LUNG OR ALONG THE INTER-LOBAR FISSURES. 'DORSAL' - PREDOMINANTLY POSTERIOR PORTION OF THE LUNG. 'VENTRAL'- PREDOMINANTLY ANTERIOR PORTION OF THE LUNG. 'DEPENDENT' - PREDOMINANTLY POSTERIOR PORTION OF THE LUNG WITH A RELATIVELY HORIZONTAL ANTERIOR MARGIN. 53
TABLE 3.2 DEFINITION OF TERMS FOR ABNORMALITIES ON HRCT. ALL FROM FLEISCHNER GLOSSARY OF TERMS (186) APART FROM ASTERISKED (185). 54
TABLE 3.3 (A) SCORING SYSTEM FOR HRCT ABNORMALITIES, AND (B) DEFINITION OF 'ZONES' USED TO MEASURE EXTENT. GGO – GROUND GLASS ABNORMALITIES. NODULES THAT WERE SUB-1MM WERE CLASSIFIED UNDER GGO. 54
TABLE 3.4 DEMOGRAPHIC DATA FOR PATIENT COHORT. PULMONARY FUNCTION TEST REFERS TO % PREDICTED FOR AGE, GENDER AND HEIGHT. UNLESS STATED, % IN PARENTHESES REFER TO % OF COHORT. 56
TABLE 3.5 RANGE OF CT ABNORMALITY EXTENT SCORES IN PATIENT COHORT 57
TABLE 3.6 CORRELATION BETWEEN DISEASE ACTIVITY SURROGATES AND INDIVIDUAL CT ABNORMALITIES. R AND P VALUE CALCULATED USING SPEARMAN RANK CORRELATION TEST. SIGNIFICANT CORRELATIONS ARE SHOWN IN RED. 58
TABLE 4.1 SCORES ATTRIBUTED TO SERUM/CELLULAR BIOMARKERS, TAKEN TO REFLECT DIFFERING PATHWAYS OF ACTIVE DISEASE. SCORE IS 0 IF RESULTS ARE WITHIN NORMAL LIMITS, 1 IF ABNORMAL. SACE IS TIERED THAT A SCORE OF 2 IS GIVEN IF LEVELS ARE VERY HIGH (>100). (IGG - IMMUNOGLOBULIN G; SACE – SERUM ANGIOTENSIN CONVERTING ENZYME). 75
TABLE 4.2 DEMOGRAPHICS AND CLINICAL DATA OF PULMONARY SARCOIDOSIS PATIENT COHORT FOR SCAS V CTAS 79
TABLE 4.3 DEMOGRAPHICS TABLE OF ALL VALIDATION COHORTS FOR SCAS. 81
TABLE 4.4 SPEARMAN RANK CORRELATION OF SCORED/MEASURABLE CLINICAL PARAMETERS, ALONE AND IN COMBINATION, VERSUS CTAS. SIGNIFICANT CORRELATIONS SHOWN IN RED. SCAS CONSISTS OF CCAS + RCAS IN COMBINATION. 82
TABLE 4.5 CORRELATION OF SCAS TO CELL PHENOTYPE, CYTOKINE PRODUCTION AND SERUM PROTEINS. 85
TABLE 5.1 DEMOGRAPHICS OF PATIENTS ENROLLED IN MICROARRAY SCAS COHORT. PBMC - PERIPHERAL BLOOD MONONUCLEAR CELL SAMPLES, BAL CELL - BRONCHOALVEOLAR LAVAGE CELL SAMPLES. OTHER ORGAN INVOLVEMENT INCLUDES CARDIAC/NEURO/BONE/SPLEEN/LIVER. 105
TABLE 5.2 TABLE SHOWING SAMPLES UNDERGOING WHOLE-GENOME GENE EXPRESSION DIRECT HYBRIDIZATION ASSAY (ILLUMINA) AND ASSOCIATED RIN. (OTHER SAMPLES ALSO WERE ALSO RUN WITH THESE (ALL WITH RIN >7), BUT WERE EXCLUDED AFTER THE NORMALIZATION PROCESS AS THEY WERE EITHER OUTLIERS OR WERE FOUND NOT TO MEET INCLUSION/EXCLUSION CRITERIA.) 107
TABLE 5.3 TOP TEN POSITIVELY ENRICHED GENE SETS IDENTIFIED USING GSEA ANALYSIS OF WHOLE GENE EXPRESSION FROM BAL CELLS CORRELATED TO INCREASING SCAS. 110
TABLE 5.4 ALL LEADING EDGE GENES CONTRIBUTING TO THE ENRICHMENT OF THE TOP GENE SET, IMMUNE RESPONSE, ARE SHOWN WITH THEIR RANKING IN THE TOTAL GENE LIST (OF 15534 GENES) 111
TABLE 5.5 TOP TEN NEGATIVELY ENRICHED GENE SETS IDENTIFIED USING GSEA ANALYSIS OF WHOLE GENE EXPRESSION FROM BAL CELLS CORRELATED TO INCREASING SCAS. 114
TABLE 5.6 ALL LEADING EDGE GENES CONTRIBUTING TO THE ENRICHMENT OF THE MOST NEGATIVELY ENRICHED GENE SET, TGFB RECEPTOR SIGNALING PATHWAY, ARE SHOWN WITH THEIR RANKING IN THE TOTAL GENE LIST (OF 15534 GENES), IN THIS CASE THE HIGHER THE NUMBER, THE STRONGER THE NEGATIVE CORRELATION. 115
TABLE 5.7 TOP 10 GO PROCESSES DERIVED FROM GENE LIST (N=947) THAT SHOWS SIGNIFICANT POSITIVE CORRELATION TO SCAS (IN BAL) WHEN ANALYSED THROUGH METACORETM 117
TABLE 5.8 ALL GENES IN BAL HIGH ACTIVITY DATASET CONTRIBUTING TO TOP GO PROCESS – IMMUNE SYSTEM PROCESS - FROM METACORE ANALYSIS. 118

TABLE 5.9 TOP 10 PATHWAYS DERIVED FROM GENE LIST (N=947) THAT SHOW SIGNIFICANT POSITIVE CORRELATION TO SCAS (IN BAL) WHEN ANALYSED THROUGH METACORETM. 119
TABLE 5.10 TOP 10 PROCESS NETWORKS DERIVED FROM GENE LIST (N=947) THAT SHOWS SIGNIFICANT POSITIVE CORRELATION TO SCAS (IN BAL) WHEN ANALYSED THROUGH METACORETM. 120
TABLE 5.11 TOP 10 GO PROCESSES DERIVED FROM GENE LIST (N=1369) THAT SHOWS SIGNIFICANT NEGATIVE CORRELATION TO SCAS (IN BAL) WHEN ANALYSED THROUGH METACORETM. 121
TABLE 5.12 TOP 10 PATHWAYS DERIVED FROM GENE LIST (N=1369) THAT SHOWS SIGNIFICANT NEGATIVE CORRELATION TO SCAS (IN BAL) WHEN ANALYSED THROUGH METACORETM. 121
TABLE 5.13 TOP 10 PROCESS NETWORKS DERIVED FROM GENE LIST (N=1369) THAT SHOWS SIGNIFICANT NEGATIVE CORRELATION TO SCAS (IN BAL) WHEN ANALYSED THROUGH METACORETM. 122
TABLE 5.14 TOP TEN POSITIVELY ENRICHED GENE SETS IDENTIFIED USING GSEA ANALYSIS OF WHOLE GENE EXPRESSION FROM PBMCS CORRELATED TO INCREASING SCAS. 123
TABLE 5.15 LEADING EDGE GENES AND THEIR RANKING IN THE ENTIRE GENE LIST, FROM THE AMINO ACID CATABOLIC PROCESSES GENE SET OF PBMC WHOLE GENE EXPRESSION. 125
TABLE 5.16 LEADING EDGE GENES AND THEIR RANKING IN THE ENTIRE GENE LIST, FROM THE DEFENSE RESPONSE GENE SET OF PBMC WHOLE GENE EXPRESSION 125
TABLE 5.17 TOP 10 GO PROCESSES DERIVED FROM GENE LIST (N=413) THAT SHOWS SIGNIFICANT POSITIVE CORRELATION TO SCAS (IN BLOOD) WHEN ANALYSED THROUGH METACORETM. 127
TABLE 5.18 ALL GENES IN PBMC HIGH ACTIVITY DATASET CONTRIBUTING TO TOP GO PROCESS – IMMUNE SYSTEM PROCESS - FROM METACORE ANALYSIS. 128
TABLE 5.19 TOP 10 PATHWAYS DERIVED FROM GENE LIST (N=413) THAT SHOWS SIGNIFICANT POSITIVE CORRELATION TO SCAS (IN BLOOD) WHEN ANALYSED THROUGH METACORETM. 129
TABLE 5.20 ALL GENES IN PBMC HIGH ACTIVITY DATASET CONTRIBUTING TO TOP GO PROCESS – IMMUNE SYSTEM PROCESS - FROM METACORE ANALYSIS. 130
TABLE 5.21 TOP 10 PROCESS NETWORKS DERIVED FROM GENE LIST (N=413) THAT SHOWS SIGNIFICANT POSITIVE CORRELATION TO SCAS (IN BLOOD) WHEN ANALYSED THROUGH METACORETM. 130
TABLE 5.22 TOP 10 GO PROCESSES DERIVED FROM GENE LIST (N=339) THAT SHOWS SIGNIFICANT NEGATIVE CORRELATION TO SCAS (IN BLOOD) 131
TABLE 5.23 TOP 10 PATHWAYS SEEN IN SIGNIFICANTLY NEGATIVELY CORRELATED GENES TO SCAS WHEN ANALYSED THROUGH METACORETM. 131
TABLE 5.24 TABLE OF OVERLAPPING GENES BETWEEN PBMC AND BAL, WHEN CORRELATION COEFFICIENT TO INCREASING SCAS IS SET AT >0.5. 133
TABLE 5.25 METACORE ANALYSIS OF OVERLAPPING GENES BETWEEN PBMC AND BAL THAT SHOW A SIGNIFICANT CORRELATION TO INCREASING SCAS. 134
TABLE 6.1 20 GENE SIGNATURE PUBLISHED BY ZHOU ET AL DISTINGUISHING COMPLICATED SARCOIDOSIS FROM HEALTHY CONTROLS AND UNCOMPLICATED SARCOIDOSIS. 147
TABLE 6.2 RIN NUMBERS OF SAMPLES IN EACH MODALITY ENTERED INTO GENE EXPRESSION PROFILING. 152
TABLE 6.3 PATIENT DEMOGRAPHICS FOR ACTIVE FIBROTIC AND ACTIVE NON-FIBROTIC PATIENTS IN EACH SAMPLING MODALITY (LUNG, PBMC AND BAL). 156
TABLE 6.4 TOP TEN GENE SETS POSITIVELY ENRICHED IN LUNG SAMPLES FROM PATIENTS WITH ACTIVE FIBROTIC SARCOIDOSIS COMPARED TO ACTIVE NON-FIBROTIC SARCOIDOSIS. 158
TABLE 6.5 GENES CONTRIBUTING TO THE CORE ENRICHMENT OF THE TWO SIGNIFICANTLY DOWNREGULATED GENE SETS IN THE LUNG SAMPLES FROM ACTIVE FIBROTIC V ACTIVE NON-FIBROTIC PATIENTS AND THEIR RANKING IN THE ENTIRE GENE LIST (OF 1348 GENES). 164
TABLE 6.6 TOP TWENTY GENES (OR LESS) CONTRIBUTING TO THE LEADING EDGE OF EACH ENRICHMENT PLOT FROM THE IPF SIGNATURES OBTAINED FROM GSE10667(280), GSE24206 (281) AND LENG ET AL (279). 167
TABLE 6.7 TOP 20 GENES CONTRIBUTING TO LEADING EDGE OF ENRICHMENT PLOTS FOR IPF V SARCOIDOSIS SIGNATURE IN ACTIVE FIBROTIC SARCOIDOSIS COHORT. 168
TABLE 6.8 ENRICHMENT SCORES, NORMALISED ENRICHMENT SCORES, NOMINAL P VALUES AND FDRQ VALUES FOR EACH OF THE GENE SIGNATURES SHOWN IN THE ENRICHMENT PLOTS. 170
TABLE 6.9 TOP 50 LEADING EDGE GENES AND THEIR RANKING IN THE COMPLETE LIST OF 21334 GENES CONTRIBUTING TO THE ENRICHMENT OF THE CHEMOKINE SIGNALING PATHWAY GENE SET IN THE ACTIVE FIBROTIC COHORT 174
TABLE 6.10 TOP TEN GENE SETS UPREGULATED IN BLOOD SAMPLES FROM ACTIVE FIBROTIC PATIENTS COMPARED TO ACTIVE NON-FIBROTIC PATIENTS. ES = ENRICHMENT SCORE, NES =

Normalised Enrichment Score, Nom p-val = Nominal p value (sig <0.05), FDR 1-val = false discovery rate q value (sig <0.25). 175
Table 6.11 Genes and their rank in the entire gene list of 21334 genes showing core enrichment in the top gene set of antigen processing and presentation of exogenous peptide antigen via MHC class I. 176
Table 6.12 Genes involved in core enrichment of the chemokine signaling pathway gene set in the active fibrotic blood samples, and their ranking in the entire 21334 gene list 179
Table 6.13 Top ten significantly upregulated gene sets in the active fibrotic BAL cell cohort, with their associated enrichment scores, normalized enrichment scores, nominal p values and FDR q values 181
Table 6.14 Genes contributing to the core enrichment of the top upregulated gene set (Regulation of Natural Killer Cell Mediated Immunity) in active fibrotic disease in the BAL cohort, and their associated ranking in the entire gene list (of 21334 genes) 181
Table 6.15 All genes contributing to core enrichment of the M1 gene signatures in the BAL cell samples and their associated ranking amongst the entire gene list of 21334 genes. 184
Table 6.16 Table showing chemokines and their receptors seen to be enriched in Blood and their complimentary molecules seen in the lung compartment. 193

1 Introduction

1.1 Sarcoidosis

1.1.1 History

Sarcoidosis is a condition characterized by non-caseating granulomata that can develop in any organ. The disease was first described in part as a dermatological phenomenon in 1877 by Dr Jonathan Hutchinson (1) when he reported on the presentation of a coal wharf worker who, over 2 years, had developed painless, purple skin plaques on his arms and legs. Subsequently, several other dermatological descriptions came about, with Besnier publishing on an infiltrating cutaneous violaceous swelling of the nose, ears and fingers, which he named ‘lupus pernio’ in 1889 (2), and the term ‘Sarcoid’ (or Sarkoid) was coined by Professor Caesar Boeck in 1899, after undertaking histological analysis of the typical skin lesions, which showed sharp, well-defined foci of epithelioid cells with giant cells, similar in appearance to sarcoma (3).

Professor Boeck also went on to report the details of 24 patients with ‘benign miliary lupoids’ that involved the lungs, conjunctiva, bone, lymph nodes, spleen and nasal mucosa, first hinting at the multisystemic nature of the disorder, and in time reports of granulomatous lesions in other organs became more forthcoming. Kreibich was the first to report bone cysts in sarcoidosis in 1904 (4), in 1909 Heerfordt called attention to a syndrome of uveoparotid fever, and in 1915 Kuznitzky and Bittorf published details of a 27 year old man with pulmonary, skin, spleen and lymph node involvement (5). Schaumann was the first to propose a common pathological basis

to this multisystem disease (6), and in 1953 the triad of hilar adenopathy, erythema nodosum and arthritis, as a specific syndrome in sarcoidosis with a good prognosis, was described by Sven Lofgren (7), raising it's profile as a diagnosis.

In 1999, the following descriptive definition was provided in the American Thoracic Society, European Respiratory Society, and World Association of Sarcoidosis and Other Granulomatous (WASOG) Disorders Statement (8), which is still in use today: *'Sarcoidosis is a multisystem disorder of unknown cause. It commonly affects young and middle-aged adults and frequently presents with bilateral hilar adenopathy, pulmonary infiltration and ocular and skin lesions. The liver, spleen, lymph nodes, salivary glands, heart, nervous system, muscles, bones and other organs also may be involved. The diagnosis is established when clinicoradiographic findings are supported by histologic evidence of noncaseating epithelioid cell granulomas. Granulomas of unknown origin and local sarcoid reactions must be excluded. Frequently observed immunologic features are depression of cutaneous delayed-type hypersensitivity and a heightened Th-1 immune response at sites of disease. Circulating immune complexes along with signs of B-cell hyperactivity also may be found.'*

1.1.2 Epidemiology

Sarcoidosis can be found throughout the world in all ethnic groups, although large regional variations in prevalence of disease have been reported. It usually presents in patients under the age of 40, with respiratory symptoms, skin or ocular disease, and changes in the chest radiograph are commonly observed (9). The true incidence and prevalence of the disease is not easy to determine, and is likely to be

underestimated with many cases being both clinically and radiologically silent (10). Studies to address this have been variable in their design and in their methods for detection of cases. A population based study reported an increased prevalence of asymptomatic disease in immigrants and health care professionals (11), but as these patients went through health screening that included chest radiographs, the results may have been biased. Studies that have used mass screening with chest radiographs have reported varying prevalences between nations, such as the study by Pietinalho et al, which showed prevalence to be 0.2 per 100,000 in Portugal, 64 per 100,000 in Sweden, and 28.2 per 100,000 in Finland (12). The highest incidence is thought to be in Northern European countries, at 5-40 per 100,000 population (12). The incidence in the UK has been reported as 5 per 100,000 population (13). A 3.8 fold greater risk of disease has been reported in African-Americans in comparison to Caucasians (assessed amongst members of a particular health insurance plan in the metropolitan population of Detroit, Michigan) (14).

In 2001, the ACCESS study (A Case Control Etiologic Study of Sarcoidosis) established the clinical epidemiologic features in 736 patients recruited from 10 clinical centres in the United States. The study population was heterogeneous in terms of race (53% white, 44% black), sex (64% female, 36% male), and age (46% < 40 yr old, 54% > or = 40 yr old). They found that women were more likely to have eye and neurologic involvement ($p < 0.05$), have erythema nodosum ($p < 0.01$), and to be age 40 yr or over ($p < 0.02$) whereas men were more likely to be hypercalcaemic ($p < 0.01$). Black subjects were more likely to have skin involvement other than erythema nodosum ($p < 0.05$), and eye ($p < 0.0001$), liver ($p < 0.0001$),

bone marrow ($p < 0.001$), and extrathoracic lymph node involvement ($p < 0.01$). They concluded that the initial presentation of sarcoidosis is related to sex, race, and age (15). Another study using the Danish National Registry ascertained that women seemed to have two main peaks of incidence, at ages 25-29 (10.5/100,000) and 65-69 (11/100,000), whereas men were more likely to present between the ages of 30 and 34 (incidence 14.8/100,000), but with a roughly equal male to female ratio (16).

Occupational and Environmental associations have also been investigated. The aforementioned ACCESS study group conducted a case control study to test an 'a priori' hypotheses that environmental and occupational exposures are associated with sarcoidosis. They observed several positive associations with sarcoidosis risk, including exposure to insecticides, agricultural employment, and exposure to microbial bioaerosols, but none were statistically significant. Neither a positive or negative association was found between sarcoidosis and previously hypothesised risk factors such as non-occupational wood dusts, occupational metals or silica. Smoking status (active, passive and former) was strongly associated with a reduced risk of sarcoidosis (Odds Ratio 0.65, CI 0.51–0.82) (17).

In summary, the global incidence and prevalence of sarcoidosis and associated risk factors have been extensively studied, but differences in study design, case identification and diagnostic criteria, has made ascertaining the true extent of disease and risk in any given community extremely challenging.

1.1.3 Aetiology

The condition is thought to develop in genetically susceptible individuals following exposure to an environmental trigger, of which there are likely to be many (18). The environmental trigger is commonly thought to be air-borne, in view of the pre-dilection for disease in eyes, skin and lungs (19).

Many triggers have been associated with the disease over the years, and the search has been endless for an infectious agent that may be contained within the granuloma. In view of the clinical and histological overlap between sarcoidosis and infectious granulomatous diseases such as Tuberculosis, a lot of past work has focused on trying to identify a causative role for Mycobacterium tuberculosis (20, 21). Several differing laboratory techniques have been employed in this search, including PCR methods to detect the presence of DNA or RNA from mycobacterial or other organisms in granulomatous sarcoid tissue. Results from over 24 studies in mycobacterial nucleic acid detection have varied from 0% detection to greater than 80% (22, 23). Chen et al used both flow cytometry and IFNγ ELISPOT assays to asses responses to a Mycobacterium tuberculosis antigen (mKatG) in patients from two centres, and demonstrated that T cell responses to mKatG in sarcoidosis were similar across centres, occurred more frequently in sarcoidosis patients than controls, and only occurred in chronically active patients, not in those with inactive disease (24). Its role in sarcoidosis however remains inconclusive.

A role for Propionobacterium acnes has also been suggested by several independent groups. It has been seen to localize in and around sarcoid granuloma,

identified using qPCR (25) and with monoclonal antibodies (26). It has been isolated from sarcoid lesions and grown in bacterial culture (27), and reportedly was cultured from lymph node biopsies of 78% of Japanese patients with sarcoidosis, compared to just 21% of non-sarcoid controls (28). However, it is also reported to be the most common commensal bacterium in the lungs and lymph nodes of patients without sarcoidosis (29) so its actual role in the aetiology of all cases of sarcoidosis is very uncertain.

The idea of genetic susceptibility arose from the observations on prevalence and incidence amongst ethnically diverse, and globally separated populations. Greater concordance of the disease has been demonstrated in monozygotic twins compared to dizygotic twins (30), and clustering in families has also been observed, with the ACCESS trial reporting a familial relative risk of 4.7 in Caucasians (higher than that of African Americans) (18), although its mode of inheritance is not of a mendelian pattern. Several candidate genes have been linked to susceptibility for developing sarcoidosis (31, 32). Both HLA class I and II antigens have been investigated and the most consistent associations with susceptibility to sarcoidosis involve alleles of HLA-A (33), -B (34), -DQB1 (35), -DRB1 (36) and –DRB3 (37). In the case of HLA-DRB1, some alleles are seen to protect against disease, whilst others are associated with an increased risk (38). Clinical course of disease and HLA alleles have also been studied, with HLA-DRB1*03 showing association with an acute onset and spontaneous recovery phenotype, and Loefgren's syndrome (39, 40), and HLA-DRB1*15 correlating with persistent disease (41). It is likely however that many genes act in concert to promote the pathophysiology, and phenotypical variation. A number of non-HLA genes associated with sarcoidosis risk or phenotype have also

been published. A mutation in BTNL-2, a member of the immunoglobulin superfamily implicated as a co-stimulatory factor for T-cell activation, has been associated with disease susceptibility (42), as has Annexin 11 (ANXA11), which has roles in apoptosis and proliferation (43). PCR was used to study polymorphisms of CCL5 and the homozygous A allele was found to associate with disease extent of three or more organs involved, Genetic association studies of two common variants (single nucleotide polymorphisms or SNPs) of MyD88 were shown to confer disease susceptibility (44), three IL23R SNPs showed increased frequency in the sarcoidosis population studied over controls (45), a single polymorphism of MMP9 was seen to associate with sarcoidosis (46), and single nucleotide polymorphisms of TGFβ have been associated with fibrotic sarcoidosis (47, 48).

So far attempts to understand the genetic background of sarcoidosis have not resulted in categorical answers, which may be made more complex by the wide-ranging phenotypes of the diseased population, and the potential for varied environmental triggers. With the advent of gene-expression profiling and proteomic techniques, in conjunction with robust phenotyping in terms of disease activity and system involvement, we may begin to uncover the genomics in sarcoidosis and its consequences at a gene expression level.

1.1.4 Immune pathology

The key feature in sarcoidosis is the granuloma, and it is widely accepted that the interaction of CD4+ T cells with antigen-presenting cells leads to it's formation (49). The over-exuberant CD4+ T cell response that occurs in sarcoidosis has a Th1 bias (50), releasing IL-2 and IFNγ, resulting in excess T cell proliferation, macrophage

activation and finally granuloma formation (Figure 1.1).

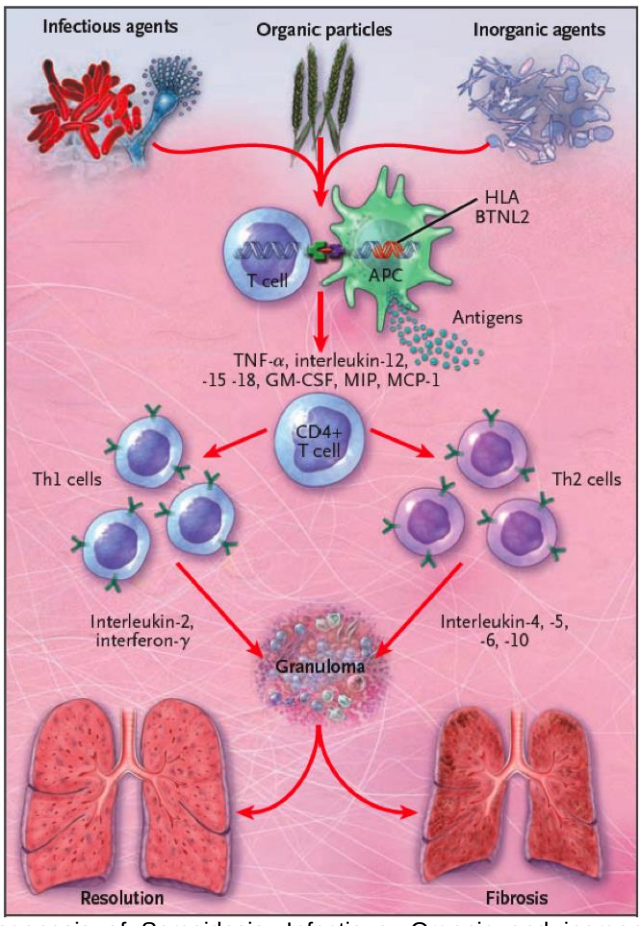


Figure 1.1 Immunopathogenesis of Sarcoidosis. Infectious, Organic and inorganic particles are possible antigens in sarcoidosis. After presentation by antigen presenting cells to CD4+ T cells, release of cytokines including TNF α, IL-12, -15, -18, MCP-1, MIP and GM-CSF occurs. The efficiency of antigen processing, presentation and cytokine release is probably under genetic governance, and macrophage HLA and BTNL2 alleles have strong associations with disease susceptibility. The activated CD4+ T cells undergo differentiation and proliferation, with a bias towards Th1 cells and release of IL-2 and IFN-γ. Granuloma formation occurs, which may persist, resolve or progress to fibrosis. From Iannuzzi MC et al (51)

The interacting Th1 cells and activated macrophages in sarcoidosis actually produce a large number of cytokines and chemokines, including TNFα, TGFβ, IFNγ, IL-1β, GM-CSF, IL-1, IL-2, IL-15, IL-12, IL-18, CCL2 (MCP-1) and CXCL10 (52-56). IL1 is produced by alveolar macrophages. It stimulates production of IL-2 from T lymphocytes, thus setting the stage for subsequent T cell proliferation and differentiation, but is also chemoattractant to T lymphocytes (57). IL-1 has been shown to be spontaneously produced by alveolar macrophages in active sarcoidosis (58). The increased IL-12 and IL-18 production enhances IFNγ, a potent immune

mediator that plays a central role in sarcoidosis by enhancing alveolar macrophage activation and suppressing apoptosis. IFNγ knockout mice were unable to produce granuloma when exposed to thermophilic bacteria in a model of hypersensitivity pneumonitis, suggesting a fairly crucial role in granuloma formation (59). Levels of IFNγ are seen to reduce with corticosteroid therapy, which in part may explain the treatment response to steroids in this condition (60). TNFα has been shown to be increased in patients with active sarcoidosis (61), is implicated as a major player in the persistent inflammatory response, reflected in other non-granulomatous acute inflammatory conditions such as sepsis and Rheumatoid Arthritis, and is also vital in the formation of granuloma (62). IL-2 and IL-15 have a proliferative and anti-apoptotic effect on T cells, contributing to their excess at the site of disease (52). CXCL10 is secreted by several cell types in response to IFNγ, and these include monocytes/macrophages, endothelial cells and fibroblasts (63). It is chemoattractant for monocytes, macrophages and T cells, again furthering recruitment to the sites of granuloma formation and therefore active disease (64). Macrophages also produce CCL2 (also known as MCP-1) and CCL5 (RANTES), both of which have been shown to have increased expression in the BAL fluid of pulmonary sarcoidosis patients in one of the largest observational studies of cytokines and chemokines (65), and are again chemoattractant for monocytes via the receptors CCR2 and CCR5 respectively. CCL2 can increase the expression of CCR2 (66), and can be both pro and anti-inflammatory. The anti-inflammatory properties are dependent on CCR2 expression on Natural Killer T cells, a cell population that has in fact been shown to be reduced in the blood and lungs of sarcoidosis patients, perhaps reflecting a diminished regulatory mechanism, since the presence of CD1d-restricted natural killer T cells would usually protect against an exaggerated Th1

host response (67). CCR2 is also expressed on macrophages, fibroblasts and endothelial cells. Macrophage produced TGFβ acts on fibroblasts to induce collagen deposition, which may play a role in the development of fibrosis in ongoing active disease (68).

The Th2 response in sarcoidosis at sites of inflammation is observed to be attenuated, with a reduction in the levels of IL4 and IL5, but interestingly not IL-13 (55, 69). Baumer et al used PCR to detect cytokines present in lung tissue, BAL fluid and blood at the Th1/Th0/Th2 and found equal Th1 and Th2 response in blood and BAL, but a Th1 predominance at the sites of granuloma formation in the lung tissue (70). However, studies using various cytokine deficient mice to investigate the Th milieu involved in fibrosis, have shown that fibrogenesis is strongly linked to a Th2 phenotype involving IL4, IL5 and IL13, which may suggest that for the development of fibrosis in sarcoidosis, a switch or increase in the presence of Th2 cells and their associated cytokines is required (71, 72). In Idiopathic pulmonary fibrosis, an increase in IL4 has been observed in the BAL fluid (73), and parenchyma (74) when compared to healthy controls. Overexpression of IL-4 or IL-13 in the lung of transgenic mice have shown that both cytokines function as profibrotic mediators by both directly and indirectly influencing the activation of myofibroblasts (75, 76). The role of Th17 cells in sarcoidosis has also been explored, as the Th17 subset has been linked to other inflammatory conditions such as inflammatory bowel disease (also a granulomatous disorder) and psoriasis (77, 78). Facco et al found an increase of Th17 cells in both blood and BAL of active sarcoidosis patients and have suggested it may have a role in progressive disease (79), and IL17A, secreted by Th17 cells has been implicated in the pathogenesis of pulmonary fibrosis (80).

The role of macrophage polarisation in sarcoidosis and fibrosis has also been intriguing. Th1 cytokines such as IFNγ trigger classical activation of macrophages (known as M1 phenotype), which regulate cellular immunity to intracellular pathogens, promoting granuloma formation, and have a pro-inflammatory role. The Th2 cytokines, IL-4 and IL-13, trigger a characteristic alternative state of activation in macrophages that is distinct from the classical Th1-type activation, and is associated with IL-10 and TGF-β (81). Alternatively activated macrophages (also called M2) are associated with chronic helminth infections and allergic disease, they participate in humoral immune responses, facilitate clearance and presentation of antigens, and regulate the important process of tissue repair. This latter function is obviously of interest when the burden of inflammation then results in healing by fibrosis, as although sarcoidosis is thought of as a Th1 disease that would ordinarily promote M1 activation of macrophages, a shift again may occur towards a Th2 and subsequent M2 phenotype contributing to fibrosis (82, 83). Finally, it was once thought that the disease of sarcoidosis was compartmentalized to sites of active disease. This idea initially came about as peripheral anergy seen in sarcoidosis, was noted not to result in immune deficiency and increased risk of infection, but rather increased T cell activity at sites of active disease. It was therefore suggested that there was a more a redistribution of lymphocytes in the body rather than a generalized up or down regulation. However, as studies have progressed, it is more obvious that there are other things going on in the periphery that accurately reflect what is going on more centrally at the site of active disease.

1.1.5 Diagnosis

The diagnosis of sarcoidosis is made on the basis of clinical and radiological findings of granulomatous disease with supporting histology as per the guidelines set out by WASOG mentioned in section 1.1.1. Patients often have had symptoms, many of which may be non-specific, for more than three months and it is not unusual that multiple consultations with health care providers occur before a specific diagnosis is made (84).

The chest radiograph is the most commonly used initial radiological imaging technique and radiographic disease staging is still based on a system proposed by Scadding in the 1960s (85). This has been modified to include stage 0: no adenopathy or infiltrates; stage 1: hilar and mediastinal adenopathy alone; stage 2: adenopathy and pulmonary infiltrates; stage 3 pulmonary infiltrates alone; and stage 4: pulmonary fibrosis. This system allows some prognostication, with patients with stage 1 radiographs at presentation having a greater than 90% chance of resolution by 2 years, compared with those with stage 3 who have less than a 30% chance of resolution. There is considerable inter-observer variability in staging, even amongst radiologists, which diminishes the clinical utility of this scale (86). In addition, there is little correlation between the chest radiographic findings and the level of dyspnoea, serum disease activity markers and physiological assessment (such as six-minute walk testing and pulmonary function testing). Computed Tomography is now often performed to confirm the clinical and chest x-ray findings, to exclude other pathologies, such as lymphoma, and to detect both pulmonary and extra-pulmonary complications

(8). In addition, it also has a use in determining the degree of parenchymal

involvement temporally, particularly prior to treatment, and as a possible indication of disease activity level and burden. The importance of reporting by a dedicated chest radiologist, preferably with an interest in interstitial lung disease however, is stressed in the British Thoracic Society (BTS) Guidelines (87). In cases where there is no apparent pulmonary involvement, F-18 fluorodeoxyglucose positron emission tomography (^{18}FDG PET) may prove useful in identifying areas with active granulomatous inflammation for biopsy (88, 89). The sensitivity and specificity is unknown and a positive result may be due to malignancy or an alternative inflammatory condition. Cardiac and neurological involvement may be better detected with ^{18}FDG PET and MRI with gadolinium than with CT (88, 90). Biopsy must demonstrate non-caseating granulomata and exclusion of infective organisms or focal particle and must be considered in conjunction with other clinical evidence for multisystem disease. In patients who present with Löefgren's syndrome (bilateral hilar lymphadenopathy, erythema nodosum and arthritis), a clinical diagnosis is usually sufficient and a biopsy may not be required. These patients generally have a good prognosis.

Transbronchial biopsy has a diagnostic yield of 40-90% and it is suggested that 4-5 biopsies are required to maximise the diagnostic yield (91). The yield can be increased by performing endobronchial biopsy, in conjunction with transbronchial biopsy, giving a rate closer to 90% (92). The yield can be increased yet further if endobronchial ultrasound is used to access smaller nodes and those not accessible by mediastinoscopy (92, 93). Examination of bronchoalveolar lavage fluid for lymphocytes (specifically the CD4/CD8 ratio) can be used as a complementary diagnostic test. If the fluid contains more than 15% lymphocytes, the sensitivity for

diagnosing sarcoidosis approaches 90%; the specificity however is low (94) .

Angiotensin converting enzyme (ACE) is produced in the epithelioid cells of the granuloma and may reflect the systemic granuloma burden (95). Serum ACE is an insensitive test for the diagnosis of sarcoidosis (sensitivity 57%, specificity 90%, NPV 60%, PPV 90%) (96). An elevated serum ACE is not sufficient to confirm a diagnosis of sarcoidosis and therefore is not useful as a screening diagnostic test. Serial ACE measurements have been proposed to monitor the course of the disease, which may be useful in individual patients in whom the level is elevated at presentation and falls with remission.

1.1.6 Treatment

For most patients with pulmonary sarcoidosis, a conservative approach with close observation is all that is initially required. The BTS guidelines, published in 2008 (87), specifically state that in asymptomatic disease, with only mild abnormalities in lung function, and stage I, II or II scadding chest radiograph stages, treatment is not indicated as high rates of remission are observed in this disease. However, the natural clinical course of the disease is not fully understood, and varies markedly from patient to patient. When disease progression is evident on either serial lung function testing or on radiology suggesting end organ dysfunction, when there is presence of hypercalcaemia, or other extra-pulmonary disease requiring treatment, or in rare cases, severe symptoms, treatment, usually in the form of corticosteroids, should be initiated. Again, BTS guidelines recommend treatment in the form of Prednisolone, 0.5mg/kg/day for 4 weeks, reducing gradually after this to the lowest maintenance dose that controls symptoms or stabilizes disease, which should then

continue from 6 to 24 months depending on the individual case, apparent level of disease activity and response to treatment. At the time of writing these guidelines, little evidence was available on the use of other immunosuppressants/steroid sparing agents in sarcoidosis, but when corticosteroids alone are not controlling the disease, or the side effects of steroids are intolerable, then methotrexate was advised as the alternative treatment of choice. Since the publication of these guidelines however, steroid sparing agents have been prescribed more frequently, especially for those patients that display a relapsing remitting course, requiring intermittent high dose steroids for flare of their disease, failure to discontinue steroids after several attempts, or ongoing progression despite corticosteroids. Methotrexate, Azathioprine, Mycophenolate Mofetil, and Anti-TNF alpha have all been used with varying success, but the evidence for use still remains limited (97).

Methotrexate has anti-inflammatory properties, and is known to inhibit cellular proliferation by stimulating adenosine release (98). It has been shown to be equivalent to corticosteroid treatment, in a study by Goljan-Geremek (99), where 50 patients with sarcoidosis were treated with either 10 or 15mg of methotrexate weekly for 6-24 months. A significant improvement in both TLCO and the six minute walk test was noted, which correlated to methotrexate dose (i.e. greater improvement in the higher dose of 15mg weekly). Baughman et al also reported that following twelve months of methotrexate use patients required less corticosteroid treatment than the control group (100). Reported adverse effects of methotrexate however include fatigue, nausea, alopecia, skin rashes, interstitial pneumonitis and bone marrow suppression.

Azathioprine inhibits lymphocytes by blocking the synthesis of RNA. Only case series and open label studies have been reported on its use in sarcoidosis, there are no randomized control trials, and numbers are obviously very small. Pacheco et al reported on 10 patients with corticosteroid resistant disease treated with 150mg Azathioprine, and found that 70% of patients showed clinical or radiographic improvement (101). More recently Vorselaars compared methotrexate to azathioprine in a retrospective study, and both were found to have similar steroid sparing agents and side effect profiles, although azathioprine had a higher rate of infections (102). Close monitoring of full blood count and liver function tests are required, especially during initiation of treatment, as hepatotoxicity and bone marrow depression are reported side effects. In addition, nausea, vomiting, diarrhea, rash, myalgia and fever has been reported, and an increased risk of infection.

Mycophenolate Mofetil (MMF), a drug that also inhibits lymphocyte proliferation, has been used in pulmonary sarcoidosis, in addition to prednisolone, and has been shown to induce stability or improvement in lung function, clinical symptoms or radiographic appearance in a small cohort of patients (103). MMF is well tolerated, with usually only self limiting mild nausea and/or diarrhoea reported. Infections and bone marrow suppression can occur but are less common than other immunosuppressive agents.

Finally, the use of anti-TNFα agents has been studied briefly in sarcoidosis. Macrophages, which are key players in granuloma formation, and therefore active sarcoidosis, produce TNFα, and therefore the use of anti- TNFα agents in sarcoidosis treatment has raised curiosity. Infliximab, a monoclonal antibody against

TNFα, has undergone two randomised controlled trials in patients with chronic pulmonary sarcoidosis (104, 105), and significant improvements in FVC were observed on treatment, with added benefit on extrapulmonary signs and symptoms. Its use however has not been widespread, and it remains unclear as to who would benefit most from this course of treatment.

1.2 Disease Phenotype

Sarcoidosis is a disease of diverse clinical presentation, course and outcome. Some patients present acutely with the classic syndrome of Löefgren's, which as mentioned earlier involves the triad of arthralgia, bihilar lymphadenopathy and erythema nodosum (106), whereas others present after many years of suffering mild or slowly progressive symptoms. Some patients have diagnoses made incidentally, others present with specific signs/symptoms of the affected organ e.g. cough or shortness of breath with lung involvement, skin lesions with cutaneous involvement, nerve palsies, meningeal disease, or even seizures with central nervous system involvement, or palpitations with cardiac involvement. While sarcoidosis can affect any organ, those involved are usually established early in the course of the disease, with less than 25% of patients (more commonly those with initial extrapulmonary involvement) presenting with new organ involvement at two year follow-up (107).

The clinical course also appears to be extremely variable, with some patients experiencing a gradual improvement towards disease resolution, some remaining or becoming asymptomatic despite ongoing serological or radiological evidence of disease persistence, those that run a relapsing remitting course of symptoms, and others with prominent symptoms, which may or may not progress to end organ

dysfunction. The effects of immunosuppressive treatment on course and outcome is difficult to assess, as treatment is initiated in some in an attempt to dampen the inflammatory component and ensuing damage of the disease that may or may not become irreversible, but ethically, it would not be possible to ascertain the natural history of that particular disease manifestation by withholding immunosuppression. Finally, the prognosis of these patients in terms of disease resolution, persistence or progression is also highly variable, with up to 30 percent of patients developing irreversible end organ damage (108), such as pulmonary fibrosis, sometimes despite immunosuppressive therapy. Apart from the need for a high index of suspicion to make the diagnosis of sarcoidosis, the heterogeneity of the disease makes it difficult to ascertain any pattern that may predict course or outcome. With the assumption that the characteristics of a particular phenotype can be used to more accurately predict a patient's eventual outcome, Baughman et al (109) reviewed the clinical features at presentation of 215 biopsy confirmed sarcoidosis patients, a cohort of the ACCESS study (107), and correlated these with need for treatment at 18-24 months. They found that only 'report of dyspnea', or immunosuppressive therapy already in place at initial assessment, correlated with being on treatment at follow-up (suggesting this group represents a poorer outcome), but also noted that there was no set protocol for treatment in these patients. Neville et al (110) studied the disease manifestations in 818 patients with sarcoidosis and confirmed a good prognosis for patient's with Löefgren's triad, but a poor/chronic prognosis (i.e. persistent disease evidence on chest radiograph at 2 years) in patients who also had extra-pulmonary disease manifestations. Chest radiograph at presentation was assessed using Scadding's criteria (85), and each stage from one to three was found to have a worsening clinical and radiographic

prognosis. A study by Hillerday et al that involved follow-up of pulmonary sarcoidosis patients over 15 years, demonstrated that 80% of patients with CXR scadding stage 1 had a normal chest x-ray at 5 years from presentation, compared with 70% with stage 2, 40% with stage 3 and none with stage 4 (111). These studies along with several others (107, 112, 113), have looked at the condition of sarcoidosis as a whole, without employing any phenotypical criteria to ensure like with like comparison.

The World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) task group have attempted to define clinical outcome as a clinical phenotype that relates to disease course at 5 years following disease presentation/onset in sarcoidosis (114). Nine clinical phenotypes were proposed, that involved disease activity at 5 years being first defined as resolved, minimal or persistent. Persistent disease was defined as remnant shadows on chest radiographs >25% of initial radiograph, whereas minimal activities are defined as <25% of the initial activity. Further subdivisions were then made depending on the treatment received. Four hundred patients from 8 different geographic areas worldwide were allocated to these phenotypes. They found that although 40% of the patients were receiving treatment, only 8% of them progressed despite treatment. Again, however, note must be made that no set protocol was used for treatment in these centres, and that sarcoidosis patients were viewed as a whole, rather than those with just pulmonary, or predominantly cardiac etc. Also the assessment of more or less than 25% radiographic activity could be argued to have an element of subjectivity.

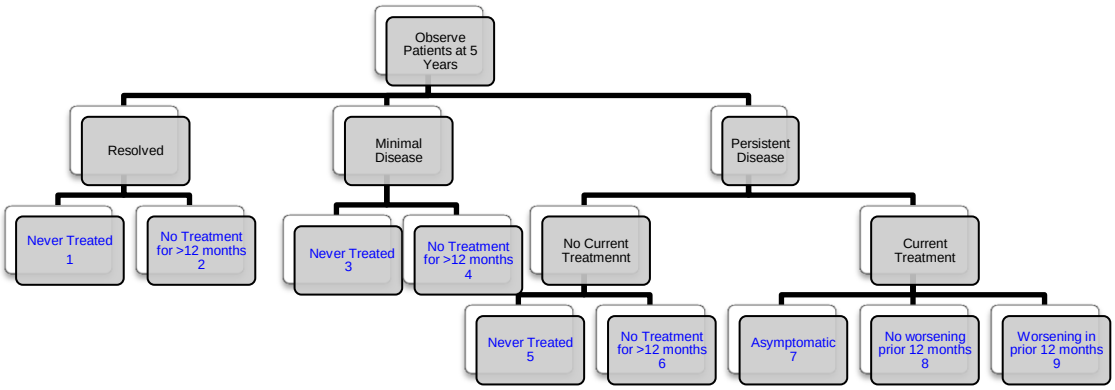


Figure 1.2 The nine clinical phenotypes relating to outcome as proposed by Judson et al (114)

One of the biggest difficulties in trying to categorize disease is whether to include parameters for subjective measurements such as dyspnea (in the absence of deteriorating lung function) and lethargy/fatigue, another prominent feature of this illness and one frequently reported as being the most troublesome (115). It seems wrong to exclude a patient's opinion on their state of health and symptoms, but it is a very variable entity that cannot be measured accurately. There are quality of life questionnaires (116, 117) and respiratory questionnaires (118) that standardize the description of symptoms, but what one person will report as severe fatigue, the next may report just minor lethargy. However, to give further weight to including such subjective measures, TLCO has been shown to correlate with the Quality of Life Domain of independence (119), so it's inclusion in a categorization process may still be important.

Alongside phenotyping disease course, it may be that the clinical heterogeneity of organ involvement described earlier should also be classified and phenotyped, as again, what occurs in neurosarcoidosis, may have a subtly different set of triggers or immune pathology when compared to cutaneous sarcoidosis or cardiac sarcoidosis. Our group has therefore tried to categorise patients at presentation based on their clinical phenotype of organ involvement, with focus for our research being on those with prominent pulmonary disease, so that direct comparisons can be made more easily.

1.3 Disease Activity

For the clinical management of patients with sarcoidosis, and for the stratification of recruits to research studies and clinical trials, the evaluation of disease activity is of great importance. To date there is no standardized reference to define disease activity in sarcoidosis, and generally a label of active sarcoidosis is down to each individual clinician or researcher's judgment. Following a consensus conference in 1994, The World Association for Sarcoidosis and Other Granulomatous diseases (WASOG) defined activity as one or more of i) worsening of sarcoidosis attributed symptoms, pulmonary function tests or chest radiographic changes; ii) Evidence of hypermetabolism on function imaging such as PET or Gallium scanning; and iii) Serum ACE >100. This definition however is still susceptible to varied interpretation (120).

For over three decades various efforts have been made to establish reliable and non-invasive means for the assessment of activity in pulmonary sarcoidosis, but still

we are no further along in our definition. The gathering of information that may reflect disease activity in the follow up for patients with sarcoidosis includes history, pulmonary function testing, serum biomarkers and chest imaging (usually in the form of chest radiographs), and with experience the treating clinician will correlate changes in these parameters with response to treatment.

Reported symptoms can be difficult to interpret, with lethargy and tiredness being commonplace, alongside cough and shortness of breath. An improvement in pulmonary function testing may represent either a positive response to treatment and/or a decrease in pulmonary inflammation and hence decreased disease activity. However, persistently abnormal results may not necessarily reflect ongoing disease activity or poor response to treatment, but irreversible fibrotic changes, despite disease quiescence.

Serum Angiotensin Converting Enzyme (ACE) – the first reported biomarker in Sarcoidosis (121), has been used as a diagnostic and prognostic marker, and has been seen to reduce with corticosteroid treatment. It is an acid glycoprotein that converts angiotensin I into angiotensin II. It is secreted by monocytes and macrophages, and pulmonary endothelial cells and is considered a marker of granuloma formation and monocyte/macrophage activity in the disease. However, it is found only to be raised in 60% of patients with a confirmed diagnosis of sarcoidosis (122), it's variability thought to be due to ACE gene polymorphism, and it may also be elevated in other diseases including miliary tuberculosis(123), diabetes mellitus (124) and thyroid disease(125), further complicating its utility as a solitary marker of activity. Serum soluble IL-2 Receptor (sIL-2R), thought to

represent and quantify T cell activation, has also been reported as a marker of sarcoidosis activity (although again is not specific for sarcoidosis and can be seen in several autoimmune diseases), and was validated using positive 67Ga scanning as a correlate of active pulmonary disease (126). Interestingly, soluble IL2-R did not correlate with Serum ACE results. Serum KL-6 has also been reported as an activity marker (127) (128) and correlated with parenchymal infiltration on chest radiographs, but neither this nor soluble IL2-R are currently measured routinely in clinical practice. Lymphopaenia has been reported to represent severe disease (129), and may be seen due to the upregulation of homing markers at sites of inflammation that result in trafficking of the lymphocytes to the site of active disease/inflammation/granuloma, and a relative decrease measured in the periphery. Lymphopaenia has also been correlated as a marker of a less favourable outcome and the development of chronic disease, along with a likelihood of requiring immunosuppressive treatment (130). Immunoglobulin G, a non-specific marker of systemic inflammation, has been well documented to be raised in sarcoidosis (131). The mechanism is unclear but could partly be explained by T-Helper cell activation and the subsequent influence on B cells and their differentiation to plasma cells. The presence and intensity of the active component of airways disease, “alveolitis", in sarcoidosis may be indicated by the level of IgG, and decreases in IgG have been shown to correlate with improvement in lung function (132). However, if other inflammatory diseases are present in the same patient it’s use would have to be modified, as it would not be specific for sarcoidosis activity in that circumstance. Immunoglobulin has also been shown to localize in pulmonary sarcoid granulomas alongside plasma cells(133), so the higher the activity or disease burden, the more granulomas, then theoretically the higher the

IgG measurement will be. Other granulomatous diseases, such as Gaucher's disease, have reported other serological markers of active granuloma and disease burden which have also been investigated in sarcoidosis, such as CCL18 (134) and chitotriosdiase(135, 136), but again lack firm results. Bronchoalveolar lavage (BAL) has been assessed for its cytokine profile, and for its cellular content, based on the concept of a T cell alveolitis being evident as a lymphocytosis in BAL in active disease (110). The CD4:CD8 ratio in BAL fluid has been seen to change with corticosteroid therapy and the same has been seen in induced sputum, a potentially less invasive indices of activity (137).

Various modalities are available to image the chest, and all have been assessed on their ability to reflect active disease over the years, but all have their limitations. Of note of course, is that as there is no agreed upon marker of activity, not all have been correlated against the same parameters. Firstly, the chest radiograph. Historically, a descriptive 'scoring system' put in place by Scadding, defines the changes observed on the chest radiograph (85). Correlation of this scoring system to various putative markers of activity including serum ACE (138), Bronchoalveolar lavage (139), Gallium scanning (140, 141) and clinical correlates have however not demonstrated a linear pattern in disease activity based on the Scadding staging of chest radiographs. In a study by Klech et al, 60 patients were assigned a label of either **active** disease (n=35), based on clinical symptoms (fever, cough, dyspnea etc) with or without a pathological chest radiograph, or progressive radiographic deterioration (even if asymptomatic) or recent new developing skin lesions, ocular inflammation or lymphadenopathy; or assigned **inactive**, suggested by the absence of clinical symptoms together with negative or non-progressive chest-x-rays (n=25).

Findings showed that when compared to 67Ga scanning, chest radiographs failed in 32% of cases to document activity, particularly in patients with peripheral pulmonary involvement (141). However, Baumann et al showed that worsening serial chest radiographs correlated with an increase in disease activity induced by withdrawal of corticosteroid therapy (chi-square = 4.05, $p<0.05$), validated by deteriorating FVC, and also in some cases pre-dated the clinical relapse. The sensitivity and specificity of the chest radiograph for detecting a clinical relapse in this series was 70% and 100% respectively (142). Müller, using a chest x-ray scoring system adapted from the ILO classification of pneumoconiosis to quantitate infiltrative lung diseases (143), demonstrated that although extent of disease at CT showed a significant correlation with clinical and functional impairment, the correlation is similar to that with the profusion of chest radiography (CT and CXR v dyspnea r= 0.61, 0.58 respectively p= <0.001, CT and CXR v TLCO r= -0.62, -0.52 respectively p= <0.01, CT and CXR v FVC r= -0.49, -0.51 respectively p= <0.01). Furthermore, it appeared that irregular linear opacities, rather than nodular lesions correlated more with functional impairment ($p<0.05$)(144). Gallium scanning has previously been reported as being the most sensitive method (94% sensitivity) for detecting disease activity, this time correlating to number of BAL lymphocytes, thought to reflect the extent of sarcoidosis alveolitis in active pulmonary disease (145). Baughman et al studied the correlation of Broncholaveolar lavage lymphocyte numbers, Serum ACE level and Gallium uptake pre-treatment with corticosteroid therapy, against change in Forced Vital Capacity (FVC) following therapy, and found a close correlation with gallium scanning and CD4 cell count in BAL ($r=0.95$ $p<0.01$, $r=0.92$ $p<0.01$ respectively), a positive correlation between CD4:CD8 ratio (r 0.62 $p<0.05$), a low correlation between pre-treatment ACE levels (r 0.368 $p>0.05$) and no relationship between the

percentage of lymphocytes at lavage versus change in FVC following therapy (146). The limitation of gallium scanning is that it obviously involves a significant dose of radiation (30mSv compared to 0.1mSv for a chest radiograph), preventing repeated and regular use as an assessment for disease activity and response to treatment. The value of High Resolution Computed Tomography (HRCT) scanning in assessing the extent and type of parenchymal changes in sarcoidosis is well established. Obertsein et al investigated the utility of the findings on HRCT thorax as a measure of disease activity by correlating these with clinical examination, lung function tests, bronchoscopy for bronchoalveolar lavage (BAL) and transbronchial biopsy. They found the comparison of the HRCT findings, like pathologic appearance of the bronchovascular bundle and intraparenchymal nodules, with serologic and BAL-parameters yielded high correlation coefficients with the total cell count in BAL and sIL-2R, and moderate correlations with the lavage lymphocyte count, T4/T8 ratio, IL-2R and HLA-DR expression, and concluded that as a non invasive method, HRCT depicts pathologic findings of the lung parenchyma which are associated with the inflammatory activity of sarcoidosis (147). Similarly Leung et al correlated HRCT to 67Ga scan, BAL cell number/ratios, and Serum ACE (SACE) (n=29, with all measurements performed within one month of each other) and found that the extent of nodules and consolidation correlated with the intensity of lung gallium uptake ($r = 0.46$, $p < 0.02$), BAL lymphocyte percentage ($r = 0.50$, $p < 0.01$), and SACE levels ($r = 0.38$, $p < 0.05$). No significant correlation was found between extent of ground-glass attenuation or linear opacities with any indexes of disease activity (148). CT has also been correlated to the histopathological findings at open lung biopsy and autopsy (149, 150). Thickened bronchovascular bundles have been seen to correspond histologically to granulomas, either with or without

perigranulomatous fibrosis in the connective tissue sheath around the pulmonary vessels and airways. Conglomerate granulomas are represented on CT by high-attenuation nodules, and the ground-glass opacities are caused by the summation of a number of small granulomas in the interstitium – also proving that granuloma burden and therefore extent of disease can be visualized accurately on HRCT. If this could be quantified, then this may also represent a method of measuring disease activity, but again the amount of radiation involved in HRCT would preclude regular imaging in this form. Finally, the role of PET CT has also been explored in measuring disease activity. In a retrospective study, 77 newly diagnosed pulmonary sarcoidosis patients underwent BAL and (18)F-FDG PET to assess whether metabolic activity imaged by (18)F-FDG PET represents signs of disease activity as reflected by BAL cell profile. Percentage of BAL lymphocytes, CD4/CD8 ratio, the CD4+ Tcell subgroup CD103(+)CD4(+)/CD4(+) ratio, and percentage of BAL neutrophils were compared with the extent of metabolic activity expressed as the maximum standardized uptake value (SUV(max)). They concluded that (18)F-FDG PET correlates with the CD4/CD8 ratio and neutrophils, suggesting that (18)F-FDG PET represents this specific cell profile in BAL, and that high SUV(max) values of the lung parenchyma may correlate with more severe parenchymal involvement, and therefore (18)F-FDG PET may be an alternative non-invasive measure of disease activity when compared to Bronchoscopy and BAL. (151). The same group published a follow-up paper in 2014 that showed FDG PET also correlates well with serum biomarkers such as soluble interleukin-2 receptor in symptomatic patients, and even visualizes active lesions in the context of normal serum biomarkers. Moreover, FDG PET activity in lung parenchyma correlates with decrease of lung function values over time (152). Sobic-Saranovic et al studied 90 sarcoidosis

patients with persistent symptoms who were referred for 18F-FDG PET/CT evaluation to assess the extent of inflammation. They also underwent CT scanning and measurement of serum ACE level. After follow- up, the clinical status and changes in therapy were analyzed. The therapy was initiated or changed in 73 out of 90 patients (81%). Both univariate and multivariate logistic regression analyses indicated that positive 18F-FDG PET/CT results were significantly (*p* 0.001) associated with changes in therapy, with no contribution from age, sex, ACE level, CT results, or previous therapy, concluding that 18F-FDG PET/CT is a useful adjunct to other diagnostic methods for detecting active inflammatory sites in chronic sarcoidosis patients with persistent symptoms, especially those with normal ACE levels, and that 18F-FDG PET/CT proved advantageous for determining the spread of active disease throughout the body and influenced the decision to adjust therapy (153).

Yet again though, the amount of radiation involved (24mSv) would preclude regular monitoring with this modality, and would not provide a quick clinic based tool to allow speedy management decisions at the time of review.

To summarise, there is not one agreed upon individual method for establishing the level of disease activity in sarcoidosis, and some of those purported involve invasive testing or high dose radiation. My studies have therefore been focused on developing a robust and reproducible method to evaluate disease activity in pulmonary sarcoidosis, firstly based on high resolution CT chest imaging, but with a view to developing a clinic based tool, that will allow a graded, objective measure of disease activity, using parameters routinely measured in the clinical setting.

1.4 Lung fibrosis

Pulmonary fibrosis is a feature or complication of many different diseases. Patients with autoimmune or connective tissue diseases such as Rheumatoid Arthritis and Scleroderma can develop respiratory disease, often inflammatory driven, that if not controlled results in the repair process healing by fibrosis. It can be induced by airborne particles, such as asbestos (Asbestosis), but also moulds (Farmers lung) and bird feathers (Bird Fancier's Lung), where the immune response to these particles causes the damage. It can be drug-induced, with chemotherapy agents, certain antibiotics (nitrofurantion), heart drugs (amiodarone), and even immunosuppressant drugs themselves (methotrexate) being implicated. In many situations no underlying cause for the pulmonary fibrosis is found, and the disease is termed Idiopathic Pulmonary Fibrosis (IPF). Whilst the inciting pathological event may be very different across all of these cases of pulmonary fibrosis, the repair process and subsequent scarring may share many similarities, and may provide useful direction to the study of fibrotic sarcoidosis.

As previously mentioned, the majority of patients with pulmonary sarcoidosis have a fairly good prognosis, and whilst immunosuppression may be required in some to dampen down the inflammatory effects of the disease resulting in end-organ dysfunction (as measured by lung function), only a proportion of these patients will go on to develop fibrosis (approximately 20%). Moreover, there appears to be varying radiological patterns of the fibrosis seen in pulmonary sarcoidosis, ranging from small areas of focal fibrosis, likely formed in an area of significant but focused granulomatous inflammation, perihilar fibrosis, which is a more frequently observed finding that follows on from the granulomatous conglomeration of bihilar nodes and

surrounding inflammatory interstitium, or more diffuse fibrosis, usually with an upper zone pre-dilection. The pathophysiology of fibrotic sarcoidosis is poorly understood, but a change in macrophage phenotype, transition from Th1 to Th2 immune responses, and increased production of TGFβ have all been implicated in its development (72, 82, 154, 155). Fibrogenesis is often defined as an 'out of control' wound healing response, and the mechanisms thought to be involved are depicted in Figure 1.3 that demonstrate the distinct phases of wound repair (156).

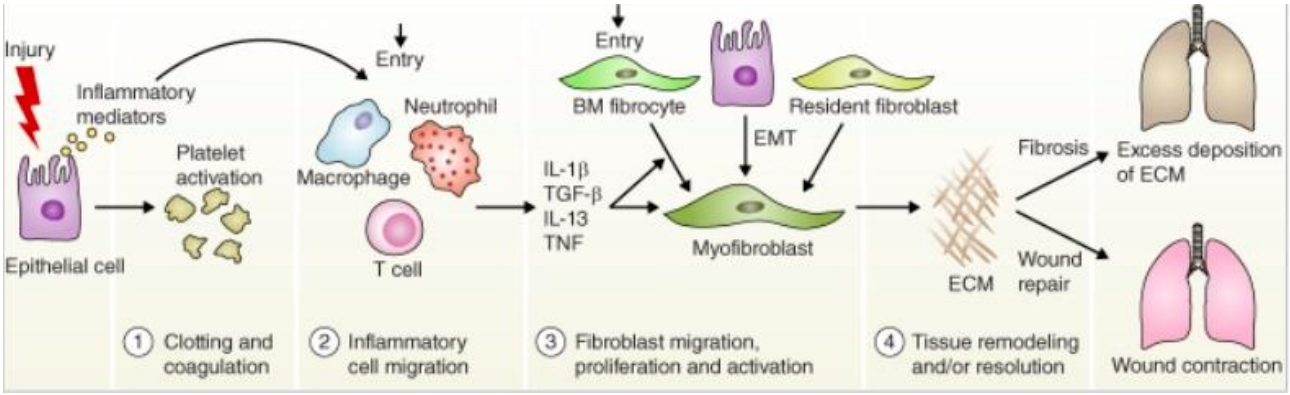


Figure 1.3 Phases of wound repair and the potential mechanisms for fibrogenesis After lung injury, epithelial cells release inflammatory mediators that initiate an antifibrinolytic coagulation cascade, which triggers platelet activation and blood clot formation. This is followed by entry of leukocytes (e.g., neutrophils, macrophages, and T cells). The recruited leukocytes secrete profibrotic cytokines such as IL-1β, TNF, IL-13, and TGF-β. The activated macrophages and neutrophils also remove dead cells and eliminate any invading organisms. In the subsequent phase, fibrocytes from the bone marrow and resident fibroblasts proliferate and differentiate into myofibroblasts, which release extracellular matric (ECM) components. Fibroblasts and myofibroblasts may also be derived from epithelial cells undergoing epithelial-mesenchymal transition (EMT). In the final remodeling and resolution phase, activated myofibroblasts can promote wound repair, leading to wound contraction and restoration of blood vessels. However, fibrosis often develops if any stage in the tissue repair program is dysregulated or when the lung-damaging stimulus persists. From Wynn et al (156).

Macrophages are key in all stages of the fibrotic process, serving as regulators of fibroblast recruitment, proliferation, and activation (157). They promote fibrosis by secreting chemokines and specific matrix metalloproteinases that degrade extracellular matrix (ECM) components, and facilitate the recruitment of inflammatory cells to sites of tissue injury (158, 159). They also produce several

profibrotic mediators, including TGF-β1 and PDGF that induce the proliferation and activation of collagen-secreting myofibroblasts (160).

In sarcoidosis, it is reasonable to expect foci of fibrosis around sites of uncontrolled granulomatous inflammation, which will often follow the same lymphatic distribution as the active disease (161), but the extent to which this happens obviously impacts on the clinical outcome, and in some cases of fibrotic sarcoidosis there appears to be a more exuberant and pathological fibrotic response. The histological pattern of disease does appear to be distinct from that of Usual Interstitial Pneumonia (UIP) pattern fibrosis seen in IPF, and although fibroblastic foci may occasionally be seen (162), bronchiolar honeycombing is not an observed feature in sarcoidosis, and granulomas are not observed in UIP (163). The mechanism of fibrogenesis in sarcoidosis is a grossly understudied area. Again, the lack of phenotyping for studies and the ability to determine the presence of ongoing active disease, especially in cases of fibrotic sarcoidosis, has likely hampered this research area. We are still unclear whether fibrosis only occurs in pre-disposed individuals, in which cases the fibrotic process may be activated at disease onset, or whether it has the potential to evolve in anyone as a wound-healing response if the inflammation remains uncontrolled. Any advance in understanding the pathophysiology of fibrosis in sarcoidosis will move research closer to targeted therapy for this cohort of patients that suffer significant morbidity, and any overlap with other fibrotic lung diseases that are observed may also open up treatment options, such as the recently introduced anti-fibrotic agents in IPF.

1.5 Gene Expression Profiling

Gene expression profiling is a technique for measuring the expression of thousands of genes at once, to create a global picture of cellular function or active pathways. Messenger RNA (mRNA) is extracted from biological samples, reverse transcriptase is employed to create cDNA (complimentary DNA) from this template, and then DNA microarray technology is employed to 'probe' for a specific DNA sequence that relates to a specific gene, thus allowing the intensity of that gene presence to be calculated. The relative amounts of gene expression between two distinct populations can then be observed, and further interrogated for their role in distinguishing the two.

A parallel can be observed in the investigation of cancer phenotypes and sarcoidosis phenotypes. It is well recognized that patients with the same cancer diagnosis are also observed to have markedly different clinical courses and treatment responses, which likely reflects molecular heterogeneity within a diagnostic category. Cancer research has led the way in demonstrating how gene expression microarray can be used to predict prognosis or establish pathological pathways suitable for targeted therapy. Using gene expression profiling, the identification of groups of genes (signatures) that can predict a good or poor prognosis in patients with lymph-node-negative breast cancer (164), led to the development of two gene-expression-based technologies which are currently available to patients and clinicians as a prognostic indicator. Although the overall influence of these prognostic tests on the actual treatment of breast cancer remains to be seen, the successful translation of microarray data to a clinical test provides a

strong rationale that similar approaches might lead to improved care of patients in other disciplines, including sarcoidosis.

Despite great focus on the identification of biomarkers for immune mediated diseases such as rheumatoid arthritis, SLE and now sarcoidosis, using gene expression and proteomic approaches, very little progress has been made so far to translate these biomarker data into novel therapeutic agents and diagnostic tests. Other fields of medicine, however, again most notably in cancer research, have benefited from biomarker-driven research (37). In the late 1980s, researchers found that 30% of patients with metastatic breast cancer had amplifications in the *HER2* gene, which encodes human epidermal growth factor receptor 2 (165). *HER2* overexpression in tumors is associated with diminished survival of patients. Following confirmation at the protein level (166) a humanized mono- clonal antibody was generated that interferes with Her2 signaling. Trastuzumab (Herceptin®; Genentech, san Francisco, Ca) was approved for the treatment of Her2-positive breast cancer in 1998 and provides a striking example of a biomarker discovery that led to a targeted therapeutic intervention.

Gene expression profiling has been used for the investigation of a variety of fibrotic lung diseases. Selman et al (167) compared gene expression patterns from patients with idiopathic pulmonary fibrosis (IPF), hypersensitivity pneumonitis (HP) and nonspecific interstitial pneumonia (NSIP) using custom oligonucleotide microarrays, and found distinct signatures that differentiated HP from IPF. The HP gene expression signature was interrogated and found to be enriched for genes that are functionally associated with inflammation, T-cell activation and immune

responses, whereas the IPF signature was characterized by the expression of tissue remodeling, epithelial and myofibroblast genes. Zuo et al (158) used gene expression profiling and identified a significant enrichment of MMP7 in IPF patients compared to controls. They went on further to demonstrate that MMP7 knockout mice were dramatically protected from the development of pulmonary fibrosis in a bleomycin induced model, thus paving the way for potential targeted therapy in this condition. Gene expression profiling in sarcoidosis has identified ADAMDEC1 and MMP12 as being highly overexpressed in sarcoidosis lung tissue compared to controls, the levels of which were also seen to correlate to disease severity (168). Published work from our group (169) examined microarray expression of 26,626 genes in transbronchial biopsies of granulomatous areas in the lungs sampled at first diagnosis of sarcoidosis, with longitudinal follow-up of their outcome at two years. Patients were then categorized as either self-limiting disease, or progressive-fibrotic, on the basis of persistence of symptoms or changes in chest radiograph. Three hundred and thirty-four genes were seen to be differentially expressed between these two groups, with gene set enrichment analysis revealing gene sets related to host immune activation, proliferation and defense among those genes upregulated in the progressive fibrotic group.

These examples illustrate the power of global gene expression analysis of human tissue samples to identify molecular pathways and individual genes involved in clinical disease. This in turn may lead to targeted therapies, greater understanding of the pathogenesis of this heterogenic condition, and potentially a way to determine not only who is at risk of developing fibrotic sarcoidosis, but how best to treat them.

1.6 Project Aims

Our understanding of the heterogeneity of sarcoidosis is limited. Whilst the active granulomatous disease process is clearly defined as a hyperactive Th1 driven response, the variation in disease course and outcome suggests that the immune system molds over time to either resolve, persist, or develop into more complex and progressive disease such as fibrotic sarcoidosis. Fibrotic pulmonary sarcoidosis to date has been poorly defined and understudied, with research for sarcoidosis as a whole centering on sarcoidosis versus controls or other diseases. Moreover, defining what constitutes ongoing disease activity, particularly in the fibrotic cohort where it can be difficult to ascertain whether disease is 'burnt out' or at risk of further progression, has been a challenge.

In this thesis, having phenotyped my cohort as predominant pulmonary sarcoidosis patients, I first focus on the challenge of defining and measuring disease activity. Without this, the decisions on who to treat, and measurement of the response to treatment is rather grey. Starting with data from high resolution computed tomography, that has previously been correlated to pathological specimens, the aim is to address which features correlate with active granulomatous/inflammatory disease, the hypothesis being that disease activity should be quantifiable by the presence of and extent of disease on CT. Following on from this, the development of a surrogate for the CT scoring system, that will allow robust and reproducible clinic-based assessment of activity, using frequently measured and readily available parameters, becomes the focus. Using this scoring system, I can then hypothesise that a difference in the level of immune activation, in particular chemokine activity, will be observed in high activity when compared to low disease activity, and a

transcriptomic approach is employed to address this and gain further understanding of what promotes and enhances this inflammatory condition. For the final chapter, I address what immune mechanisms are at play in active fibrotic sarcoidosis, when compared to active disease without fibrosis, one hypothesis being that a more prominent M2 phenotype of gene expression may be observed in fibrotic disease, as it has been in other fibrotic interstitial lung conditions.

A more complete understanding of the pathways, processes and individual genes involved in the promotion of disease activity in sarcoidosis, and those specifically involved in the pathogenesis of fibrotic sarcoidosis, may illuminate new therapeutic targets, or pathways for disease modulation.

2 Materials and Methods

2.1 Patient Recruitment

The Oxford Sarcoidosis Service is a tertiary referral centre for the surrounding areas of Oxfordshire, Buckinghamshire, Berkshire, Wiltshire, Gloucestershire and Northamptonshire, and sees approximately 192 new referrals and 768 follow-up patients per year. A local protocol developed by the service is in place for investigation, follow-up and initiation of treatment in pulmonary sarcoidosis. A high threshold for starting immune suppression is employed, given the natural history of resolution of disease in the majority of cases. This protocol dictates that patients are followed up in the first 2 years at 3 to 6 monthly intervals, dependent on level of activity. All undergo pulmonary function testing, clinical evaluation of symptoms, and blood tests to look for serological and haematological markers of disease activity at the clinic visit. Chest radiographs are performed yearly, unless otherwise indicated. All patients with a chest x-ray abnormality at presentation have an initial High Resolution Thoracic CT (HRCT). Attempt is made on all patients to obtain tissue diagnosis, either via bronchoscopic sampling, or the sampling of other involved organs. Exclusion of infectious agents is also established via bronchsocopic evaluation. Treatment is only initiated if there is evidence of moderate to severe end-organ dysfunction (TLCO<60% predicted, FEV1 <60% predicted) (170), persistent mild – moderate end-organ dysfunction over two follow-up visits three months apart (TLCO 60 – 79% predicted, FEV1 60 -79% predicted) (170), evidence of fibrosis on chest imaging and persistent disease activity, or in rare cases

progressive symptoms despite topical therapy (inhalers) with evidence of progressive or high burden of disease activity on chest radiographs and blood tests.

For the studies herein, consecutive outpatients referred to the Oxford Sarcoidosis-ILD Service, with a diagnosis of pulmonary sarcoidosis as defined by WASOG criteria (120), with evidence of pulmonary granuloma on transbronchial biopsy, or evidence of typical parenchymal disease on thoracic HRCT. Exclusion criteria included patients not meeting the above definition of pulmonary sarcoidosis; those with more than one other major organ involvement; those with evidence of other concomitant active inflammatory disease; those on Angiotensin Converting Enzyme Inhibitors (ACEi). Recruited patients were phenotyped as per the flow chart in Figure 2.1. Only those following the phenotyping pathway highlighted in green were recruited.

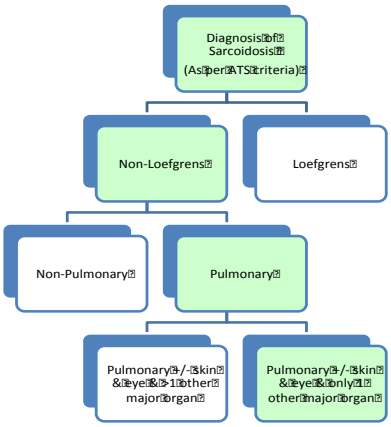


Figure 2.1 Phenotyping algorithm used to recruit sarcoidosis patients for study. Green pathway highlights those qualifying for recruitment.

The studies were approved by the South Central National Research Ethics Service committee.

2.2 Radiology Methods

2.2.1 HRCT Protocol

In patients under the age of 45 years, an inspiratory HRCT was acquired using 1.25mm slices every 10mm and an expiratory HRCT using 1.25mm slices every 30mm; images were reconstructed using a high spatial resolution algorithm. In patients over the age of 45 years, a volumetric scan was performed with 0.625mm slice thickness at an interval of 0.625mm. 2.5mm axial and 3mm coronal and sagittal images were reconstructed using a soft tissue kernel. 1.25mm HRCT images were also reconstructed from the raw data; 1.25mm expiratory HRCT images were acquired in the same way as in patients under 45 years. Intravenous contrast was not routinely used but was administered in several patients in whom malignancy was considered clinically at the time of the scan. This protocol has been developed by the Thoracic Radiologists at Oxford University Hospital NHS Trust, and it's use is routine for all patients undergoing HRCT, not specifically those with sarcoidosis (R Benamore, F Gleeson, V St. Noble).

2.3 Laboratory Methods

2.3.1 Isolation of human peripheral blood mononuclear cells

Blood was collected into tubes containing sodium heparin (≥1U/ml) as an anticoagulant (Greiner Bio-One) or into 50ml polypropylene tubes (BD Falcon) prepared with 100µl (1000U/ml) of sodium heparin (Wockhardt). Peripheral blood

mononuclear cells (PBMC) were separated from whole blood using standard density gradient centrifugation. Whole blood was diluted 1:1 with R10 and a maximum of 25mls of the diluted blood was layered onto 15mls of Lymphoprep^{TM} solution (AXIS-SHIELD). This was centrifuged at 2000 rpm for 20 minutes with no brake and the interface containing PBMCs harvested with a Pasteur pipette. The cells were washed in R10 (RPMI-1640 (Sigma) + 2mM 1-glutamine, + 50µg/ml penicillin (PAA) + 10% sterile heat inactivated Fetal Calf Serum (Sigma)) and centrifuged at 1600 rpm for 15 minutes. One further wash was then performed followed by centrifugation at 1200 RPM for 10 minutes. PBMC were counted and then frozen immediately after isolation in FCS + 10% DMSO (Sigma), and stored in liquid nitrogen. PBMCs were thawed by placing cryovials in a water bath at 37°C for rapid thawing. They were then transferred to a 50ml Falcon tube (BD) and washed twice in R10 prior to RNA extraction or staining with monoclonal antibodies and flow cytometry analysis.

2.3.2 Isolation of human bronchoalveolar lavage cells

Bronchoalveolar lavage (BAL) fluid was collected at the time of bronchoscopy, by wedging the scope either in the right middle lobe or left lingula orifice, and washing twice with 50mls of normal saline. The amount of BAL returned was recorded, and collected into 50ml polypropylene tubes (BD Falcon). BAL fluid was then passed twice through a 70µm cell strainer, then centrifuged at 1600 RPM for 8 minutes. The BAL fluid was then removed and stored, and the cell pellet re-suspended in 2ml Red Cell Lysis Solution (Qiagen), and incubated at room temperature for 10 minutes. Cells were then centrifuged again at 1600 RPM for 8 minutes, and the process repeated if the presence of red cells was still evident. Cells were then re-suspended in 1ml of PBS, counted, and immediately frozen in FCS + 10% DMSO (Sigma), and

stored in liquid nitrogen until analysis. BAL cells were thawed as per PBMCs at the time of analysis/RNA extraction.

Dilution of BAL fluid was calculated in conjunction with the Biochemistry department at Oxford University Hospitals NHS Trust (Natalie Mullee and Tim James), by measuring paired samples of serum with the BAL for total urea, given that the two should be equal.

2.3.3 Isolation of RNA from transbronchial lung tissue

Transbronchial lung tissue was collected at the time of bronchoscopy and immediately placed in to RNA later solution (Thermo Fisher Scientific), which is an aqueous, nontoxic tissue storage reagent that rapidly permeates tissues to stabilize and protect cellular RNA. Tissue samples were then homogenized manually within 24 hours of sample collection, using glass homogenisers, in RLT buffer (Qiagen) and RNA extraction performed (see section 2.3.8).

2.3.4 Serum Collection

Whole blood was collected from patients at the time of bronchoscopy, into tubes containing clot activator and gel for serum separation (BD Vacutainer Blood Collection Tubes – gold top; www.bd.com/vacutainer). Samples were centrifuged at 3000 RPM for 5 minutes at 4°C. Serum was then aliquoted into 1.5ml eppendorfs, and frozen at -80°C until the time of analysis. Serum samples were kept on ice until the point of freezing.

2.3.5 Flow Cytometry

Antibodies to CD14, CD16, CD3, CD4, CD8a, IL-2, IL-10, IL-17A, TNF-a, and IFNg were obtained from eBioscience. Antibodies against CCL18 were obtained from R&D systems and BD Pharmingen. Table 2.1 lists the antibodies used in this project. For surface staining (phenotyping), cells were incubated for 20 minutes at 4°C in the dark with fluorochrome conjugated monoclonal antibodies (mAb) in 96 well plates at a concentration of $0.25x10^6$ cells in FACS buffer (10% fetal calf serum, 1% bovine serum albumin in PBS). All mAbs were optimised for flow cytometry, appropriate concentrations of antibodies were determined by staining $0.25x10^6$ cells with different concentrations of mAb. Cells were washed twice and resuspended in FACS buffer, and fixed with cell fixative (1.25% Formaldehyde (Sigma) in PBS). For Intracellular cytokine staining, surface antigens were stained and fixed as described above. Cells were then permeabilized by washing and resuspending in Saponin FACS buffer (0.5% Saponin (sigma) + FACS buffer as described above), before conjugation with monoclonal antibodies for cytokines of interest. Again cells were incubated with mAb for 20- 30 mins at 4°C (depending on manufacturers instruction), then washed twice and fixed with cell fixative. Cells were then re-suspended in FACS buffer prior to flow cytometry analysis.

Antibody	Fluorochrome	Clone	Isotype	Supplier	Catalogue No
CD14	eFluor 450	61D3	Mouse IgG1 kappa	eBioscience	48-0149
CD14	FITC	61D3	Mouse IgG1 kappa	eBioscience	11-0149
CD14	APC-eFluor 780	61D3	Mouse IgG1 kappa	eBioscience	47-0149
CD14	PE-Cy7	61D3	Mouse IgG1 kappa	eBioscience	25-0149
CD14	APC	61D3	Mouse IgG1 kappa	eBioscience	17-0149
CD16	FITC	eBioCB16	Mouse IgG1 kappa	eBioscience	11-0168
CD16	PE	B73.1	Mouse IgG1 kappa	eBioscience	12-0167
CD16	eFluor 450	eBioCB16	Mouse IgG1 kappa	eBioscience	48-0168
CD3	PE-Cyanine7	UCHT1	Mouse IgG1 kappa	eBioscience	25-0038
CD3	eFluor 450	OKT3	Mouse IgG2a kappa	eBioscience	48-0037
CD3	APC	SK7	Mouse IgG1 kappa	eBioscience	17-0036
CD3	PE-Cy5	UCHT1	Mouse IgG1 kappa	eBioscience	15-0038
CD3	APC-eFluor 780	SK7	Mouse IgG1 kappa	eBioscience	47-0036
CD4	FITC	OKT4	Mouse IgG2b kappa	eBioscience	11-0048
CD8a	APC-eFluor 780	SK1	Mouse IgG1 kappa	eBioscience	47-0087
CD8a	FITC	SK1	Mouse IgG1 kappa	eBioscience	11-0087
CD8a	eFluor 450	SK1	Mouse IgG1 kappa	eBioscience	48-0087
CD8a	PE-Cy7	SK1	Mouse IgG1 kappa	eBioscience	25-0087
CD8a	APC	SK1	Mouse IgG1 kappa	eBioscience	17-0087
IL-2	FITC	MQ1-17H12	Rat IgG2a kappa	eBioscience	11-7029
IL-2	PE	MQ1-17H12	Rat IgG2a kappa	eBioscience	12-7029
IL-17A	APC-eFluor 780	eBio64DEC17	Mouse IgG1 kappa	eBioscience	47-7179
IL-17A	PE	eBio64DEC17	Mouse IgG1 kappa	eBioscience	12-7179
IL-17A	FITC	eBio64DEC17	Mouse IgG1 kappa	eBioscience	11-7179
IL-10	eFluor 660	JES3-9D7	Rat IgG1 kappa	eBioscience	50-7108
RatIgG1 K Isotype control	eFluor 660		Rat IgG1 kappa	eBioscience	50-4301
TNF-a	APC	MAb11	Mouse IgG1 kappa	eBioscience	17-7349
TNF-a	eFluor 450	MAb11	Mouse IgG1 kappa	eBioscience	48-7349
IFN-g	APC	4S.B3	Mouse IgG1 kappa	eBioscience	17-7319
CCL18	PE	AZN-CK18	Mouse IgG1	BD Pharmingen	561179
CCL18	APC	64507	Mouse IgG	R&D systems	IC394A
7-AAD	PE-Cy5 equivalent			eBioscience	00-6993
RatIgG1 K Isotype control	eFluor 660		Rat IgG1 kappa	eBioscience	50-4301
Mouse IgG1 K Isotype control	PE-Cy7		Mouse IgG1 kappa	eBioscience	25-4714
Mouse IgG1 K Isotype control	APC-eFluor 780		Mouse IgG1 kappa	eBioscience	47-4714
Mouse IgG1 K Isotype control	FITC		Mouse IgG1 kappa	eBioscience	11-4714
Mouse IgM Isotype control	APC		Mouse IgM lambda	eBioscience	17-4752

Table 2.1 Table of antibodies and isotype controls used in flow cytometry analysis

2.3.6 Multiplex-cytokine detection from serum

Custom made multiplex-cytokine detection kit for human IL-2R, IL-10, IL-13, IL-17A, IFNγ and TNFα was obtained from BioRad. The assay was performed according to the manufacturer's instructions (http://www.bio-rad.com/webroot/web/pdf/lsr/literature/10014905.pdf). Briefly, wells were pre-wet and 50µl of the cytokine bead mix (which have the capture antibodies of the cytokines of interest, mixed at bench prior to use) were added to each well and washed twice. Standards were prepared by four fold dilutions of a premixed standard. 50µl of standards, samples or blanks of media were added to appropriate wells and incubated on a plate shaker at room temperature in the dark for two hours. The plate was then washed three times and incubated for another thirty minutes with 50µl of detection antibody. After washing, plates were incubated for 10 minutes with 50µl of streptavidin- HRP. A final wash step was performed and the plates read immediately on the Bio-PlexTM 200 System (Bio-Rad) using Bio-Plex manager software version 5.0. Files were exported to Microsoft Excel for analysis.

2.3.7 Enzyme-linked immunosorbent assay (ELISA)

ELISA kit for CCL18 was obtained from R&D systems and the assay performed as per manufacturers instructions:- http://www.rndsystems.com/pdf/DCL180.pdf. Briefly, samples were diluted 3-fold with calibrator diluent, 100µl of assay diluent was added to each well, which had been pre-coated with CCL18 monoclonal antibodies, then 50µl of standard, control or sample were added. The plate was incubated for 2 hours at room temperature on a horizontal orbital microplate shaker set at 500 RPM. Wells were then washed five times using provided wash buffer. Conjugate antibody (200µl) was then added to each well, and again incubated on the shaker for a further 2 hours at room temperature. The plate was again washed

as described above, followed by addition of substrate solution. This was incubated in the dark at 30 minutes room temperature, 50µl of stop solution added, and the plates read immediately on the Bio-PlexTM 200 System (Bio-Rad) using Bio-Plex manager software version 5.0. Files were exported to Microsoft Excel for analysis.

2.3.8 RNA extraction

Transbronchial biopsies samples were homogenised in RLT buffer and underwent RNA extraction within 24 hours of collection. Bronchoalveolar lavage (BAL) cells and PBMCs were thawed as described above, counted, and immediately underwent RNA extraction. This was performed using the RNeasy mini kit (Qiagen) according to manufacturer's instructions: http://www.qiagen.com/us/products/catalog/sample-technologies/rna-sample-technologies/total-rna/rneasy-mini-kit/#resources. All buffers used were supplied with the kit. Thawed cells, or homogenized tissues were suspended in RLT buffer, and passed through a QIAshredder spin column (centrifuged at 13000 rpm for two minutes). Following centrifugation, 350µl of 70% ethanol was added to the sample before transfer to an RNeasy spin column. This was centrifuged for fifteen seconds at 10000 rpm and the flow-through discarded. The column membrane was then washed with 350µl RW1 wash buffer during a fifteen second centrifugation at 1000 rpm. DNAse incubation mix was prepared by adding 10µl DNase solution, prepared with RNA free water (according to manufacturer's instructions) to 70ul RDD buffer. 80µl DNase I incubation mix was then transferred directly onto the RNeasy spin column membrane and incubated at room temperature for fifteen minutes. One further wash with 350µl RW1 buffer with fifteen seconds centrifugation at 10000 rpm was conducted and immediately followed by two further washes with 500µl RPE buffer (the first centrifugation for

fifteen seconds, the second for two minutes, both at 10000 rpm). The column was then dried by centrifugation at 13000 rpm for two minutes. 30µl of RNAse free water was transferred directly onto the column membrane and centrifuged for one minute at 10000rpm. The flow-through was again transferred to the column membrane and the centrifugation repeated. To ensure all samples put forward for whole genome gene expression assays were of high quality, RNA samples were initially screened using the nanodrop 2000 spectrophotometer (Thermo Scientific) for their 260/280 readings, then further assessed through the "RNA integrity number" (RIN) of the RNA samples using Agilent technology. RIN is an algorithm that evaluates the electrophoretic trace of RNA and determines the RNA integrity. A score of 1 to 10 is assigned to the sample, where 10 represents completely intact RNA. A stringent cut-off of RIN >7 was employed for samples suitable for microarray assay. Samples were then stored at -80°C prior to batch hybridisation to Illumina HT12 V4 Beadarrays by Karene Argoud at the Wellcome Trust Centre for Human Genetics, Oxford.

2.4 Microarray Analysis

RNA from lung tissue, PBMC and BAL cells obtained as described above, was normalized for concentration (300ng in 11µl of RNAse free water) in a 96 well plate (an RIN cut-off of >7 was employed to ensure only good quality RNA was used in the assay). cRNA synthesis and gene chip hybridisation was performed by Dr. Karene Argoud at the Wellcome Trust Centre for Human Genetics, in a single batch using an Illumina HumanHT-12 v4 Expression Bead Chip Kit. After initial quality control steps, data was RMA normalised and a mean intensity was computed for

each probeset. For the study on disease activity, I performed correlation (Spearman Rank) of the gene intensity to increasing SCAS. I then rank ordered this correlation using R program (*www.**r-project**.org/)* and further analysed it using Gene-Set Enrichment Analysis (GSEA) (www.broad.mit.edu/gsea/), to interrogate the biological meaning of those genes correlating with increasing disease activity. For the study on fibrotic sarcoidosis, I computed differential gene expression using TMEV (171) between those with active fibrotic sarcoidosis, and those with active sarcoidosis with no fibrosis, and again rank ordered using R program and further analysed using Gene Set Enrichment Analysis. Dr Eleni Giannoulatou at the Computational Biology Research Group, Weatherall Institute of Molecular Medicine and Dr. Helen Lockstone, Bioinformatics and Statistical Genetics Core, Oxford Genomics Centre, Wellcome Trust Centre for Human Genetics, supported me with scripts for R program.

2.5 Statistical analysis

Statistical analysis was performed using R-program (172), GraphPad Prism 7.0a (GraphPad, La Jolla, USA) and TMEV (171), and the relevant methods used are described in detail for each chapter.

3 Measuring disease activity using CT scans – CTAS

3.1 Introduction

There is currently no established method for quantifying disease activity in sarcoidosis. This contributes to difficulties in deciding when treatment should be started or discontinued, particularly in fibrotic disease (173). It also hampers the evaluation of new treatments such as biologics or cytokine modulators, as there is no agreed upon measurement to assess response. The most commonly reported end point in clinical trials for pulmonary sarcoidosis is change in lung function measurements. However, this does not detect reduction in disease activity after treatment, which, in sarcoidosis, can occur without change in lung function.

Several disease activity measures have been proposed. The more established ones are serum angiotensin converting enzyme (sACE), neopterin, chitotriosidase, soluble interleukin-2 receptor (IL-2Ra), ^{18}F-FDG positron emission tomography (PET) and gallium-67 scans (96, 126, 141, 174-181). However, with the exception of gallium and PET scans, these methods measure activity from specific cell types – for example, serum levels of IL-2Ra, a T-cell surface receptor, represent T lymphocyte activity (126, 175), and neopterin (126, 175) and chitotriosidase (135, 136, 176, 179) reflect macrophage activity. Other factors to be taken into account are that sACE is increased in only 60% of patients and is influenced by functional polymorphism in the ACE gene (96, 182), and the

sensitivity and specificity of gallium-67 scans in sarcoidosis are not yet established (180, 181) and difficult to quantify (141, 174). Sarcoidosis is a complex disease – several immune defects are likely to contribute to disease pathogenesis at any one time – e.g. high T cell activity, abnormally activated macrophages, increased monocyte levels and disturbance in regulatory immune cells (183). These defects may dominate at different points of the disease (154), and therefore an assay that primarily reflects one disease pathway is unlikely to be useful throughout the disease course.

High resolution computed tomographic (HRCT) scans of the lung could assimilate all forms of disease activity pathways, and has the advantage of being performed in nearly all patients for diagnostic purposes (180). Further, CT-pathological studies have shown that typical abnormalities like broncho-vascular nodularity, ground glass opacity and consolidation, correlate with presence of granuloma on open lung biopsies (148-150). Nodularity also associates strongly with bronchoalveolar lavage (BAL) cell count and serum IL-2Ra levels (149) and the profusion of interlobular septal thickening correlates with BAL lymphocytosis (147). Taken together, these studies suggest that systematically measuring these HRCT abnormalities may allow quantification of disease activity in pulmonary sarcoidosis. Indeed, unlike gallium-67 and FDG-PET scanning, HRCT also offers an opportunity to measure the extent or volume of active disease.

In this study, a visual scoring method was devised for extent of each of the CT abnormalities found in pulmonary sarcoidosis, and tested against selected markers of disease activities - sACE, sIL-2Ra and change in lung function. The

CT abnormalities correlating with cellular disease markers were used to make a composite CT scoring system that reflects both the presence and volume of disease activity. This composite score was then validated against response to treatment at one year after commencement or increase in treatment.

3.2 Aim

The aim of this chapter is to determine a method for accurately measuring disease activity for pulmonary sarcoidosis patients, encompassing the requirement for accuracy, the reality of multiple disease pathways, and ensuring accessibility for clinicians. The hypothesis is that disease activity should be quantifiable by the presence and extent of disease on CT.

3.3 Methods

3.3.1 Patient selection and study design

100 consecutive patients with pulmonary sarcoidosis from the Oxford Sarcoidosis Service who had a HRCT scan performed from 2011, and fulfilled the diagnostic criteria proposed by WASOG (161) were recruited. All patients were followed up for one year to record change in treatment and lung function.

Thoracic HRCT scanning and full pulmonary function testing were performed within 4 weeks of each other for all patients. In an unselected proportion of patients - serum soluble IL-2Ra levels (n=21), sACE (n=72) and lung function at baseline and 12 months after CT scan (regardless of treatment) (n=40) were also

performed. Those on ACE inhibitors were excluded due to impact of this on sACE levels.

3.3.2 HRCT protocol, lung function and serum markers

HRCT scans were acquired using a 64-detector row CT scanner (LightSpeed VCT XT; GE Medical Systems, Milwaukee, WI, USA). The protocol is further described in methods section 2.2.1. Pulmonary function tests were performed in one standard pulmonary function laboratory. Serum ACE levels were measured on the Konelab 30 chemistry analyser using a Buhlmann kit (Basel, Switzerland) in a clinical laboratory. Serum IL-2Ra levels were measured using the Bio-plex™ Luminex immunoassay (Bio-Rad, Hemel Hampstead, UK) according to the manufacturer's instructions.

3.3.3 Scoring methods for HRCT abnormalities

Each HRCT was reported systematically by a thoracic radiologist (details below), on a comprehensive proforma that included all CT abnormalities (including fibrotic changes) (Table 3.1). These scans were then scored for the presence, character and extent of sarcoidosis-related HRCT abnormalities (ground glass opacification [GGO], interlobular septal thickening [IST], nodularity, conglomeration, consolidation and intra-thoracic lymphadenopathy) using pre-defined criteria from either the Fleischner Society glossary of terms or pre-defined criteria from previous peer-reviewed publications (Table 3.2) (184-186). Specifically, conglomeration, a CT feature over-represented in pulmonary sarcoidosis is not a Fleischner term but defined as "a large opacity >3cm surrounding and encompassing hilar bronchi and vessels" (185). Extent was estimated visually in

each of the anatomically defined zones on the HRCT (Table 3.3); and measured by quartiles (25%) for GGO and consolidation, 2.5 cm intervals of short-axis diameter for conglomeration, number of nodules per zone (0-25 / 26-50 / >50), and numbers of IST (</=5 or >5) per image for each zone (187). Sub-1mm nodules were scored as GGO since the distinction between innumerable sub-1mm nodules and GGO was demanding with current HRCT resolution.

ABNORMALITY		DESCRIPTION	
Ground glass opacity (GGO)	Y / N	Axial distribution in lung: central / peripheral / dorsal / ventral / diffuse / dependent / patchy Relation to secondary pulmonary lobule: random / peribronchovascular / subpleural / centrilobular	Extent (to nearest 25%) RUZ LUZ RMZ LMZ RLZ LLZ
Consolidation	Y / N	Axial distribution in lung: central / peripheral / dorsal / ventral / diffuse / dependent / patchy Relation to secondary pulmonary lobule: random / peribronchovascular / subpleural / centrilobular Cavitation Y / N if yes: location (lobe)	Extent (to nearest 25%) RUZ LUZ RMZ LMZ RLZ LLZ
Interlobular septal thickening (IST)	Y / N	Smooth / nodular	<=5 / >5 per image (if striking feature) RUZ LUZ RMZ LMZ RLZ LLZ
Nodules	Y / N	Predominant size: <1mm / 1-3mm / 4-10mm, / >10mm Axial distribution in lung: central / peripheral / dorsal / ventral / diffuse / dependent Relation to secondary pulmonary lobule: random / bronchocentric / subpleural / centrilobular / tree in bud / perilobular margins: smooth / irregular cavitation: Y / N if yes: location (lobe)	Number of nodules per zone (0-25 / 26-50 / >50) RUZ LUZ RMZ LMZ RLZ LLZ Extent (to nearest 25%) RUZ LUZ RMZ LMZ RLZ LLZ
Conglomeration	Y / N	In continuity with R hilum? Y / N L hilum? Y / N short axis dimension R: 0-2.5cm / 2.6-5cm / >5cm L: 0-2.5cm / 2.6-5cm / >5cm	
Lymphadenopathy	Y / N	Largest hilar: R (size+slice no) L (size+slice no) Largest mediastinal: R (size+slice no) L (size+slice no) calcification? Y / N If yes: amorphous / punctate / uniform / eggshell / popcorn (circle all that apply) associated bronchial narrowing? Y / N	if bronchial narrowing: RUL LUL RML Lingula RLL LLL

Table 3.1 Proforma of all relevant CT abnormalities for pulmonary sarcoidosis, and extent of the abnormalities. Full definitions are found in Appendix 8.1. Full definition of zones are found in Appendix 8.2. 'Central'- predominantly occupying inner 2/3 of the lung. 'Peripheral' - predominantly occupying outer 1/3 of the lung or along the inter-lobar fissures. 'Dorsal' - predominantly posterior portion of the lung. 'Ventral'- predominantly anterior portion of the lung. 'Dependent' -predominantly posterior portion of the lung with a relatively horizontal anterior margin.

Terms	Definition
Ground glass opacity	Hazy increased lung opacity, with preserved bronchial and vascular margins.
Consolidation	Homogeneous increase in pulmonary parenchymal attenuation which obscures vessels and airway walls.
Interlobular septal thickening	Thickening of the septa between lobules.
Traction bronchiectasis	Irregular bronchial dilatation due to surrounding fibrosis
***Conglomeration**	A large opacity >3cm surrounding and encompassing hilar bronch and vessels

Table 3.2 Definition of terms for abnormalities on HRCT. All from Fleischner Glossary of Terms (186) apart from asterisked (185).

A.

Abnormalities	Score for extent
GGO (or nodules <1mm)	One zone 1-25% = 1 point; one zone 26-50% =2; One zone 51-75%=3; one zone 76-100% = 4
Consolidation	One zone 1-25% = 1 point; one zone 26-50% =2;one zone 51-75%=3;one zone 76-100% = 4
Interlobular septal thickening	One zone, up to 5= 1 point; one zone >5 = 2
Nodules 1 mm or greater	One zone 1-25 nodules = 1 point; one zone 26-50 nodules = 2 one zone >50 nodules = 3
Conglomeration	Yes =1 point; no =0; each 2.5cm dimension (Right or Left) = 1
Lymphadenopathy	Yes =1 point; no =0

B.

Definition of zones on HRCT	
Upper zone	Predominantly above the carina in the cranio-caudal plane
Lower zone	Predominantly below inferior pulmonary veins in the cranio-cauda plane
Mid zone	Predominantly between UZ and LZ

Table 3.3 (**A**) Scoring system for HRCT abnormalities, and (**B**) definition of 'zones' used to measure extent. GGO – ground glass abnormalities. Nodules that were sub-1mm were classified under GGO.

Scoring of the HRCT scans was performed by a Fellowship-trained Consultant Thoracic Radiologist based at a regional Thoracic Imaging Centre, with ten years experience in interpreting thoracic CT. To evaluate reproducibility of composite markers, the first 85 scans were selected for scoring by a second radiologist, who is also Fellowship-trained, with seven years experience, and based in a district general hospital. The HRCT scans were reviewed in random order, without knowledge of clinical details other than the patients had a diagnosis of sarcoidosis. Images were reviewed on standard window settings (1500WW, -700WL for lung windows, 400WW, 40WL for soft tissue windows).

3.3.4 Statistical methods

Correlations between extent of CT abnormalities and lung function changes or disease activity variables were performed using Spearman's Ranked Correlation test. The predictive capability of the CT scores for treatment response was modeled with linear regression analysis. These statistical methods are shown within the text of the Results. Analysis was performed in conjunction with a bio-statistician, using the R statistical programming environment (172) or GraphPad Prism. Agreement between radiologists was measured using Pearson's correlation and systematic discrepancies between the two scores were examined with Bland-Altman test.

3.4 Results

3.4.1 Patients demographics

97 patients had complete CT analyses. Patients were aged between 24 and 79 years, with an FVC range between 37-135% predicted. 91% were Caucasian and all but 8 had had tissue confirmation of non-caseating granuloma. In those who did not have histological confirmation, the clinical and radiographic findings were typical, and bronchoalveolar lavage (BAL) showed lymphocytosis, with no evidence of infectious agents on BAL or bronchial washes. None of the patients presented with Loefgren's syndrome, although this was not excluded intentionally. Demographic data are shown in Table 3.4.

n	**97**
Age range (years)	24-79
Age mean (S.D.)	51 (13)
FEV1 Range	32-124%
FEV1 Mean (S.D.)	85 (20)%
FVC Range	37-135%
FVC Mean (S.D.)	98 (19)%
DLCO Range	26-139%
DLCO Mean (S.D.)	80 (21)%
Gender (F)	42 (44%)
Current smoker	8 (8%)
Afro-carribean	6 (6%)
Asian (sub-Indian continent)	3 (3%)
Scadding Stage 1/2/3/4	20/49/9/19
Lung involvement	97 (100%)
Skin involvement	8 (8%)
Liver involvement	7 (7%)
Cardiac involvement	7 (7%)
CNS involvement	2 (2%)
Bone involvement	2 (2%)

Table 3.4 Demographic data for patient cohort. Pulmonary function test refers to % predicted for age, gender and height. Unless stated, % in parentheses refer to % of cohort.

3.4.3 HRCT findings in patient cohort

All patients demonstrated typical changes of sarcoidosis. Of the six abnormalities examined, 100% of patients had intra-thoracic lymphadenopathy, 33% showed presence of GGO, 73% nodularity, 34% IST, 24% conglomeration and 20% consolidation. The range and median of the scores are shown in Table 3.5. Thirty-four (35%) patients showed evidence of fibrosis (defined as presence of any honeycombing, traction bronchiectasis, reticulation or intra-lobular linear opacities with or without architectural distortion or lobar volume loss). Four patients had evidence of pulmonary hypertension on CT scan (increase pulmonary artery diameter), confirmed on right heart pressure measurements on echocardiography. One of these patients was on treatment for secondary pulmonary hypertension (Sildanefil).

	Nodularity	GGO	Consolidation	ILT	Conglomeration
Range	0-18	0-24	0-6	0-12	0-4
Median (IQR)	6 (0-14)	0(0-2)	0(0-0)	0(0-2)	0(0-1)

Table 3.5 Range of CT abnormality extent scores in patient cohort

3.4.4 Correlation of individual CT abnormality score with markers of disease activity and burden

I first examined whether the extent scores for the six CT abnormalities of GGO, nodularity, IST, conglomeration, consolidation and lymphadenopathy showed an association with disease activity. Serum ACE, IL-2Ra and change in FVC within a year (regardless of treatment) were used to reflect active macrophage and granuloma, T cell activity, and overall disease activity respectively. As there was no single gold standard disease activity marker, these parameters were selected to represent the range -granuloma burden (sACE), T cell activity (IL-2Ra) and

composite physiological changes (FVC change). FVC changes were measured as % change over a year regardless of treatment to represent disease reversibility and a surrogate of activity. Each CT abnormality score was compared to each of these disease activity markers using Spearman's rank correlation test. Graphical representation of these analyses are shown in Appendix 8.3.

I found that the extent score for nodularity, IST, consolidation and GGO but not conglomeration correlated with at least one of the disease activity markers (**Table 3.6**). Notably, no significant correlation was found between conglomeration and lymphadenopathy, with any of the disease activity markers or lung function tests.

Correlation r (p value)	Nodularity	GGO	Consolidation	ILT	Conglomeration
sACE	0 (0.70)	0.3 (0.01)	0.1 (0.52)	0.1 (0.40)	0 (0.91)
sIL2-R	0.6 (0.01)	-0.4 (0.08)	-0.1 (0.77)	0.2 (0.49)	-0.3 (0.88)
Change in FVC	0.1 (0.40)	0.3 (0.03)	0.3 (0.03)	0.3 (0.03)	0.3 (0.09)

Table 3.6 Correlation between disease activity surrogates and individual CT abnormalities. R and p value calculated using Spearman Rank Correlation test. Significant correlations are shown in red.

The number of nodules may not reflect the true extent of nodularity as this may be confounded by nodule size (e.g. twenty 10-mm nodules may reflect more extensive active disease than fifty 2-mm nodules). I therefore questioned if the extent of nodularity scored by 25% portions of involvement rather than number of nodules (as above) showed better correlation with our selected disease activity marker and lung function tests. A random subset of scans was examined (n=50). These correlated strongly with each other ($r=0.94$; $p<0.0001$; Spearman rank sum test); and the original method of enumerating nodules per zone was used.

Based on the above findings, I excluded conglomeration and lymphadenopathy and took the sum of the scores of GGO, IST, consolidation and nodularity as the composite score for disease activity. I termed this the 'CT Activity Score' or 'CTAS'. The CTAS score in our cohort ranged from 0 to 58, with a median of 10 (IQR 0-19) (Figure 3.1 A). Fibrotic patients had higher CTAS compared to non-fibrotic [(median of 15.8 (IQR 8.5-22.0) vs 9.8 (4.0-15.0); p<0.001, Mann-Whitney test] (Figure 3.1 B).

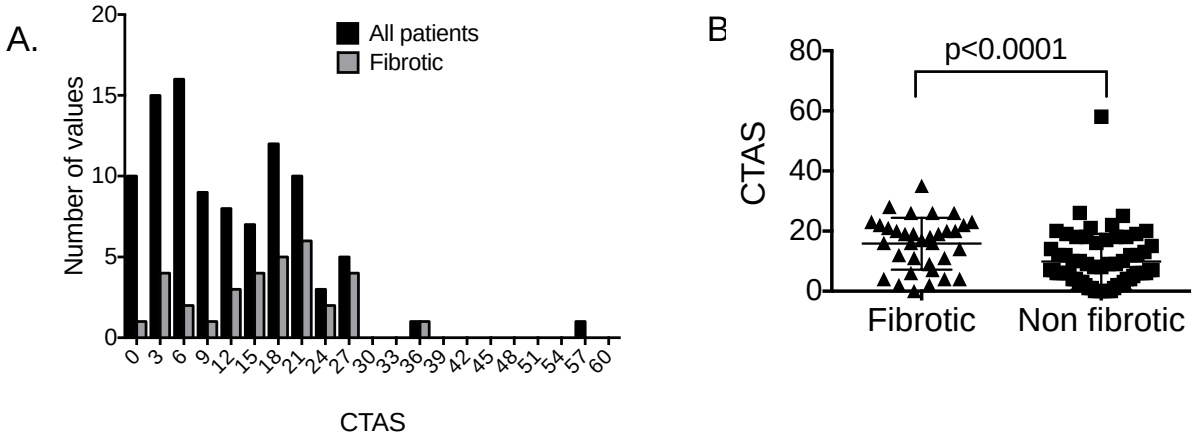


Figure 3.1 A. Distribution of CTAS for the patient cohort (n=97). **B.** CTAS in non-fibrotic and fibrotic pulmonary sarcoidosis. p value measured by Mann-Whitney test.

CTAS showed a significant negative correlation with lung function:- FEV1 (r= -0.45; p<0.001), DLCO(r= -0.36; p<0.001), FVC (r= -0.30; p<0.001); thus the more severe the lung disease, the more active the disease (Figure 3.2 A-C). The four patients with pulmonary hypertension had a CTAS of 0,6,9 and 35.

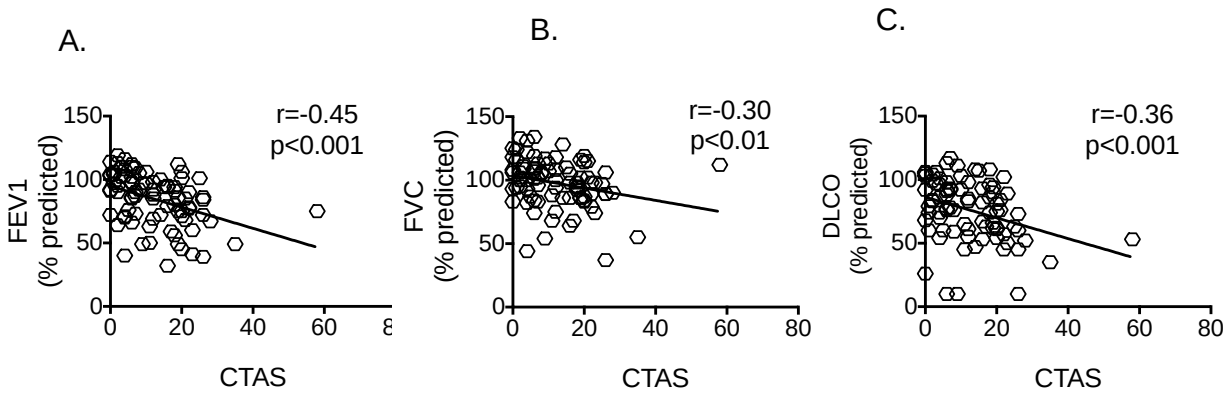


Figure 3.2 A-C Correlation of CTAS with contemporaneous lung function in all patients. r and p value measured by Spearman correlation test.

3.4.5 CTAS predicts response to treatment

To provide a validation of the relationship of CTAS to disease activity, I examined if CTAS predicted response to treatment. I reasoned that if CTAS reflected disease activity, then, the higher the CTAS value, the greater the ability of lung function (FVC) to improve with treatment, since non-active, 'burnt out' disease would show little improvement in FVC. For the first of these analyses, I extracted patients who, at the point of HRCT scanning, went on to start or increase treatment specifically for their lung disease (n=31). Two patients were excluded, one because of the development of pulmonary embolism and a new diagnosis of lung cancer during the follow up period, and one because commencement of treatment was for neuro-sarcoidosis, leaving 29 patients. All patients received either Prednisolone at 0.4-0.5 mg/kg (n=28), or Azathioprine (n=1). FVC values at 12 months (± 8 weeks) after treatment were compared to pre-treatment. I found that CTAS predicted the size of FVC response to treatment at one year (b=0.42 [C.I. 0.05-0.79]; r2=0.17,p=0.03) (Figure **3.3** A). This was also observed for the fibrotic subgroup, where 20 of the 34 patients received treatment (b=0.32 [C.I. 0.0-0.6]; r2=0.20,p=0.04) (Figure **3.3** B).

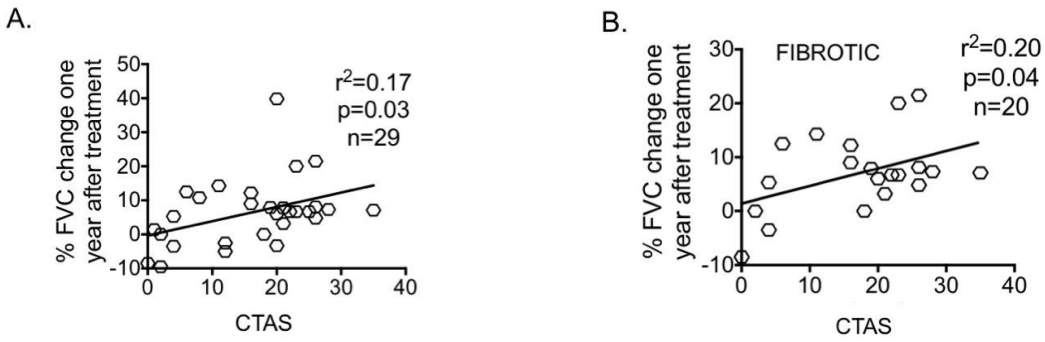


Figure 3.3 Association between CTAS and changes in FVC at 12 months in all patients (**A**) and in the fibrotic subset (**B**) who received new or increased treatment after CT scan. For all patients r and p values shown are from linear regression calculations. Spearman Correlation values are (**A-B**) r=0.40, p=0.03 and r=0.42, p=0.02 respectively.

Comparing the CTAS to any one CT parameter (GGO, nodularity, consolidation and IST), I found only GGO and consolidation to predict treatment response (p=0.04 and 0.05 respectively), and at a weaker certainty than CTAS (p=0.03). Therefore the composite measure of CTAS provides the strongest correlation with FVC response after treatment.

3.4.6 Reproducibility of CTAS between 2 radiologists

Finally, I questioned if CTAS measurements were reproducible between two radiologists. I found a strong correlation for the CTAS (r = 0.91; p<0.0001; Pearson's correlation test; n=85) (Figure 3.4 A), and Bland-Altman analysis confirmed that there was no systematic deviation in measurements over the range of values (Figure 3.4 B). There was no difference in aCTAS between the two radiologists (r = 1.0; n=85).

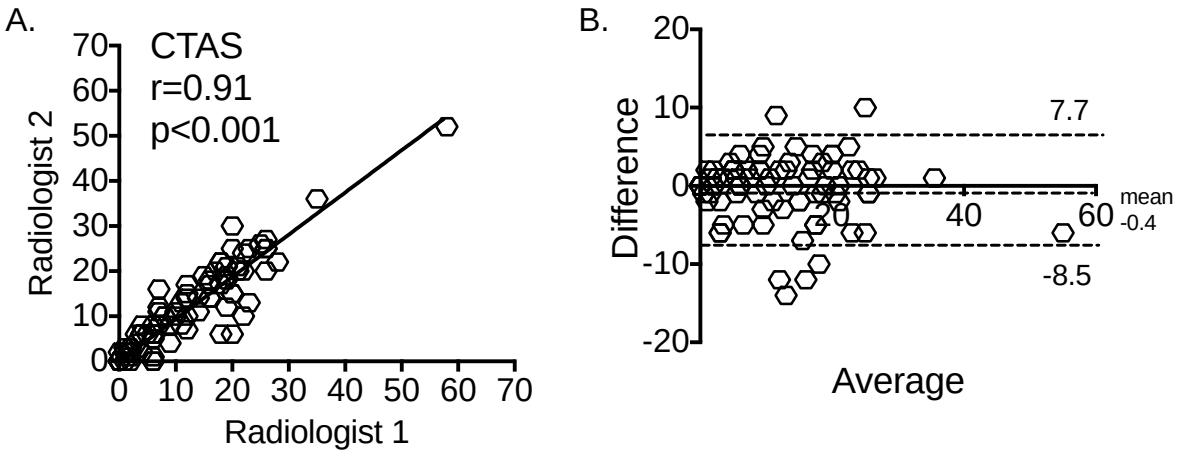


Figure 3.4 A. Agreement between two radiologists for CTAS. Pearson correlation analysis (n=85), and **B.** Bland Altman test. Mean differences (bias) of -0.4, and 95% limit of agreements (precision) (-8.5 to 7.7) are shown.

3.5 Discussion

In this study, I describe a systematic method for scoring disease activity in pulmonary sarcoidosis, using thoracic HRCT scans. Six HRCT abnormalities, which are typical for sarcoidosis and which have also been shown to correlate with granuloma or T cell infiltrates, were first selected as potential components for a composite activity score. These were tested against a range of surrogate disease activity markers, and those CT features that did not correlate with any markers were excluded. The sum of the remaining four CT extent scores (GGO, nodularity, IST and consolidation) was used as the composite lung disease activity score (CTAS). I show that CTAS correlated with the extent of improvement in lung function in pulmonary sarcoidosis, including the fibrotic subset. This composite score was superior to any single CT abnormality in its correlation to FVC response to treatment.

CTAS provides a promising concept for the measurement of disease activity in pulmonary sarcoidosis. In addition to being highly accessible, it is also reproducible

since both radiologists registered near unity in the correlation tests despite differences in experience and expertise. However, several limitations exist. As there is no universally accepted marker for disease activity against which a new marker can be developed, I was reliant on current best measurements. To enhance this, I started with CT features that were directly correlated to relevant pathology in CT-pathology and post mortem tissue studies (147, 149, 150) providing the closest gold standard disease activity marker. From these six pathology-correlated CT features, I narrowed this down to those that showed significant positive association with contemporaneous cellular and physiological markers of activity. Change in FVC (better or worse) was used regardless of treatment to reflect activity, with improvement indicating reduction in activity and worsening, persistent active disease. For validation of the composite marker, I examined the ability of the CTAS to predict improvement with treatment. Here only positive change (i.e. improvement in FVC; although I accepted -10% within the limit of variability) was used and I only selected those who received new or increase in treatment compared to baseline. I reasoned that a response to treatment specifically indicated reversible disease activity. Arguably, a worsening of disease despite treatment could also indicate disease activity but this is less specific, since incipient pulmonary hypertension or undiagnosed cardiac failure could also worsen disease. The strongest marker of disease activity is likely to be the response of FVC to treatment in fibrotic pulmonary sarcoidosis since this indicates that there is reversible and therefore a non-fibrotic, active component. Here, despite the small numbers, I showed that CTAS correlated with the size of improvement in FVC after treatment.

The study was limited in size. Although the starting numbers were large relative to most other studies on sarcoidosis, the correlation studies were smaller due to restrictions in synchronizing collection of samples with CT scanning. The prospective study on response to treatment was necessarily limited by patients within the cohort who went on to receive treatment or an increase in treatment after the CT scan. Despite these limitations, this is the first description of a numerical CT score for extent of disease activity in pulmonary sarcoidosis. Improving on previous CT-related studies in sarcoidosis (147, 148, 188-190), the CT abnormalities were defined strictly using radiological terms from an established glossary (the Fleischner Society glossary of terms for pulmonary imaging) in order to reduce subjective interpretation. This is reflected in the excellent agreement between the two radiologists despite their different hospital settings. I also refined the measurement of extent of the activity by defining the degree of involvement in each anatomically defined zone.

Measurements of disease severity on HRCT (but not activity) have been previously published (188, 190, 191). These variably incorporate different HRCT abnormalities, functional impairment and end-organ damage. CTAS is specifically designed to reflect the extent of disease activity regardless of the level of lung function impairment and fibrosis. Indeed, CTAS would be particularly useful in aiding decisions on whether to treat fibrotic sarcoidosis. A CT scan showing volume loss and established fibrosis has often been termed ‘burnt out’ fibrosis but may contain areas of active disease that may be driving the fibrotic process (169). A larger number of patients in a confirmatory cohort will strengthen this activity

score and provide an opportunity to derive numerical cut off values for activity using receiver operating characteristic (ROC) curves**.**

It is acknowledged that there are pros and cons to studying patients recruited from a specialist clinic, as is the case here. The cohort is likely to comprise more complex patients, and may be less representative of those found in the community. However, the patients would have been rigorously diagnosed, management patterns are consistent and more severe patients could be captured. These types of patients are also those most likely to be treated and in whom activity scores would be most useful. It is also unlikely that a sizeable study could be performed in a single center without such a setting.

The use of a single radiologist's score with specified experience and expertise in interstitial lung disease radiology was intentional. The reporting of CT abnormalities like GGO and nodularity is well established and studies have previously shown consistently low inter-observer variation for these abnormalities. A second radiologist's CTAS scores were used to interrogate reproducibility of this score, and found to be highly reproducible.

Finally, the finding that CTAS levels predicted treatment response is noteworthy from a disease stratification point. In clinical trials, diseases like sarcoidosis are seldom stratified by activity despite the likelihood that non-active disease may well not respond to treatment. This therefore suggests that CTAS could potentially be used to refine the selection and classification of patients in clinical trials, in addition to its use as an endpoint.

3.6 Conclusion

In conclusion, the study presents an objective and reproducible HRCT scoring system (CTAS) that measures the extent of disease activity in patients with pulmonary sarcoidosis. This utilizes only four CT abnormalities with established CT terms and anatomical definitions, and shows excellent agreement between two radiologists. CTAS could be used to guide treatment decisions in patients with pulmonary sarcoidosis, and potentially to measure response to established and novel treatments. In this thesis, it provides the basis for measuring disease activity in pulmonary sarcoidosis.

4 Sarcoidosis Clinical Activity Score (SCAS) – Design, development and validation of a clinic based tool

4.1 Introduction

The development and validation of the CT activity score (CTAS - Chapter 3), provided a basis for objectively quantifying disease activity in pulmonary sarcoidosis. However, regular HRCT scanning is not justified due to the high radiation dose involved (~5mSv), and is therefore only done at strategic clinically indicated points in a patient's clinical pathway. There is therefore a need for a clinic-based tool, using information already routinely collected by the attending physician.

Without a formal scoring system for activity, the experienced clinician gathers information based on history, examination, pulmonary function testing, serum biomarkers and chest radiography, and a clinical judgment on the level of active disease is made.

Several serological tests have been put forward as potential markers of disease activity. Serum Angiotensin Converting Enzyme (sACE) – the first reported biomarker in sarcoidosis (121), has been used as a diagnostic and prognostic marker, and has been seen to reduce with corticosteroid treatment. It is an acid glycoprotein that converts angiotensin I into angiotensin II. It is secreted by monocytes and macrophages, and pulmonary endothelial cells and is considered a marker of granuloma formation and monocyte/macrophage activity in the disease. However, it is found only to be raised in 60% of patients with a confirmed diagnosis

of sarcoidosis (122), it's variability thought to be due to ACE gene polymorphism, and it may also be elevated in other diseases including miliary tuberculosis(123), diabetes mellitus (124) and thyroid disease(125), further complicating its utility as a solitary marker of activity.

Another potential marker of activity is the level of circulating lymphocytes. Paradoxically, although lymphocytes are activated and total lymphocyte counts are increased, the numbers circulating in the periphery are often seen to be reduced. Lymphopaenia has been reported to represent severe disease (129), and may be seen due to the upregulation of homing markers at sites of inflammation that result in trafficking of the activated lymphocytes (in particular CD4 T cells) (192). The resulting sequestration in affected tissues is balanced by a relative decrease measured in the periphery. Lymphopaenia has also been correlated as a marker of a less favourable outcome and the development of chronic disease, along with a likelihood of requiring immunosuppressive treatment (130).

Yet another possible, clinically validated blood test that might reflect activity is circulating immunoglobulin levels (specifically IgG). Immunoglobulin G, a non-specific marker of systemic inflammation, has been well documented to be raised in sarcoidosis (131). The mechanism is unclear but could partly be explained by T-Helper cell activation and the subsequent influence on B cells and their differentiation to plasma cells. The presence and intensity of the active component of airways disease, "alveolitis", in sarcoidosis may be indicated by the level of IgG, and decreases in IgG have been shown to correlate with improvement in lung function (132). However, if other inflammatory diseases are present in the same

patient its use would have to be modified, as it would not be specific for sarcoidosis activity in that circumstance. Immunoglobulin has also been shown to localize in pulmonary sarcoid granulomas alongside plasma cells (133), so the higher the activity or disease burden, the more granulomas, then theoretically the higher the IgG measurement will be.

It is not unexpected that several immune pathways are taking place during active disease, but it may be that one pathway can be more dominant at any one time-point. There are many other cytokines, not routinely tested for in clinical practice, that have been reported to relate to 'active' sarcoidosis, including for example, TNFα (61), IFN gamma (193), and IL-10 (134, 194). Serum soluble IL-2 Receptor, thought to represent and quantify T cell activation, has also been reported as a marker of sarcoidosis activity (although again is not specific for sarcoidosis and can be seen in several autoimmune diseases), and was validated using positive 67Ga scanning as a correlate of active pulmonary disease (126). Serum chitiotriosidase, produced by activated macrophages found in granuloma has also been suggested as a useful marker of sarcoidosis (135, 195), as has serum Neopterin (196). However, none of the aforementioned markers are readily available in mainstream hospital laboratories. It can be postulated that these may be interesting validation points for an activity score and its associated parameters.

Whilst lung function is routinely assessed by the attending physician to guide on disease progress, disease activity may not necessarily be reflected in persistently abnormal lung function results. Instead, irreversible fibrotic changes, despite disease quiescence, may account for such abnormalities. Change in lung function

may be more representative of changes in disease activity, but may not be a clinically significant (>10% FVC or >15% TLCO) or detectable change.

Finally, addressing the use of symptoms as a measure of active disease, as often used in other activity scoring systems for inflammatory conditions such as Rheumatoid arthiritis and SLE, it must be noted in sarcoidosis that reported symptoms can be difficult to interpret, with lethargy and tiredness being commonplace, alongside cough and shortness of breath. Dyspnoea has been shown to have a strong correlation to low scores in quality of life questionnaires but no correlation to FVC or TLCO (197). It has also been demonstrated that defining active disease using clinical symptoms (e.g. cough, breathlessness, fatigue), results in mis-labelling of up to a third of patients when activity is assessed using 67-Gallium scanning (141).

In this chapter, I examine the use of several readily available blood measurements (sACE, IgG and lymphocyte count) and chest radiography, in order to formalise the clinical judgement usually made from these parameters, into an objective and quantifiable disease activity index in the outpatient setting. This may allow future standardisation in decisions to treat, or change management, and can be performed by the treating physician.

4.2 Aim

To develop a simple to use clinic based tool (Sarcoidosis Clinical Activity Score – SCAS), to facilitate a graded, objective measure of disease activity in pulmonary sarcoidosis, using parameters routinely measured in the clinical setting.

4.3 Methods

A score based on the summation of that attributed to sACE, IgG, lymphocyte count, and CXR score (SCAS) was tested against CTAS derived from thoracic HRCT performed contemporaneously to determine their validity as disease activity measures. In a subcohort, this score and its components were further tested against immunologically relevant molecular and cellular measures that depict disease activity.

4.3.1 Patient selection

This is a retrospective cohort study, where consecutive patients with a diagnosis of pulmonary sarcoidosis as defined by WASOG criteria, were recruited from the Oxford Sarcoidosis Service (n= 65). All patients either had evidence of pulmonary granuloma on transbronchial biopsy (n=53), or in the absence of positive histology, evidence of typical sarcoidosis with parenchymal disease on thoracic HRCT (n=12). Patients not meeting the above definition of pulmonary sarcoidosis; those with predominant disease in other major organs, those with evidence of other concomitant active inflammatory disease, and those on ACE inhibitor

antihypertensive medication were excluded. Patients receiving immune suppression therapy at the time of analysis were permitted (n= 11).

For the initial cohort all patients had sACE, Immunglobulin G, circulating lymphocyte count, a chest radiograph (CXR) and HRCT within 90 days of each other (n= 65). For the CXR and HRCT, standardized scoring was performed, CXR as described in this chapter (section 4.3.3), and HRCT as described in Chapter 3 (section 3.3.3). A subcohort reviewed measurements made within 35 days (n=34).

The following details were recorded for each patient (corresponding to the time of HRCT [or SCAS parameters in further validation cohorts]):- age, sex, smoking status, treatment, pulmonary function tests, presence of symptoms, organ involvement, serum ACE level (IU/l), Serum Immunoglobulin G (IgG) (g/l), Peripheral Lymphocyte count (1.0 x 10^9/l), a chest radiograph score and a CT activity score.

Validation subcohorts had the composite marker of sACE, IgG, lymphocyte count +/- chest x-ray tested against immunologically relevant molecular and cellular events thought to relate to disease activity. The only difference in inclusion/exclusion criteria was that they were not required to have also had a HRCT within 35-90 days of other parameters, but research samples had to be within 35-90 days of the clinical measurements (i.e. laboratory and radiological collection data).

4.3.2 HRCT Scanning protocol and clinical parameter measurements.

HRCT scans were acquired using a 64-detector row CT scanner (LightSpeed VCT XT; GE Medical Systems, Milwaukee, WI, USA). Full protocol for scanning is provided in section 2.2.1.

All cellular parameters were measured in the clinical laboratory at Oxford University Hospitals NHS Trust. Serum ACE levels were measured on the Konelab 30 chemistry analyser using a Buhlmann kit (Basel, Switzerland). Lymphocyte counts were measured using a single platform flow cytometric method on a FACScantoII flow cytometer (Becton Dickinson). Immunoglobulins were measured by nephelometry on a BNII machine (Siemens).

Standard PA chest radiographs were obtained by trained clinical radiographers at the Churchill Hospital, Oxford.

Pulmonary function tests were performed in one standard pulmonary function laboratory (using a medisoft machine by Vitalograph).

4.3.3 Scoring method for chest radiograph (RCAS)

A chest x-ray score was designed to assess both the type of abnormality and the extent of parenchymal involvement, the latter taken to represent the burden of granulomatous deposits (Figure 4.1). First an abnormality score was applied to the radiograph, with a score of 0 for a normal CXR, 1 for bilateral hilar lymphadenopathy (BHL) or reticulation alone, and 2 for presence of nodules or consolidation with or without BHL and reticulation. The extent of nodularity/consolidation over 6 zones was also measured, scoring one point for each zone. Each zone is defined as per

conventional radiological methods (144, 198, 199), described as follows:- upper zone = above inferior margin of 2nd rib anteriorly, middle zone = above inferior margin 4th rib to 2nd rib anteriorly, lower zone = below inferior margin of 4th rib anteriorly. A maximum score of 8 can be achieved from the chest radiograph. This score does not require a thoracic radiologist for interpretation. This score is further referenced as the Radiographic Clinical Activity Score (**RCAS**).

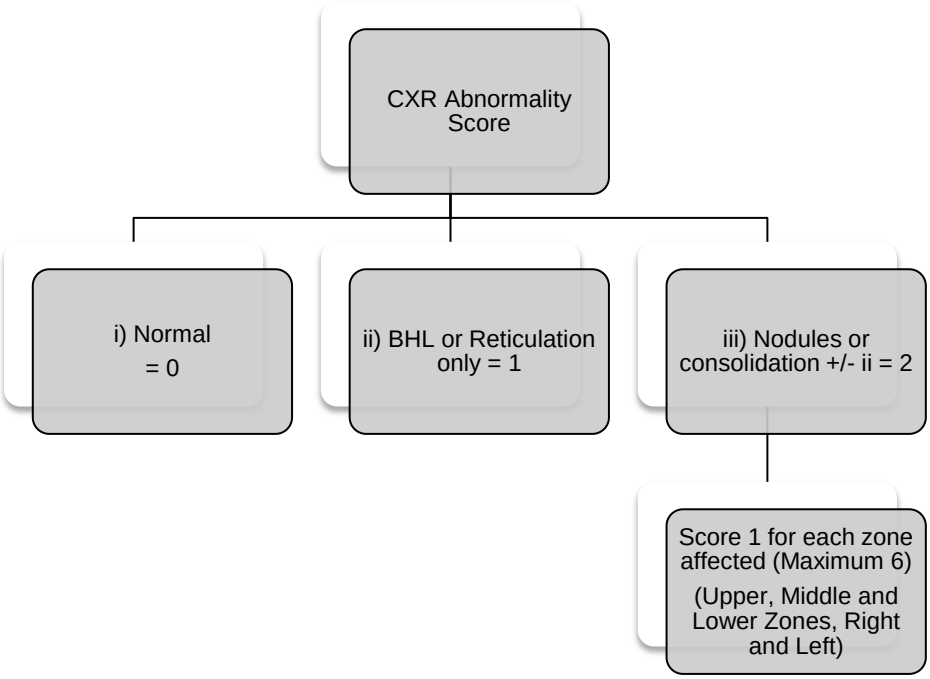


Figure 4.1 Flowchart depicting scoring system for chest radiographs. First a score for abnormality is given (range 0-2). If nodules or consolidation is present (abnormality score 2), then a further score is applied to account for extent. Each lung (Left and Right) is divided into three zones, and for each zone that the abnormality is present in a further additional point is given (range 1-6). A maximum score of 8 can be achieved with the chest radiograph score.

4.3.4 Scoring for serum and cellular biomarkers (CCAS)

For each of the commonly measured biomarkers, a score was attributed, defined by the reference ranges in our laboratory (http://www.ouh.nhs.uk/services/departments/laboratory-medicine/default.aspx), for normal versus abnormal (Table 4.1). Serum ACE was given an extra tier for scoring to reflect the bimodal distribution of abnormal levels (Figure 4.2).

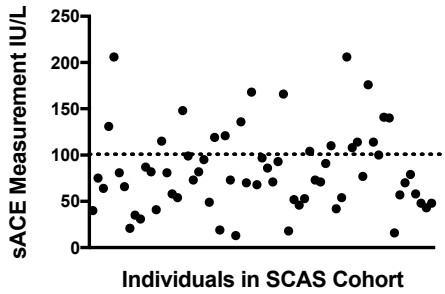


Figure 4.2 Range of Serum ACE values (IU/L) seen in the 65 recruited patients. Each dot represents an individual patient.

Parameter	Score=0	Score=1	Score=2
IgG	≤13 g/l	>13 g/l	
Lymphocyte count	>1x10^9/l	≤1x10^9/l	
sACE	≤55 IU/l	56-100 IU/l	>100 IU/l

Table 4.1 Scores attributed to serum/cellular biomarkers, taken to reflect differing pathways of active disease. Score is 0 if results are within normal limits, 1 if abnormal. sACE is tiered that a score of 2 is given if levels are very high (>100). (IgG - Immunoglobulin G; sACE – Serum Angiotensin Converting Enzyme).

This score is further referenced as the Cellular Clinical Activity Score (**CCAS**).

4.3.5 Sarcoidosis Clinical Activity Score (SCAS)

The Sarcoidosis Clinical Activity Score, or SCAS, is comprised of the addition of the CCAS and RCAS scores as defined above.

4.3.6 CTAS score

A CT activity score (CTAS), which evaluates the extent of active disease, was attributed to each HRCT by two thoracic radiologists (200) (as described in detail in chapter 3). In brief, thoracic HRCT scans were scored for presence and extent over both lungs, of nodules, consolidation, interlobular septal thickening and ground glass opacification, as a representation of disease activity.

4.3.7 Laboratory methods used in validation subsets

All laboratory methods used in the subset analysis for SCAS are described in detail in Chapter 2 Materials and Methods (Section 2.3).

4.3.7.1 PBMC and Serum samples

In brief, peripheral blood mononuclear cells (PBMC) were isolated from whole blood samples obtained at the point of recruitment using standard density gradient centrifugation, and stored in liquid nitrogen until the point of analysis (section 2.3.1). Serum was separated from whole blood (via centrifugation of BD Vacutainer Blood Collection Tubes – gold top; www.bd.com/vacutainer) from patients at the time of first presentation to the Oxford Sarcoidosis service, and frozen at -80°C until the time of analysis. Serum samples were kept on ice until the point of freezing (section 2.3.4).

4.3.7.2 Flow cytometry and serum cytokine quantification

For flow cytometry, antibodies to CD14, CD16, CD3, CD8, IL-2, IL-17, TNF-a, IFNg (ebioscience) and CCL18 (BD Pharmingen and R&D systems) were used for surface and intracellular cytokine staining (described in detail in section 2.3.5). The protocol for surface and intracellular staining is detailed in section 2.3.5. A custom made multiplex-cytokine detection kit for levels of human IL-2R, IL-10, IL-13, IL-17A, IFNγ and TNFα (BioRad) was performed according to the manufacturer's instructions:-http://www.biorad.com/webroot/web/pdf/lsr/literature/10014905.pdf (detailed further in section 2.3.6). Serum was measured for CCL18 using an ELISA kit (R&D systems) according to manufacturers instructions:-http://www.rndsystems.com/pdf/DCL180.pdf (section 2.3.7). The plates for both

multiplex and ELISA assays were read on the Bio-PlexTM 200 System (Bio-Rad) using Bio-Plex manager software version 5.0.

4.3.8 Statistical Analysis

Spearman Rank correlation, Linear Regression analysis, Pearson's correlation and ANOVA analysis were used in the statistical analysis. The use of each test is described in the relevant results section.

4.4 Results

4.4.1 Patient demographics

Sixty-five consecutive outpatients with pulmonary sarcoidosis fulfilling the inclusion criteria were recruited from the Oxford Sarcoidosis Service outpatient clinic at the Churchill Hospital between 2011 and 2014. These patients had initially presented between June 2007 and July 2012, consequently their HRCT and contemporaneous biomarker/radiographs were analysed from the initial presentation (Table 4.2). Thirty-four patients (52%) had contemporaneous measurements within 35 days. The frequency of SCAS distribution in the cohorts is displayed in Figure 4.3.

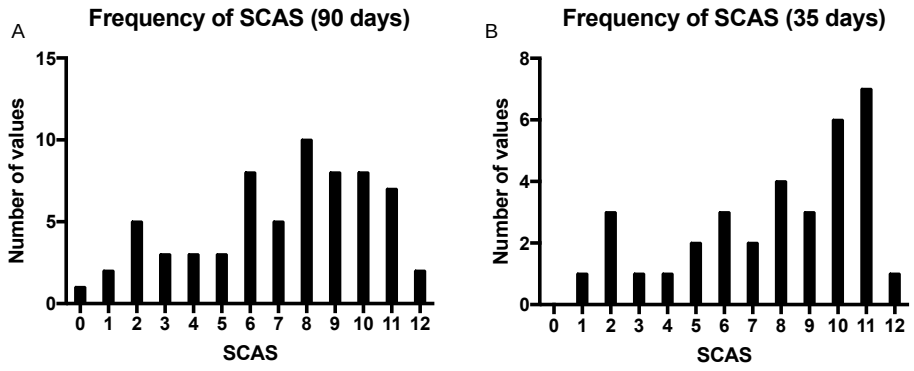


Figure 4.3 Distribution of SCAS across the two cohorts. A – within 90 days of parameter measurements; B – within 35 days of parameter measurements.

n	**65 (within 90 days)**	**34 (within 35 days)**
Age range (years)	24-78	24-78
Age mean (S.D.)	48 (12)	49 (13)
FEV1 (% Predicted) Range	32-119	32 -119
FEV1 (% Predicted) Mean (S.D.)	86.4 (20.43)	84.5 (22.63)
FVC (% Predicted) Range	37-134	37 - 133
FVC (% Predicted) Mean (S.D.)	100.2 (16.35)	98.1 (18.92)
TLCO (% Predicted) Range	26-117	26 - 117
TLCO (% Predicted) Mean (S.D.)	79.8 (20.96)	76.9 (21.34)
Gender M	39 (60%)	21(60%)
Gender F	26 (40%)	13 (40%)
Current smoker	6 (9%)	2 (5%)
Ethnicity - Afro-Caribbean	4 (6%)	1 (3%)
Ethnicity - Asian	3 (5%)	1 (3%)
Ethnicity - Caucasian	58 (89%)	32 (94%)
sACE (IU/L) Range (mean)	13 - 206 (83)	13 - 206 (84)
IgG (g/l) Range (mean)	6.6 - 50 (13)	6.6 - 22 (12.2)
Lymphocyte Count (1.0 x 10^9/l) Range (mean)	0.32 - 2.34 (1.15)	0.32 - 2.34 (1.23)
CXR Score Range (mean)	0-8 (5)	1-8 (6)
Respiratory symptoms	50 (77%)	27 (79%)
On immunosuppression	11 (17%)	5 (15%)
Granuloma on biopsy	60 (92%)	32 (94%)

Table 4.2 Demographics and clinical data of pulmonary sarcoidosis patient cohort for SCAS v CTAS

At the point of presentation, 71% had a raised serum ACE, 35% had lymphopaenia, and 43% had a raised IgG (Figure 4.4).

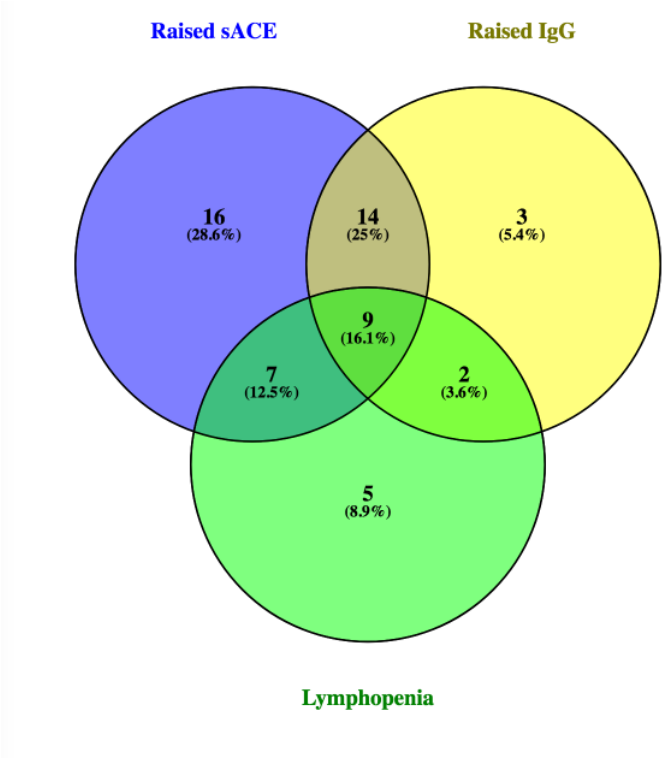


Figure 4.4 Number of patients from n=65 cohort that have a raised serum ACE, raised immunoglobulin G or lymphopaenia.

Following the initial CTAS v SCAS analysis, forty of the recruited patients underwent FACS analysis of PBMCs (isolated at the time of recruitment), to assess both circulating cell phenotype and inflammatory cytokine profile (including TNF alpha, IFN gamma, IL-17, CCL18, IL-2) (as described in methods section 2.3.5). Thirty-two patients had serological analysis for soluble IL2-receptor, and forty-four patients had serological measurements of the chemokine CCL18 (as described in section 2.3.6 and section 2.3.7 respectively). Finally, twenty of the patients in the cohort had longitudinal SCAS data collected following initiation of treatment. Demographic and clinical data for each cohort is shown in Table 4.3.

	CD14+/TNFα	sIL2Ra	CCL18	Longitudinal SCAS
n	**40**	**32**	**44**	**20**
Age range (years)	29-78	29-69	27-69	30-65
Age mean (S.D.)	48 (11)	49 (10)	49 (10.4)	49 (10.7)
FEV1 (% Predicted) Range	41-124	41 - 122	41-122	32-106
FEV1 (% Predicted) Mean (S.D.)	90 (17.4)	90 (17.9)	92 (17.7)	77 (20.3)
FVC (% Predicted) Range	72-133	73-133	72-133	64-118
FVC (% Predicted) Mean (S.D.)	101 (15.8)	102 (15.2)	101 (15.7)	96 (12.8)
TLCO (% Predicted) Range	26-139	47-139	47-139	47-103
TLCO (% Predicted) Mean (S.D.)	83 (20.2)	85 (17.9)	85 (17)	72 (18.1)
Gender M	23 (57%)	19 (59%)	26 (59)	11 (55%)
Gender F	17 (43%)	13 (41%)	18 (41)	9 (45%)
Current smoker	0	0	0	3 (15%)
Ethnicity - Afro-Caribbean	3 (8%)	2 (6%)	3 (7)	0
Ethnicity – Asian	2 (5%)	2 (6%)	2 (5)	2 (10%)
Ethnicity – Caucasian	35 (88%)	28 (88%)	39 (89)	18 (90%)
sACE (IU/L) Range (mean)	19-224 (80)	19 -166 (76)	17-176 (72)	17-206 (99)
IgG (g/l) Range (mean)	6.3-23.9 (13.4)	8.8-28.5 (14.3)	6.3-28.5 (13.2)	6.6-17.3 (12.2)
Lymphocyte Count (1.0 x 10^9/l) Range (mean	0.43-2.22 (1.21)	0.43-2.17 (1.15)	0.43-2.25 (1.2)	0.55-2.17 (1.15)
CXR Score Range (mean)	0-8 (4.8)	1-8 (5)	0--8 (5)	1-8 (6.5)
Respiratory symptoms	28 (70%)	22 (69%)	33 (75)	18 (90%)
On immunosuppression at recruitment	2 (5%)	0	2 (5)	0
Granuloma on biopsy	35 (88%)	29 (91%)	35 (80)	20 (100%)

Table 4.3 Demographics table of all validation cohorts for SCAS.

4.4.2 Correlation of measured parameters to CTAS

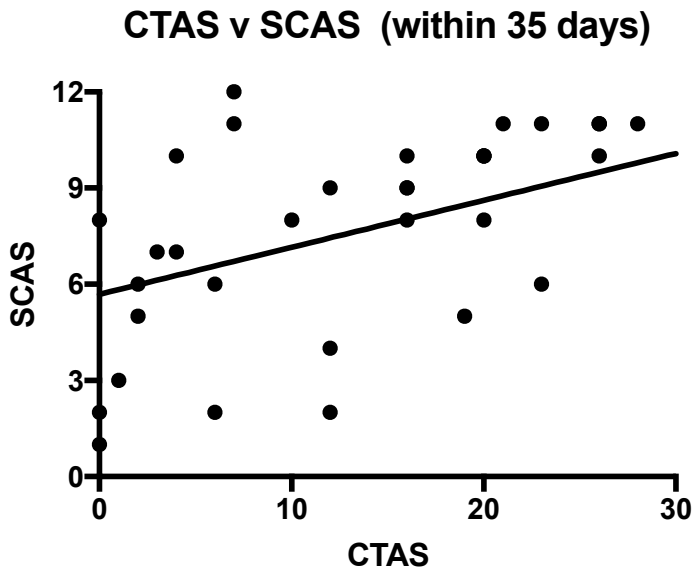
Thirty-four patients had undergone thoracic high-resolution computed tomography (HRCT) within 35 days of measured biomarkers and chest radiography. Individual parameters, combinations of the parameters, biomarker absolute values and scoring systems were correlated to CTAS to develop a Sarcoidosis Clinical Activity Score (SCAS). This was also performed for the larger cohort of sixty-five patients with measurements obtained within 90 days (Table 4.4).

Spearman Rank Correlation to CTAS	Within 35 days Cohort r (p value)	Within 90 days Cohort r (p value)
sACE (absolute values IU/L)	0.34 (0.049)	0.22 (0.07)
IgG (absolute values g/L)	0.22 (0.22)	0.19 (0.14)
Lymphocyte Count (absolute values 1.0 x 10^9/L)	-0.43 (0.01)	-0.34 (0.006)
(sACE + IgG) – Lymphocyte count (absolute numbers)	0.36 (0.04)	0.23 (0.06)
CCAS	0.5 (0.003)	0.34 (0.006)
RCAS	0.62 (<0.0001)	0.71 (<0.0001)
SCAS (CCAS + RCAS)	0.64 (<0.0001)	0.71 (<0.0001)

Table 4.4 Spearman rank correlation of scored/measurable clinical parameters, alone and in combination, versus CTAS. Significant correlations shown in red. SCAS consists of CCAS + RCAS in combination.

In the 'within 35 days' cohort, the use of the previously described scoring system for normal or abnormal serum and cellular biomarkers (CCAS), revealed a stronger correlation (r^2=0.5, p=0.003 Spearman Rank) compared to any of the individual biomarkers in their absolute values. The scoring system for chest radiographic abnormalities (RCAS) to CTAS is also strong (r^2=0.62, p<0.0001 Spearman Rank). The strongest significant correlation in this cohort was observed between SCAS and

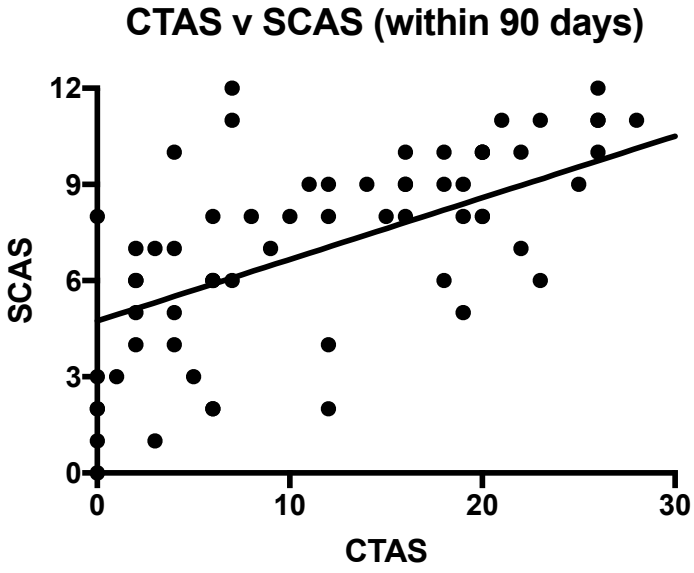
CTAS (r^2=0.64, p<0.0001 Spearman Rank; Linear regression p < 0.0001) (Figure 4.5).



Linear regression p = 0.0009
Spearman Rank correlation p <0.0001; r^2 = 0.64

Figure 4.5 CTAS versus SCAS when parameters are measured within 35 days of each other.

This was preserved, and in fact stronger, in the larger 'within 90 days' cohort (Figure 4.6). This suggests that activity quantified by granuloma burden and disease extent on thoracic HRCT can be reflected in a combination of parameters easily measured in the clinical setting, without the need for repeated high dose radiation exposure through CT scanning. The insidious development and slow progressive/resolving nature of the disease is reflected by there being little change in measured parameters over 90 days.



Linear regression $p < 0.0001$
Spearman Rank correlation $p < 0.0001$; $r^2 = 0.71$

Figure 4.6 CTAS versus SCAS when parameters are measured within 90 days of each other.

Whilst it can be observed that the strong correlation between CTAS and SCAS relates in the main to the radiological scoring of disease extent (RCAS), the SCAS as a whole can be used as a surrogate for activity measured by CTAS, and further work in this chapter illustrates it's likely correlation to activity in it's own right.

4.4.3 Correlation of SCAS to cell phenotype, cytokine production, and serum proteins

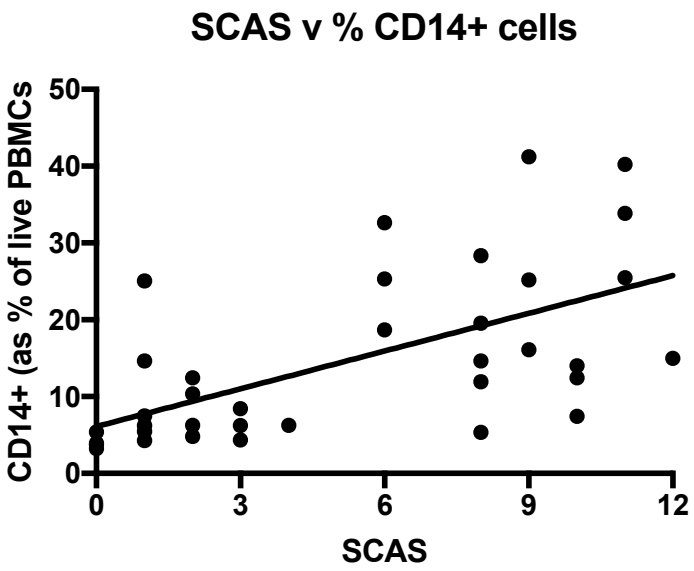
Table 4.5 shows the significant correlations observed between SCAS and cell phenotype, cytokine production and serum proteins.

SCAS v …..	**% CD14+ cells**	**%CD4+TNFa+ cells**	**Serum sIL2Ra**	**Serum CCL18**
Spearman Rank Correlation r^2 (p value)	0.68 (p<0.001)	0.46 (p = 0.003)	0.38 (p = 0.034)	0.46 (p = 0.002)
Linear Regression	p < 0.001	p = 0.019	p = 0.017	P = 0.002

Table 4.5 Correlation of SCAS to cell phenotype, cytokine production and serum proteins.

The significant positive correlation observed between the percentage of peripheral CD14+ monocytes and SCAS may indicate a potentially important role for monocytes in the persistence and progression of disease activity, but also an increase in circulating immune cells corresponding with the increase in disease activity (Figure 4.7).

A



Linear regression $p < 0.0001$
Spearman Rank correlation $p < 0.0001$; $r^2 = 0.67$

B C

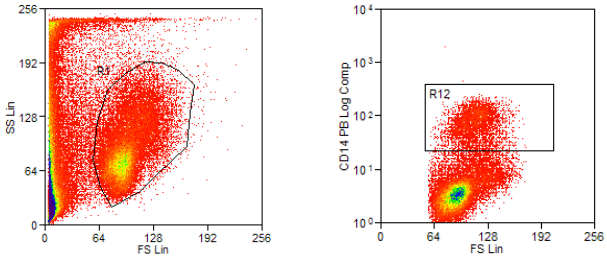
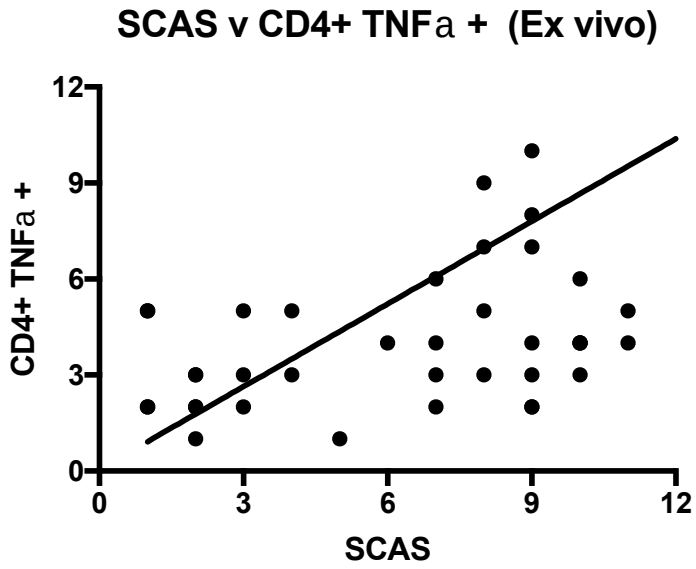


Figure 4.7 A - Graph depicts strong correlation of CD14+ monocytes as a percentage of live PBMCs versus SCAS. FACS plots above show gating strategy for CD14+ cells; B:- First all PBMCs are gated; C:- CD14+ cells are gated and percentage of all calculated.

Figure 4.8 shows the significant positive correlation between the percentage of TNFα expressing unstimulated CD4+ cells and SCAS. The presence of a typical increase in inflammatory effectors/mediators as the activity in pulmonary sarcoidosis

increases helps validate this score, and again reflects that this can be measured in the periphery.

A



Linear regression p = 0.019
Spearman Rank correlation p = 0.003; $r^2 = 0.46$

B C D E

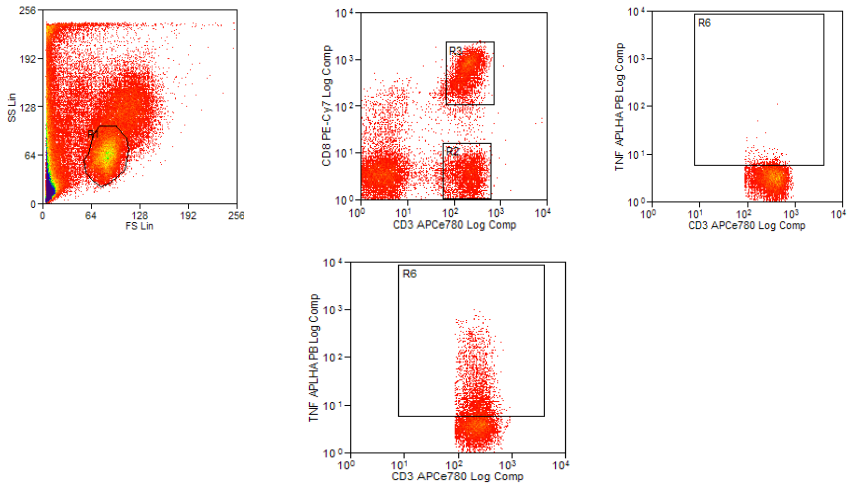
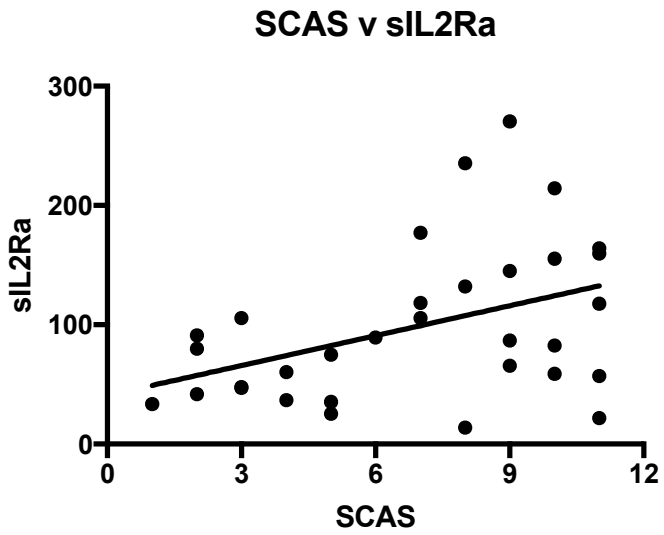


Figure 4.8. A - Significant positive correlation between percentage of TNFα expressing unstimulated CD4+ cells and SCAS. FACS plots above show gating strategy for CD4+ cells – B:- Lymphocyte population gated, C:- CD8 +ve and –ve populations gated, the CD8-ve population is then used when assessing TNF alpha production without (C) and with (D) PMA/Ionomycin stimulation.

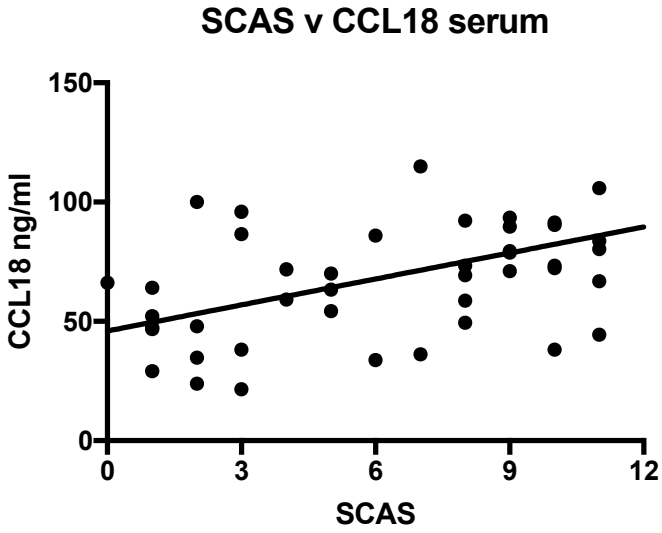
A moderate correlation between SCAS and serum sIL2R (measured by ELISA), a previously reported marker of disease activity, is also observed (Figure 4.9). This obviously also reflects the fact that the correlation between SCAS and CTAS is strong, as CTAS also correlates well with sIL2Ra levels (See Chapter 3).



Linear regression p = 0.0174
Spearman Rank correlation p = 0.03; $r^2 = 0.38$

Figure 4.9 – Strong positive correlation of SCAS with serum sIL2-R.

Finally, Figure 4.10 shows the moderate correlation of SCAS to serum CCL18 levels (measured by ELISA). This may represent active granuloma producing CCL-18 in the setting of active sarcoidosis, as it has been shown in Gaucher's disease.



Linear regression p = 0.002
Spearman Rank correlation p = 0.002; $r^2 = 0.46$

Figure 4.10 Correlation between SCAS and Serum CCL18 as measured by ELISA

4.4.4 Effects of treatment on SCAS

Longitudinal data shows that with treatment, SCAS was seen to significantly reduce ($p = 0.004$ ANOVA) when SCAS was repeated (between 81 and 279 days later) following initiation of immunosuppression (high dose Prednisolone 30mg daily [except for one patient that began with 3 days IV methylprednisolone]) (Figure 4.11).

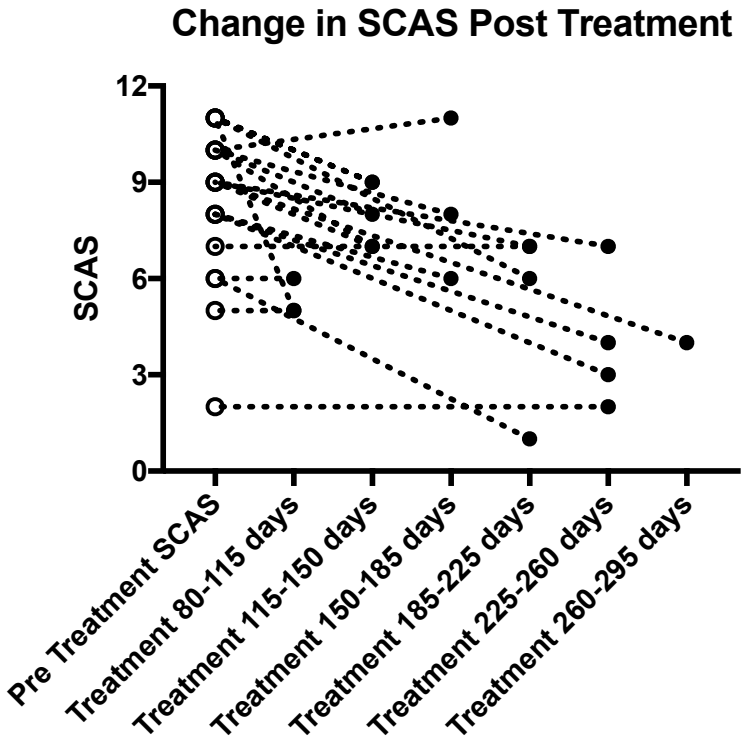
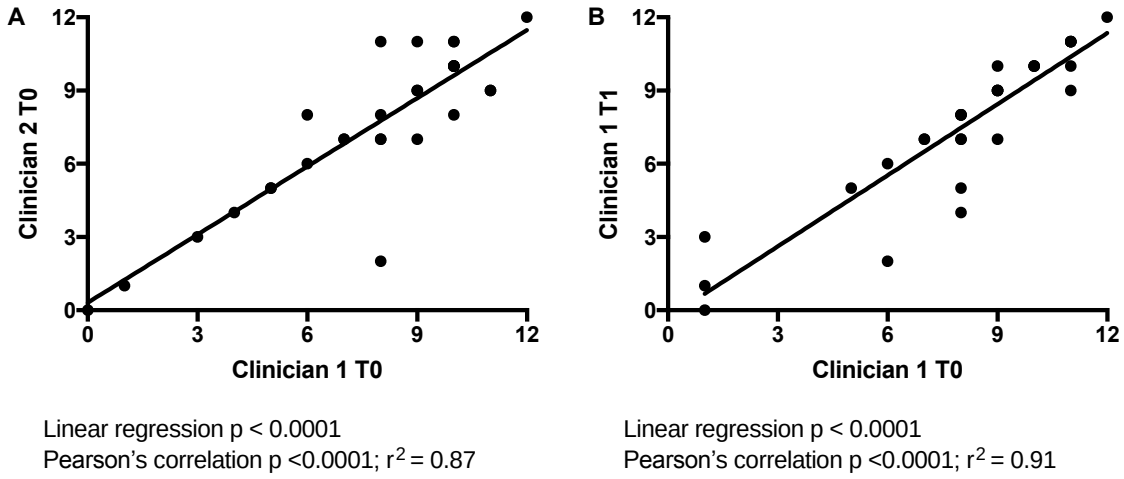


Figure 4.11 Change in SCAS following treatment between two clinic visits (81 to 279 days apart)($p = 0.004$ ANOVA)

4.4.5 Inter- and Intra-observer correlation for SCAS

To calculate inter-observer correlation for SCAS, two clinicians involved in the Oxford sarcoidosis service scored thirty patients for SCAS at the same time point, with written guidance on how to calculate the score. A significant correlation was observed between the two reporting clinicians ($r^2=0.81$; $p<0.0001$ Pearson's), confirming the reproducibility of this as a clinic based tool (Figure 4.12 A). Intra-

observer correlation was calculated by asking one clinician to score the SCAS for 30 patients, and then to repeat the calculation 1 year later (same time point for the patients enrolled). This also showed a strong correlation and confirmed reproducibility (r^2=0.93, p<0.0001 Pearson's) (Figure 4.12 B).



Linear regression p < 0.0001
Pearson's correlation p <0.0001; r^2 = 0.87

Linear regression p < 0.0001
Pearson's correlation p <0.0001; r^2 = 0.91

Figure 4.12 A – Inter-observer agreement between two clinicians scoring SCAS for n=30 patients; **B** – Intra-observer agreement from the same clinician scoring the same patients at presentation 1 year apart.

4.5 Discussion

In this chapter I have presented a clinic based, easy to use scoring system that provides an objective measure of disease activity in pulmonary sarcoidosis. The Sarcoidosis Clinical Activity Score (SCAS), is a detailed and discriminant scoring system, incorporating abnormalities seen on CXR with cellular markers of immune activity, which grades the level of disease activity from zero to twelve. It has content validity and shows good variation when monitoring response to treatment. Two independent Respiratory Physicians subsequently verified the clinical applicability of the scoring system (SCAS) and demonstrated both ease of use and reproducibility.

Sarcoidosis as a disease is extremely heterogeneous, and in this study I aimed to limit the application of the activity score to patients with prominent pulmonary sarcoidosis (even if other organs were involved). However, this scoring system may be adapted to look at disease in other organs, and may also lead to a more extensive scoring system incorporating evidence of granuloma burden and disease activity in other affected organs. Other complicated immunological based diseases such as Rheumatoid Arthritis (201), Wegener's Granulomatosis (202), and Systemic Lupus Erythematosus (SLE) (203) have faced the same difficulties in developing accurate scoring systems for disease activity, due to both the heterogeneity of disease, and the lack of a 'gold standard' to measure any scoring system against. Different methods for validation have been employed, and the more successful scoring systems that have made it in to mainstream use display ease of use and reproducibility. I have highlighted the following journeys for scoring systems, as they have influenced the development of the SCAS.

The presentation and disease course of Rheumatoid Arthritis (RA), as in sarcoidosis, is highly variable over time, between and within individuals. The symptoms and signs of RA may vary from joint complaints including pain, stiffness, swelling and functional impairment, to more constitutional complaints like fatigue and loss of general health. All of these can also be reported in sarcoidosis, as well as all the other organ specific symptoms, that may occur, e.g. neurological involvement causing neuropathy, palsy, visual impairment; Cardiac involvement leading to conduction defects, arrhythmias and systolic impairment; Liver disease leading to pruritis; Renal disease resulting in renal impairment and uraemia, to name just a few organs. Because of the variety in disease expression in RA, a large

number of variables have been used in past decades to evaluate status and course of RA disease activity and its consequences, but these systems have often been extremely cumbersome and difficult to use in the clinical setting, and during the design of our scoring systems the aim was to avoid complexity. In 1992, Van der Heijde developed the Disease Activity Score (DAS) (204, 205), which combines information from swollen joints, tender joints, acute phase response parameters and patient self-reporting of general health, into one continuous measure of RA disease activity. This correlated well against changes in disease-modifying anti-rheumatic drug (DMARD) therapy by rheumatologists. The components in this score were investigated alongside other disease activity measures commonly used in RA trials in an effort to standardize the assessment of RA and render study results interchangeable. A core data set has been put forward by the American College of Rheumatology, that should be included in clinical trials for accurate cross trial comparison (206). If a similar core set for sarcoidosis could be established, treatment outcomes in sarcoidosis could be evaluated more accurately and progress treating this condition may take a leap forward.

Both ANCA-associated Vasculitides (AAV) and SLE have even more similarity to sarcoidosis in terms of problems related to developing an activity score, as the disease affects many different organs, and quantification of total extent of disease activity in it's various locations needs to be measured. The Birmingham Vasculitis Activity Score (BVAS), one of several scores available, was developed for AAV by Luqmani et al (207), and comprises a scoring sheet, weighted towards disease parameters with more clinical significance, and covers nine systems, including systemic, cutaneous, mucous membranes/eyes, ENT, chest, cardiovascular,

abdominal, renal and nervous system. The score was validated against Physician's Global Assessment (PGA) – a common validation tool employed in diseases such as these where a gold standard marker of disease activity is absent – and involves physicians scoring a patient, usually between 0 and 10, on the level of overall activity they are deemed to have using their expert clinical judgment. The score was also compared to some serological markers of immunological activity (including CRP), and previously reported scoring systems in the same condition. In SLE, the 'SLAM'(208) and the 'SLEDAI'(209) are two of many published scoring systems for this complex systemic disease. The former includes both objective and subjective measures, the latter is a global index that evaluates disease activity over the previous ten days, and includes 24 items collecting data on specific manifestations in nine organ systems (Neurological, Musculoskeletal, Renal, Mucocutaneous, General, Heart, Respiratory, Vascular and Haematology). Both scores have been modified several times since their first publication.

There has been very little published on combined scoring systems in sarcoidosis. One study by Staton et al (210) published in 1986, assessed a combined score from n=50 patients, comprising degree of dyspnea, chest radiographic stage, forced vital capacity, and clinical assessment of disease activity (range normal = 0, maximum = 16), but found no correlation to serum angiotensin converting enzyme (SACE) levels, and only a weak correlation to Gallium-67 lung uptake (measured as total lung to background ratio), and percentage of lymphocytes in bronchoalveolar lavage. This scoring system has not made it into mainstream use. Rosenbach et el in 2013 published a scoring system for activity in a cutaneous sarcoidosis phenotype (211). They developed the Cutaneous Sarcoidosis Activity and

Morphology Instrument (CSAMI) for evaluating cutaneous sarcoidosis outcomes. Their score was again validated using the Physicians Global Assessment (PGA as previously described in the BVAS score for AAV) in a dedicated cutaneous sarcoidosis clinic. Information on the location of the skin lesion, then properties relating to the degree of inflammation, induration or depression, surface change and area were scored. Inter-rater and intra-rater reliability was good to excellent, and its correlation to PGA was statistically significant. Baughman et al in 2008 developed the SASI (Sarcoidosis Activity and Severity Index), which scored facial lesions in cutaneous sarcoidosis (n=25) (212). Twenty-four of these patients also had lung involvement. They concluded that the SASI was a reproducible scoring system for facial cutaneous sarcoidosis, which in turn may also guide activity level of other organ involvement if present. However, limitations of both these latter two scores, as with the SCAS presented here, is that they require a particular phenotype of the disease to make it valid. Rohatgi et al utilized a computer assisted quantitative analysis of gallium scans to obtain a gallium score, which was validated against serum ACE and objective improvement in clinical status. Again, this has not been taken up by the research community as a way to standardize the description or level of disease activity globally and in all phenotypes.

The SCAS may provide a platform to further our understanding of the mechanisms that drive disease activity, and by serving as a standard for testing, alternative serological markers could also be assessed. It also provides an opportunity to investigate disease modulation by targeting specific pathways in the immunological process. For example, as will be discussed in more detail in chapter 5, in the microarray analysis of both peripheral blood mononuclear cells and bronchoalveolar

lavage cells, CCR2 was the leading edge gene from both sampling modalities, which correlated with increasing activity. CCR2 is the receptor for MCP-1, a chemokine that specifically mediates monocyte chemotaxis. We have also shown in our results the correlation between SCAS and percentage of circulating monocytes, highlighting the importance of monocyte activity in active sarcoidosis. The potential for therapeutic intervention, especially in view of the recent phase 2 trial of a CCR2 antagonist(213), is ever nearer, and response to treatment could be ascertained by the SCAS.

4.6 Conclusion

The SCAS allows standardization of the assessment of current clinical disease activity in pulmonary sarcoidosis. It is presented here as a useful tool, derived from tests immediately available in the outpatient clinic, for assessing disease activity in pulmonary sarcoidosis. I hope that with publication it can be critiqued and optimized further for mainstream use in the clinical setting. A web based calculator for SCAS is in development.

5 Investigation of genes upregulated in high disease activity in pulmonary sarcoidosis.

5.1 Introduction

Having established the Sarcoidosis Clinical Activity Score (SCAS), a surrogate for measuring the presence and level of disease activity on CT chest (CTAS), I have used SCAS to examine the genes and processes that contribute to disease activity in a prominent pulmonary sarcoidosis phenotype.

Many studies have examined the genes associated with aetiology (18, 30, 214, 215), susceptibility (43, 216-218), and disease phenotype (40, 48, 219) in sarcoidosis, but very few have focused on profiling genes directly related to disease activity. One major barrier to this has been the lack of disease activity biomarkers for the condition. The mechanisms that contribute to the induction, persistence and resolution of disease in sarcoidosis, and therefore disease activity, remains incompletely understood. Genome wide association studies (GWAS) have been used to assess human leukocyte antigens (HLA), known to be important in antigen presentation and immunoregulation. Associations between the class I gene HLA B8 and acute sarcoidosis have been repeatedly reported (39, 220, 221). Numerous studies on HLA class II genes, in particular HLA-DRB1 and DQB1 associations, have identified several alleles segregating with susceptibility (222), disease course (41) (223), and clinical phenotype (224-226). Polymorphisms of genes involved in immune cell activation such as BTNL2 (42), TNF (227), IL23R(45), and those involved in regulation of apoptosis e.g. ANXA11(42) and XAF1 (214), have also

been identified as key players in sarcoidosis susceptibility. Gene expression profiling of lung tissue was used by Crouser et al (168), and MMP12 and ADAMDEC1, both macrophage-derived proteases, were identified as likely mediators of lung damage and/or remodeling. It was also felt that they may serve as markers of disease activity due to their correlation with severity. Zhou et al (228) identified a universal twenty gene signature to differentiate sarcoidosis patients from healthy controls, and to distinguish complicated sarcoidosis (defined as pulmonary-FVC<50%, cardiac, or neuro- sarcoidosis) from uncomplicated sarcoidosis.

In other heterogeneous conditions such as Systemic Lupus Erythematosus (SLE), gene expression profiling has demonstrated granulopoiesis-related and interferon (IFN)-induced genes correlating with disease activity according to the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) (229).

As genetic associations in sarcoidosis have been shown to be partly phenotype specific, subclassifying disease phenotype prior to investigation is likely to provide clearer insight into genetically determined mechanisms that steer the disease pathophysiology towards persistent or high levels of activity. Furthermore, surveying cell populations isolated from different organs/systems, e.g. peripheral blood and lung, may enhance these discoveries and allow comparison of potential biomarkers in each sampling modality.

Using gene expression analysis, dominant processes and individual genes involved in disease activity may be identified. This in turn may identify novel targets in the treatment of the disease, and the development of gene signatures, which could be

used to monitor disease activity, response to therapy, and outcomes, as it has done for SLE (230).

5.2 Aim

To determine genes and immune processes involved in disease activity in pulmonary sarcoidosis, using genome-wide gene expression profiling of peripheral blood mononuclear cells and bronchoalveolar lavage cells, and its correlation to SCAS.

5.3 Methods

5.3.1 Patient selection

Sequential patients with a diagnosis of pulmonary sarcoidosis, as defined by WASOG criteria (120), and as described in Chapter 4, were recruited for the study over a defined 2 year period. All patients had lung involvement. Concomittant cutaneous or eye involvement was also permitted, any other systemic involvement either came later or did not require treatment. Patients without lung parenchymal involvement on high resolution CT scanning, those on immune suppressive therapy at the point of sampling, those with evidence of other concomitant active inflammatory disease, smokers, and those on ACE inhibitor medication were excluded.

5.3.2 Experimental design

Serum ACE, IgG and lymphocyte count, chest radiograph, peripheral blood mononuclear cell (PBMC) isolation and bronchoalveolar lavage were performed within 90 days of each other.

Blood for PBMC isolation was collected directly into tubes containing sodium heparin, and PBMC extraction using standard density gradient centrifugation was performed within 3 hours of collection (methods section 2.3.1). PBMCs were then stored in FCS (with 10% DMSO) in liquid nitrogen, until batched RNA extraction. Bronchoalveolar lavage (BAL) samples were collected into polypropylene tubes. BAL was passed through a cell strainer within 3 hours of collection, BAL fluid was removed, and cells were re-suspended in FCS (with 10% DMSO) and stored as per PBMCs in liquid nitrogen until RNA extraction (methods section 2.3.2).

RNA from PBMC and BAL cells were obtained using the RNeasy mini kit (Qiagen) according to manufacturer's instructions (described in methods sections 2.3.8). RNA quantity was assessed on a NanoDrop photospectrometer and quality assessed using a high-resolution electrophoresis bioanalyser system (Agilent Technologies). All RNA was stored at -80°C. RNA samples were normalized for concentration (300ng in 11µl of RNAse free water) in a 96 well plate. cRNA synthesis and gene chip hybridisation was performed by Dr. Karene Argoud at the Wellcome Trust Centre for Human Genetics, in a single batch, using an Illumina HumanHT-12 v4 Expression Bead Chip Kit.

The SCAS for each patient was calculated from the radiological and serological information as described in Chapter 4. Demographics for patients were also collected (Table 5.1).

5.3.3 Data analysis of gene expression profiles

Expression of each gene was first correlated to numerical values of SCAS using Spearman Rank Correlation, incorporating multiple gene comparisons. Each gene was then rank-ordered based on the correlation coefficient, and further analysed using Gene-Set Enrichment Analysis (GSEA) (www.broad.mit.edu/gsea/). GSEA tests the biological meaning of a ranked list of all genes in a sample (231). It is one of several Functional Class Scoring (FCS) methods, that alleviates the need to select significant genes in a dataset as a first step. The ranked list is ordered such that genes showing higher expression in one phenotype (in this case, highly correlated with high activity as measured by high SCAS) are at the top, and those showing higher expression in the other phenotype (highly correlated with low activity as measure by low SCAS) are at the bottom, and unchanged genes are in the middle. Pre-defined gene-sets are then tested for enrichment in one of the phenotypes, and the significance of the enrichment is assessed by a permutation method. In this study, GSEA was used in Pre-Ranked mode (231), where, as already mentioned, the list of genes was ranked according to the correlation coefficient. Gene-sets comprising all biological process categories (C5bp) defined by the Gene Ontology Consortium (232) and based on GO terms (www.geneontology.org) and their associations to human genes, were tested. Biological processes are a recognized series of events (interactions or biochemical

reactions) accomplished by one or more ordered assemblies of molecular functions with a defined beginning and end.

MetacoreTM (MetaCore from Thomson Reuters), an integrated software suite for functional analysis of microarray, was also employed to interrogate the gene expression data. MetaCore is based on a high-quality, manually-curated database of: 1)transcription factors, receptors, ligands, kinases, drugs, and endogenous metabolites as well as other molecular classes; 2) species-specific directional interactions between protein-protein, protein-DNA and protein-RNA interactions, drug targeting, and bioactive molecules and their effects; 3) signalling and metabolic pathways represented on maps and networks; and 4) ontologies for diseases and processes with hierarchical or graphic output. MetaCore creates a list of possible networks, ranked according to how many objects in the network correspond to the user's list of genes, how many nodes are in the network and how many nodes are in each smaller network. Network size varies widely (from less than 10 to over 2000). Correlation coefficients of >0.5 or <-0.5, with corresponding significant p values (<0.05) from the correlated gene lists were used to input into metacore, i.e., only genes identified as significant based on correlation were used for analysis, in contrast to GSEA.

Dr Eleni Giannoulatou and Dr Steven Taylor at the Computational Biology Research Group, Weatherall Institute of Molecular Medicine, and Dr. Helen Lockstone, Bioinformatics and Statistical Genetics Core, Oxford Genomics Centre, Wellcome Trust Centre for Human Genetics, provided support with scripts for R program for creating rank lists for use in GSEA, and training in using the software.

5.3.4 Data Interpretation

In Gene Set Enrichment Analysis, the enrichment score (ES) reflects the degree to which a gene set is overrepresented at the top or bottom of a ranked list of genes. A positive ES indicates gene set enrichment at the top of the ranked list; a negative ES indicates gene set enrichment at the bottom of the ranked list. The normalised enrichment score (NES) is the primary statistic for examining gene set enrichment results. It normalises the enrichment score for gene set size and correlations between gene sets and the expression dataset. The NES can be used to compare analysis results across gene sets. The nominal p value estimates the statistical significance of the enrichment score for a single gene set. The false discovery rate (FDR) is the estimated probability that a gene set with a given NES represents a false positive finding. For example, an FDR of 0.25 indicates that the result is likely to be valid 3 out of 4 times. The “leading-edge subset” in a gene set are those genes that account for the gene set enrichment signal. Leading edge analysis is a further function of the GSEA software, which allows comparison of the leading edge genes from different gene sets, and visualisation of the overlap between the selected leading edge subsets (i.e. how many of the analysed leading edge subsets does a single gene appear in).

In Metacore, pathway maps (equivalent to canonical pathways) are derived from an inputted list of genes. Pathway analysis is key in determining the relevance of genes or proteins that alone may not be significantly altered, but if they exist in a signalling pathway with a set of other altered genes or proteins, then collectively they have a more profound biological outcome. Maps illustrate the multiple sequential steps of

interactions, defining a signalling mechanism. Each step in the interaction is accepted by the field as being experimentally validated and well defined. Each gene in a pathway may also be involved in a number of other pathways. This can be displayed as a process network, which also describes biological function, but in greater detail than in individual pathway analysis. A pathway map could be thought of as a small section of a process network.

5.4 Results

5.4.1 Patient demographics

The demographic details of patients in each sampling modality are shown in Table 5.1. Eight patients had both PBMC and BAL samples, 21 patients had PBMC samples alone and four had BAL only (Figure 5.1).

Sampling modality	PBMC	BAL Cell
n	29	12
Age range (years)	29 - 69	30 – 69
Age mean [S.D.]	48 [9.67]	50 [13]
SCAS mean [S.D.]	6.9 [3.13]	7.9 [2.75]
FEV1 (% Predicted) Range	65-124	68 – 116
FEV1 (% Predicted) Mean [S.D.]	93 [15.70]	90 [11.62]
FVC (% Predicted) Range	72 - 133	83 - 122
FVC (% Predicted) Mean [S.D.]	103 [16.33]	98 [13.54]
TLCO (% Predicted) Range	62 - 139	63 - 97
TLCO Mean [S.D.]	87 [1.50]	81 [10.40]
Gender M	20	7
Gender F	9	5
Ethnicity - Afro-Caribbean	3	2
Ethnicity - Asian	2	1
Ethnicity - Caucasian	24	9
Fibrosis on CT	4	2
Proceeded to treatment within 1 yr	8	6
Other organ involvement	8	6
Scadding 1/2/3/4	1/5/4/2	7/14/4/4

Table 5.1 Demographics of patients enrolled in microarray SCAS cohort. PBMC - Peripheral Blood Mononuclear Cell samples, BAL Cell - bronchoalveolar lavage cell samples. Other organ involvement includes Cardiac/Neuro/Bone/Spleen/Liver.

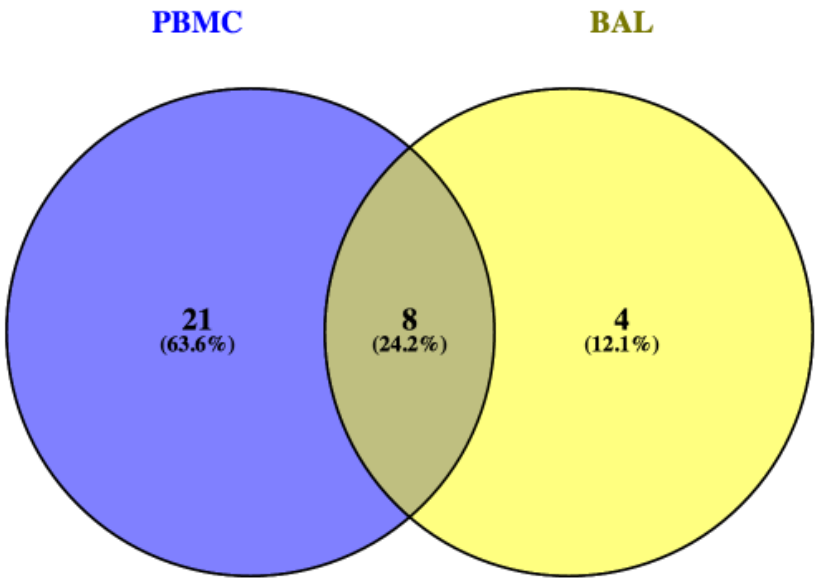


Figure 5.1 Venn diagram showing overlap in PBMC and BAL sampling groups.

The range of SCAS in each sampling modality is depicted in Figure 5.2.

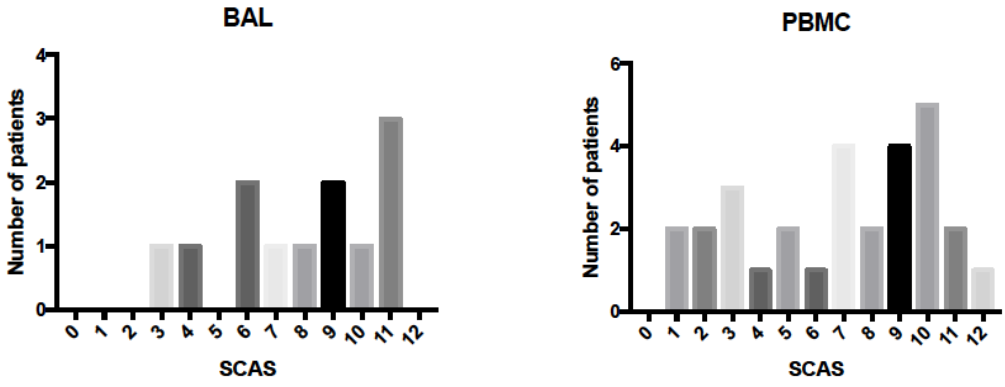


Figure 5.2 The range of SCAS values in the BAL and PBMC cohort respectively.

5.4.2 Quality control of microarray samples

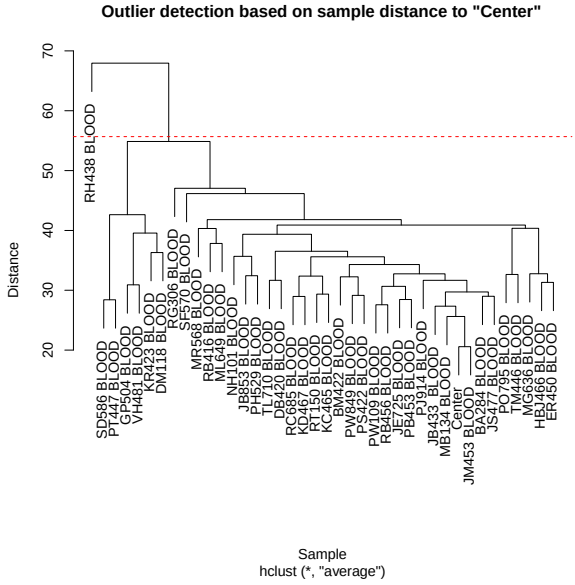
RNA quantity and quality (260/280 readings) was initially screened using the nanodrop 2000 spectrophotometer (Thermo Scientific). RNA quality was then further assessed through the "RNA integrity number" (RIN) of the RNA samples. RIN is an algorithm that evaluates the electrophoretic trace of RNA and determines the RNA integrity. A score of 1 to 10 is assigned to the sample, where 10 represents completely intact RNA. A stringent cut-off of RIN >7 was employed to ensure all samples put forward for whole genome gene expression assays were of high quality (Table 5.2).

PBMC Sample ID	RIN	BAL Sample ID	RIN
RB416 Blood	9.7	ML649 Bal	8.2
HBJ466 Blood	9.5	HBJ466 Bal	9.2
SD586 Blood	9.3	SH895 Bal	8.7
ML649 Blood	9.5	KR423 Bal	9.4
MR568 Blood	9.2	JB853 Bal	9.3
BA284 Blood	9.6	PW109 Bal	8.8
PT447 Blood	9.1	CS468 Bal	8.7
GP504 Blood	9.3	DM118 Bal	9.0
PW109 Blood	9.1	MG636 Bal	9.3
PO795 Blood	9.4	LH335 Bal	9.2
MG636 Blood	9.3	AS163 Bal	9.1
JB853 Blood	9.6	KC465 Bal	8.7
JE725 Blood	8.7		
BM422 Blood	9.6		
MB134 Blood	9.5		
VH481 Blood	9.4		
DB420 Blood	9.4		
KR423 Blood	9.3		
PB453 Blood	9.5		
DM118 Blood	9.3		
PW849 Blood	9.9		
PH529 Blood	9.3		
RT150 Blood	9.3		
KC465 Blood	9.6		
JM453 Blood	9.3		
RG306 Blood	9.3		
RC685 Blood	9.2		
KD467 Blood	9.1		
JB433 Blood	9.0		

Table 5.2 Table showing samples undergoing whole-genome gene expression direct hybridization assay (Illumina) and associated RIN. (Other samples also were also run with these (all with RIN >7), but were excluded after the normalization process as they were either outliers or were found not to meet inclusion/exclusion criteria.)

Standard quality control steps, which consist of correlation between mean gene intensity in each group, assessment for systematic difference in the distribution of fluorescence intensity on the gene chips used, and batch effect testing, were also performed. After these initial quality control steps, the hybridization process was completed, the resultant data was assessed and outliers were removed based on 'sample distance to centre' (Figure 5.3).

A



B

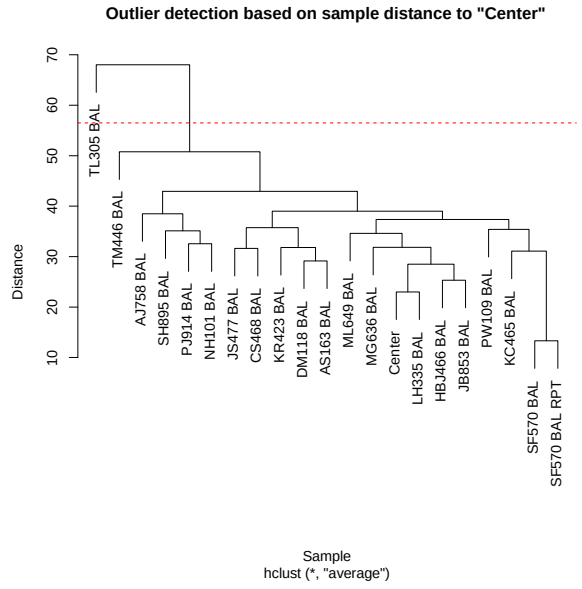


Figure 5.3 Dendogram showing clustering of samples in each sampling modality (A = PBMC, B= BAL cell), and identification of outliers prior to RMA normalization (RH438 in Blood and TL305 in BAL).

After excluding outliers based on sample distance to centre (RH438 in Blood and TL305 in BAL as shown in Figure 5.3), data was then normalised using Robust Multi-array Average Normalisation (RMA). Finally, a mean intensity was computed for each probeset.

5.4.3 Correlation of gene expression profile of BAL cells to disease activity

5.4.3.1 GSEA analysis

Analysis of whole gene expression from cells isolated from bronchoalveolar lavage identified 19561 genes, and level of expression was correlated to activity (SCAS). As described in the methods section of this chapter (5.3.3), the genes were then ranked according to their correlation coefficient, and this pre-ranked list was analyzed using GSEA. Gene-sets representing all biological processes (C5bp - as described in section 5.3.3) were tested. A portion of the 19561 are either uncharacterized genes without GO associated terms, or they have missing gene information, therefore, 15534 of the 19561 expressed genes were analysed for these gene sets.

In total 221 of 464 gene sets were enriched in the high activity group, of which 58 gene sets were significant with an FDR <25%. Table 5.3 shows the top ten positively enriched gene sets (the description for the definition of nominal p value, ES, NES and FDR is given in section 5.3.4), several of which are noted to be immune related.

No	Gene Set	ES	NES	NOM p-val	FDR q-val
1	Immune Response	0.39	2.16	0.000	0.021
2	Chemical Homeostasis	0.43	2.12	0.000	0.016
3	Locomotory Behaviour	0.45	2.10	0.000	0.015
4	Defence Response	0.36	2.05	0.000	0.019
5	Homeostatic Process	0.37	1.99	0.000	0.026
6	Response to wounding	0.38	1.97	0.000	0.027
7	Immune system process	0.34	1.95	0.000	0.029
8	Ion Homeostasis	0.41	1.94	0.000	0.027
9	Cellular Homeostasis	0.40	1.93	0.000	0.029
10	Behaviour	0.39	1.92	0.000	0.028

Table 5.3 Top ten positively enriched gene sets identified using GSEA analysis of whole gene expression from BAL cells correlated to increasing SCAS.

Within the 'leading edge subset' (i.e. genes that contribute most to the enrichment score) of the top gene set 'Immune Response', were several genes of interest in pulmonary sarcoidosis, for example C-C Chemokine Receptor 2 (CCR2), Triggering Receptor Expressed on Myeloid Cells -2 (TREM2), IL6, CCL18, CCL2, IL10RB, CCR6, IL15, and CCL20. The enrichment plot for this gene set is shown in **Figure 5.4,** whilst **Table 5.4** shows the list of leading edge genes and the position of the gene in the initial ranked list (the smaller values being more strongly correlated on an individual basis to increasing activity).

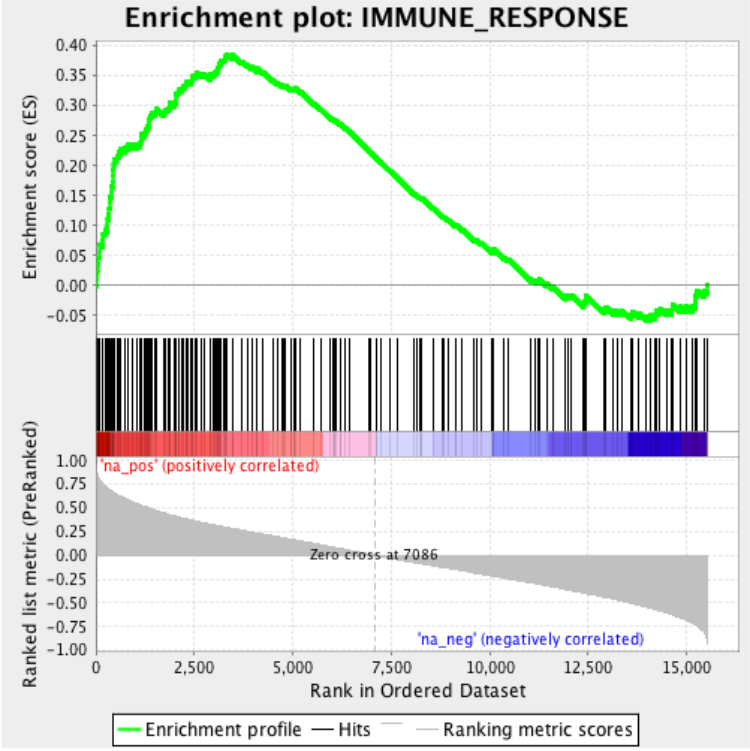


Figure 5.4 Enrichment plot of top gene set revealed by GSEA analysis of BAL gene expression correlated to increasing SCAS. The enrichment plot illustrates increasing signal with a left shift indicating enrichment as the activity increases. The red-blue strip represents the 15534 genes expressed in the samples, ranked according to correlation of expression to increasing SCAS. Each line in the 'bar code' represents presence of a gene from the GO-defined gene-set (e.g. immune response gene set) within the ranked genes. The enrichment score (ES) is a cumulative statistic, which increases when the gene in that particular gene-set is found in the ranked gene list and decreases otherwise. The changes are weighted such that the magnitude is greater if there is a strong correlation of the gene from the defined gene-set with phenotype (i.e. found towards the extreme ends of the ranked list). The leading-edge subset is defined as genes appearing at or before the cumulative sum reaches maximum deviation from zero.

Gene	Rank	Gene	Rank	Gene	Rank	Gene	Rank
GPR65	21	CCL2	451	ST6GAL1	1536	FCGR1A	2536
FYB	33	IL10RB	471	GBP2	1724	SEMA4D	2552
AIM2	36	CCR6	529	CCL4	1754	C2	2676
HAMP	66	IL15	584	BNIP3	1827	CD40LG	2756
CMKLR1	89	IL18BP	635	HLA-DRB3	1865	CTLA4	2907
TREM2	92	TCF12	737	SLA2	1882	CTSS	2908
IL6	164	APOBEC3F	809	TAPBP	1980	IL7R	2970
PTPRC	182	APOBEC3G	914	BCL2	2025	AQP9	3007
MR1	230	CXCR4	1048	CADM1	2028	KIR2DL3	3055
LILRB2	292	CD96	1128	CCR2	2029	C5AR1	3090
PTGER4	298	PSMB10	1142	CD74	2087	CST7	3095
RGS1	312	FCGR3B	1152	TRAT1	2116	SEMA3C	3119
CD86	336	CD22	1247	NFIL3	2177	IL16	3140
CD83	354	CXCL12	1272	IL6R	2201	CCL23	3172
FCN1	358	CD79B	1293	CCL20	2277	IKBKG	3186
LCP2	385	IGSF6	1352	FYN	2281	CX3CL1	3259
TNFRSF14	390	VTN	1360	GZMA	2346	PTAFR	3288
XBP1	423	DPP4	1368	CD7	2425	ZAP70	3299
MS4A2	429	TARBP2	1428	LY75	2436	CCL5	3310
CCL18	433	CTSC	1513	IL32	2464		

Table 5.4 All leading edge genes contributing to the enrichment of the top gene set, immune response, are shown with their ranking in the total gene list (of 15534 genes).

All functions of these genes were identified using GeneCard (www.genecards.org), which is a searchable, integrated, database of human genes that provides concise genomic related information, on all known and predicted human genes.

I then applied the 'Leading edge analysis' function of the GSEA software, which allows comparison of the leading edge genes from different gene sets and visualisation of the overlap between the selected leading edge subsets. Leading edge analysis of the top ten gene sets (as shown in Table 5.3) has revealed CCR2 as the most frequently occurring leading edge gene, present in all of the top ten enriched gene sets. CCR2 is ranked at 2029 in the whole gene list (of 15534 genes) in this series, with a correlation coefficient of 0.41. Other leading edge genes and their frequency in the top ten gene sets are shown in Figure 5.5.

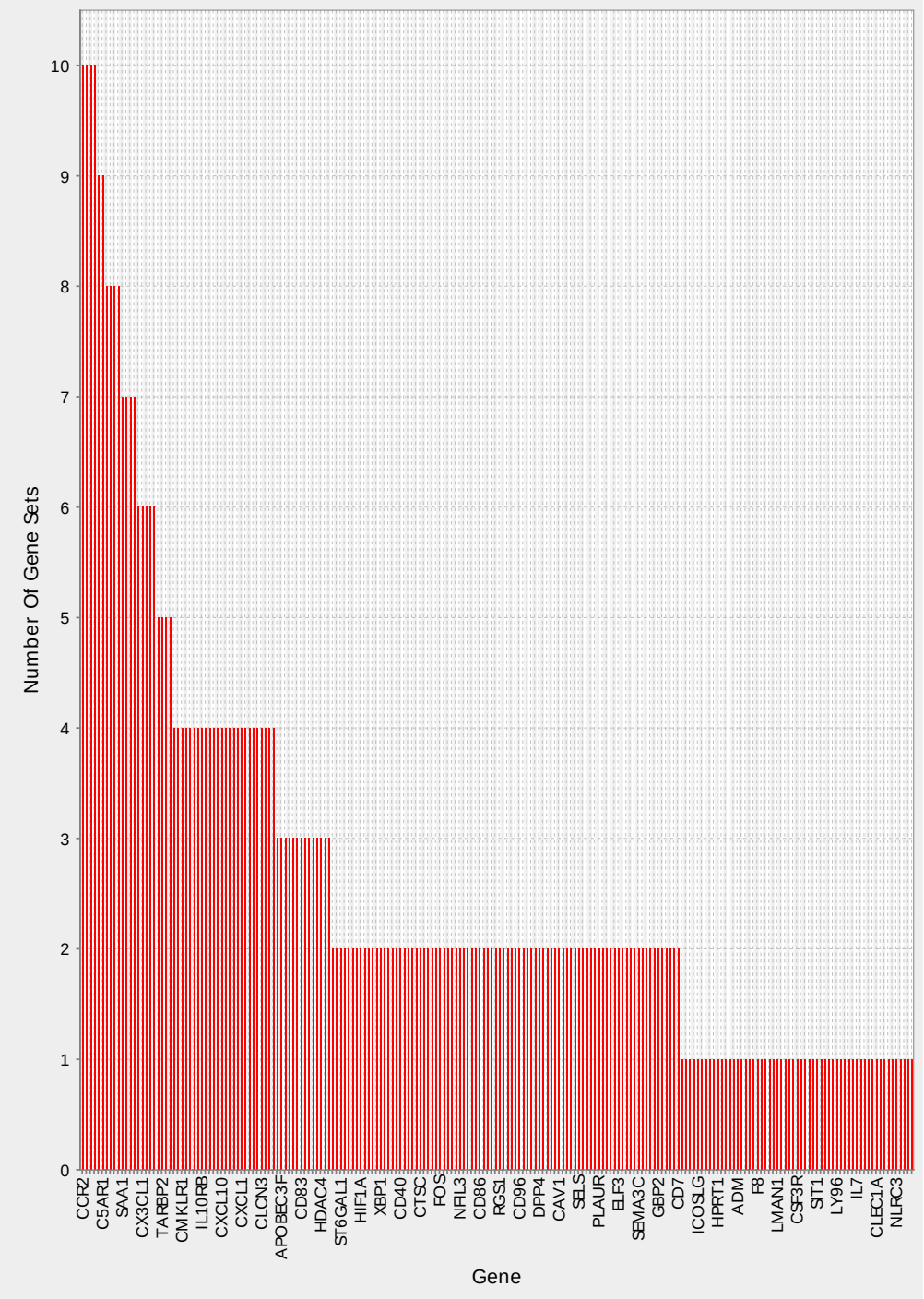


Figure 5.5 Leading edge analysis of top ten enriched gene sets in GSEA of ranked BAL gene expression correlated to SCAS. Every 6th gene, ranked by the number of gene sets they are found in, is shown on the x-axis. The individual genes which were represented in 4 or more of the 10 gene sets analysed are shown alongside the histogram.

A gene that is in many of the leading-edge subsets is more likely to be of interest than a gene that is in only a few of the leading-edge subsets, as it is contributing more to the enrichment signals seen in GSEA.

For the genes that correlated negatively with increasing SCAS, 243 of 464 gene sets were found to be negatively enriched. Twenty-four gene sets were significant at FDR < 0.25. The top ten negatively enriched gene sets are displayed in Table 5.5.

No	Gene Set	ES	NES	NOM p-val	FDR q-val
1	TGF beta receptor signalling pathway	-0.63	-2.12	0.000	0.010
2	Transmembrane receptor protein serine threonine kinase signalling pathway	-0.58	-2.09	0.000	0.008
3	DNA replication	-0.41	-1.83	0.000	0.130
4	DNA metabolic process	-0.35	-1.79	0.000	0.134
5	DNA repair	-0.37	-1.74	0.004	0.181
6	Positive regulation of transcription DNA dependent	-0.38	-1.73	0.003	0.174
7	Glucose metabolic process	-0.52	-1.72	0.010	0.162
8	Steroid hormone receptor signalling pathway	-0.54	-1.71	0.012	0.153
9	Positive regulation of transcription	-0.37	-1.71	0.000	0.137
10	Response to DNA damage stimulus	-0.35	-1.71	0.000	0.124

Table 5.5 Top ten negatively enriched gene sets identified using GSEA analysis of whole gene expression from BAL cells correlated to increasing SCAS.

The enrichment plot for the top negatively enriched gene set, TGFβ receptor signaling pathway, is shown in Figure 5.6, with the leading edge genes for this gene set shown in Table 5.6 (i.e. genes that increase as the activity levels decrease).

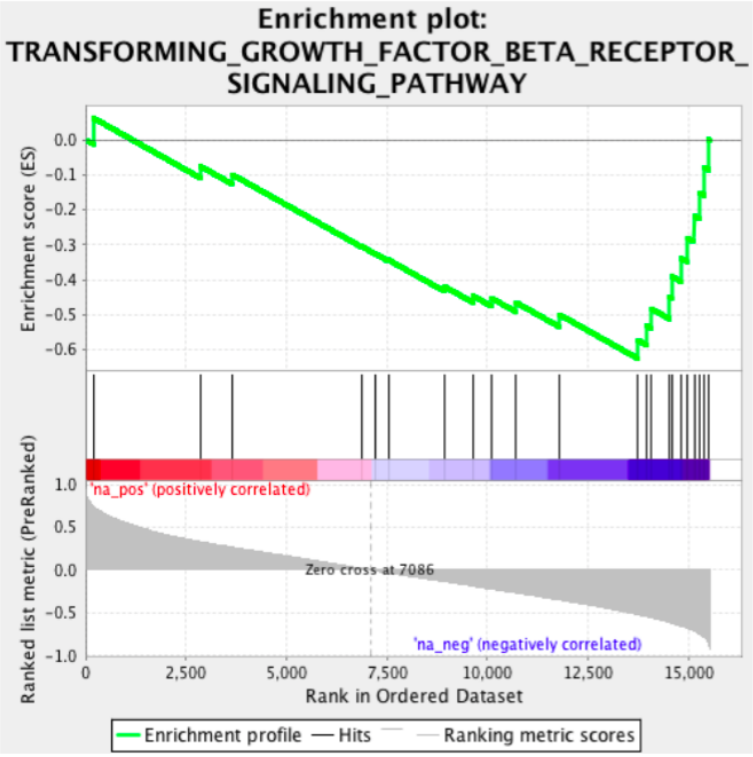


Figure 5.6 Enrichment plot of top negatively enriched gene set revealed by GSEA analysis of BAL gene expression correlated to increasing SCAS. The enrichment plot illustrates enrichment with a right shift as the activity decreases (i.e. negative correlation to increasing SCAS). The red-blue strip represents the 15534 genes expressed in the samples, ranked according to correlation of expression to increasing SCAS as described before. The leading-edge subset is defined as genes appearing at or before the cumulative sum reaches maximum deviation from zero.

Gene	Rank
SMAD4	13750
ENG	13962
STUB1	14094
SMAD7	14539
ACVRL1	14594
HPGD	14819
CDKN2B	14993
BMPR1A	15147
MAP3K7	15272
FNTA	15404
SMAD2	15509

Table 5.6 All leading edge genes contributing to the enrichment of the most negatively enriched gene set, TGFβ Receptor signaling pathway, are shown with their ranking in the total gene list (of 15534 genes), in this case the higher the number, the stronger the negative correlation.

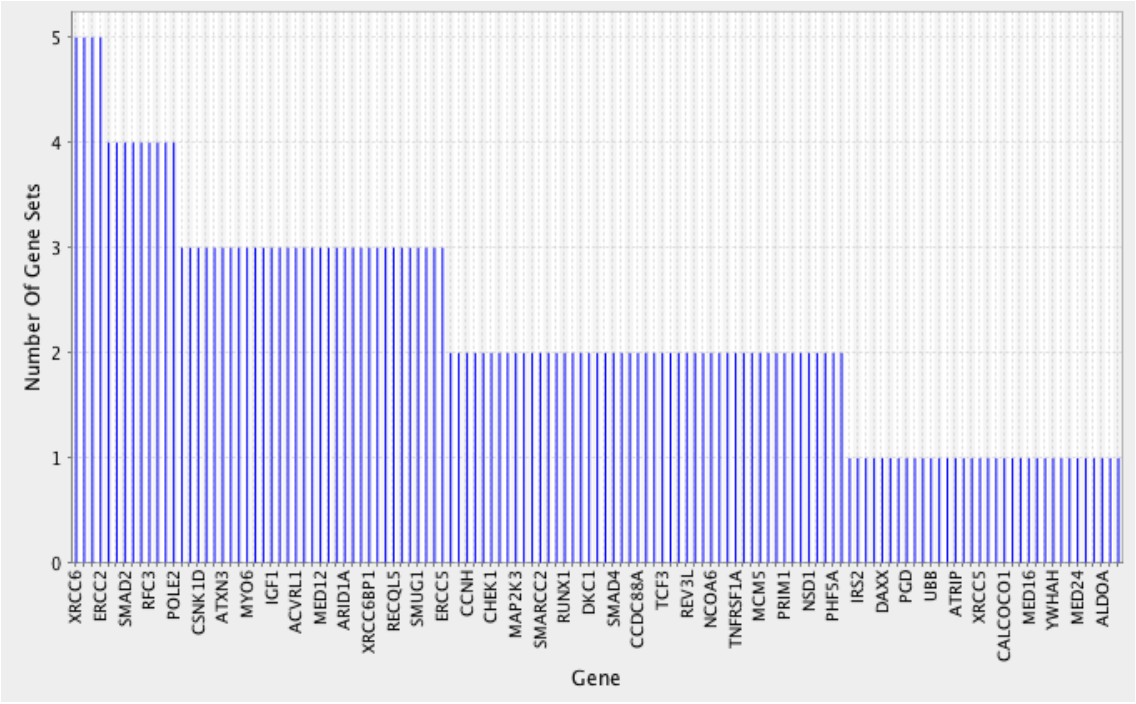


Figure 5.7 Leading edge analysis of top ten negatively enriched gene sets in GSEA of ranked BAL gene expression correlated to SCAS. Every 6th gene, ranked by the number of gene sets they are found in, are shown on the x-axis. The individual genes which were represented in 4 or more of the 10 gene sets analysed are shown alongside the histogram.

XRCC6, the top leading edge gene present in five of the top ten negatively enriched gene sets, is a negative regulator of transcription, and is involved in DNA repair. It's downregulation in high disease activity, suggests that transcription is therefore increased, which would be the case in inflammatory and granulomatous inflammation and T cell proliferation. Figure 5.7 shows the frequency of individual genes contributing to the top ten negative enrichment scores in high disease activity.

5.4.3.2 Metacore Analysis

While GSEA looks at the entire set of genes in rank order, metacore analysis allows us to narrow down the functional assessment of the genes found at the most significant end of the correlated list. A correlation co-efficient cut-off of r >0.5

(Spearman Rank) for correlation with increasing activity, or r <-0.5 for correlation with decreasing activity (p-value of <0.05 for both), when SCAS was correlated to BAL cell gene expression, was employed to create separate positive and negative significant gene lists for BAL microarray analysis for use in metacore. For ease of description, the first group is labeled 'high activity genes group' and the second group, the 'low activity genes group'. For each analysis, (i) enrichment by GO processes, (ii) Pathway analysis and (iii) Process networks, as described in section 5.3.3 were analysed. It is expected that (i) will show similar output to the GSEA as this was also used to interrogate for GO biological processes.

In the 'high activity genes' group, 947 genes were identified. Enrichment in GO processes are seen to be related exclusively to the immune system. Table 5.7 shows the top 10 GO processes derived from this gene list. 'Immune system process', the top GO process, includes ADAMDEC1, Chitotriosidase, CCL18, TREM2, CCL2 and IL10RB. The full list is provided in Table 5.8.

No	Enrichment by GO Processes	p-value	FDR	Genes in dataset
1	Immune system process	1.807E-31	7.492E-28	238
2	Immune response	1.992E-31	7.492E-28	181
3	Regulation of immune system process	2.392E-30	5.996E-27	186
4	Response to stress	5.472E-28	1.029E-24	315
5	Positive regulation of immune system process	6.760E-27	1.017E-23	132
6	Defence response	1.286E-26	1.612E-23	187
7	Single-organism metabolic process	1.277E-23	1.372E-20	354
8	Regulation of response to stress	1.397E-22	1.313E-19	158
9	Inflammatory response	2.133E-22	1.783E-19	86
10	Response to external stimulus	3.471E-22	2.611E-19	221

Table 5.7 Top 10 GO processes derived from gene list (n=947) that shows significant positive correlation to SCAS (in BAL) when analysed through MetacoreTM.

Gene	Gene	Gene	Gene	Gene	Gene
NUB1	CD83	ARPC1	IL-23 receptor	C1qRp	Slp76
BAFF(TNFSF13B)	Pleckstrin	Gas6	ILT3	SPP lyase	SHMT2
Dematin	Rictor	FMIP	HLA-E	Fc epsilon RI beta	SHMT
IL21R	IL10RB	XBP1	SBDS	PLCE	PTPN2
STAT1	CCL2	VCAM1	TLR6	GCH1	Galpha(i)-specific EDG GPCRs
MHC class II alpha chain	PSMC1	Alpha-synuclein	ILT4	NKp30	Galpha(q)-specific EDG GPCRs
MHC class II	NF-kB	HMG1,2	Tropomodulin-2	AIM2	CCL14
RING6	NF-kB2 (p100)	HMG2	CCL13	TAP1 (PSF1)	IRp60
Rab-27A	GBP1	Semaphorin 4A	APOBEC3	MIG	PRK1
CD84	GBP4	DC-SIGNR	APOBEC3B	CCL18	SOD2
Cyclin D	APG7	PP2A regulatory	HLA-H	AHSP	GBP3
CD36	CLEC10A	CCL3L1	IL-12RB1	A2M	DNAJC3
Presenilin	NALP3	C1s	PLA2	TIM-3	CD45
Presenilin 1	MIST	PAPD4	Ferroportin 1	C/EBP	K(+) channel
HLADOB	Cathepsin K	E2A/HLF fusion protein	SCAP2	C/EBPdelta	subfamily J
MHC class II beta chain	HLA-DRB1	E2A	P2Y14	TREM2	hIAN5
TCF12	HLA-DRB	Adenylate cyclase type III	Galpha(i)-specific class A orphan/other GPCRs	CD1E	Myosin X
CD86	Adenylate cyclase VIII	Lrmp	AP1S2	Sphingomyelinase	ARID4A
Rab-32	Adenylate cyclase	Chitotriosidase	IRP	UBCH8	SAHH
MMP-9	MAPBPIP	MafB	IRP2	ATF/CREB	PBX
DR4(TNFRSF10A)	Hepcidin	LILRB5	FR-beta	Cathepsin B	RBP-J kappa (CBF1)
CD20	IL-15	HLAHP	ANO6	SAA1	Bcl-XL
Myc/Max	TISB	CD40(TNFRSF5)	NAIP	SIN1	ZFR2
c-Myc	Caspase-3	p90Rsk	Factor H	MR1	NF-kB1 (p105)
Galpha(i)-specific nucleotide-like GPCRs	HLA-DM	p90RSK3(RPS6KA2)	Clusterin	IL-26	TLR10
Galpha(q)-specific nucleotide-like GPCRs	c-Kit	GBP6	MER	CCL4L1	Calreticulin
Adenosine A3 receptor	IL-6	NGAL	I-kB	CCL4L2	BTN3A2
Galpha(s)-specific prostanoid GPCRs	p21	Integrin	NFKBIA	MIP-1-beta	GRO-3
PGE2R4	RPS27A	ITGB2	PTEN	BATF	Sec24
ATP6V0A	APOBEC3F	nAChR alpha	GRO-2	Siglec-10	Sec24D
G6PT1	UBC	CD229	GPR65	PEA15	RGS1
CSF2RA	Ubiquitin	Oncostatin M	HLA-DRB5	Aiolos	GADD45 gamma
WWP1	Galpha(q)-specific peptide GPCRs	PSMB8(LMP7)	HLA-DRB1-1	PDE4	DR5(TNFRSF10B)
Galpha(q)-specific Class A Orphan/other GPCRs	A20	HLA-DQA1	IP10	PDE	SLAP-130(ADAP)
TAP	IL-18BP	AP1S1A	Rb protein	PDE4B	CalDAG-GEFIII
TAP2 (PSF2)	EBI2	AP-1 sigma subunits	Ficolin-1	ADAMDEC1	SWAP-70
RIPK2	CMKLR1	CD161	Ficolin	DMP4	MEF2
PSME1	WaspIP	CaMK II delta	Cyclophilin 60	Raftlin	MEF2C
Peroxiredoxin	CD209	CCR6	PAK1	CRACC	ABL2
HIF1A	ARPC1B	IL23R	Legumain	TR2(TNFRSF14)	

Table 5.8 All genes in BAL high activity dataset contributing to top GO process – Immune System Process - from metacore analysis.

Pathway analysis of the 947 genes identified in the high activity genes group showed several pathways related to immune response, including TLR signalling, apoptosis and survival, MAPK related pathways, and IL10 signalling pathways (Table 5.9).

No	Pathway Map	p-value	FDR	Genes in dataset
1	Transcription – Sirtuin6 regulation and function	1.617E-07	1.218E-04	13/64
2	Rheumatoid arthritis (general schema)	6.301E-07	2.372E-04	11/50
3	Immune response – Substance P stimulated expression of pro-inflammatory cytokines via MAPKs	1.191E-06	2.990E-04	10/43
4	Apoptosis and survival – APRIL and BAFF signalling	4.418E-06	8.316E-04	9/39
5	Development – Role of IL-8 in angiogenesis	9.601E-06	1.446E-03	11/65
6	Immune response TLR2 and TLR4 signalling pathways	1.763E-05	1.896E-03	10/57
7	Role of B cells in SLE	1.763E-05	1.896E-03	10/57
8	PDE4 regulation of cyto/chemokine expression in arthritis	3.180E-05	2.846E-03	9/49
9	PDE4 regulation of cyto/chemokine expression in inflammatory skin diseases	3.765E-05	2.846E-03	9/50
10	Immune response IL-10 signalling pathway	3.779E-05	2.846E-03	10/62

Table 5.9 Top 10 Pathways derived from gene list (n=947) that show significant positive correlation to SCAS (in BAL) when analysed through MetacoreTM.

While pathway analysis highlights single sequential links between genes, one gene may be involved in multiple pathways. As such, a process network, which takes this into account, provides a more complete, “three-dimensional” visualisation of these interactions. Table 5.10 shows the process networks for the BAL cell dataset. This has revealed ‘Antigen Presentation’ as part of the Immune response network, as the leading network.

No	Process Network	p-value	FDR	Genes in dataset
1	Immune response – Antigen presentation	2.856E-07	4.485E-05	28/197
2	Cell adhesion – Leucocyte chemotaxis	5.951E-07	4.671E-05	28/204
3	Inflammation – MIF signalling	3.836E-06	2.007E-04	21/140
4	Inflammation – IL-10 anti-inflammatory response	1.752E-05	6.878E-04	15/87
5	Immune response – Th17-derived cytokines	7.436E-05	2.335E-03	15/98
6	Chemotaxis	1.145E-04	2.995E-03	18/137
7	Apoptosis – Anti-apoptosis mediated by external signals via NF-kB15/115	3.096E-04	6.944E-03	15/111
8	Inflammation – IL-4 signalling	4.567E-04	8.963E-03	15/115
9	Inflammation IL-6 signalling	6.599E-04	1.151E-02	15/119
10	Inflammation – Protein C signalling	7.480E-04	1.174E-02	14/108

Table 5.10 Top 10 Process Networks derived from gene list (n=947) that shows significant positive correlation to SCAS (in BAL) when analysed through MetacoreTM.

In fact there are many networks of interest in these results, confirming the role of inflammation, immune response, chemotaxis and apoptosis in increasing disease activity in pulmonary sarcoidosis.

1369 genes were identified in the 'low activity genes' group. The top ten GO processes, Pathway maps, and Process networks, as analysed in metacore, are displayed in Table 5.11, Table 5.12 and Table 5.13 respectively.

No	Enrichment by GO Processes	p-value	FDR	Genes in dataset
1	Cellular macromolecule metabolic process	1.700E-27	1.325E-23	626
2	Metabolic process	7.511E-27	2.928E-23	831
3	Cellular metabolic process	1.656E-26	4.305E-23	739
4	Nucleic acid metabolic process	3.961E-26	7.721E-23	404
5	Heterocyte metabolic process	5.752E-26	8.970E-23	444
6	Macromolecule metabolic process	3.011E-25	3.913E-22	456
7	Primary metabolic process	2.004E-24	2.132E-21	665
8	Organic substance metabolic process	2.187E-24	2.132E-21	749
9	Organic substance metabolic process	3.375E-24	2.924E-21	767
10	Cellular aromatic compound metabolic process	4.185E-24	3.263E-21	454

Table 5.11 Top 10 GO processes derived from gene list (n=1369) that shows significant negative correlation to SCAS (in BAL) when analysed through MetacoreTM.

No	Pathway Map	p-value	FDR	Genes in dataset
1	Cytoskeleton remodelling – TGF, WNT and cytoskeletal remodelling	5.937E-10	4.643E-07	21/111
2	Neurophysiological process – ACM regulation of nerve impulse	6.795E-07	2.657E-04	11/46
3	Regulation of GSK3 beta in bipolar disorder	5.584E-06	1.456E-03	10/46
4	Apoptosis and survival – Role of PKR in stress-induced apoptosis	2.117E-05	2.974E-03	10/53
5	Development – Differentiation of white adipocytes	2.117E-05	2.974E-03	10/53
6	Cytoskeleton remodelling – Cytoskeleton remodelling	2.326E-05	2.974E-03	14/102
7	Transcription- Role of the non-genomic action of retinoic acid and phosphorylation of retinoic acid receptors in the initiation of transcription	2.662E-05	2.974E-03	8/34
8	Transcription – Androgen receptor nuclear signalling	3.374E-05	3.041E-03	9/45
9	Airway smooth muscle contraction in asthma	3.500E-05	3.041E-03	10/56
10	Development – Positive regulation of STK3/4 (Hippo) pathway and negative regulation of YAP/TAZ function	4.943E-05	3.865E-03	11/70

Table 5.12 Top 10 Pathways derived from gene list (n=1369) that shows significant negative correlation to SCAS (in BAL) when analysed through MetacoreTM.

No	Process Network	p-value	FDR	Genes in dataset
1	DNA damage – BER-NER repair	1.787E-05	1.611E-03	19/109
2	Transcription – mRNA processing	2.066E-05	1.611E-03	24/160
3	Cell cycle G2-M	7.443E-05	3.870E-03	27/206
4	Cell adhesion Cadherins	1.425E-04	5.559E-03	24/180
5	Signal transduction – WNT signalling	2.874E-04	8.965E-03	23/177
6	Protein folding – Protein folding nucleus	5.006E-04	1.302E-02	11/58
7	Signal transduction – TGF-beta, GDF and activin signalling	7.122E-04	1.587E-02	20/154
8	Signal transduction – BMP and GDF signalling	8.460E-04	1.650E-02	14/91
9	Translation - Regulation of initiation	1.253E-03	1.833E-02	17/127
10	Transcription – Chromatin modification	1.253E-03	1.833E-02	17/127

Table 5.13 Top 10 Process Networks derived from gene list (n=1369) that shows significant negative correlation to SCAS (in BAL) when analysed through MetacoreTM.

A decrease in WNT signaling is noted from analysis of the negatively correlated gene lists.

5.4.4 Correlation of gene expression profile of PBMCs to disease activity

5.4.4.1 GSEA analysis

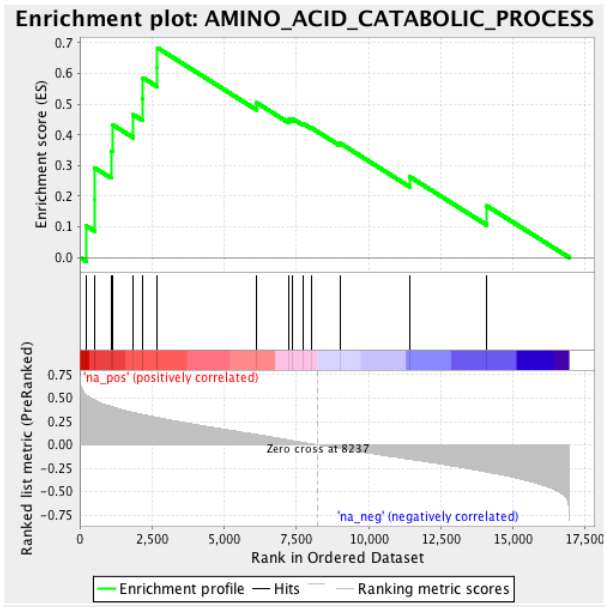
Analysis of whole gene expression from cells isolated from PBMCs identified 21373 genes and level of expression was correlated with activity (SCAS). A total of 16954 genes had GO associated terms, and as per BAL analysis, these genes were ranked according to increasing expression with increasing activity, and Gene-sets representing all biological processes (C5bp - as described in section 5.3.3) were tested.

In total 364 of 474 gene sets were enriched in the high activity group. 156 gene sets were significant at FDR < 0.25. The top ten significantly ($p < 0.05$) enriched gene sets in the high activity group are listed in Table 5.14. The enrichment plot for the top gene set is shown in Figure 5.8A. The 'Defense Response' gene set features in the top ten, as it did in the BAL cell GSEA analysis (Figure 5.8B).

No	Gene Set	ES	NES	NOM p-val	FDR q-val
1	Amino Acid Catabolic Process	0.684	2.279	0.0	0.001
2	Cofactor Metabolic Process	0.519	2.184	0.0	0.006
3	Nitrogen Compound Catabolic Process	0.625	2.183	0.0	0.004
4	Amine Catabolic Process	0.630	2.169	0.0	0.004
5	Defense Response	0.382	2.098	0.0	0.010
6	Cellular Defence Response	0.489	2.048	0.0	0.013
7	Heterocycle Metabolic Process	0.574	2.030	0.0	0.015
8	Pigment Metabolic Process	0.632	2.007	0.0	0.016
9	Cellular Respiration	0.606	2.002	0.0	0.016
10	Organic Acid Metabolic Process	0.384	2.000	0.0	0.015

Table 5.14 Top ten positively enriched gene sets identified using GSEA analysis of whole gene expression from PBMCs correlated to increasing SCAS.

A



B

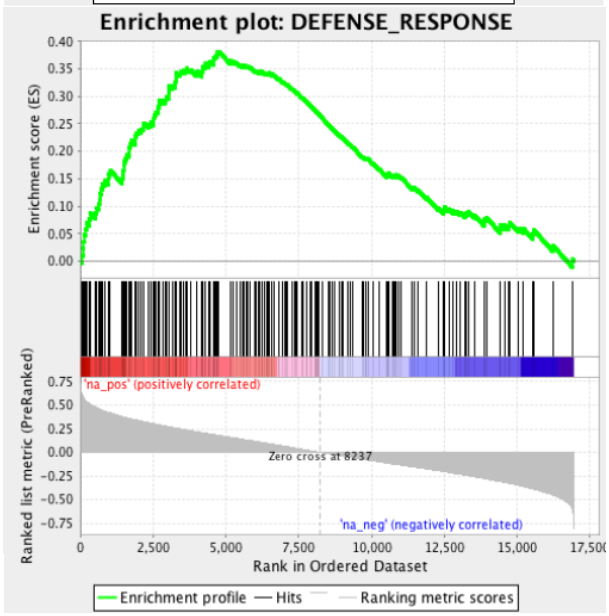


Figure 5.8 A) Enrichment plot for the top gene set in PBMC GSEA analysis – Amino Acid Catabolic Process. **B**) Enrichment plot for the Defense Response gene set, which is also present in the BAL cell GSEA analysis.

Leading edge genes of the amino acid catabolic process, and defense response gene sets are shown in Table 5.15 and Table 5.16 respectively.

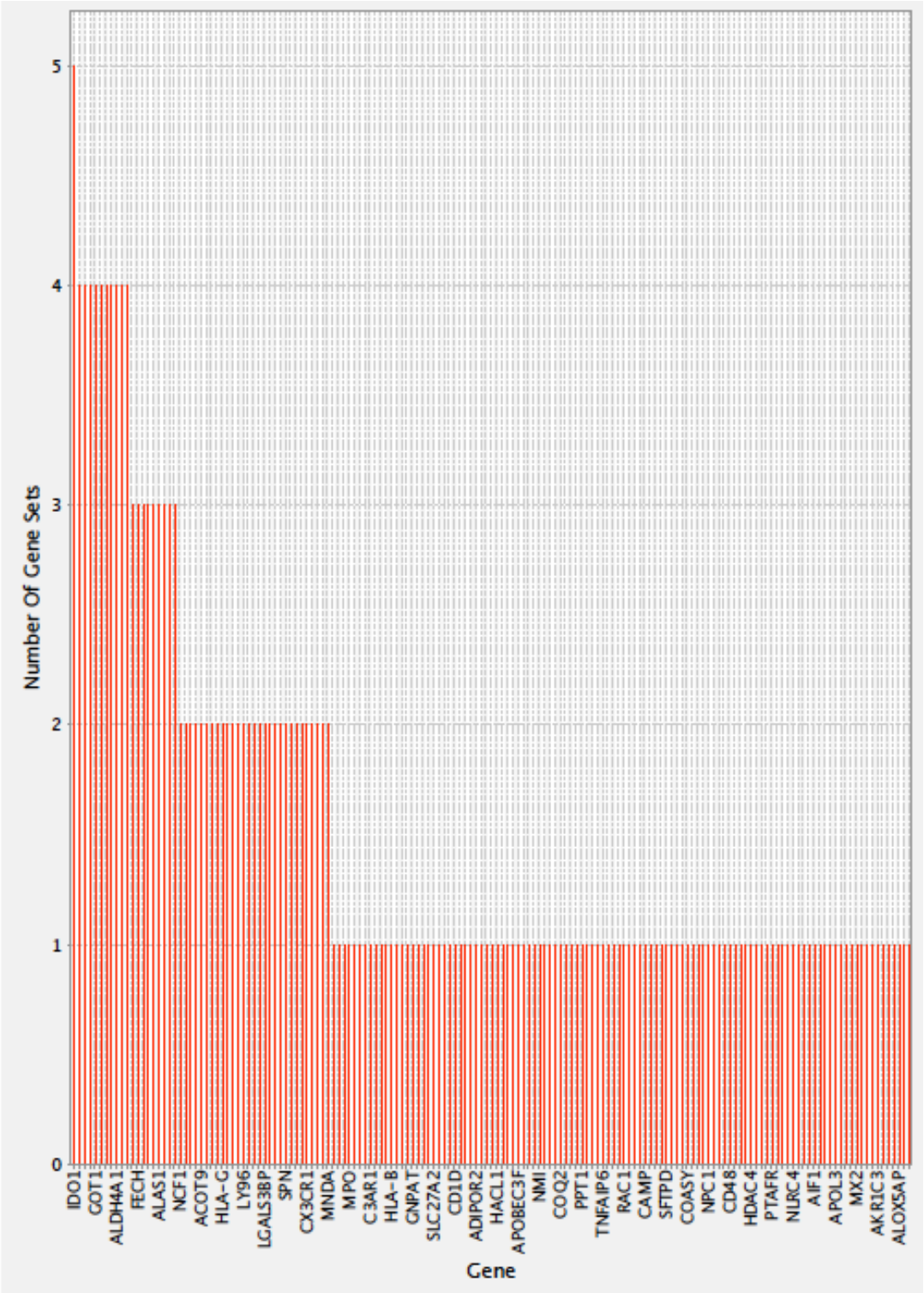
Gene	Rank
MCCC2	219
GOT1	511
GLUD1	513
ALDH4A1	1084
IDO1	1137
GSTZ1	1856
ASRGL1	2167
BCKDHA	2185
BCKDK	2689
GLDC	2692

Table 5.15 Leading edge genes and their ranking in the entire gene list, from the amino acid catabolic processes gene set of PBMC whole gene expression.

Gene	Rank	Gene	Rank	Gene	Rank	Gene	Rank
APOBEC3F	63	NCF1	1493	HLA-B	2674	KLRC2	3992
CD97	90	LILRA1	1517	MNDA	2705	CD48	4018
CXCL9	113	MICB	1560	NMI	2711	ALOX5AP	4183
TLR7	127	MEFV	1610	HCP5	2826	CD1D	4184
CXCL10	194	AIF1	1675	WAS	2925	SP140	4208
CX3CR1	211	LY96	1678	LGALS3BP	2944	S1PR3	4221
TCIRG1	317	NCF2	1706	PRF1	3029	IL17C	4274
TLR8	344	NLRC4	1755	CD300C	3031	NOD2	4279
C2	521	SELE	1898	PLA2G7	3040	LSP1	4402
CCL23	555	HLA-G	1900	MPO	3053	MX1	4421
BECN1	625	LYST	1941	SFTPD	3151	LILRB5	4459
HDAC4	661	VPS45	1999	APOL3	3250	MX2	4565
RAC1	679	LILRB2	2092	PTPRC	3264	PGLYRP2	4566
APOBEC3G	712	TLR6	2156	CYSLTR1	3279	FOS	4568
CYBB	756	PSG3	2188	TNFRSF1A	3285	CAMP	4640
HDAC7	830	SPN	2350	PTAFR	3289	ANXA1	4656
IL10RB	953	C3AR1	2417	LILRB3	3412	GNLY	4695
LILRA2	966	NCR1	2519	CEBPB	3440	BCL10	4698
RSAD2	1032	KIR2DS3	2525	TPSAB1	3530	CEBPG	4762
CCR2	1429	CD40	2559	ADORA3	3607		
FPR2	1461	PARP4	2587	TNFAIP6	3650		
NFE2L1	1476	TYROBP	2655	C5	3797		

Table 5.16 Leading edge genes and their ranking in the entire gene list, from the defense response gene set of PBMC whole gene expression.

Leading edge analysis of the top 10 gene sets reveals IDO1 as the top leading edge gene, present in 5 of the top ten upregulated gene sets (Figure 5.9). IDO1 is ranked at 1137 in the whole gene list (of 16954 genes) in this series in order of correlation to SCAS, with a correlation coefficient of 0.4 ($p = 0.02$).



IDO1
GLDC
MCCC2
ASRGL1
GOT1
GLUD1
NFE2L1
BCKDHA
ALDH4A1
GST21
BCKDK

Figure 5.9 Leading edge analysis of top ten enriched gene sets in GSEA, revealing IDO1 as the top leading edge gene. Leading edge genes present in four or more gene sets are shown alongside the histogram.

No gene sets were enriched in the low activity group, and are therefore not shown.

5.4.4.2 Metacore analysis

As for BAL cells, using SCAS correlation to PBMC gene expression, a correlation co-efficient cut-off of r >0.5 (Spearman rank) for correlation with increasing activity, and r <-0.5 for correlation with decreasing activity (p-value of <0.05 for both), was employed to create separate positive and negative significant gene lists from PBMC microarray analysis for use in Metacore. As before, for ease of description, the first group is labeled 'high activity genes group' and the second group, the 'low activity genes group'.

413 genes were identified in the high activity genes group. Enrichment in GO processes again revealed many of the upregulated processes as being related to the immune system (Table 5.17). Annexin II (also known as ANXA II) features in the genes in the top dataset for GO processes, which has been previously identified as a susceptibility gene in sarcoidosis by Hoffman et al (43). All genes from PBMC gene expression present in the top dataset are shown in Table 5.18.

No	Enrichment by GO Processes	p-value	FDR	Genes in dataset
1	Immune system process	6.244E-19	3.239E-15	120
2	Antigen processing and presentation of peptide antigen via MHC class I	1.149E-17	2.981E-14	26
3	Immune response	3.457E-17	5.977E-14	89
4	Protein complex subunit organization	1.046E-14	1.357E-11	86
5	Innate immune response	1.613E-14	1.674E-11	67
6	Single-organism metabolic process	3.038E-14	2.627E-11	174
7	Macromolecular complex subunit organisation	5.823E-14	4.315E-11	106
8	Antigen processing and presentation of exogenous peptide antigen – MHC class I	7.825E-14	4.680E-11	19
9	Immune effector process	8.120E-14	4.680E-11	45
10	Antigen processing and presentation of exogenous peptide antigen - MHC class I	2.375E-13	1.232E-10	19

Table 5.17 Top 10 GO processes derived from gene list (n=413) that shows significant positive correlation to SCAS (in blood) when analysed through MetacoreTM.

Gene	Gene	Gene	Gene	Gene	Gene
STAT1	TAP2 (PSF2)	FOXO3A	XAF1	JAK2	DOK2
Tcf (Lef)	Deoxyribonuclease II	GBP1	DYNC1LI	Annexin 11	TLR8
ERK1/2	14-3-3 beta/alpha	GBP4	PSMA5	LMO2	DHRS2
ERK1 (MAPK3)	14-3-3	Apo-2L (TNFSF10)	Sec23	IRp60	PSMB8 (LMP7)
Perforin	ITGA4	EFP	IGFRB	LAIR1	IKK-gamma
HLA-F	PSME1	SAMD9L	PSMB10	PRK1	Sec24
APOBEC3	PSME2	ARP2	UBCH8	PRDX3	Sec24D
APOBEC3B	Fc gamma R1	ARPC2	Tapasin	E2A/HLF Fusion protein	ATP11C
TLR7	Peroxiresoxin	ADAM10	SAR1	E2A	IFI27
ERK2 (MAPK1)	Actin cytoskeletal	GCH1	SAR1B	GBP3	TWSG1
ATP6VDA	Actin	DNA ligase III	APOBEC3F	PSMD1	TRIM14
LMO4	ACTB	ADA	MR1	SAMHD1	
PSMA4	CX3CR1	FAM111A	Plastin	GBP5	
PSMB9	Galpha (I)-specific peptide GPCRs	Beta 2-microglobulin	MOV-10	K(+) channel	
SCAP2	PSMD10 (Gankyrin)	LyGDI	Siglec-10	Subfamily J	
Cathepsin C	Caspase-1	CD97	ADRM1	C1TC	
CD38	IP10	TAP1 (PSF1)	RC3H2	C6orf150	
CD164	Presenilin	MIG	CRACC	GBP6	
BCAP	Presenilin 2	P22-phox	2’5’-oligoadenylate synthetase	Caspase-10	
GBP2	CAPZA1	ACTR3	OAS2	PDE	
UBE2A	CAPZA	ARP3	TRIM5	PRDX1	
TAP	PAK1	CALCOCO2	PSMB2	TRAF3	

Table 5.18 All genes in PBMC high activity dataset contributing to top GO process – Immune System Process - from metacore analysis.

Pathway analysis reveals several interesting and likely relevant immune pathways upregulated in this dataset (Table 5.19).

No	Pathway Map	p-value	FDR	Genes in dataset
1	Immune response – Antigen presentation by MHC class I	4.352E-09	2.502E-06	8/28
2	Cytoskeleton remodelling – TGF,WNT and cytoskeleton remodeling	5.005E-06	1.258E-03	10/111
3	LRRK2 in neurons in Parkinson's disease	7.274E-06	1.258E-03	6/33
4	Development of EGFR signalling pathway	8.749E-06	1.258E-03	8/71
5	Cytoskeleton remodelling – Regulation of actin cytoeketon by Rho GTPases	1.744E-05	2.006E-03	5/23
6	Transport- Macropinocytosis regulation by growth factors	3.571E-05	3.422E-03	7/63
7	Immune response – IL-4 signaling pathway	6.891E-05	5.661E-03	8/94
8	Immune response CCR3 signaling in eosinophils	1.308E-04	7.879E-03	7/77
9	Development – EPO-induced Jak-STAT pathway	1.460E-04	7.879E-03	5/35
10	Role of Tissue factor in cancer independent coagulation protease signaling	1.460E-04	7.879E-03	5/35

Table 5.19 Top 10 Pathways derived from gene list (n=413) that shows significant positive correlation to SCAS (in blood) when analysed through MetacoreTM.

The top pathway identified relates to Antigen presentation by MHC class I.

Table 5.20 lists the genes present from the high activity list contributing to this pathway, which include TAP1 and TAP2. TAP genes are membrane associated proteins involved in the transport of antigen from the cytoplasm to the endoplasmic reticulum for association with MHC class I molecules (233). Expression of TAP is downregulated by Epstein Barr Virus vIL-10 protein, again lending thought to an escape mechanism of TAP expression as a result of low IL10 production in sarcoidosis.

Gene
TAP2
PSME-1
PSME-2
B-2 Microglobulin
TAP1
Tapasin
PSMD1
LMP7

Table 5.20 All genes in PBMC high activity dataset contributing to top GO process – Immune System Process - from metacore analysis.

The top 10 process networks from the high activity PBMC gene list are shown in Table 5.21.

No	Process Network	p-value	FDR	Genes in dataset
1	Immune response – Phagosome in antigen presentation	5.643E-08	4.570E-06	21/243
2	Cytoskeleton – Regulation of cytoskeleton rearrangement	7.618E-08	4.570E-06	18/183
3	Inflammation – Interferon signaling	9.456E-08	4.570E-06	14/110
4	Cytoskeleton – Actin filaments	2.342E-07	6.823E-06	17/176
5	Immune response – Antigen presentation	2.353E-07	6.823E-06	18/197
6	Proteolysis – Ubiquitin-proteasomal proteolysis	5.584E-07	1.349E-05	16/166
7	Immune response – Innate immune response to RNA viral infection	8.971E-05	1.858E-03	9/85
8	Cell adhesion – Leucocyte chemotaxis	1.375E-04	2.493E-03	14/204
9	Inflammation – Amphoterin signalling	1.056E-03	1.701E-02	9/118
10	Signal transduction – ERBB-family signalling	1.186E-03	1.720E-02	7/75

Table 5.21 Top 10 Process Networks derived from gene list (n=413) that shows significant positive correlation to SCAS (in blood) when analysed through MetacoreTM.

The top ten GO processes and pathways in the low activity gene group in PBMC analysis (n=339) are shown in Table 5.22 and Table 5.23 respectively.

No	Enrichment by GO Processes	p-value	FDR	Genes in dataset
1	Single organism intracellular transport	3.551E-08	1.599E-04	43
2	Establishment of protein localization to membrane	1.200E-07	1.929E-04	17
3	Protein localization to membrane	1.285E-07	1.929E-04	21
4	Intracellular protein transport	2.530E-07	2.390E-04	29
5	Intracellular transport	2.654E-07	2.390E-04	44
6	Cellular protein localization	3.583E-07	2.689E-04	38
7	Cellular macromolecule localization	4.316E-07	2.777E-04	38
8	Protein targeting to ER	6.315E-07	3.554E-04	10
9	Protein targeting to membrane	7.961E-07	3.672E-04	13
10	Establishment of localization in cell	8.267E-07	3.672E-04	50

Table 5.22 Top 10 GO processes derived from gene list (n=339) that shows significant negative correlation to SCAS (in blood).

No	Pathway Map	p-value	FDR	Genes in dataset
1	LRRK2 in neurons in Parkinson's disease	6.621E-10	1.695E-07	7/33
2	Transport Alpha-2 adrenergic receptor regulation of ion channels	2.660E-03	2.729E-01	3/47
3	Regulation of degradation of wtCFTR	6.124E-03	2.729E-01	2/20
4	Immune response IL15 signalling via JAK-STAT cascade	7.389E-03	2.729E-01	2/22
5	Development – Glucocorticoid receptor signaling	9.488E-03	2.729E-01	2/25
6	Transcription regulation of aminoacid metabolism	9.488E-03	2.729E-01	2/25
7	Colorectal cancer	1.350E-02	2.729E-01	2/30
8	Muscle contraction S1P2 receptor-mediated smooth muscle contraction	1.350E-02	2.729E-01	2/30
9	Signal transduction PTMs in IL23 signalling pathway	1.438E-02	2.729E-01	2/31
10	Development – Transcription regulation of granulocyte development	1.528E-02	2.729E-01	2/32

Table 5.23 Top 10 pathways seen in significantly negatively correlated genes to SCAS when analysed through MetacoreTM.

Overall, analysis of GO processes, pathways and process networks prominent in these 339 negatively enriched genes did not reveal anything of obvious clinical relevance to the study (process networks not shown).

5.4.5 Comparison of significant genes, gene sets, pathways and processes from different sampling modalities.

Comparison of the results obtained from correlation to activity in GSEA and Metacore, between the two sampling modalities, was performed to interrogate whether certain genes and their associated pathways that are upregulated in the lungs are reflected in the periphery. This may be useful for further work on therapeutic targets, understanding the pathogenesis of the active disease process in sarcoidosis, or as a useful peripheral marker of organ specific disease activity.

Genes with a correlation co-efficient cut-off of ≥0.5 when correlated to SCAS (i.e. positive correlation to increasing activity) were compared between sampling modalities (BAL and PBMC) (Figure 5.10). Fifty-nine genes were seen to be present in both, and are listed in Table 5.24.

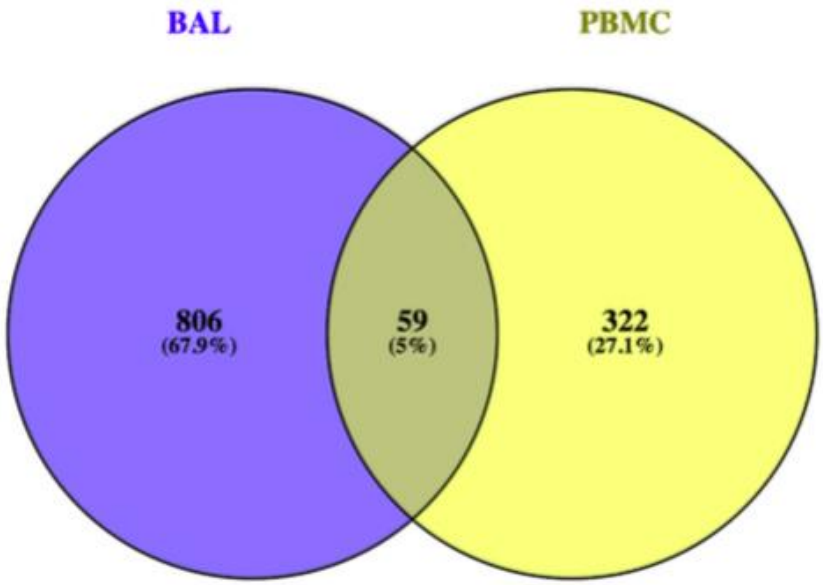


Figure 5.10 Overlap between genes with correlation coefficient >0.5 and p value <0.05 from the two sampling modalities, BAL and PBMC (i.e. positive correlation as SCAS increases).

PBMC & BAL		
SEC24D	ASCL2	C11ORF48
PSTPIP2	KCNJ2	DCAF6
ABI3	BOLA2	PLEKHB2
SLAMF7	GIYD1	PRDX4
VAMP5	PAK1	GLB1
LOC728368	BRP44L	CD300A
CXCL9	ETNK1	SIGLEC10
LOC400759	GBP1	IKBIP
IDH2	ME2	GBA2
SFT2D2	WDR41	SKAP2
FLJ38482	GCH1	BCAP29
PSMB8	CECR1	CXCL10
C10ORF58	PIGN	TAP1
C19ORF12	TAP2	G3BP1
MTHFD2	THEM2	INPP1
MR1	DBI	PSME1
PHKB	ECHS1	APOBEC3F
STAT1	PALLD	

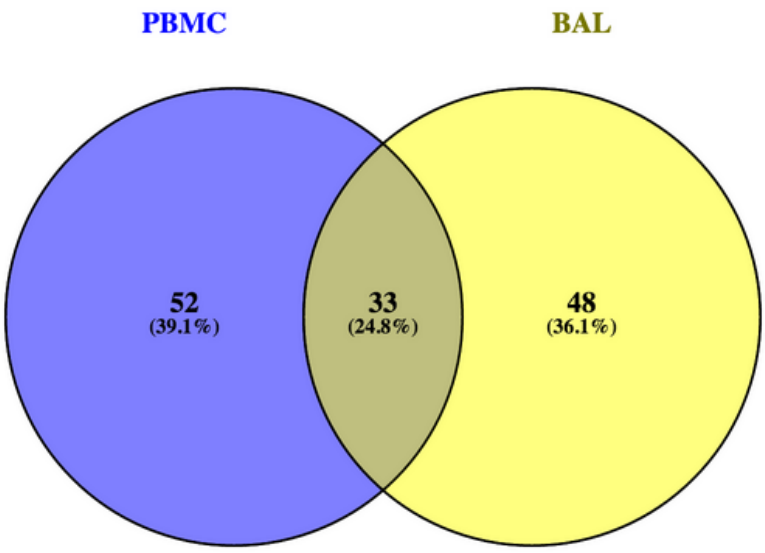
Table 5.24 Table of overlapping genes between PBMC and BAL, when correlation coefficient to increasing SCAS is set at >0.5.

Metacore analysis of these 59 overlapping genes (between BAL cell and PBMC) reveal a significant predominant immune process taking place (Table 5.25).

No	Enrichment by GO Processes	p-value	FDR	Genes in dataset
1	Immune System Process	5.793E-10	1.284E-06	30
2	Response to other organism	1.427E-08	8.321E-06	19
3	Response to external biotic stimulus	1.459E-08	8.321E-06	19
4	Response to bacterium	1.661E-08	8.321E-06	16
5	Defence response to virus	1.924E-08	8.321E-06	9
6	Cytosol to ER transport	2.411E-08	8.321E-06	3
7	Response to biotic stimulus	2.627E-08	8.321E-06	19
8	Antigen processing and presentation of peptide antigen via MHC class I	3.403E-08	8.759E-06	8
9	Immune response	3.556E-08	8.759E-06	22
10	Defence response	9.203E-08	2.040E-05	23

Table 5.25 Metacore analysis of overlapping genes between PBMC and BAL that show a significant correlation to increasing SCAS.

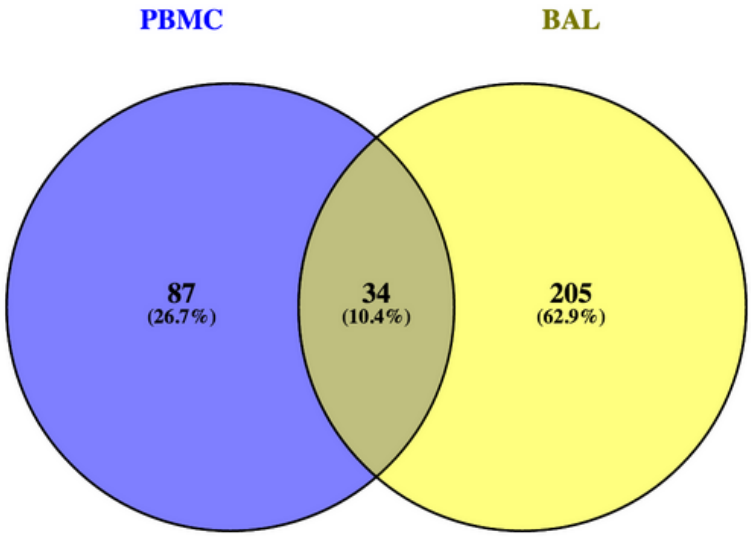
The gene set of 'Defense Response' was enriched in increasing SCAS in both BAL (4/10) and PBMC (5/10) using GSEA. Reviewing overlapping genes involved in this gene set from the two modalities reveals some genes that are of interest, having previously been associated with sarcoidosis e.g. CXCL10 (234) and CCR2 (235) (Figure 5.11).



APOBEC3F	C3AR1
CXCL9	KIR2DS3
CXCL10	CD40
CX3CR1	HCP5
TCIRG1	PLA2G7
C2	APOL3
CCL23	PTPRC
BECN1	PTAFR
HDAC4	LILRB3
APOBEC3G	TPSAB1
IL10RB	ADORA3
CCR2	TNFAIP6
NCF1	CD48
LY96	IL17C
HLA-G	LILRB5
LILRB2	FOS
TLR6	

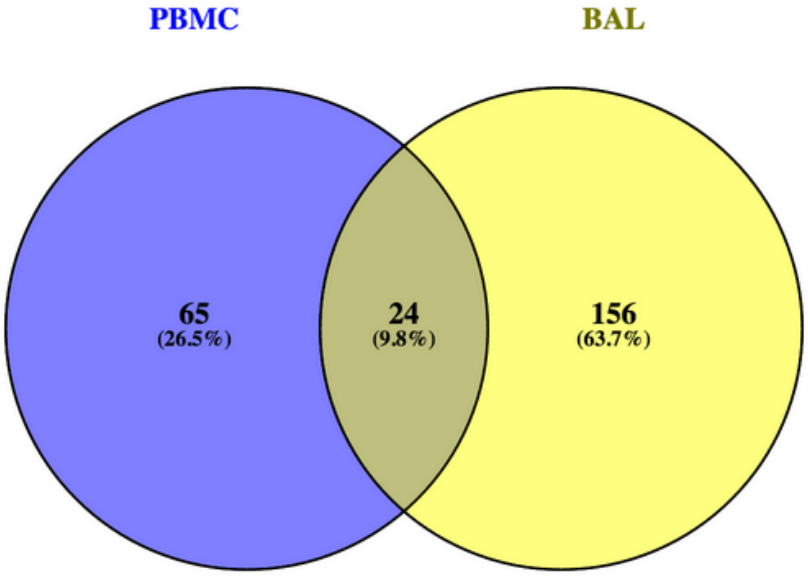
Figure 5.11 Overlap of genes from 'Defense Response' gene set in PBMC and BAL. The 33 genes that overlap are shown next to venn diagram.

Both BAL and PBMC reveal 'Immune system processes' and 'Immune response' as two of the top ten GO processes in the metacore analysis, of the high activity genes group. The overlap of genes in these sets, between the two sampling modalities are depicted in Figure 5.12 and Figure 5.13 respectively. CXCL10 (IP-10) again appears in this list, but TAP1 and TAP2 also feature.



STAT1
APOBEC3
APOBEC3B
ATP6V0A
SCAP2
TAP
TAP2 (PSF2)
PSME1
Peroxiredoxin
IP10
Presenilin
PAK1
GBP1
GBP4
GCH1
TAP1 (PSF1)
MIG
UBCH8
APOBEC3F
MR1
Siglec-10
CRACC
IRp60
PRK1
E2A/HLF fusion protein
E2A
GBP3
K(+) channel subfamily J
GBP6
PDE
PSMB8(LMP7)
Sec24
Sec24D

Figure 5.12 Overlap of genes from GO - Immune system process (Metacore) between BAL and PBMC. The 34 genes that overlap are shown next to venn diagram.



STAT1
APOBEC3
APOBEC3B
ATP6V0A
TAP
TAP2 (PSF2)
PSME1
Peroxiredoxin
IP10
Presenilin
PAK1
GBP1
GBP4
GCH1
TAP1 (PSF1)
MIG
UBCH8
APOBEC3F
MR1
IRp60
GBP3
GBP6
PDE
PSMB8(LMP7)

Figure 5.13 Overlap of genes from GO - Immune response (Metacore) between PBMC and BAL. The 24 genes that overlap are shown next to venn diagram.

5.5 Discussion

This chapter examined genes from two different sampling modalities, namely PBMC and BAL cells, that relate to disease activity as measured by the Sarcoidosis Clinical Activity Score (SCAS) (described in detail in Chapter 4).

The most informative results were observed in BAL cell gene expression. This is expected in these patients that have lung dominant disease, as it is the site of disease activity. Gene set enrichment analysis revealed several immune related gene sets, the strongest being 'Immune Response'. Within this gene set, CCR2, TREM2, IL6, CCL18, CCL2, IL10RB, IL15, CCR6 and CCL20 contribute to the enrichment.

C-C Chemokine Receptor 2 (CCR2) was present in all top ten positively enriched gene sets in the leading edge analysis of BAL gene set enrichment results. It was also present in the 'Defense Response' gene set that overlapped between BAL cell and PBMC. CCR2 encodes the receptor for monocyte chemoattractant protein-1 (CCL2 or MCP-1), a chemokine that specifically mediates monocyte chemotaxis. The protein is expressed primarily on the surface of macrophages and monocytes, but is also seen on dendritic cells and T lymphocytes. CCR2 is required for infiltration of monocytes to the site of inflammation (236) in many conditions (e.g. tumour (237, 238), infection (239), inflammatory diseases (239, 240)). An association between CCR2 single nucleotide polymorphisms and Löefgrens syndrome has previously been described (235), and several other studies have also investigated the presence of CCR2 gene variants in both susceptibility to (241), or protection from (242, 243) sarcoidosis. CCR2 may be involved in fine-tuning the

granulomatous immune reaction. CCR2 knockout mice are unable to initiate immune control of the infectious granulomatous disease *Mycobacterium tuberculosis*, as the absence of CCR2 causes an early and persistent defect in macrophage recruitment to the lungs. In addition, the numbers of macrophages and dentritic cells in the mediastinal lymph nodes are reduced, delaying T lymphocyte priming and deficiency in CD4 and CD8 T cell accumulation in the lungs (244). High levels of CCR2 expression suggest that CCR2 expressing monocytes are a major cellular player in disease activity in sarcoidosis; and may be involved in an over-exuberant granulomatous process in the lungs.

With regards to CCR2 gene polymorphisms, some are known to result in non-functioning receptors, so it is understandable that the type of CCR2 polymorphism may play an important role in promoting disease. However, an increase in CCR2 overall, may still be significant and may still represent a pathway that could be targeted to reduce disease activity.

Inhibition of CCR2 has been considered as a target for multiple diseases including autoimmune disease, atherosclerosis, pain, and metabolic disease, based in part on the critical role this receptor plays on monocyte migration. Blockade of CCR2, or of it's primary ligand CCL2, has proven efficacious in a diverse array of inflammatory disease murine models (245-249). Two combined CCR2/CCR5 drugs have reached phase 2 clinical trials, namely PF-04634817 (Pfizer) and Cenicriviroc (or TAK-652) (Takeda Pharmaceuticals). The former is being trialled in the treatment of diabetic macular oedema, and the latter in HIV infection (where CCR5 is best known for it's role as a co-receptor for entry of HIV into cells). Several other CCR2 antagonists

have also been identified, but progression to use in clinical practice has been limited (250). The potential for a new drug soon to be released that blocks CCR2, may pave the way to a new clinical trial in the treatment of pulmonary sarcoidosis.

Of the other genes contributing to the enrichment of the top gene set 'immune response' (Table 5.4), CCL2, IL-6, IL15, CCR6 and CCL20 have previously been attributed to sarcoidosis pathogenesis.

CCL2 (MCP-1), the ligand for CCR2, has also been attributed to pulmonary sarcoidosis. In a study by Palchevskiy et al, which looked at BALF chemokine levels in 72 patients affected by sarcoidosis, CCL2 and CCL5 levels were found to be elevated, and subgroup analysis showed higher levels of both chemokines in all stages of pulmonary sarcoidosis (as per scadding CXR stage) (65). IL6 has been reported as having a role in granuloma formation, and Seitzer et al found that there were significant modulatory effects of interleukin-6 on the cell number per in vitro granuloma and on the morphology of the cells involved. They concluded that elevated interleukin-6 levels may modulate granuloma formation with respect to the number of cells involved, and influence distinct cell populations involved in granuloma formation (251). Also, a significant correlation ($r=0.70$, $p<0.05$) between IL6 production from the alveolar macrophages of patients with pulmonary sarcoidosis, and the ratio of CD4+/CD8+ in BAL fluid, has suggested involvement in the initiation and maintenance of alveolitis by activating and inducing the proliferation of T cells (252). IL15, a cytokine that regulates T and natural killer cell activation and proliferation, and induces the release of IFN-γ from T cells, has previously been reported as being highly expressed in alveolar macrophages

derived from patients with sarcoidosis (168, 253, 254). CCL20 (MIP-3α), the ligand of CCR6 (also seen to contribute to the enrichment of high activity gene sets), is involved in homing and trafficking of lymphocytes into secondary lymphoid organs, and in the maturation of leukocytes (255). CCL20 is also known to stimulate the migration of B cells (256), immature dendritic cells (257) and a subset of memory CCR6 expressing T cells (258). Facco et al hypothesised that CCL20 could contribute to the development of the CD4+ T cell alveolitis, characterising the initial phase of pulmonary sarcoidosis, and to the building of granulomatous structures. In particular, this study evaluated whether Th1 cells, accumulating in the lungs of patients with sarcoidosis, express CCR6, and the eventual production and release of CCL20 by sarcoid alveolar macrophages. Of note, this paper also demonstrated a significant difference in expression/detection of CCL20 and CCR6 between active (BAL lymphocytosis +/- lung CD4:CD8 ratio >4 +/- positive gallium scan) and chronic sarcoidosis (259).

There were however, other genes in this same set that have not received as much attention in sarcoidosis, but may be worth pursuing as their potential role in the pathogenesis of pulmonary sarcoidosis could be easily hypothesized. Specifically, looking again at the correlated gene expression in BAL cells, it has been shown that as SCAS increases, TREM-2 (Triggering Receptor Expressed on Myeloid Cells- 2) expression increases. This is seen in both the Immune Response Gene Set and the Immune System Process Gene Set. TREM2 is expressed by myeloid cells, osteoclasts and microglia, and is associated with DAP12, a critical step required for surface expression and signaling by TREM-2. DAP12 mediates downstream signaling through a cytoplasmic ITAM domain, which can recruit Syk and activate

PI3K, phospholipase C, and Vav signaling cascades. Its ligand(s) are not known. The TREM-2/DAP12 pathway is associated with the differentiation of macrophages to their fusogenic state and creation of large multinucleated giant cells typical for granulomas in sarcoidosis (260-262). Phagocytosis is also a TREM-2 dependent process (263, 264). It is therefore possible that increased TREM2 expression is related to the development of granuloma formation and active disease in sarcoidosis. Suchankova et al observed increased TREM1 and TREM2 BAL cell surface expression in patients with pulmonary sarcoidosis (265), but there is very little else in the literature on the role of TREM2 in this disease.

Another putative gene involved in disease activity is CCL18 (PARC; DCCK1; AMAC-1; MIP-4). This is expressed by a broad range of monocytes/macrophages and dendritic cells, and was shown in vitro, to induce chemotaxis in T cells, B cells, DC, haematopoetic progenitor cells, fibroblasts and potentially monocytes/macrophages, but not in neutrophils (266). Its receptor has yet to be elucidated. Functionally, CCL18-stimulated macrophages have been shown to have a high capacity for unspecific phagocytosis and for pinocytosis, without an associated oxidative burst. The chemokines that are produced in response to CCL18 are angiogenic and attract various leucocyte populations, which sustain inflammation. However, the capacity of these cells to remove cellular debris without causing oxidative damage, and the production by monocytes of the anti-inflammatory cytokine IL10 will initiate termination of the inflammatory response, suggesting that CCL18-activated macrophages are pivotal in determining persistence of inflammation or its resolution (267). IL-10 in combination with IL-4 or IL-13 enhances CCL18 production by Macrophages, and has been shown to cause

the acquisition of M2 traits. Although CCL18 is constitutively expressed, enhanced CCL18 production has been described in a number of inflammatory disease conditions, including Gaucher's disease (a lysosomal storage disorder characterized by the accumulation of glycolipids in macrophages and granuloma formation) (134), rheumatoid arthritis (268), Sjogren's syndrome (269), and Hypersensitivity Pneumonitis and IPF (270, 271). Although the role CCL18 may take in active sarcoidosis has not yet been investigated, it has been reported as a potential marker of granulomatous activity in the disease. Having been identified along with chitotriosidase in the granulomatous condition of 'Gauchers disease', it was found in sarcoidosis that an increase in plasma chitotriosidase correlated with the stage of disease, being highest in active sarcoidosis with extrapulmonary involvement. In addition, therapy with steroids resulted in clear reduction of plasma chitotriosidase and CCL18 and relapse of disease activity was preceded by increases in these parameters. The group concluded that sarcoid macrophages secrete high quantities of chitotriosidase and CCL18 and that determination of plasma chitotriosidase and CCL18 may be useful to monitor changes in granulomatous macrophages during the course of sarcoidosis (136).

Finally, our group has previously shown an inverse relationship between IL10 expression and SCAS (194). IL10 has distinct T cell suppressive effects (272) and it's absence could play a role in the marked TH1 CD4+ T cell overactivity of sarcoidosis. It could be postulated that an increase expression of IL10RB is related to the lack of negative feedback from the low IL10 production in sarcoidosis patients.

The Metacore analysis of BAL cell gene expression data confirmed the findings from GSEA, and also revealed ADAMDEC1 and Chitotriosidase as genes present in it's top GO process - 'Immune system process'. Where GSEA uses all expressed genes ranked in order of correlation to activity, and in my analysis interrogates this with 4436 gene sets of GO biological processes, the metacore analysis focuses only on genes with a significant correlation of >0.5 correlation coefficient, and the 'immune system' process comprises 3132 genes, which include ADAMDEC1 and Chitotriosidase (Neither ADAMDEC1 or Chitotriosidase feature in the immune system process gene set on GSEA based on Gene Ontology biological processes, hence were not identified as contributing to enrichment). ADAMDEC1 has previously been reported by Crouser et al as a potential pathogenic mediator in sarcoidosis (168), and Chitotriosidase, a marker of activity in the granulomatous condition of Gaucher's disease, as already mentioned, has also been linked to activity in sarcoidosis (135, 176, 273).

Comparing BAL and PBMC GSEA results, the 'Defense Response' gene set is seen to be enriched in both BAL and PBMC, and assessment of the overlapping genes involved in this set in both sampling modalities again highlight CCR2, and IL10RB (Figure 5.11). CXCL10 (IP-10) and CXCL9 were also identified in the overlapping list between PBMC and BAL cell GSEA output. CXCL10 is a chemokine secreted by several cell types including monocytes, endothelial cells and fibroblasts in response to IFN-γ. It stimulates the directional migration of activated T lymphocytes and has been shown to play a crucial role in Th1 cell migration in the lungs of patients affected by sarcoidosis (64). In addition to this, a positive relationship between BAL levels of CXCL10 and the degree of CD4 alveolitis denoting active disease,

enforces its likely role in sarcoid CD4 T cell activation and the development of sarcoid granuloma and active disease (64). It's potential role as a marker of disease activity has also previously been reported (274) (234). This is important as it may suggest a peripheral marker of disease activity, reflecting the processes occurring in the lung at the site of activity. CXCL9 (Chemokine [C-X-C motif] ligand 9) is a small cytokine belonging to the CXC chemokine family that is also known as Monokine induced by gamma interferon MIG). CXCL9 is a T-cell chemoattractant, which is induced by IFN gamma and also fits with the theory of recruitment of cells from the periphery to the lung in ongoing active states of disease.

Finally, from the PBMC analysis alone, leading edge analysis of the top ten gene sets has revealed IDO1 as the top gene. IDO1 has a role in a variety of pathophysiological processes such as antimicrobial and antitumor defense, neuropathology, immunoregulation, and antioxidant activity. It is expressed on dendritic cells, monocytes, and macrophages, and is an enzyme thought to modulate T-cell behavior by catabolising the essential amino acid tryptophan. Tryptophan catabolism is believed to be a powerful mechanism of peripheral immune tolerance, and maintains homeostasis by preventing the autoimmunity or immunopathology that would result from uncontrolled and overreacting immune responses. Increased IDO-1 expression may protect the 'trigger' or initial pathogen from immune clearance, by inducing tolerance in the catabolisation of Tryptophan. Finally GO processes revealed by Metacore analysis of PBMC data identified several immune related processes as prominent in the dataset. The top GO process was 'Immune system process', and found to be enriched in this set was Annexin 11. This has also previously been attributed to susceptibility in sarcoidosis (43).

The limitations of this study include the fact that the numbers of patients assigned to each level of activity are small, there is as yet no validation to these microarray findings, and although the patients are pulmonary disease dominant, the phenotypes have not been further categorized to allow for fibrotic versus non-fibrotic in these cohorts. The strengths of this study are enhanced by good patient selection, with phenotypical grouping allowing the cohort to be relatively homogeneous from an organ involvement perspective, and the exclusion of confounding factors, such as inflammatory comorbidities.

5.6 Conclusion

This chapter shows that SCAS levels positively correlate with genes that have established links to disease activity, strengthening the role of SCAS as a marker of disease activity. By comparing gene expression in cells derived from BAL and blood, it also highlights several genes that could be used as markers of disease activity for example CXCL9 and CXCL10; and the importance of monocytes to the disease activity. Additionally, genes which have not previously been linked to disease activity, in particular CCR2, CCL18 and TREM2, have been identified, and may provide a platform for future analysis in the pathogenesis and promotion of disease activity in pulmonary sarcoidosis. Detail of future planned experiments is shown in future work (section 7.3).

6 Investigation of the processes, pathways and genes involved in fibrotic sarcoidosis.

6.1 Introduction

As discussed in previous chapters, sarcoidosis is known to be an immunological disorder centring around the dysregulation of T cell proliferation and activity resulting in the formation of granuloma. These granulomata can be deposited throughout the body but more than ninety percent of patients will have pulmonary involvement at some point in their clinical course. Patients with pulmonary sarcoidosis generally have a good prognosis, but a subset of patients have severe disease resulting in lung fibrosis. It is unclear what causes or drives this fibrosis, whether the fibrotic processes are active from the outset in pre-disposed individuals, or whether persistent granulomatous inflammation promotes a pro-fibrotic state in response. Late in the disease course, tissue samples often don't reveal granulomas (162), yet the fibrotic changes are often found along bronchovascular bundles and septal lines, paralleling the distribution of granulomatous inflammation. Histopathologically, the fibrotic changes are seen to begin at the periphery of granuloma, and extend centrally (275). There has been little progress in development of new drugs for this subset of patients, and when they are treated with currently available medicines, it is unclear when treatment should be stopped.

Due to the lack of objective markers for active disease, it has also been difficult to ascertain which patients with fibrosis have ongoing disease activity, potentially

putting them at risk of further progression of their fibrosis and providing the impetus to treat this group more aggressively. It is possible that targeted immunosuppression in those with signs of both fibrosis and active granulomatous inflammation, may alter the long-term disease trajectory and associated clinical outcomes, but this has yet to be proven, as first consensus needs to be able to identify such patients for trials.

Several studies have tried to investigate the gene profiles of sarcoidosis patients, and in a previous study by our research group (Lockstone et al (169)), and from which this work builds on, it was noted that gene expression patterns were distinct between patients with self limiting disease, and in samples from patient with active progressive disease.

Zhou et al analysed the differential gene expression from the peripheral blood of patients with sarcoidosis versus healthy controls, establishing a 20 gene signature that distinguished the two groups, but also distinguished complicated sarcoidosis, defined as pulmonary involvement with FVC < 50% (suggesting fibrosis), cardiac, or neurologic sarcoidosis from uncomplicated sarcoidosis (228). Genes identified for this 20 gene signature are shown in Table 6.1.

HBEGF	LOC100287290	CRIP1	FITM2
SAP30	KKBP1A	CX3CR1	MEI1
LOC100132356	RBM12B	APOBEC3D	ZNF671
SERTAD1	ZNF614	SESN3	TSHZ2
KIAA1147	KLRB1	NOG	ZNF540

Table 6.1 20 Gene signature published by Zhou et al distinguishing complicated sarcoidosis from healthy controls and uncomplicated sarcoidosis.

Crouser et al (168) investigated the differential gene expression in the lung tissue of treatment naïve sarcoidosis patients at the time of diagnosis to that of healthy controls. The patients in the sarcoidosis cohort were not phenotyped, and we do not know what their disease course has been subsequent to this study. Furthermore, a selection of genes and proteins that were identified to be over-expressed in the sarcoidosis cohort were analysed in bronchoalveolar lavage samples in an independent series of sarcoidosis patients and controls. Their findings focus on ADAMDEC-1 and MMP-12, having shown for these genes >25 fold greater expression in sarcoidosis lung tissue samples. They were also increased in the bronchoalveolar lavage fluid (BALF) of patients with sarcoidosis. In addition to this, it was noted that when sarcoidosis patients were stratified into severe and non-severe phenotypes, based on whether there was an indication for systemic treatment at the time of bronchoscopy or any time during follow-up (In brief, the decision to treat for pulmonary disease was predicated upon the demonstration of deteriorating pulmonary physiology (i.e., 10% decline in serial FVC) or substantial pulmonary symptoms in the presence of significantly abnormal pulmonary function tests. All but one of the patients in the severe group had advanced Scadding chest X-ray (CXR) stages (II–IV)), the levels of both ADAMDEC-1 and MMP-12 in BALF appeared to correlate with increasing severity of disease.

Sarcoidosis traditionally has been described as a Th1 mediated disease (56, 276), the cytokine milieu of which promotes classical pro-inflammatory M1 macrophage polarisation and activation. However, alternative activation of macrophages to an M2 polarisation has been demonstrated in fibrotic lung diseases including fibrotic sarcoidosis (82), and this occurs in the setting of a Th2 cytokine milieu.

Macrophages are obviously key in the formation of the sarcoid granuloma, and may play a pivotal role in the transition from an inflammatory to a fibrotic state in terms of their M1 or M2 state of polarisiation (277, 278).

The clinical heterogeneity of sarcoidosis is such that studying patients collectively is likely to obscure the detection of signals important for specific phenotypes, such as fibrotic sarcoidosis. In this thesis, I first examined the use of high resolution CT scans to measure disease activity, then a set of readily available blood markers and radiographic score as a surrogate for this. Using these indices of activity, I have identified patients with active fibrosis from the microarray cohort and have used a transcriptomic approach to determine the genes and immunological pathways found in active fibrotic but not in active non-fibrotic patients. In addition to this, I have also investigated the overlap between gene signatures for Idiopathic Pulmonary Fibrosis in the active fibrotic sarcoidosis cohort, dominance of M1 and M2 gene signatures, Th1/Th2/Th17 patterns of expression, and upregulation of a chemokine profile, particularly between lung and the periphery, in these cohorts.

6.2 Aim

To assess genes, pathways and processes differentially expressed in patients with active fibrotic sarcoidosis, compared to active non-fibrotic disease, to further understand the mechanisms involved in promoting fibrosis.

6.3 Methods

6.3.1 Patient selection

Sequential patients with a diagnosis of pulmonary sarcoidosis, as defined by WASOG criteria (120), and as described in Chapter 4, were recruited for the study over a defined 2 year period. All patients had lung involvement. Concomitant cutaneous or eye involvement was also permitted, any other systemic involvement either came later or did not require treatment. Patients without lung parenchymal involvement on high resolution CT scanning, those on immune suppressive therapy at the point of sampling, those with evidence of other concomitant active inflammatory disease, smokers, and those on ACE inhibitor medication were excluded.

Of those labelled fibrotic sarcoidosis, all had evidence of either perihilar fibrosis, or more extensive fibrosis with a mid and upper zone predominance.

6.3.2 Experimental design

Serum ACE, IgG and lymphocyte count, chest radiograph, peripheral blood mononuclear cell (PBMC) isolation, bronchoalveolar lavage, and transbronchial biopsy were performed within 90 days of each other.

Blood for PBMC isolation was collected directly into tubes containing sodium heparin, and PBMC extraction using standard density gradient centrifugation was performed within 3 hours of collection (methods section 2.3.1). PBMCs were then stored in FCS (with 10% DMSO) in liquid nitrogen, until batched RNA extraction.

Bronchoalveolar lavage (BAL) samples were collected into polypropylene tubes. BAL was passed through a cell strainer within 3 hours of collection, BAL fluid was removed, and cells were re-suspended in FCS (with 10% DMSO) and stored as per PBMCs in liquid nitrogen until RNA extraction (methods section 2.3.2). Transbronchial lung tissue was collected at the time of bronchoscopy and immediately placed in to RNA later solution (Thermo Fisher Scientific), an aqueous, nontoxic tissue storage reagent that rapidly permeates tissues to stabilize and protect cellular RNA. Tissue samples were then homogenized manually within 24 hours of sample collection, using glass homogenisers, in RLT buffer (Qiagen) and RNA extraction performed.

RNA from PBMC BAL cells, and Transbronchial lung biopsy tissue (TBB) were obtained using the RNeasy mini kit (Qiagen) according to manufacturer's instructions (described in methods sections 2.3.8). RNA quantity was assessed on a NanoDrop photospectrometer and quality assessed using a high-resolution electrophoresis bioanalyser system (Agilent Technologies). All RNA was stored at -80°C. RNA samples were normalized for concentration (300ng in 11µl of RNAse free water) in a 96 well plate. cRNA synthesis and gene chip hybridisation was performed by Dr. Karene Argoud at the Wellcome Trust Centre for Human Genetics, in a single batch, using an Illumina HumanHT-12 v4 Expression Bead Chip Kit.

The SCAS for each patient was calculated from the radiological and serological information as described in Chapter 4. Demographics for patients were also collected (Table 6.3).

6.3.3 Quality control of microarray samples

RNA quantity and quality (260/280 readings) was initially screened using the nanodrop 2000 spectrophotometer (Thermo Scientific). RNA quality was then further assessed through the "RNA integrity number" (RIN) of the RNA samples. RIN is an algorithm that evaluates the electrophoretic trace of RNA and determines the RNA integrity. A score of 1 to 10 is assigned to the sample, where 10 represents completely intact RNA. A stringent cut-off of RIN >7 was employed to ensure all samples put forward for whole genome gene expression assays were of high quality (Table 6.2).

Lung Sample ID	RIN	PBMC Sample ID	RIN	BAL Sample ID	RIN
JR412 Lung	7.3	RB416 Blood	9.7	ML649 Bal	8.2
CW429 Lung	8.3	HBJ466 Blood	9.5	HBJ466 Bal	9.2
JS477 Lung	8.8	SD586 Blood	9.3	SH895 Bal	8.7
KP114 Lung	8.4	ML649 Blood	9.5	KR423 Bal	9.4
BA284 Lung	7.2	MR568 Blood	9.2	JB853 Bal	9.3
JB853 Lung	8.0	BA284 Blood	9.6	PW109 Bal	8.8
JB433 Lung	7.6	PT447 Blood	9.1	CS468 Bal	8.7
LH335 Lung	7.3	GP504 Blood	9.3	DM118 Bal	9.0
RH438 Lung	7.8	PW109 Blood	9.1	MG636 Bal	9.3
GP504 Lung	7.8	PO795 Blood	9.4	LH335 Bal	9.2
TL305 Lung	7.0	MG636 Blood	9.3	AS163 Bal	9.1
VH481 Lung	9.0	JB853 Blood	9.6	KC465 Bal	8.7
PT447 Lung	8.8	JE725 Blood	8.7		
CS468 Lung	9.0	BM422 Blood	9.6		
TL710 Lung	8.0	MB134 Blood	9.5		
		VH481 Blood	9.4		
		DB420 Blood	9.4		
		KR423 Blood	9.3		
		PB453 Blood	9.5		
		DM118 Blood	9.3		
		PW849 Blood	9.9		
		PH529 Blood	9.3		
		RT150 Blood	9.3		
		KC465 Blood	9.6		
		JM453 Blood	9.3		
		RG306 Blood	9.3		
		RC685 Blood	9.2		
		KD467 Blood	9.1		
		JB433 Blood	9.0		

Table 6.2 RIN numbers of samples in each modality entered into gene expression profiling.

Standard quality control steps, which consist of correlation between mean gene intensity in each group, assessment for systematic difference in the distribution of fluorescence intensity on the gene chips used, and batch effect testing, were also performed. After these initial quality control steps, the hybridization process was completed, the resultant data was assessed and outliers were removed based on 'sample distance to centre'. Data was then normalised using Robust Multi-array Average Normalisation (RMA). Finally, a mean intensity was computed for each probeset.

6.3.4 Data analysis of gene expression profiles

Transbronchial Lung Biopsies, Bronchoalveolar lavage (BAL) cells, and Peripheral Blood mononuclear cells of patients with two sarcoidosis phenotypes:- active fibrotic and active non-fibrotic, were compared in a gene expression profile study. Principal component analysis, a technique used to emphasize variation and bring out strong patterns in a dataset, was used to statistically verify the association between differentially expressed genes and the two phenotypes analysed. Pathway and functional analysis by Gene Set Enrichment Analysis software was performed to identify genes involved in the different phenotypes.

Using a-priori analysis, differential expression of each gene between the two groups, fibrosis v no fibrosis, was assessed in TMEV (171), and further analysed using Gene-Set Enrichment Analysis (GSEA) (www.broad.mit.edu/gsea/). GSEA tests the biological meaning of a ranked list of all genes in a sample (231), this has been described in Chapter 2 and Chapter 5 in detail. The ranked list is ordered such that genes showing higher expression in one phenotype (in this case, active fibrotic

disease) are at the top, and those showing higher expression in the other phenotype (active non-fibrotic disease) are at the bottom, and unchanged genes are in the middle. Pre-defined gene-sets are then tested for enrichment in one of the phenotypes, and the significance of the enrichment is assessed by a permutation method. In this study, GSEA was used in Pre-Ranked mode (231), and the list of genes was ranked according to fold change of the differential expression between the two groups. Gene-sets comprising all biological process categories (C5bp) defined by the Gene Ontology Consortium (232) and based on GO terms (www.geneontology.org) and their associations to human genes, were tested. In addition to this, gene sets were established from various GEO sources to assess M1 and M2 phenotype in the two groups, Th1, Th2 and Th17 predominance, overlap with genes specific to IPF (Appendix 8.4), and cytokines/chemokines (Appendix 8.5).

These analyses were performed in conjunction with Dr Eleni Giannoulatou and Dr Steven Taylor at the Computational Biology Research Group, Weatherall Institute of Molecular Medicine, and Dr. Helen Lockstone, Bioinformatics and Statistical Genetics Core, Oxford Genomics Centre, Wellcome Trust Centre for Human Genetics.

6.3.5 Data Interpretation

In Gene Set Enrichment Analysis, the enrichment score (ES) reflects the degree to which a gene set is overrepresented at the top or bottom of a ranked list of genes. A positive ES indicates gene set enrichment at the top of the ranked list; a negative ES indicates gene set enrichment at the bottom of the ranked list. The normalised

enrichment score (NES) is the primary statistic for examining gene set enrichment results. It normalises the enrichment score for gene set size and correlations between gene sets and the expression dataset. The NES can be used to compare analysis results across gene sets. The nominal p value estimates the statistical significance of the enrichment score for a single gene set. The false discovery rate (FDR) is the estimated probability that a gene set with a given NES represents a false positive finding. An FDR of <0.25 indicates that the result is likely to be valid 3 out of 4 times. The “leading-edge subset” in a gene set are those genes that account for the gene set enrichment signal. Leading edge analysis is a further function of the GSEA software, which allows comparison of the leading edge genes from different gene sets, and visualisation of the overlap between the selected leading edge subsets (i.e. how many of the analysed leading edge subsets does a single gene appear in).

6.4 Results

6.4.1 Patient demographics

Sampling modality	Fibrosis TBB	Non-Fibrosis TBB	Fibrosis PBMC	Non-Fibrosis PBMC	Fibrosis BAL Cell	Non-Fibrosis BAL Cell
n	3	12	5	14	3	5
Age range (years)	47-61	34-71	39-59	34-69	45-56	36-69
Age mean [S.D.]	54 (7)	48 (13.3)	49 (7)	48 (10)	49 (6)	54 (12)
SCAS mean [S.D.]	10 (1)	6.5 (2.9)	9.2 (1.1)	8.2 (3)	9.7 (0.6)	8.6 (1.8)
FEV1 (% Predicted) Range	41-77	72-124	65-109	71-124	77-89	84-124
FEV1 (% Predicted) Mean [S.D.]	63 (20)	92 (14.7)	84 (16.1)	95 (17.4)	83 (6)	98 (15.5)
FVC (% Predicted) Range	57-86	86-132	57-115	72-132	57-93	86-135
FVC (% Predicted) Mean [S.D.]	72 (15)	104 (14.2)	90 (23.2)	103 (19.1)	78 (18.7)	109 (19)
TLCO (% Predicted) Range	47-89	76-107	62-93	62-110	63-89	78-97
TLCO Mean [S.D.]	62 (23)	91 (9.7)	77 (14.3)	86 (13.4)	80 (14.7)	87 (7.6)
Gender M	0	8	4	7	1	3
Gender F	3	4	1	7	2	2
Ethnicity - Afro-Caribbean	0	0	0	2	0	1
Ethnicity - Asian	0	0	0	1	0	0
Ethnicity - Caucasian	3	12	5	11	3	4

Table 6.3 Patient Demographics for active fibrotic and active non-fibrotic patients in each sampling modality (lung, PBMC and BAL).

6.4.2 Analysis of pathways, processes and genes contributing to active fibrotic sarcoidosis in lung samples

6.4.2.1 Principal Component Analysis

Principal Component Analysis (PCA) of the samples based on gene expression was performed in TMEV (171). The active lung fibrotic patients are seen to cluster together well, based on their total gene expression profiles (Figure 6.1).

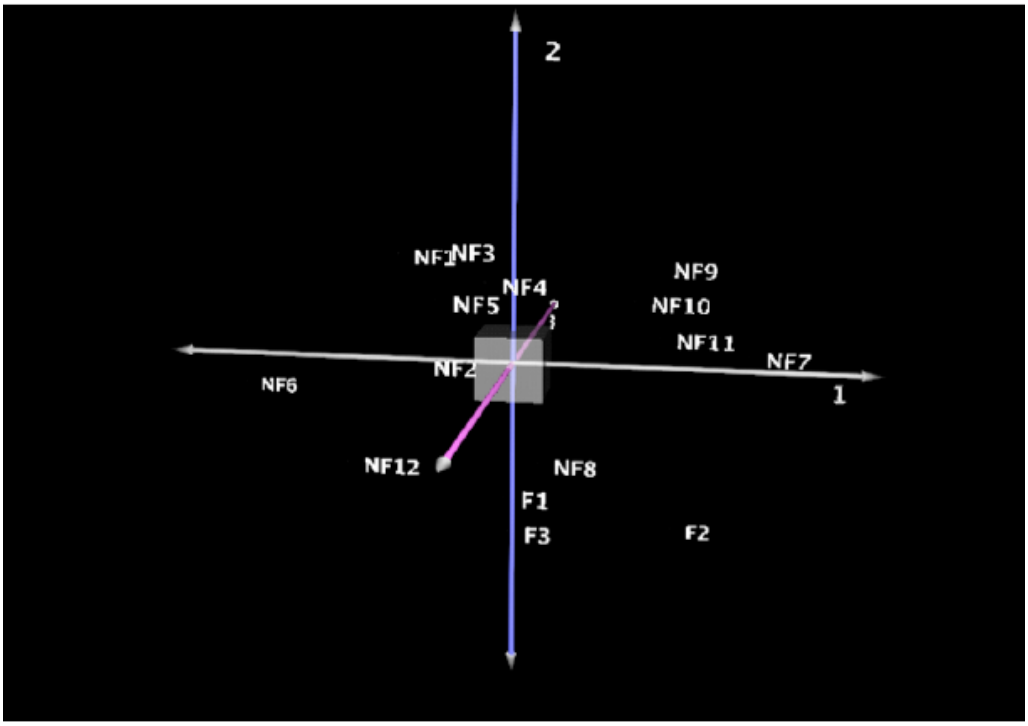


Figure 6.1 Principal Component Analysis of gene expression in lung tissue for fibrotic versus non-fibrotic samples. This shows strong similarities between the three active fibrotic patients (F1-3) in the cohort.

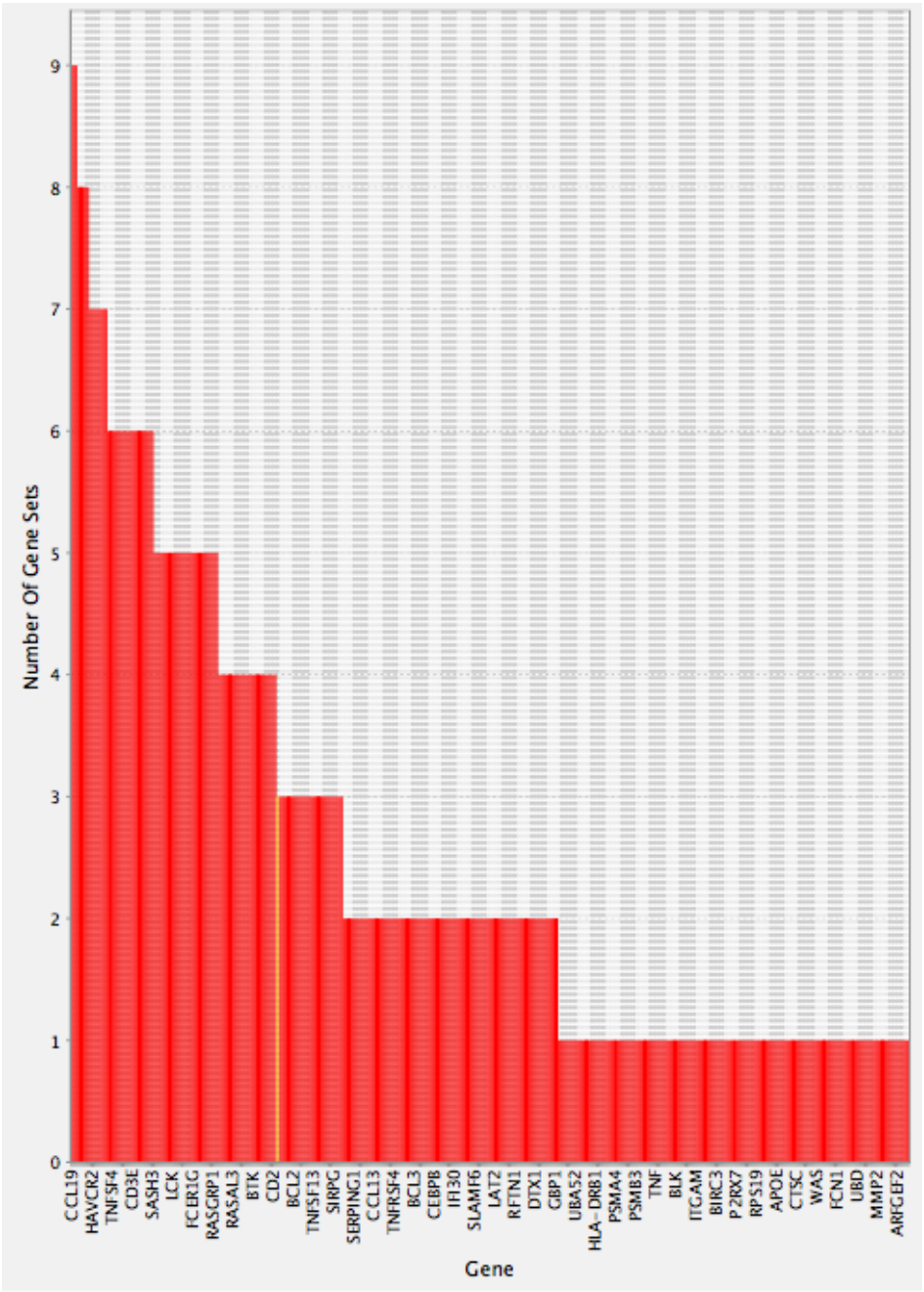
This clustering may simply reflect the underlying active fibrotic process that is occurring specifically in this group, but this in itself may provide information on the processes involved and potential targets for treatment to slow down or prevent fibrosis. Note is made however of the non-fibrotic patient (NF8) that also clusters with this group. This patient has complex sarcoidosis with cardiac involvement. At the time of sampling he did not have features of fibrotic lung disease.

6.4.2.2 Gene Set Enrichment Analysis

GSEA, again using C5bp GO terms (all biological processes), was employed on a pre-ranked gene list to further investigate which gene sets showed enrichment in the active fibrotic patients in comparison to the active non-fibrotic patients. The gene list was ranked according to fold change of the differential expression between the two groups weighted towards the active fibrosis group. In total 21334 genes were analysed. 1904 of 3085 gene sets were enriched in the active fibrotic group, of which 1034 gene sets were significant with an FDR <25%. Table 6.4 shows the top ten positively enriched gene sets (the description for the definition of nominal p value, ES, NES and FDR is given in section 6.3.5).

No	Gene Set	ES	NES	NOM p-val	FDR q-val
1	Adaptive Immune Response	0.61	3.03	0.0	0.0
2	Response to Interferon gamma	0.61	2.87	0.0	0.0
3	Positive Regulation of Cell Activation	0.56	2.84	0.0	0.0
4	Cellular Response to Interferon gamma	0.61	2.80	0.0	0.0
5	Lymphocyte Co-stimulation	0.67	2.78	0.0	0.0
6	Regulation of Cell Activation	0.52	2.77	0.0	0.0
7	Positive Regulation of Cell Activation	0.63	2.73	0.0	0.0
8	Interferon gamma Mediated Signalling Pathway	0.65	2.73	0.0	0.0
9	Activation of Immune Response	0.52	2.72	0.0	0.0
10	Innate Immune Response	0.50	2.72	0.0	0.0

Table 6.4 Top ten gene sets positively enriched in lung samples from patients with active fibrotic sarcoidosis compared to active non-fibrotic sarcoidosis.



CCL19	ICOSLG
CD86	SLAMF1
IFN γ	CD28
HLA DRB4	CD3E
HLA DRB3	CCR2
HLA DPA1	SYK
HLA DQB1	AIF1
HLA DQA1	CD40LG
HAVCR2	CD80
RIPK2	CARD11
HLA DQB2	VCAM1
IL12RB1	ZAP70
HLADRA	ZP3
CCL5	SASH3
HLA-A	NCKAPIL
CD4	JAK3
PRKCD	HLA DMB
LYN	CCL21
TNFSF4	ADAM8
CTLA4	BTLA
JAK2	TNFRSF14
PTAFR	PTPRC
PYCARD	NCK1
CSK	LCK
IL23R	SRC

Figure 6.2 Leading edge analysis of top ten enriched gene sets in GSEA of Lung tissue gene expression ranked according to fold change of differential expression in the active fibrotic group compared to the active non-fibrotic group. Every 12th gene, ranked by the number of gene sets they are found in, is shown on the x-axis. The top 50 genes are shown alongside the histogram.

CCL19 is the most frequent gene contributing to nine of the top ten gene sets. Of note CCR2 and CCL5 again feature, as they did in the activity chapter. There is a

difference in mean SCAS noted between the two groups of Fibrosis and no Fibrosis for the lung sample cohort, which may confound results, but CCL19 did not feature in the activity genes which goes against this as a confounder.

On reviewing the gene sets that were significantly downregulated in the fibrotic group, they involve cilium morphogenesis, movement and organisation, the significance of which is not clear, but may again be a feature of fibrosis and distorted architecture, and has previously been reported in IPF gene expression analysis (279).

To focus more on the genes that showed a differential expression of at least +/- 2 fold change (FC), these were extracted out from the list and further analysed. The top 50 upregulated genes in active fibrotic sarcoidosis are displayed in Figure 6.3 alongside the heatmap displaying supervised clustering of the fibrotic and non-fibrotic samples.

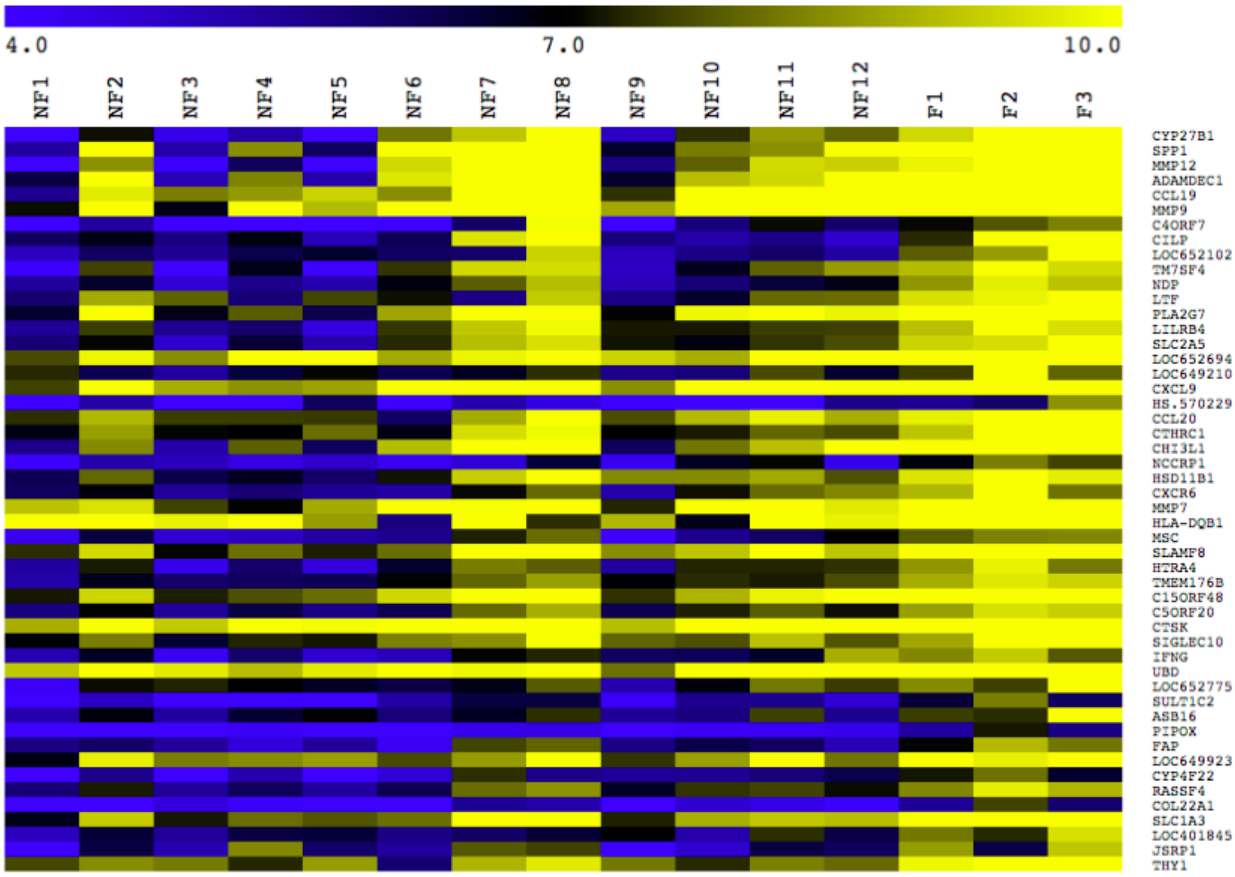


Figure 6.3 Heatmap displaying top 50 upregulated differentially expressed genes in the active fibrosis cohort in lung samples. Note ADAMDEC1 and MMP12 are 3rd and 4th in the list. (F1-3 Active Fibrosis samples, NF1-12 Active Non Fibrosis samples).

Genes that have been previously identified from other research groups as being involved in complex sarcoidosis, but not specifically fibrotic pulmonary sarcoidosis – in particular ADAMDEC1 and MMP12 (168) are again seen (Figure 6.4).

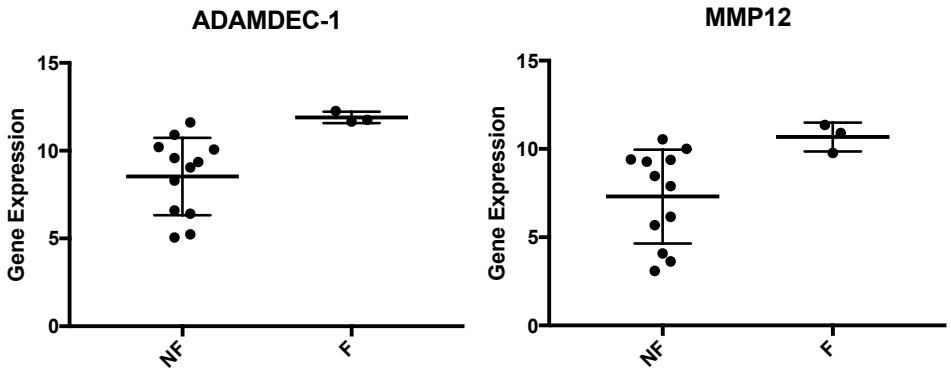


Figure 6.4 Gene expression data from lung samples plotted for ADAMDEC1 (Mann-Whitney test p = 0.004) and MMP12 (Mann-Whitney test p = 0.018) displaying the difference between active fibrosis (F) and active no fibrosis (NF).

The top 50 downregulated differentially expressed genes are displayed in the heatmap below (Figure 6.5)

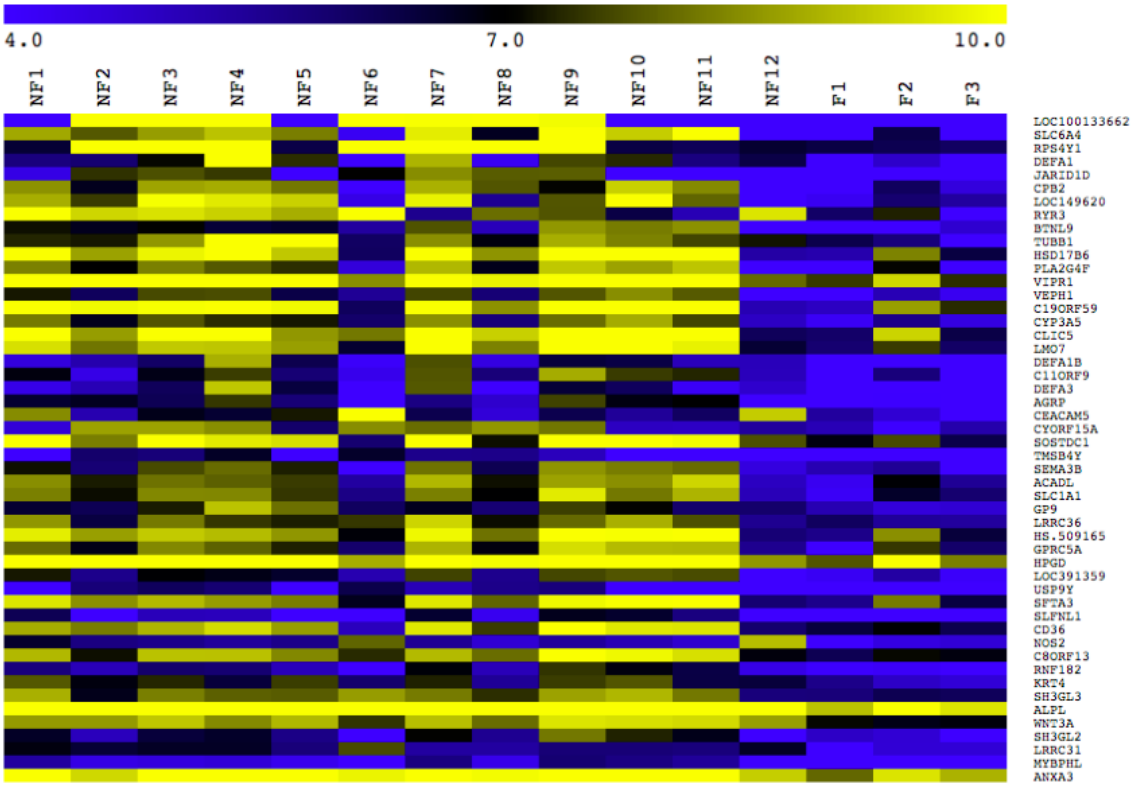


Figure 6.5 Heatmap displaying top 50 down-regulated differentially expressed genes in the active fibrosis cohort in lung samples. (F1-3 Active Fibrosis samples, NF1-12 Active Non Fibrosis samples).

Again, and not unexpectedly, the most prominent gene sets in this subset of >2FC or <-2FC differentially expressed genes were immune related. Heat maps for the genes involved in the top 5 upregulated gene sets are shown in Figure 6.6.

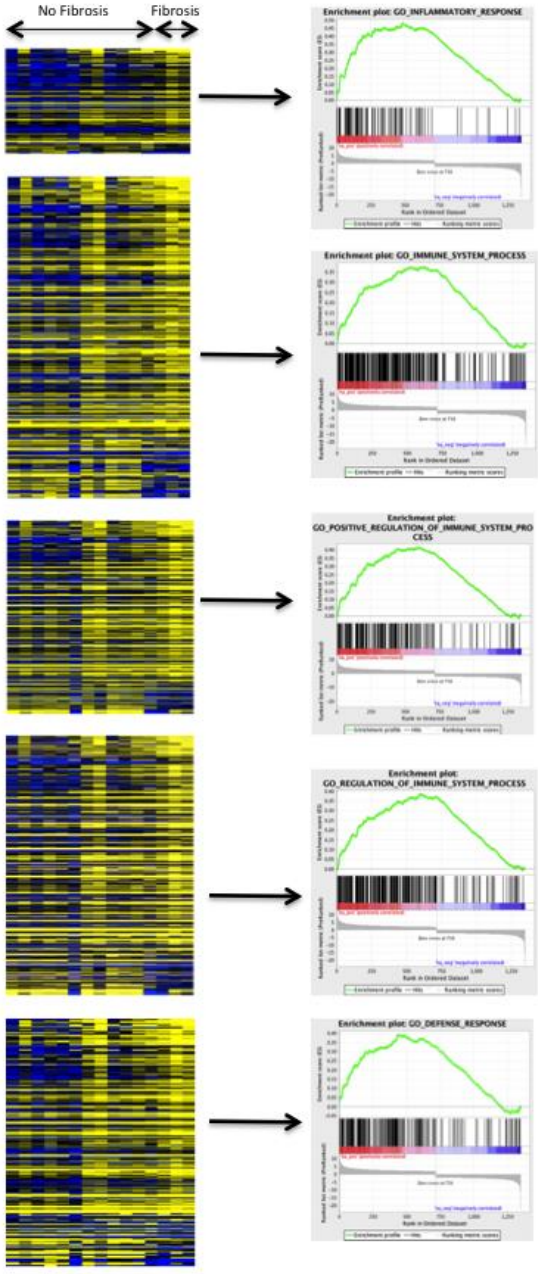


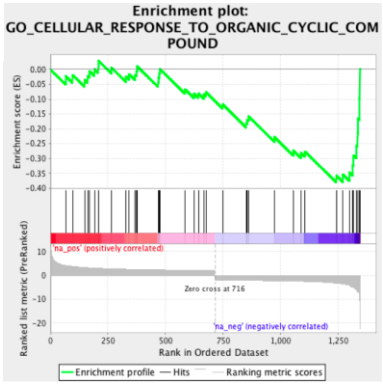
Figure 6.6 Top 5 Gene sets in >2FC or <-2FC differentially expressed gene list in active fibrotic sarcoidosis (n=3) when compared to active non-fibrotic sarcoidosis (n=12). Yellow – high expression, Blue – low expression. For the enrichment plots the left hand leading edge reflects those genes highly expressed in the active fibrotic cohort (i.e. those displayed in yellow on the heatmap).

Only two gene sets were significant in the down regulated list (with FDR q <0.25), and they were 'Cellular Response to Organic Cyclic Compound' (ES-0.38, NES-2.21, Nom p 0.00, FDRq 0.22), and 'Peptide Metabolic Process' (ES -0.51, NES -

A

2.11, Nom p 0.00, FDRq 0.20) GO gene sets. Enrichment plots for these gene sets are shown in (Figure 6.7A and B respectively), with a list of the genes involved in core enrichment in these sets, and their ranking in the entire gene list, shown in Table 6.5.

A



B

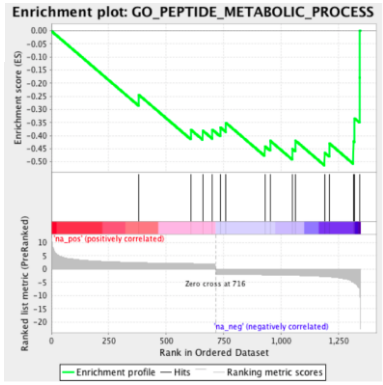


Figure 6.7 Enrichment plots showing the two significantly negatively enriched gene sets in the active fibrotic cohort from lung samples.

GENE **Cellular Response to Organic Cyclic Compound**	**RANK**	**GENE** **Peptide Metabolic Process**	**RANK**
HMGCS2	1243	GGTLC1	1191
ALPL	1272	MME	1212
ANKRD1	1300	RPS4Y2	1316
DEFA3	1315	CLIC5	1322
DEFA1B	1318	EIF1AY	1343
PLA2G4F	1331	RPS4Y1	1345
AGER	1336		
RYR3	1338		
DEFA1	1344		
SLC6A4	1346		

Table 6.5 Genes contributing to the core enrichment of the two significantly downregulated gene sets in the lung samples from active fibrotic v active non-fibrotic patients and their ranking in the entire gene list (of 1348 genes).

Idiopathic Pulmonary Fibrosis (IPF) is another interstitial lung disease that has been the focus of much recent research. Like Sarcoidosis, its aetiology is unknown, and both diseases can result in significant morbidity and mortality with end-organ lung damage secondary to fibrosis. Clinical microarray datasets are available from the Gene Expression Omnibus (GEO) published by other groups (http://www.ncbi.nlm.nih.gov/projects/geo/). Following extensive review of the datasets available, I selected two validated studies which paralleled my sarcoidosis

study by using peripheral blood or lung biopsy samples, in IPF patients; one by Konishi et al (GSE10667) (280), and one by Meltzer et al (GSE24206) (281). Alongside these, a gene signature generated from a meta-analysis of several GSE datasets of IPF samples (279), was also used, to determine if previously defined gene signatures for IPF, consisting of either significantly up or downregulated expression in biopsies from IPF patients compared to normal, were also upregulated in the active fibrotic sarcoidosis cohort, compared to the active non-fibrotic sarcoidosis cohort. The enrichment plots, their associated enrichment scores and FDR q values, and a summary of the top 20 genes involved in the leading edge of each enrichment plot is displayed in Figure 6.8 and Table 6.6 respectively.

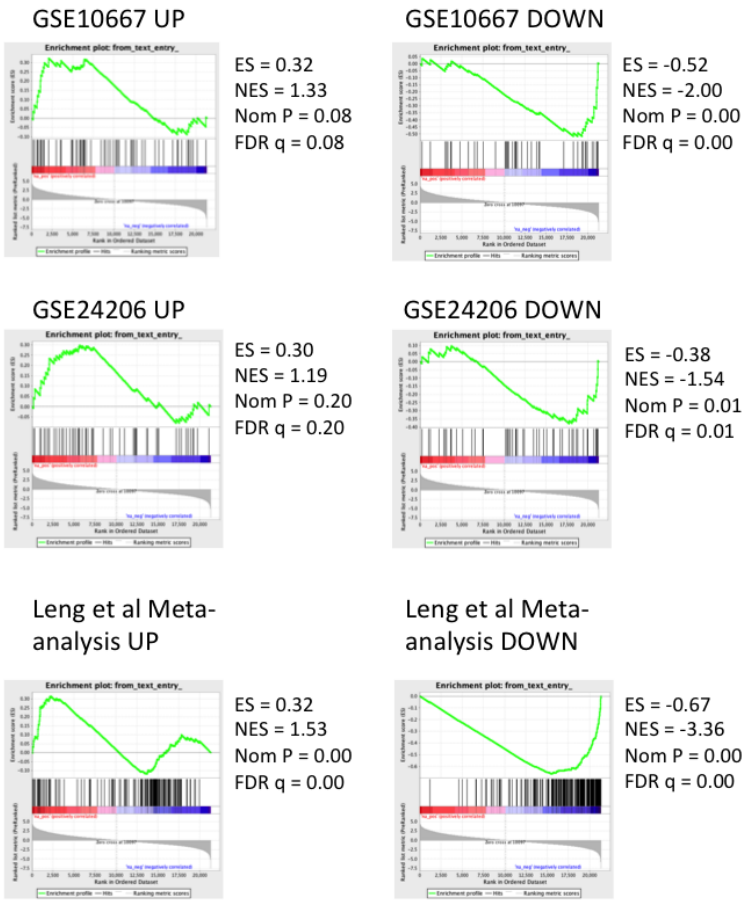


Figure 6.8 IPF signatures obtained from GEO from various sources and their enrichment in the active fibrotic sarcoidosis lung cohort. The left hand side of the enrichment plot represents upregulation in the active fibrotic group.

GSE10667 UP	GSE10667 DOWN	GSE24206 UP	GSE24206 DOWN	Leng et al UP	Leng et al DOWN
COL8A2	BCL6	COL8A2	MT1M	THY1	HEY1
RGS10	TMTC1	RGS10	SIK1	LTF	SEMA5A
COL1A1	AFF2	HLA-DOA	BCL6	IGF1	RNF128
COL3A1	ZNF33A	CDCA7	AXUD1	TDO2	KLF4
ASPM	PFKFB2	ENC1	PIGA	NPTX2	STXBP6
CDCA7	FMO5	ZSCAN2	HNRPDL	GREM1	CD55
RPS27L	SDR16C5	MDH1	TMTC1	COL1A1	FAM107A
ABCC3	PDE4D	PHLDA3	S100A12	COL5A2	EPAS1
DDB2	MAOA	FNDC1	KIAA0146	PLA2G7	PRKCZ
ENC1	IDI1	RPS15A	ZEB1	COL3A1	RAB17
TYMS	ACPP	VKORC1	SDR16C5	CXCL14	BEX2
PHLDA3	LDLR	RBM15	ERRFI1	MMP7	CDKL2
FNDC1	MTSS1	PSD3	MAOA	POSTN	TUFT1
	CEBPD	KPNB1	TACC2	CTSK	SLCO4C1
	FAM107A	LOC100131096	LDLR	SULF1	LNX2
	REPS2	DNAJC4	CEBPD	CCL19	TMEM125
		VWA1	FAM107A	CILP	PIP5K1B
		ZNF562	FLJ43663	MMP11	SELENBP1
		SDHALP1		POU2AF1	MICALCL
		LAPTM4B		RARRES1	GOLT1A

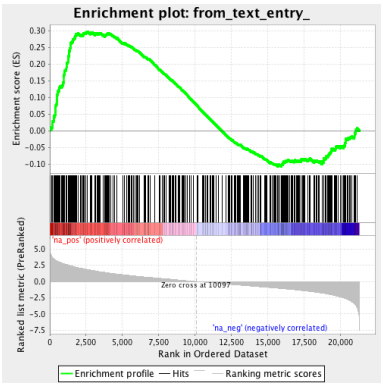
Table 6.6 Top twenty genes (or less) contributing to the leading edge of each enrichment plot from the IPF signatures obtained from GSE10667(280), GSE24206 (281) and Leng et al (279).

Various collagen related genes can be seen to be upregulated in the active fibrotic sarcoidosis group, overlapping with the IPF group, in particular COL8A2, COL1A1, COL3A1 and COL5A2. In addition, CCL19 and MMP7 are also present in the IPF signature derived from meta-analysis, and are enriched in the active fibrotic sarcoidosis group. Both of these genes had already been noted to be significantly differentially expressed between active fibrotic and active non-fibrotic lung samples in the earlier analysis.

Of further interest, in their meta-analysis, Leng et al also derived a signature to differentiate IPF from sarcoidosis (Appendices 8.6 and 8.7). Of note, the sarcoidosis patients used to define these signatures were not stratified into predominantly pulmonary sarcoidosis as they have been for this study, or note made of whether/how many of these patients had active lung fibrosis at the time of sampling. This makes it all the more interesting when the signatures are tested in the active

lung fibrosis cohort, as it is the IPF upregulated signature that is significantly enriched (ES 0.30, NES 1.57, Nom P 0.0, FDRq 0.0) in the fibrotic group (Figure 6.9A), and the IPF downregulated signature that is significantly enriched (ES -0.52, NES -2.67, Nom P 0.0, FDRq 0.0) in the non-fibrotic group (Figure 6.9B). The top twenty genes contributing to the leading edges of each enrichment plot are detailed in Table 6.7.

A



B

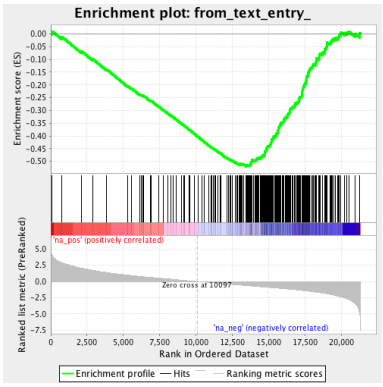


Figure 6.9 A) IPF v Sarcoidosis upregulated gene signature showing significant enrichment in the active fibrotic sarcoidosis cohort. B) IPF v Sarcoidosis downregulated gene signature showing significant enrichment in the active non-fibrotic sarcoidosis cohort.

IPF v Sarc UP	IPF v Sarc DOWN
CLEC4E	TTC12
SUCNR1	FSIP1
FCER1G	FZD3
CXCR6	POF1B
CCL5	ELMOD1
SOD2	KNDC1
CD2	DNALI1
FAP	AGBL2
RASGRP1	AKR1B10
GPR68	PPP1R14C
IL2RB	ZNF214
TLR8	GAS8
IL21R	IL20RA
TNFAIP6	FGF14
PDCD1LG2	CEACAM5
HSD11B1	GRAMD1C
IFNG	KLK10
TNFSF13B	PRSS12
LYZ	KLK11
HCK	SCNN1B

Table 6.7 Top 20 genes contributing to leading edge of enrichment plots for IPF v Sarcoidosis Signature in Active Fibrotic Sarcoidosis cohort.

It is generally thought that M1 macrophages (or classically activated), have a pro-inflammatory function, acting mainly to promote the pathology, while M2 macrophages (or alternatively activated) show regulatory activity, control inflammation and also remodel the repairing process. The balance of M1 and M2 polarisation may play a role in the development of fibrosis. Several gene sets for M1 and M2 are available via the Gene ontology consortium (http://www.ncbi.nlm.nih.gov/projects/geo/) (GSE35495, GSE18686, GSE5099, GSE35449), and were assessed in this dataset of pre-ranked active fibrotic patients v active non-fibrotic patients. M1 was seen to be more significantly upregulated than M2, and there were also common genes to both lists that actually contributed to up to half of the M2 gene signatures, suggesting an overall M1 predominance, but as previously mentioned, it may not simply be a predominance, but the balance of these macrophage phenotypes that drives the fibrosis. The enrichment plots for the aforementioned M1 and M2 gene signatures are displayed in Figure 6.10. The relevant Enrichment Scores (ES), Normalised Enrichment Scores (NES), Nominal p values and FDR q values are shown in Table 6.8.

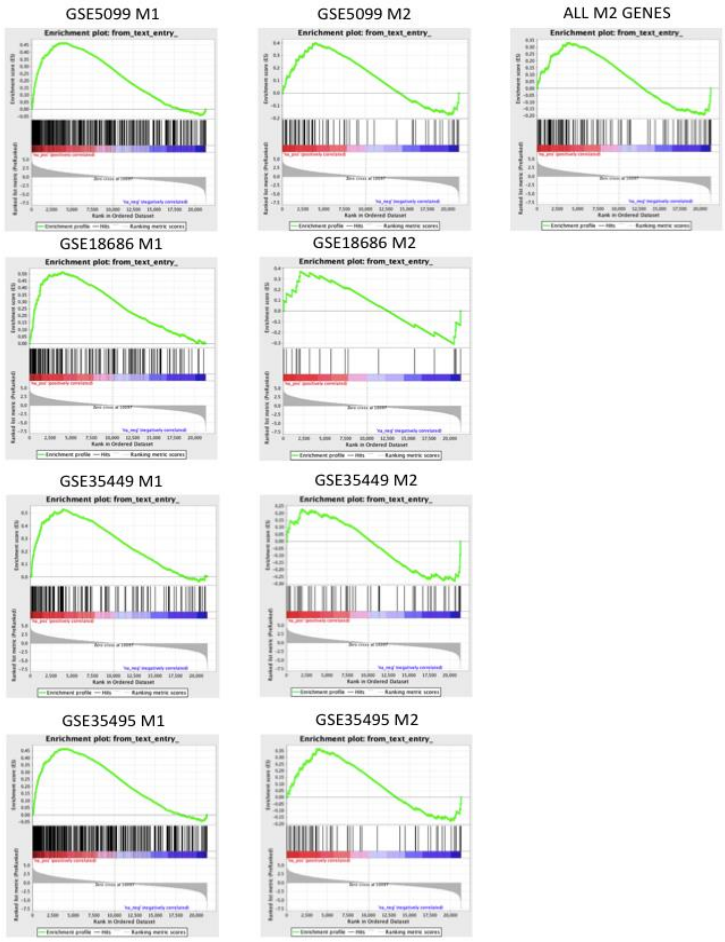


Figure 6.10 Enrichment plots for various M1 and M2 gene signatures obtained from the GEO, showing enrichment for both macrophage phenotypes, but more prominence and significance in the M1 signatures. NB 'All M2 genes' is a composite list of all genes from the four contributing GEO M2 signatures shown above.

Signature Reference	ES	NES	Nom. p	FDRq
GSE5099M1	0.47	0.47	0.00	0.00
GSE5099M2	0.40	0.40	0.00	0.00
GSE18686M1	0.51	0.51	0.00	0.00
GSE18686M2	0.37	0.37	0.30	0.30
GSE35449M1	0.53	0.53	0.00	0.00
GSE35449M2	-0.28	-0.28	0.20	0.20
GSE35495M1	0.47	0.47	0.00	0.00
GSE35495M2	0.37	0.37	0.03	0.03
ALL M2	0.33	0.33	0.01	0.01

Table 6.8 Enrichment scores, Normalised enrichment scores, nominal p values and FDRq values for each of the gene signatures shown in the enrichment plots.

Zhou's gene signature for complicated sarcoidosis, as mentioned in the introduction, was also tested in the pre-ranked list (ranked as previously mentioned) (Figure 6.11).

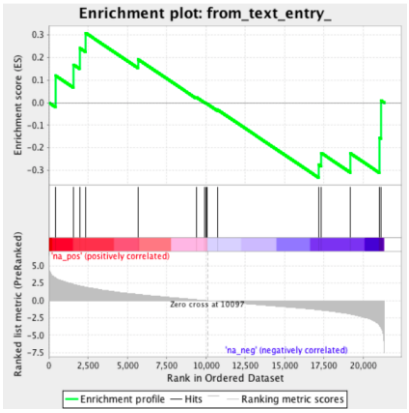
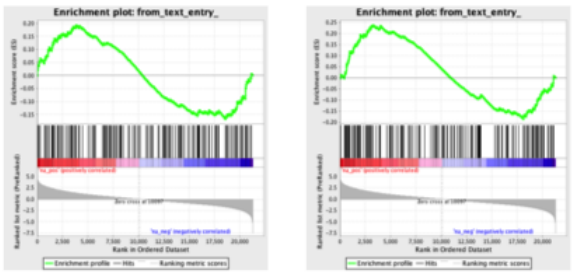


Figure 6.11 Gene signature for complicated sarcoidosis as described by Zhou et al. No enrichment is seen in the active fibrotic cohort.

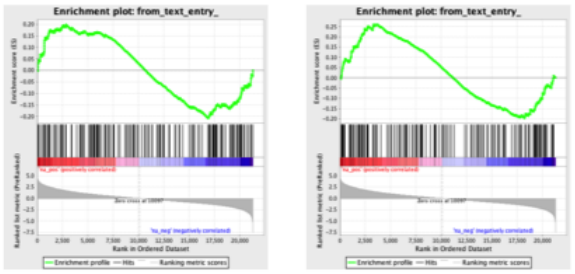
In fact only three genes were represented in the negative leading edge of the gene set, namely MEI1, KLRB1, and APOBEC3D (i.e. they were downregulated in the active fibrotic cohort). The FDRq value was not significant.

Sarcoidosis has traditionally been described as a Th1 predominant immune disorder, but again this is brought into question when the phenotype between active fibrotic and active non-fibrotic sarcoidosis is further assessed. Again, lists obtained from the Gene Expression Omnibus provide signatures for Th2 v Th1 (GSE14308), Th1 v Th2 (GSE11924), Th1 v Th17 (GSE14026) and Th2 v TH17 (GSE14308), all of which have again been assessed in this active fibrotic sarcoidosis cohort. There is no particular predominance of Th1, Th2 or Th17 gene signatures in this cohort, and no significant enrichment was observed. Previous publications looking at gene expression signatures of Th1, Th2 and Th0 have found more of a shift towards Th1 in pulmonary parenchyma, but again the sarcoidosis cohort has not been phenotyped as in this study, and may reflect the T cell milieu associated more with

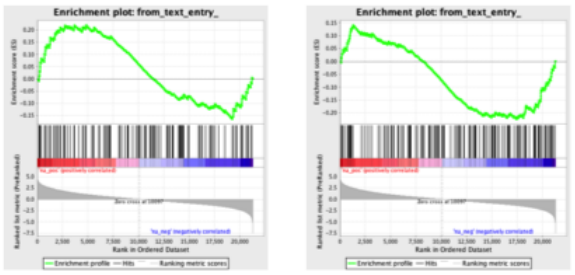
active granulomatous infiltration rather than active fibrogenesis (70). The enrichment plots are shown in Figure 6.12.



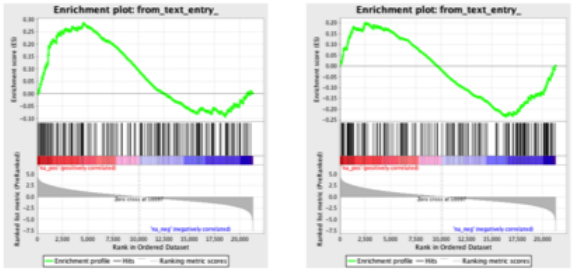
GSE11924 Th1 v Th2	ES	NES	Nom. p	FDR q
UP	0.19	0.92	0.68	0.68
DOWN	0.24	1.15	0.16	0.16



GSE14308 Th2 v Th1	ES	NES	Nom. p	FDR q
UP	-0.21	-1.01	0.43	0.43
DOWN	0.26	1.27	0.05	0.05



GSE14026 Th1 v Th17	ES	NES	Nom. p	FDR q
UP	0.22	1.02	0.38	0.38
DOWN	-0.22	-1.07	0.29	0.29



GSE14308 Th2 v Th17	ES	NES	Nom. p	FDR q
UP	0.29	1.37	0.02	0.02
DOWN	-0.24	-1.15	0.18	0.18

Figure 6.12 Enrichment plots for Th1, Th2 and Th17 gene signatures, showing no specific Th predominance in the active fibrotic sarcoidosis lung cohort. Alongside the tables show the enrichment scores, normalized enrichment scores, nominal p values and FDR q values for each signature tested.

Finally, in the lung sample modality, again ranked as before, according to fold change of the differential expression between the two groups with the highest fold change in the active fibrosis group at the top of the list, a chemokine profile was assessed, to ascertain whether there was an upregulation of chemokines in the lungs of active fibrotic patients compared to active non-fibrotic, that enhances and perpetuates a signal to the periphery and increases the homing of inflammatory cells to the lungs, thus also potentially contributing to the fibrotic process. The same chemokine profile is addressed later in the chapter in the Blood cohort. A list of chemokines (KEGG_CHEMOKINE_SIGNALING_PATHWAY) was obtained via the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/kegg1.html) (282-284), and enrichment of this set of genes was assessed as before in GSEA (Figure 6.13) (ES 0.42, NES 2.01, Nom. P 0.0, FDRq 0.0).

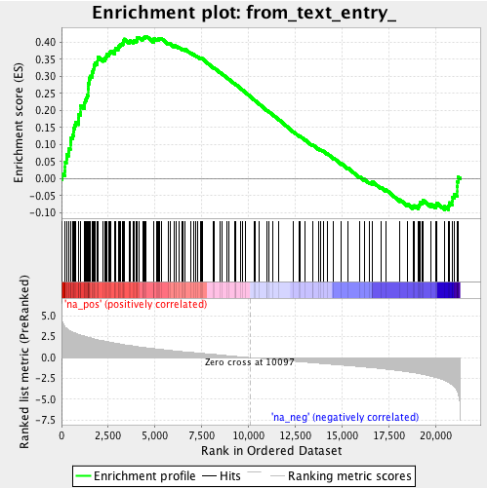


Figure 6.13 Enrichment of the chemokine signaling pathway (KEGG) gene set is demonstrated in the active fibrotic cohort compared to active no fibrosis. 70/161 genes showed core enrichment.

Seventy-four of 161 genes in this gene set were enriched in the active fibrotic cohort, and the top 50 genes and their ranking in the complete 21334 gene list is shown in Table 6.9.

GENE	RANK	GENE	RANK	GENE	RANK	GENE	RANK	GENE	RANK
TIAM2	43	HCK	510	CXCR4	1018	CCL18	1429	PLCB2	1910
CCL20	170	GRB2	602	DOCK2	1020	PRKCD	1463	CCR1	1926
CXCR6	193	NFKBIB	636	CCL4L2	1224	ADCY3	1477	GNGT2	1968
CCL5	199	RAC2	661	CCL4L1	1235	CCL3	1545	CXCL13	2230
SHC1	257	PIK3CG	687	STAT1	1250	JAK2	1632	RAP1B	2337
CXCR3	272	CXCL14	739	CXCL9	1343	FGR	1656	ADCY7	2395
NCF1	365	CCL19	899	WAS	1365	CCR7	1693	PRKX	2423
PIK3CD	406	CSK	915	CCL3L3	1410	CCL22	1719	PRKACB	2533
PTK2B	469	CCL3L1	923	ITK	1411	CCR2	1745	VAV1	2563
CXCL16	475	GRK6	1011	GNG10	1424	ARRB2	1791	LYN	2645

Table 6.9 Top 50 leading edge genes and their ranking in the complete list of 21334 genes contributing to the enrichment of the chemokine signaling pathway gene set in the active fibrotic cohort.

6.4.3 Analysis of pathways, processes and genes contributing to active fibrotic sarcoidosis in PBMC samples

6.4.3.1 Principal Component Analysis

To assess the similarity between PBMC (Blood) samples from sarcoidosis patients with both active fibrosis and active no fibrosis, as performed for the lung cohort, Principal Component Analysis (PCA) was again performed in TMEV(171).

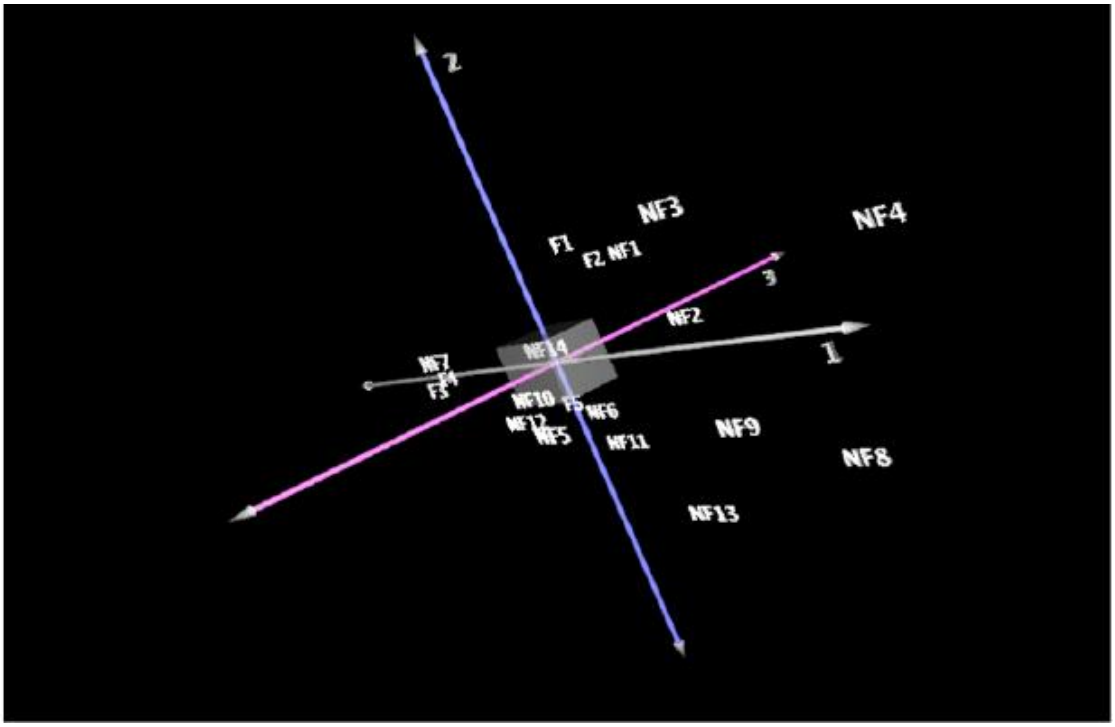


Figure 6.14 Principal Component Analysis of gene expression in PBMC for active fibrotic versus active non-fibrotic samples. (F1-5 Active Fibrosis; NF1-14 Active No Fibrosis)

The clustering in this cohort is not as distinct as for lung tissue samples, as shown in Figure 6.14, highlighting more marked variation in gene expression in peripheral sampling, confounded by multiple cell types of varying concentrations contributing to the gene expression profile of each person.

6.4.3.2 Gene Set Enrichment Analysis

GSEA, again initially using C5bp GO terms (all biological processes), was employed on a pre-ranked gene list to further investigate which gene sets showed enrichment in the active fibrotic patients in comparison to the active non-fibrotic patients. The gene list was ranked as before, according to fold change of the differential expression between the two groups, such that genes showing higher expression in active fibrosis are at the top, and those showing higher expression in the active no fibrosis, are at the bottom, and unchanged genes are in the middle.

No	Gene Set	ES	NES	NOM p-val	FDR q-val
1	Antigen Processing and Presentation of Exogenous Peptide Antigen Via MHC Class I	0.55	2.44	0.00	0.00
2	Mitochondrial Respiratory Chain Complex I Assembly	0.55	2.34	0.00	0.00
3	Mitochondrial Electron Transport NADH to Ubiquinone	0.58	2.34	0.00	0.00
4	NADH Dehydrogenase Complex Assembly	0.55	2.33	0.00	0.00
5	Oxidative Phosphorylation	0.50	2.31	0.00	0.00
6	ATP Generation from ADP	0.61	2.30	0.00	0.00
7	Response to Ammonium Ion	0.59	2.30	0.00	0.00
8	Antigen Processing and Presentation of Peptide Antigen via MHC Class I	0.48	2.28	0.00	0.00
9	Mitochondrial Respiratory Chain Complex I Biogenesis	0.55	2.28	0.00	0.00
10	Detection of Biotic Stimulus	0.66	2.28	0.00	0.00

Table 6.10 Top ten gene sets upregulated in blood samples from active fibrotic patients compared to active non-fibrotic patients. ES = enrichment score, NES = Normalised Enrichment Score, Nom p-val = Nominal p value (sig <0.05), FDR 1-val = false discovery rate q value (sig <0.25).

Table 6.10 shows the top ten gene sets upregulated in blood samples from active fibrotic patients compared to active non-fibrotic patients. All have significant FDR-q and p values. Genes contributing to core enrichment in the top gene set of Antigen Processing and Presentation of Exogenous Peptide Antigen Via MHC Class I, are displayed in **Table 6.11**. The significantly downregulated gene sets are all related to methylation and RNA modification, the significance of which is not immediately clear, and doesn't prompt keen interest (results not shown) in light of the other positive findings.

GENE	RANK	GENE	RANK	GENE	RANK	GENE	RANK
HLA-G	340	NCF2	1506	PSMA1	2462	PSMA5	3992
PSMB10	401	PSMA6	1572	PSMB3	2607	PSMC3	4174
PSMD9	513	NCF4	1611	FCGR1B	2637	PSMA7	4210
PSME2	711	PSMB1	1655	FCGR1A	2656	PSMB7	4460
IFI30	714	PSME1	1718	PSMD4	2772	PSMD6	4506
NCF1	728	CD36	1739	PSMA3	3027	HLA-B	4602
PSMB6	1070	FCER1G	1893	CD207	3122	PSMB8	4617
HLA-H	1266	CYBA	1947	PSMB2	3464	PSMD11	4820
PSMA4	1345	PSMC1	1992	PSMB9	3525	PSMA5	3992
CYBB	1497	ITGB5	2363	PSMA2	3662	PSMC3	4174

Table 6.11 Genes and their rank in the entire gene list of 21334 genes showing core enrichment in the top gene set of antigen processing and presentation of exogenous peptide antigen via MHC class I.

The top 50 most differentially expressed upregulated genes and the top 50 most differentially expressed downregulated genes are shown in Figure 6.15. The genes are labeled next to each row. Each column obviously represents individual samples, with the prefix F referring to the fibrotic patients, and NF referring to the non-fibrotic. All have active disease. Yellow is high expression, and Blue low expression.

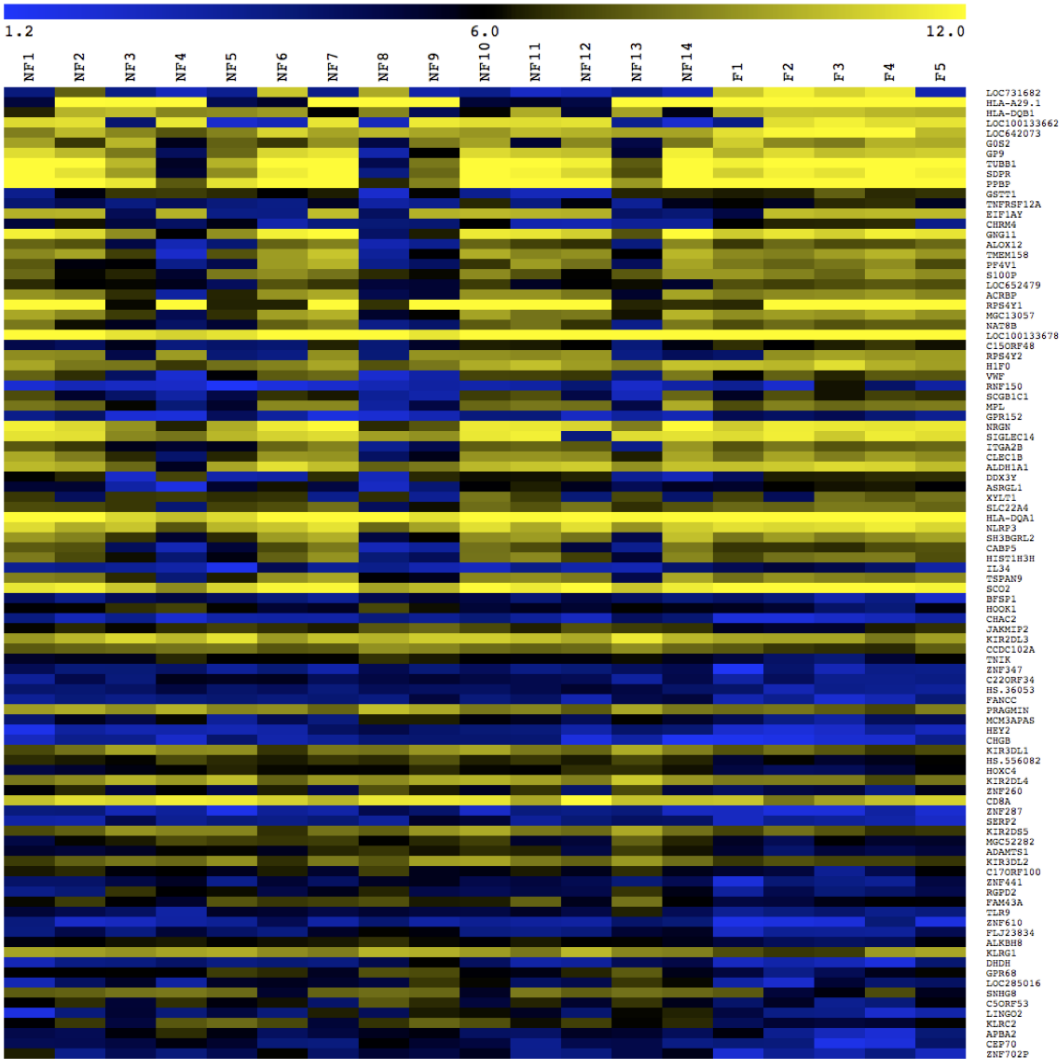


Figure 6.15 Heatmap depicting top 50 most positively and negatively differentially expressed genes in the active fibrotic cohort compared to the active non-fibrotic cohort in Blood samples from sarcoidosis patients. Yellow – high expression, blue – low expression. F = Fibrotic patients, NF = Non-fibrotic patients.

Although technically found in tissues, the macrophage gene sets used in the lung cohort to assess classical (M1) versus alternative (M2) activation were also assessed in the blood cohort, since monocytes are their pre-cursor. There was no predominance of M1 or M2 gene sets in either active fibrotic or active non-fibrotic samples (Enrichment plots not shown).

In addition, the Th1/Th2/Th17 gene signatures previously described were used to interrogate the blood cohort. The only significant finding was of an upregulated Th17 signature, when compared to Th2 or Th1, in the active non-fibrotic group. The enrichment plot for Th2 v Th17 downregulated gene set (GSE14308) is shown in Figure 6.16A (ES -0.32, NES -1.46, Nom P 0.01, FDRq 0.01), and that for the Th1 v Th17 downregulated gene set (GSE14026) is shown in Figure 6.16B (ES -0.32, NES -1.43, Nom P 0.01, FDRq 0.01).

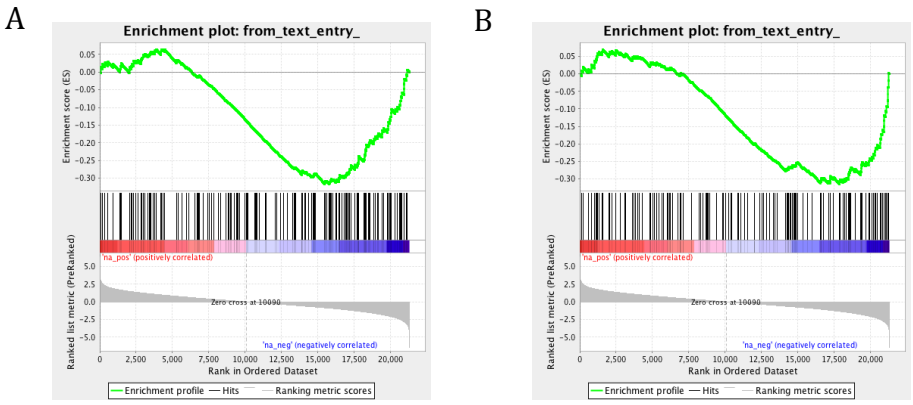


Figure 6.16 A) Enrichment plot of signature Th2 v Th17 in blood cohort ranked according to fold change of differential expression in active fibrosis versus active no fibrosis and B) Enrichment plot of signature Th1 v Th17 in blood cohort ranked according to fold change of differential expression in active fibrosis versus active no fibrosis. Both show Th17 genes to be significantly upregulated in the active no-fibrosis cohort.

Finally, and of advanced interest, especially in an attempt to answer whether or not an upregulation of chemokine signaling pathways acting on the periphery may drive the pathogenesis of fibrotic sarcoidosis, the KEGG gene set of chemokine signaling pathways as previously described for the lung cohort was examined. This has revealed significant core enrichment of 50 out of 161 genes in the gene set in the active fibrotic group. The enrichment plot is shown in Figure 6.17 and the 50 genes involved in core enrichment and their rank in the total gene list of 21334 genes is shown in Table 6.12.

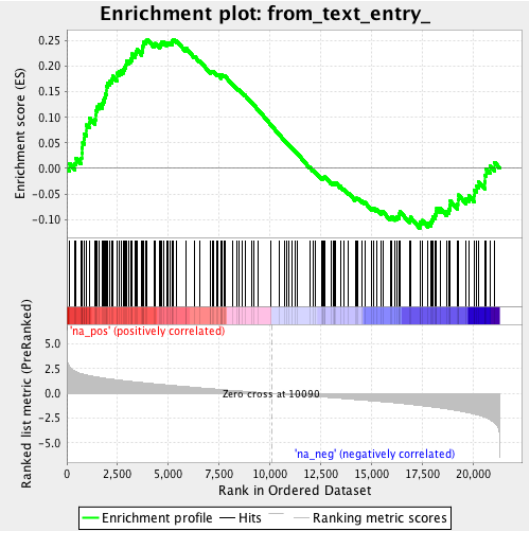


Figure 6.17 Enrichment plot of chemokine signaling pathway gene set obtained from KEGG showing upregulation in the blood samples of active fibrotic patients.

GENE	RANK	GENE	RANK	GENE	RANK	GENE	RANK	GENE	RANK
CCL14	125	CXCL16	967	PF4V1	1863	FOXO3	2563	GNAI2	3359
ARRB2	396	GRB2	1101	CXCL1	1901	PLCB1	2690	STAT2	3395
RHOA	451	CCL2	1115	FGR	1936	GNG8	2794	VAV3	3447
MAPK3	691	CCL19	1396	CXCL5	1939	PIK3CB	2857	JAK2	3699
NCF1	728	CCR1	1409	WAS	1991	RAC1	2913	CCL21	3724
GNG10	739	HCK	1428	PLCB2	2119	GSK3A	2920	PRKCD	3732
GNB2	775	GSK3B	1485	PF4	2209	CSK	3041	CXCL9	3756
NRAS	843	PPBP	1577	LYN	2248	ARRB1	3130	GNB4	3799
DOCK2	866	GNG5	1740	GNB1	2460	STAT1	3208	NFKBIA	3908
CCL26	965	PRKACG	1782	GNG11	2461	ADCY4	3214	PRKCB	3959

Table 6.12 Genes involved in core enrichment of the chemokine signaling pathway gene set in the active fibrotic blood samples, and their ranking in the entire 21334 gene list.

6.4.4 Analysis of pathways, processes and genes contributing to active fibrotic sarcoidosis in BAL cell samples

6.4.4.1 Principal Component Analysis

The total gene expression profiles of BAL cell samples were assessed, as performed for blood and lung tissue, using Principal Component Analysis (PCA) (performed in TMEV(171)). Clustering of the fibrotic patients and non-fibrotic patients in BAL cell samples is again poor compared to lung (Figure 6.18), with the three fibrotic patients (F1-F3) being seen in three different vectors.

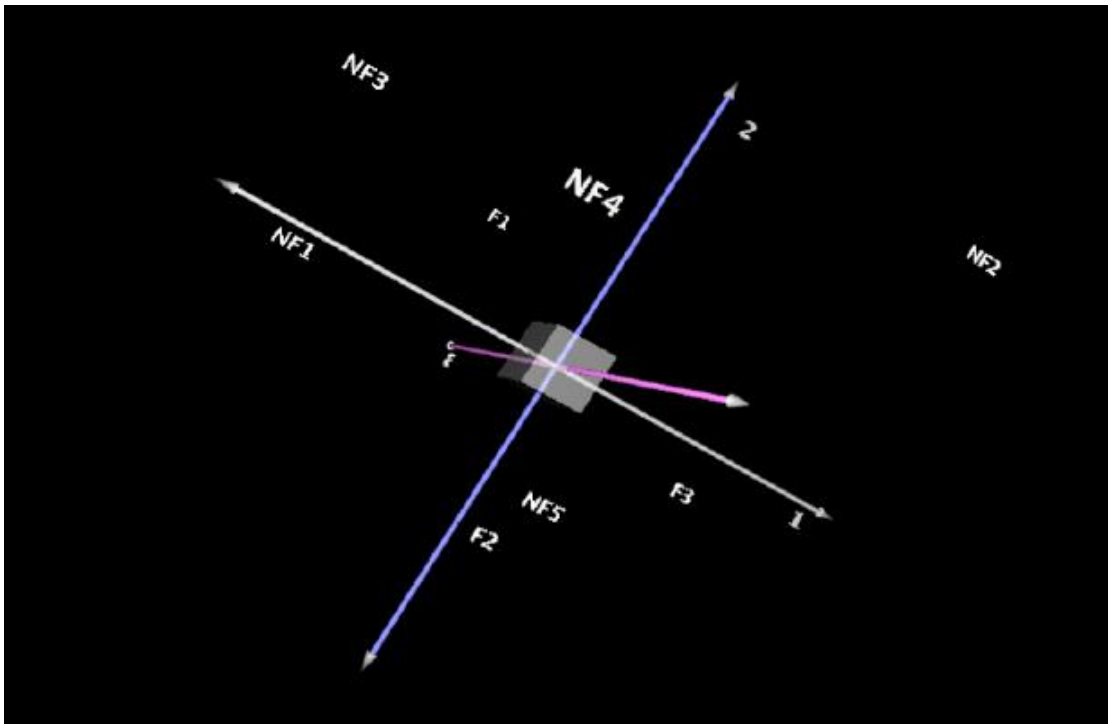


Figure 6.18 Principal Component Analysis of gene expression in BAL cells for active fibrotic (F1-F3) versus active non-fibrotic (NF1-NF5) samples.

6.4.4.2 Gene Set Enrichment Analysis

GSEA, again using C5bp GO terms (all biological processes), was employed on a pre-ranked gene list to further investigate which gene sets showed enrichment in the active fibrotic patients in comparison to the active non-fibrotic patients. The gene list was ranked, as per blood and lung tissue, according to fold change of the differential expression between the two groups, with those highest in active fibrosis at the top of the list.

No	Gene Set	ES	NES	NOM p-val	FDR q-val
1	Regulation of Natural Killer Cell Mediated Immunity	0.59	2.04	0.00	0.16
2	T cell differentiation involved in Immune Response	0.59	1.97	0.00	0.21
3	Mast Cell Mediated Immunity	0.70	1.97	0.00	0.15
4	Sterol Biosynthetic Process	0.55	1.95	0.00	0.14
5	T cell Activation involved in Immune Response	0.55	1.92	0.00	0.17
6	Leukocyte Apoptotic Process	0.60	1.89	0.00	0.19
7	Positive Regulation of Natural Killer Cell Mediated Immunity	0.66	1.88	0.00	0.19
8	Positive Regulation of Natural Killer Cell Activation	0.67	1.88	0.00	0.17
9	Sterol Metabolic Process	0.44	1.86	0.00	0.18
10	Positive Regulation of Interferon Gamma Production	0.48	1.86	0.00	0.16

Table 6.13 Top ten significantly upregulated gene sets in the active fibrotic BAL cell cohort, with their associated enrichment scores, normalized enrichment scores, nominal p values and FDR q values.

The top ten significantly upregulated gene sets in the active fibrotic BAL cohort are shown in Table 6.13. There were no significantly downregulated gene sets. The genes contributing to the core enrichment of the top gene set 'Regulation of Natural Killer Cell Mediated Immunity' are shown in Table 6.14.

GENE	RANK	GENE	RANK
PIK3R6	161	SLAMF6	2859
LEP	739	CD226	3222
NCR3	1462	LILRB1	3228
CADM1	1526	SH2D1A	3420
TAP1	1605	HAVCR2	4323
LAG3	1906	SH2D1B	4409
STAT5B	2439	SERPINB4	4990
CRTAM	2459	HLA-E	5220
SERPINB9	2462	HLA-A	5424
RASGRP1	2682		

Table 6.14 Genes contributing to the core enrichment of the top upregulated gene set (Regulation of Natural Killer Cell Mediated Immunity) in active fibrotic disease in the BAL cohort, and their associated ranking in the entire gene list (of 21334 genes).

The top 50 most differentially expressed upregulated genes and the top 50 most differentially expressed downregulated genes in the BAL cell cohort when comparing active fibrosis to active no fibrosis are shown in the heatmap below . The genes are labeled next to each row. Each column obviously represents individual samples, with the prefix F referring to the fibrotic patients, and NF referring to the non-fibrotic. All have active disease. Yellow is high expression, and Blue low expression.

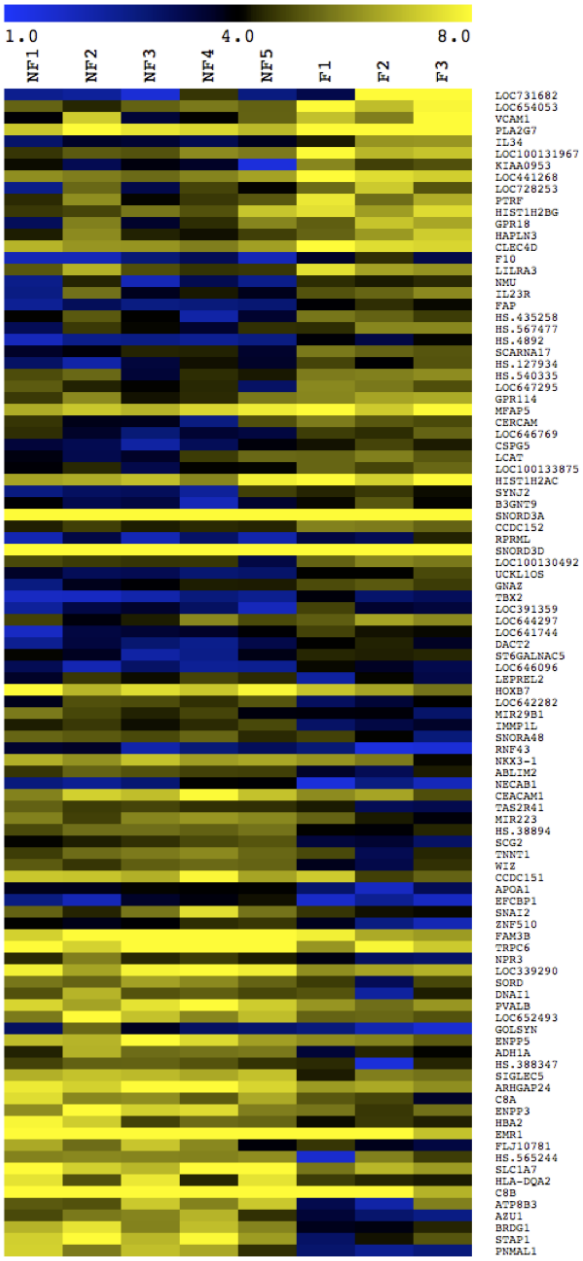


Figure 6.19 Heatmap depicting top 50 most positively and negatively differentially expressed genes in the active fibrotic cohort compared to the active non-fibrotic cohort in BAL cell samples from sarcoidosis patients. Yellow – high expression, blue – low expression. F = Fibrotic patients, NF = Non-fibrotic patients.

The M1 and M2 gene signatures derived from GEO published data, as used in the lung and blood cohorts above, were again assessed in the BAL cell cohort. This has revealed an M1 predominance and enrichment. None of the M2 signatures showed

a significant enrichment in either a positive or negative direction. The enrichment plots for the four M1 gene signatures tested are displayed in Figure 6.20, with all genes that contribute to the core enrichment of these four signatures shown in Table 6.15 alongside their ranking in the entire gene list.

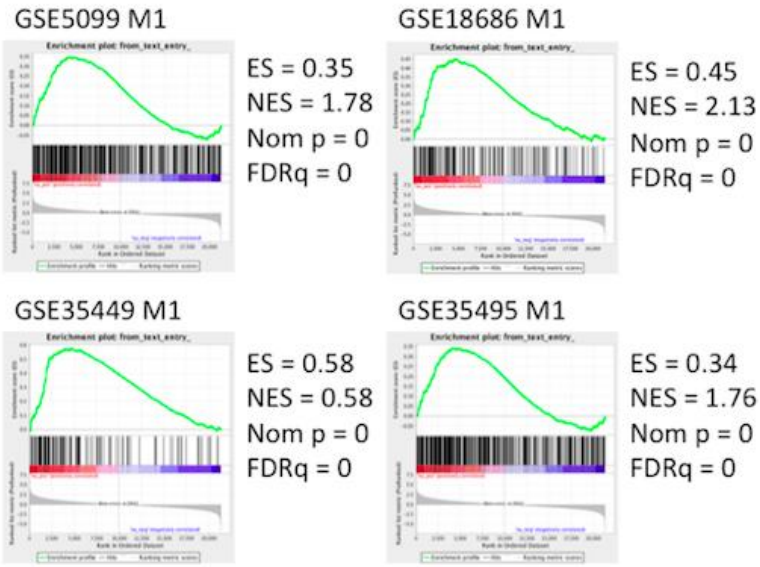


Figure 6.20 Enrichment plots for four M1 gene signatures and their associated ES, NES, Nominal p value and FDRq value, showing upregulation in the BAL cells of active fibrotic patients.

GENE	RANK	GENE	RANK	GENE	RANK	GENE	RANK	GENE	RANK
CCL22	551	DHRS2	1761	PALLD	2796	P2RY14	4456	BATF3	5363
CD1C	564	CD1B	1853	SLAMF1	3027	RNF19B	4674	FZD2	5367
MMP12	650	PTPRE	1907	SIGLEC10	3144	CCL8	4718	PTGER2	5487
GOLGA8B	660	P2RY13	2219	MOCOS	3165	ZNF366	4754	BATF3	5363
SCARB1	716	FGL2	2443	RAB33A	3229	HOMER2	4790	FZD2	5367
CD1A	856	SERPINB9	2462	ST6GAL1	3311	CMTM8	4812	PTGER2	5487
ITPRIPL1	992	CD209	2555	CCL18	3883	RAMP1	4848	BATF3	5363
ARAP2	1313	RASGRP1	2682	CD1E	4082	CCL13	4875	FZD2	5367
SYT17	1427	QPCT	2691	FOXQ1	4276	CKB	5070	PTGER2	5487
CLEC10A	1491	RAP1GAP	2760	DHRS11	4441	IL27RA	5242	BATF3	5363

Table 6.15 All genes contributing to core enrichment of the M1 gene signatures in the BAL cell samples and their associated ranking amongst the entire gene list of 21334 genes.

No enrichment was seen for Th1, Th2, Th17 gene signatures or the cytokine gene set as used earlier in this chapter in BAL cells.

6.5 Discussion

Fibrotic sarcoidosis is an important yet understudied phenotype of sarcoidosis. Much work in this thesis has focused on establishing how to assess the level of active disease, which is important in the context of fibrosis, an irreversible and potentially chronic condition, made more severe and implying more morbidity and mortality if the underlying condition and predisposition to fibrotic remodeling remains active. There has been little work to date that has phenotyped patients for the purposes of research into predominantly pulmonary sarcoidosis, and then sub-divided this group into fibrotic and non-fibrotic active disease conditions. This preparatory process is likely to provide more meaningful results in the investigation of this extremely heterogenic disease.

Having established the cohorts of prominent active pulmonary sarcoidosis with or without fibrosis, three sampling modalities, namely transbronchial lung biopsy tissue, peripheral blood mononuclear cells, and bronchoalveolar lavage cells, were interrogated for their gene expression profiles. This analysis has revealed several interesting findings.

Firstly, it was noted that the clustering of samples based on their whole gene expression was very good in lung tissue, with fibrosis patient samples lying in close proximity to each other when assessed with principal component analysis. This likely reflects the very different processes that are being observed in the lung tissue in the presence of fibrosis versus simple granulomatous inflammation. In Blood and BAL cell samples, the clustering is not strong, but when differential expression of the

two phenotypes, active fibrotic versus active non-fibrotic, is assessed, and specific gene sets and gene signatures are used to interrogate the ranked lists, again some meaningful results are observed.

Immune dominant gene sets are highly enriched in lung tissue from active fibrotic patients when compared to active non-fibrotic. Although not as strongly, this is also seen in blood samples, which focus more on antigen presentation, and in BAL cell samples, that reveal regulation of NK cell mediated immunity and activation, T cell activation and differentiation and IFN gamma production as prominent pathways. The NK cell related gene sets were an unexpected finding, as there is no indication for NK cell activation in sarcoidosis, and in fact it has previously been shown that iNKT cells are deficient in patients with sarcoidosis (67). Interrogating the genes involved in enrichment of the NK related gene sets, PIK3R6, LEP, NCR3, CADM1, TAP1, STAT5B, IL23R and IL18 are seen to contribute. Whilst the combination of these alongside others genes have flagged NK cell gene sets as enriched, these genes are not specific to NK cells, and also have roles in T cell migration, activation and proliferation at sites of inflammation. TAP1 is involved in the transport of antigens from the cytoplasm to the endoplasmic reticulum for association with MHC class I molecules, CADM1 promotes IFNγ production from CD8+ cells, STAT5B mediates cellular responses to cytokines including IL-2, NCR3 is an NK receptor, engagement of which promotes TNFα and IFNγ release by NK cells. In adaptive immunity LEP promotes switching of memory T cells to a Th1 phenotype, IL23R Binds IL23 and mediates T-cell, NK cell and macrophage/myeloid cell stimulation, and IL18 is a proinflammatory cytokine that can augment natural killer cell activity

but more importantly stimulates interferon gamma production in Th1 cells (www.genecards.org).

High scoring gene sets can be grouped on the basis of leading-edge subsets of genes that they share. Such groupings can reveal which of those gene sets correspond to the same biological processes and which represent distinct processes, and the 'Leading edge analysis' function in GSEA enables one to reveal a biologically important subset of genes within these. Of the top ten upregulated gene sets in active fibrosis from lung tissue, CCL19 was identified as the top leading edge gene, contributing to nine of the ten gene sets. Of note, CCL19 also features in the IPF v sarcoidosis signature, and the chemokine signature in lung and blood samples. CCR2 and CCL5 also feature in this leading edge list, and have been discussed in detail in the previous chapter on the gene expression analysis for activity. Yoshida et al first identified CCL19, initially termed ENI1 Ligand chemokine ECL, and also known as MIP3β, as a specific ligand for EBI1, or as it then became known, CCR7 (also known as MIP3βR) (285). It is a homeostatic chemokine that is constitutively expressed in secondary lymphoid organs, but its expression can also be induced in non-lymphoid tissues following inflammation or infection, and may help to recruit lymphocytes locally (258) to these tissues as a result of immune activation (286). It has been observed in samples of human lung tissue (287, 288), and is chemoattractant for naïve T cells and activated B cells (289), macrophage progenitor cells (290), and Natural Killer (NK) cells (291). It's role in sarcoidosis has been previously studied, with Gibejova et al identifying an upregulation of CCL19 mRNA transcripts in the BAL cells of sarcoidosis patients compared to controls, and significantly elevated CCL19 protein levels in BAL fluid in patients with chest

radiograph Scadding stage II when compared to stage I, which they propose contributes to the development of the CD4+ lymphocyte alveolitis seen in active pulmonary sarcoidosis (292). No patients with fibrosis were specifically studied however. CCL19 is part of the gene signature derived by Leng et al (279) (Appendix 8.6 and 8.7) to distinguish IPF lung biopsy samples from sarcoidosis lung biopsy samples. Of great interest, was that the IPF signature of upregulated genes was highly enriched in the fibrotic sarcoidosis group, and those seen to be downregulated in IPF (when signature derived to differentiate from sarcoidosis) were seen to be enriched in the active, non-fibrotic group. Anti-fibrotic treatment has recently been introduced for IPF, which has shown promise in slowing progression of the fibrotic process. If a significant overlap of genes associated with IPF can also be seen in fibrotic sarcoidosis, this may also prompt the suggestion that anti-fibrotics may be of benefit in this sarcoidosis subgroup. Pulmonary fibrosis is thought in general to be the result of an abnormal healing process following lung tissue damage, inflammation and immune activation. Lymphangiogenesis may play a role in the structural remodeling, and potential ensuing fibrosis, following diffuse alveolar damage, but the mechanism through which this happens is not clear. A study by Yamashita et al has shown that the CCL19-CCR7 axis likely facilitates macrophages to participate in lymphangiogenesis, by demonstrating the aggregation of CD68+ macropahges expressing CCR7 around newly formed lymphatics in a profilerative stage, and the co-expression of CCL19 on the lymphatic endothelium. This may explain the findings from this chapter of increased CCL19 expression in the lung tissue of patients with active fibrotic sarcoidosis, where the expression has been induced as a result of ongoing active inflammation. This in turn results in accumulation of macrophages at the site, which are also vital in the

healing and repair process, but may also perpetuate the homing of lymphocytes contributing therefore to ongoing alveolitis and an active disease state. The 'chicken or egg' question, is whether the patients with fibrotic sarcoidosis have a predisposed increase in CCL19 to begin the process, or whether there is a separate inciting event that begins the fibrotic process, and leads to an increase in CCL19 production. This unfortunately is beyond the scope of this study but would be of interest for future research. CCR7, the receptor for CCL19, commonly found on dendritic cells and T lymphocytes, is also seen to be enriched, and features in the top 50 leading edge genes in the chemokine signaling pathway geneset for active fibrotic sarcoidosis when analysed in the lung sample cohort, but did not feature in the PBMC cohort. CCL19 did however feature in the PBMC cohort, showing enrichment in active fibrotic disease using the chemokine signaling pathway geneset.

CCL19 and CCR7 are also seen to be a specific feature of the M1 gene sets, tested alongside M2 gene sets, in the active fibrotic cohort from lung, blood and BAL cell samples. We know that macrophages are present in all tissues in the body, and their pre-cursor, the monocyte begins in the peripheral circulation, migrates by extravasating through the endothelium, and finally differentiates into a macrophage or dendritic cell within the tissue itself. Depending on the molecular cues present in the local microenvironment, they develop distinct activation states, with specific functions. Activated macrophages are most commonly categorized into the above two broad subtypes- M1 (classically activated) and M2 (alternatively activated macrophages). M1 macrophages are pro-inflammatory, and are induced by exposure to IFN-γ or lipopolysaccharide. They defend the host from exogenous

microbes, are involved in tumour suppression, and produce large amounts of inflammatory cytokines (e.g. TNF-α, IL-6, and IL-12). In contrast, M2 macrophages are induced by IL-4 or IL-13, produce large amounts of IL10 and TGF-β, display anti-inflammatory properties and are involved in tissue repair (293). M2 macrophages have previously been implicated in the development of idiopathic pulmonary fibrosis by enhancing fibrogenesis of fibroblasts (160), favouring cell growth, promoting collagen formation, and driving tissue repair (294). In addition to this, Prasse et al. have demonstrated that a vicious circle between alveolar macrophages and fibroblasts exists, whereby the release of IL10, IL1 receptor antagonist, and CCL18, promotes collagen deposition and the fibrotic progression of IPF (277). In my gene expression analysis of lung tissue, I found that both M1 and M2 gene sets were enriched in the active fibrotic group, but there appeared to be an M1 predominance. This was also reflected in the BAL cell analysis. It may be that a combination of both M1 and M2 activated alveolar macrophages, in response to tissue damage, are required for this remodelling to take place, and that CCL19 and CCR7 promote the ongoing recruitment of M1 macrophages to the site of inflammation, while the presence of M2 macrophages provides the opportunity for tissue repair and remodelling in this highly active inflammatory milieu.

The identification of two genes in the lung tissue samples, ADAMDEC1 and MMP12, found to have more than ten fold change expression values in active fibrosis when compared to active no fibrosis, is of great interest. Both have been previously identified as likely mediators of lung damage and/or remodelling by Crouser et al. (168), and ADAMDEC1 and MMP12 expression in BAL cells were seen to correlate with severity of disease. Macrophages are the major producers of

MMP12 and ADAMDEC1. ADAMDEC1 features in the M1 gene set, whereas MMP12 is a feature of both M1 and M2. ADAMDEC1 is a secreted protein thought to be a member of the metalloproteinase family. It is upregulated during dendritic cell maturation, and may play an important role in dendritic cell function and their interactions with germinal centre T cells (https://www.ncbi.nlm.nih.gov/gene/27299). Its specific role in fibrotic sarcoidosis however is currently unclear. MMP12 is a potent elastase, that degrades soluble and insoluble elastin, is involved in the breakdown of extracellular matrix in normal physiological processes such as embryonic development, reproduction, and tissue remodeling, as well as in disease processes, such as arthritis and metastasis (295). In addition, MMP12 has been proposed to be an essential gene for the development of fibrosis following acute lung injury in a mouse model, and marked upregulation of gene expression was observed in fibrotic disease (296). Other featured genes from this highly differentially expressed gene list reveal MMP7, which has also been reported to show correlation to progression in IPF (297); IFNγ, which has been implicated for many years in the pathogenesis of sarcoidosis, known to be released spontaneously by alveolar macrophages from sarcoidosis patients, thought to relate to the level of active disease and seen to reduce with corticosteroid treatment (60); and CCR2 and CCL5, which have been already discussed in detail in Chapter 5 with regards to disease activity.

Continuing to focus on the results obtained from lung tissue, but this time on the downregulated gene sets, again overlap is seen with IPF patients. When the entire gene list is ranked according to fold change of the differential expression between the two groups, with those highest in active fibrosis at the top of the list, gene sets

for cilium morphogenesis, organisation and movement are seen to be downregulated. Leng et al have published that distinctly expressed genes distinguishing IPF from sarcoidosis are more enriched in processes of cilium biogenesis or degradation (279). Only two gene sets were seen to be significantly down-regulated in the active fibrotic cohort when the differential gene list is looked at in more detail (cut off of <-2FC), the first of which was 'Cellular Response to Organic Cyclic Compound'. The gene AGER is amongst the genes involved in the core enrichment of this gene set. AGER is a multiligand receptor that interacts with other molecules implicated in homeostasis, development, and inflammation, and this has also been seen to be consistently downregulated in IPF (280). It is produced my alveolar macrophages, and usually signifies acute lung injury (298). Its down regulation may suggest loss of this cell type in the presence of fibrosis.

The Th cell phenotype most often reported as being dominant in pulmonary sarcoidosis has been Th1 (70), but more recent work has addressed the presence of other Th subsets in various compartments (i.e. peripheral blood, BAL cell and Lung tissue). It has been suggested that in chronic sarcoidosis, possibly in response to persistent inflammation, a transition from a Th1 to a Th2 cytokine signature may occur (299, 300), and Th2 cytokines have been implicated many times in processes of tissue fibrosis (72, 301). In this chapter it was noted that Th1, Th2 and Th17 signatures showed equal prominence in the lung tissue from patients with active fibrotic sarcoidosis, as they did in BAL. Note was made however that Th17 was seen to be downregulated in blood samples from the active fibrotic group when compared to those with active disease and no-fibrosis. Th17 cells have been reported as significant components of the alveolitic/granuloma phase of sarcoidosis

in both PBMC and alveolar macrophages (79), and this is paralleled in my study, but the question of what occurs to the Th17 cells when active fibrosis is occurring has not previously been addressed. IL-17-producing cells however have been reported as a potential useful biomarker for the prognosis of sarcoidosis and may play a role in the spontaneous recovery typical of patients with Lofgren's syndrome, which again fits with the finding above, that in those patient with active disease, but no fibrosis, a Th17 phenotype is observed in the periphery, indicating a better or more acute process similar to Lofgren's syndrome (302).

Finally, looking at the chemokine signalling pathway gene sets in the periphery and lung, and comparing those seen to contribute to the enrichment, sets of complimentary chemokines and receptors are seen to be present in the opposite/receiving compartment (Table 6.16).

BLOOD	LUNG
CCL19	CCR7
CCL2	CCR2
CCR1	CCL5
CXCL9	CXCR3
CCL14	CCR1

Table 6.16 Table showing chemokines and their receptors seen to be enriched in Blood and their complimentary molecules seen in the lung compartment.

CCL19, CCR7, CCR2 and CCL5 have already been discussed in detail in this chapter and chapter 5. CXCR3 however, is the receptor for CXCL9, and both have also previously been linked to pulmonary sarcoidosis. Busuttil et al demonstrated that both CXCL9 and CXCL11, found in BAL fluid, are important mediators in recruiting CXCR3-expressing cells, and observed that both lymphocytes and cells of

monocyte lineage express CXCR3 and are involved in the formation of sarcoid lung granulomas (303). The upregulation of these genes observed in my study likely play a significant role in the chemotaxis of lymphocytes and monocytes/macrophages from the periphery to the sites of active inflammation and subsequent granuloma formation.

6.6 Conclusion

In summary, after controlling for the remarkable clinical heterogeneity in sarcoidosis, and defining the level of disease activity, I have studied the gene expression profiles of the specific groups of active fibrotic pulmonary sarcoidosis against active non-fibrotic pulmonary sarcoidosis. The results show a potentially important role for the CCL19-CCR7 axis in fibrotic sarcoidosis, alongside ADAMDEC1, MMP12, MMP7, CCR2 and CCL5. No significant difference in enrichment of Th1 or Th2, gene sets were observed in lung, BAL cell or blood gene expression profiles, and in lung, M1 and M2 gene sets were enriched with more of an M1 predominance. It may however be the balance of these immune cell phenotypes that dictate the clinical outcome of fibrosis, rather than a mere presence or absence. Th17 cell signatures were seen to be enriched in the periphery of active patients without fibrosis, which may signify a less chronic and complicated disease state. Enrichment of a chemokine and cytokine signature was realised in the lungs of the active fibrotic group, suggesting enhanced recruitment of inflammatory cells in this phenotype. Finally, despite previous comparisons of IPF to sarcoidosis showing different gene signatures, an overlap of genes significant in Idiopathic Pulmonary Fibrosis is observed when the cohort is phenotyped for fibrotic sarcoidosis. Again a prominent role for CCL19 is suggested. Details of future studies planned to investigate the role of these cytokines in fibrotic sarcoidosis are given in section 7.3.

7 Final discussion and future direction

7.1 Main Findings

The overall aim of this project was to investigate the processes, genes and immunological pathways that drive disease activity and promote pulmonary fibrosis in the extremely heterogenic condition of sarcoidosis. Studies that have gone before this have not focused on controlling for the clinical heterogeneity, both in terms of disease phenotype and organ involvement, or the level of disease activity in their cohorts. Consequently, research outcomes obtained for sarcoidosis as a whole entity may have obscured the detection of signals important for the persistence of active disease and development of fibrotic pulmonary sarcoidosis. One major barrier is that there is no agreed upon marker of disease activity, despite many years of research, although many biomarkers have been put forward. There is also no consensus opinion on how to differentiate the varied phenotypical presentations of the disease for the purposes of research.

In this thesis, I began by defining my patient group, selecting out those with biopsy proven and predominant pulmonary sarcoidosis. As both eye and skin involvement commonly co-exist within this group, these organs were also permitted, but no more than one other major organ involvement was allowed, and the primary driver for management and consideration of treatment was for pulmonary disease. Samples from recruited patients were obtained from transbronchial lung biopsies, cells from bronchoalveolar lavage, and peripheral blood. The first two sampling modalities are

obviously representative of the sites of active disease, the last a potential peripheral reflection of that occurring in affected compartments (e.g. lung).

Following on from this, I examined the use of high resolution CT scanning as a means for measuring disease activity in pulmonary sarcoidosis. Measurements of disease severity on HRCT (but not activity) have been previously published (188, 190, 191), but these variably incorporate different HRCT abnormalities, functional impairment and end-organ damage. In conjunction with experienced thoracic radiologists, HRCT abnormalities described in the literature as being typically associated with sarcoidosis, were defined, and a means to score their presence, and extent if applicable, was considered. Several of the typical radiological abnormalities had undergone pathological correlation to autopsy or lung explant specimens. They relate to various stages of pulmonary sarcoid granuloma formation, T cell infiltrates or active fibrosis, thereby strengthening the suggestion that there is an association with active disease. A final composite lung disease activity score (CTAS) was developed following correlation against various surrogate disease activity markers (soluble IL2RA, Serum ACE and change in FVC). The CTAS consisted of the following four typical radiological abnormalities:- ground glass opacification (GGO), nodularity, interlobular septal thickening (IST) and consolidation (all of which had previous pathological correlation). I then demonstrated that CTAS correlated with the extent of improvement in lung function (FVC) following treatment, thus verifying the design and development of a reproducible and systematic method for scoring disease activity in pulmonary sarcoidosis. The CTAS is specifically designed to reflect the extent of disease activity regardless of the level of lung function impairment and fibrosis, and is likely

to be particularly useful in aiding decisions on whether to treat fibrotic sarcoidosis, as it allows distinction between active granulomatous disease and residual 'burnt out' fibrosis in patients with fibrotic pulmonary sarcoidosis.

The frequency of obtaining a HRCT in patients with pulmonary sarcoidosis is obviously limited due to radiation exposure, and therefore, while the above activity score provides a means to define the presence and level of disease activity, it should only be done at strategic points in a patient's clinical course. Chest radiographs are only a fraction of the radiation dose (<0.2mSV compared to 5mSV), and tend to be performed more frequently at clinic visits. Alongside chest radiographs, readily available blood tests provide another source of information that may correlate to activity, and I therefore set about designing a clinic based tool as a surrogate marker for disease activity measured with CTAS. Combined scoring systems (210), particularly for cutaneous sarcoidosis (211, 212), have been published in the past, but none have made it into mainstream use, or focus on the level of active pulmonary disease. The focus of the design of the clinic based tool was to achieve a detailed and discriminant scoring system that could be performed quickly and easily in the clinical setting. The sarcoidosis clinical activity score (SCAS) was created by correlating abnormalities seen on CXR and cellular markers of immune activity, namely IgG, Serum ACE and Lymphocyte count, to the CT activity score (CTAS). A strong correlation to CTAS was observed with the score as a composite set of markers, over any single contributing parameter of the score, and the SCAS was subsequently defined. The SCAS grades the level of disease activity from zero to twelve, and good variability was observed when used to monitor response to treatment. Two independent respiratory physicians subsequently

verified the clinical applicability of the scoring system (SCAS) and demonstrated both ease of use and reproducibility.

The SCAS was used to investigate the processes, pathways and genes correlating to disease activity, in both the lung and the periphery. The most informative gene expression analysis results correlating to increasing activity, were observed in BAL cell gene expression. Gene set enrichment analysis revealed several immune related gene sets, the strongest being 'Immune Response', and within this gene set, CCR2, TREM2, IL6, CCL18, CCL2, IL10RB, IL15, CCR6 and CCL20 were seen to contribute its enrichment. Several of these have already had work published on their links to sarcoidosis, but not specifically for their association with disease activity. C-C Chemokine Receptor 2 (CCR2) was present in all top ten positively enriched gene sets in the leading edge analysis of BAL gene set enrichment results when correlated to increasing activity. It was also present in the 'Defense Response' gene set that overlapped between BAL cell and PBMC in this chapter, but features too in the significantly differentially expressed gene list for active fibrotic sarcoidosis patients when compared to active non-fibrotic pulmonary subjects. CCR2 encodes the receptor for monocyte chemoattractant protein-1 (CCL2 or MCP-1), a chemokine that specifically mediates monocyte chemotaxis. The protein is expressed primarily on the surface of macrophages and monocytes, and it is required for infiltration of monocytes to the site of inflammation (236). Mice deficient in its ligand, the chemokine CCL2 (MCP1), display reduced macrophage fusion during foreign body reaction in vivo and severely decreased IL-4-induced formation of multinucleated giant cells in vitro which are the primary events in granuloma formation (304). The enriched presence of CCR2 in both high disease activity and in

active fibrosis is of great interest. Inhibition of CCR2 has been considered as a target for multiple diseases including autoimmune disease, atherosclerosis, pain, and metabolic disease, based in part on the critical role this receptor plays on monocyte migration. Blockade of CCR2, or of it's primary ligand CCL2, has proven efficacious in a diverse array of inflammatory disease murine models (245-249). Two combined CCR2/CCR5 drugs have reached phase 2 clinical trials, namely PF-04634817 (Pfizer) and Cenicriviroc (or TAK-652) (Takeda Pharmaceuticals). The former is being trialled in the treatment of diabetic macular oedema, and the latter in HIV infection (where CCR5 is best known for it's role as a co-receptor for entry of HIV into cells). Several other CCR2 antagonists have also been identified, but progression to use in clinical practice has been limited (250). If a role for this gene in the persistence of high disease activity, and the subsequent development of pulmonary fibrosis, is established, the potential for a new drug that blocks CCR2, may pave the way to a new clinical trial in the treatment of pulmonary sarcoidosis, which in turn can be monitored by way of SCAS or CTAS in its response to treatment.

Several other new and potentially important findings were made in the chapter on gene expression correlated to disease activity, that warrant further mention as a platform for future study. In particular, TREM2, CCL18 , CXCL9 and CXCL10.

As SCAS increased, TREM-2 (Triggering Receptor Expressed on Myeloid Cells- 2) expression was also seen to increase, which features in both the Immune Response Gene Set and the Immune System Process Gene Set of BAL cell expression. TREM-2 was first detected on human monocyte-derived dendritic cells (DC) (305,

306), but subsequently has been shown to be expressed in osteoclast and microglia cells (262, 307), and alveolar (262), hepatic (308) and intestinal macrophages (309). The ligand(s) for TREM2 are currently unknown, but following ligand binding, TREM-2 signals through it's adaptor protein DAP12, inducing release of the protein tyrosine kinase Syk, which is mandatory for the cytokine induced fusion of macrophages and subsequent production of multinucleated giant cells, the hallmarks of granulomatous inflammation (260-262, 310, 311). Phagocytosis is also a TREM-2 dependent process (263, 264). A role for TREM-2 in lung infections has been proposed by Aoki and colleagues, where Mycobacterium bovis infection in mice was shown to increase TREM-2 expression in the lung. The same upregulation however was not observed in vitro with mycobacterial exposed isolated alveolar macrophages, which may indicate that the increase is either on another cell type present in lung tissue, or related to an increase in alveolar macrophage influx into the lung (312). This may be of relevance given the historical interest in the role of mycobacterial pathogens in the development of sarcoidosis. However, DAP12 gain-of-function transgenic mice, which display a fatal inflammatory syndrome, exhibit spontaneous formation of multinucleated giant cells in the lung even in the absence of pathogens (313), which may also help unravel the conundrum of the lack of a consistent pathogen associated with sarcoid granulomas. Finally, the role of CCL2/CCR2 associated chemotaxis, TREM-2 binding, and the development of fusion competent macrophages, shown in Figure 7.1, is a logical pathway that illustrates the role of both enriched genes in the pathogenesis of active sarcoid granulomas.

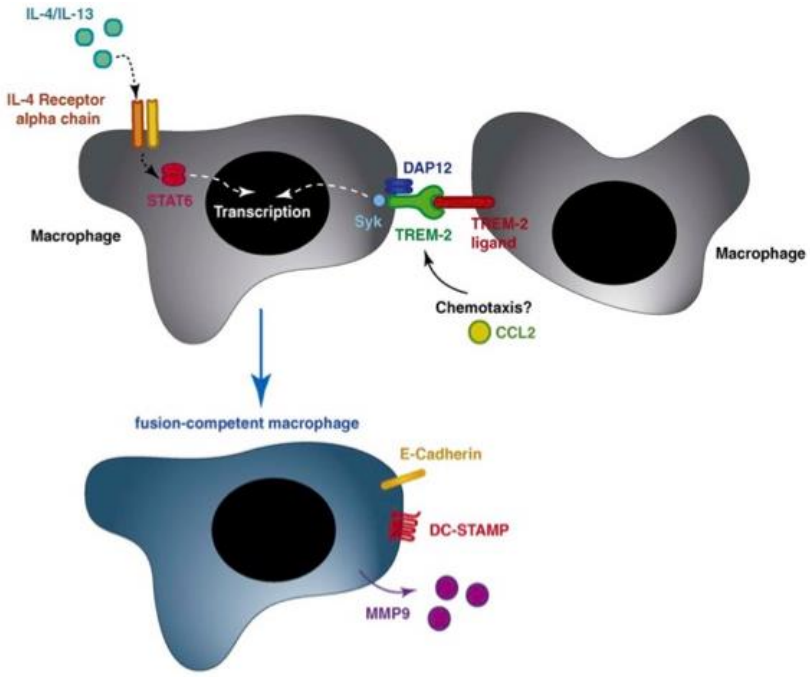


Figure 7.1 Programming of macrophages into a fusion competent status (From Helming et al 2009 (314)) Il4 and IL13 bind to their common IL4 receptor and induce a fusion-competent status through activation of the transcription factor STAT6. Fusion competency however also requires macrophage to macrophage contact, possibly dependent on CCL2 induced chemotaxis.

Following this review of what is already known of TREM-2, it is possible that increased TREM2 expression is related to the persistence of granuloma formation and active disease in sarcoidosis.

CCL18, a chemokine that is expressed by a broad range of monocytes/macrophages and dendritic cells, can induce chemotaxis in T cells, B cells, DC, haematopoetic progenitor cells, fibroblasts and potentially monocytes/macrophages (266). The chemokines that are produced in response to CCL18 are angiogenic and attract various leucocyte populations, which sustain inflammation. The capacity of these cells however, to also remove cellular debris without causing oxidative damage, and the production by monocytes of the anti-inflammatory cytokine IL10, may also initiate termination of the inflammatory response. This may suggest a pivotal role for CCL18-activated macrophages in

determining persistence of inflammation or its resolution (267), and taken in conjunction with previous work from this group, showing that monocyte IL-10 production is defective in sarcoidosis, this may be highly relevant. Enhanced CCL18 production has been described in a number of other inflammatory disease conditions, including Gaucher's disease (a lysosomal storage disorder characterized by the accumulation of glycolipids in macrophages and granuloma formation) (134), rheumatoid arthritis (268), Sjogren's syndrome (269), and Hypersensitivity Pneumonitis and IPF (270, 271). Although the specific role CCL18 may take in active sarcoidosis has not yet been investigated, it has already been reported as a potential marker of granulomatous activity in the disease (136).

CXCL9 and CXCL10 were seen to correlate with increasing disease activity in the gene expression analysis in both blood and BAL cell samples. CXCL9 (Chemokine [C-X-C motif] ligand 9) is a small cytokine belonging to the CXC chemokine family, and is induced by IFNγ. It is an important mediator in recruiting CXCR3-expressing lymphocytes and monocytes/macrophages, and the formation of sarcoid lung granulomas (303). Again, overlap for this gene was seen between the study on activity (chapter 5), and that on fibrosis (chapter 6), with the latter study revealing both an enrichment of CXCL9 in the blood and lung tissue from active fibrotic patients, and also enrichment of it's receptor, CXCR3, in the lung. CXCL10 is secreted by monocytes, endothelial cells and fibroblasts, also in response to IFNγ. It has been shown to play a crucial role in Th1 cell migration in the lungs of patients affected by sarcoidosis (64), and a positive relationship between BAL levels of CXCL10 and the degree of CD4 alveolitis, denoting active disease, has been already been reported (64), as has it's potential role as a marker of activity (274)

(234). These upregulated genes, observed both peripherally and at the site of active disease, perpetuate the notion that the disease of sarcoidosis is not completely compartmentalised, and that there may be peripheral measures of disease activity that reflect what is happening at the site of granuloma formation.

In the final part of this thesis, I addressed the question of what underlies the pathogenesis of fibrotic sarcoidosis, by comparing the gene expression profiles of active fibrotic patients with those with active disease but no fibrosis. The most noticeable results here were seen in gene expression analysis of lung tissue, with significant enrichment of several prominent immune related gene sets identified through GSEA. This is not surprising given that it is the site of major difference and the reason for the question. However, GSEA also identified a significant enrichment in antigen presentation pathways in blood, and NK and T cell activation/differentiation and Interferon gamma production related pathways in BAL cells. Assessment of macrophage phenotype in the active fibrotic group revealed enrichment of both M1 and M2 gene signatures in the lung tissue, with a slight M1 predominance; only M1 gene signature enrichment in BAL cells, and no signal in blood. Assessment of Th phenotype (Th1,Th2 and Th17) revealed an equal predominance of Th1, Th2 and Th17 in lung tissue and BAL cell in both groups (active fibrotic and active non-fibrotic), but in the blood a predominance of the Th17 cell signature was seen in those without fibrotic disease. The chemokine signaling pathway gene set revealed complimentary chemokines in the blood and the lung tissue samples, some of which are seen to overlap with those identified in the study on disease activity. This may suggest that fibrotic sarcoidosis is not necessarily a

pre-disposed phenotype, but may develop as a response to unremitting inflammation.

Further genes identified in this study on fibrosis also worthy of mention include CCL19 and its receptor CCR7, ADAMDEC1 and MMP12, all of which were seen to be enriched in the lung tissue of patients with active fibrosis compared to those with active disease but no fibrosis. Leading edge analysis of the top ten upregulated gene sets in active fibrosis from lung tissue, identified CCL19 as the top leading edge gene, contributing to nine of the ten gene sets. CCL19 is also seen in the IPF v sarcoidosis signature, the chemokine signature in lung and blood samples and the M1 gene signature. CCL19 is a homeostatic chemokine that is constitutively expressed in secondary lymphoid organs, but it's expression can be induced in non-lymphoid tissues following inflammation or infection, and may help to recruit lymphocytes locally (258) to these tissues as a result of immune activation (286). It has been observed in samples of human lung tissue (287, 288), and is chemoattractant for naïve T cells and activated B cells (289), macrophage progenitor cells (290), and NK cells (291). It's role in sarcoidosis has been previously studied, with upregulation of CCL19 mRNA transcripts in the BAL cells of sarcoidosis patients compared to controls, and significantly elevated CCL19 protein levels in BAL fluid in patients with chest radiograph Scadding stage II when compared to stage I (292). Some work has been done on the CCL19-CCR7 axis in idiopathic diffuse alveolar damage (315) and idiopathic interstitial pneumonia (316), but its role in fibrotic sarcoidosis has not been studied.

ADAMDEC1 and MMP12, found to have more than ten fold change expression values in active fibrosis when compared to active patients without fibrosis, is of great interest. Both have been previously identified as likely mediators of lung damage and/or remodelling by Crouser et al. (168) through gene expression profiling, and ADAMDEC1 and MMP12 expression in BAL cells were seen to correlate with severity of disease. Macrophages are the major producers of MMP12 and ADAMDEC1. ADAMDEC1 features in the M1 gene set, whereas MMP12 is a feature of both M1 and M2. ADAMDEC1 is a secreted protein thought to be a member of the metalloproteinase family and MMP12 is a potent elastase, that degrades soluble and insoluble elastin, is involved in the breakdown of extracellular matrix in normal physiological processes such as embryonic development, reproduction, and tissue remodelling, but also in disease processes such as arthritis and metastasis (295). MMP12 has been proposed as an essential gene for the development of fibrosis following acute lung injury in a mouse model (296). Again, the role of these genes specifically in active fibrotic pulmonary sarcoidosis has not yet been elucidated, but provides a clear platform for further investigation.

On a final note, the gene signature derived to distinguish IPF lung biopsy samples from sarcoidosis lung biopsy samples (279), revealed surprising and promising results when tested in the lung samples from the active fibrotic cohort compared to active patients without fibrosis. Despite this signature being developed to specifically differentiate IPF from sarcoidosis (Appendices 8.6 and 8.7), the upregulated IPF signature was seen to be enriched in the fibrotic sarcoidosis group, and those genes seen to be downregulated in IPF (when compared to sarcoidosis) were seen to be enriched in the active, non-fibrotic group. Anti-fibrotic treatment has recently been

introduced for IPF, which has shown promise in slowing progression of the fibrotic process. If a significant overlap of genes and immune processes/pathways associated with IPF can also be seen in fibrotic sarcoidosis, this may also prompt the suggestion that anti-fibrotics may be of benefit in this sarcoidosis subgroup. The mechanism of action of the two available antifibrotic medications, Nintedanib and Pirfenidone, are different. Nintedanib is a potent small molecule inhibitor of the receptor tyrosine kinases PDGF receptor, FGF receptor and vascular endothelial growth factor receptor, and *in vitro* studies have shown that nintedanib interferes with processes active in fibrosis such as fibroblast proliferation, migration and differentiation, and the secretion of Extra Cellular Matrix (317). The mode of action of Pirfenidone however is less clear. It is known to inhibit cell proliferation, and reduce levels of mRNA encoding Collagen I and III (318). It has been shown to inhibit transforming growth factor (TGF)-β1-induced increase in fibronectin synthesis (319).

7.2 Limitations

The numbers available for the microarray cohorts, for both the sarcoidosis activity study and the fibrotic sarcoidosis study, are relatively small for each sampling modality. However, they have provided meaningful results, some of which have already been reported and validated in sarcoidosis but not in the same context, which lends more certainty to the findings. The sample cohorts, although smaller than ideal, are highly phenotyped and selected for their similar clinical condition, and therefore high quality and highly specific results will have been obtained. Whilst the mean SCAS observed in the fibrosis study was generally higher in the fibrotic

cohorts than in the non-fibrotic cohorts, the results do not parallel those seen in the activity cohort.

The genes identified as upregulated in both disease activity and fibrotic sarcoidosis ideally would have already undergone validation by means such as qPCR. In addition to this, as gene presence does not always correlate to protein production or function, these should also be necessary steps in confirming my findings. The implications of not having done this to arrive at the conclusions displayed in the discussion could result in follow-up research embarking on a futile project. However, the parallel of certain genes identified in this thesis, with those from other published data, and the degree to which these genes play a role in already recognized mandatory processes in sarcoidosis, strongly suggests strength in their identification.

7.3 Future Work

As the clinical heterogeneity of sarcoidosis is so extraordinary, a robust and widely accepted system for phenotyping sarcoidosis patients for organ involvement, activity and disease outcome (e.g. fibrosis *versus* no fibrosis) would enhance the detection of signals important for specific phenotypes. This in turn would enable more targeted therapy and pave the way for future trials in complex subgroups such as those with fibrotic sarcoidosis. I would like to contribute to a consensus opinion amongst the sarcoidosis research community to enable such studies across different research centres, and also put forward the potential for assessing response to treatment with the use of a scoring system such as the SCAS or CTAS.

A robust platform for future work, based on the differences between high and low disease activity, and between active sarcoidosis with and without fibrotic change, has been presented in this thesis. Validation of the presence of upregulated genes seen in these different cohorts at gene, protein, and functional levels, in particular examining TREM2, CCL18 and CCR2 in high disease activity, and CCL19, CCR7, ADAMDEC1 and MMP12 in the fibrosis cohort, will be the next step in further understanding this complex disease. This in turn may reveal pathways amenable to modulation. There are already CCR2 antagonists in trials for other diseases (213), and evaluating their effects on monocytes/macrophages from sarcoidosis patients in vitro would form an interesting experiment.

Insight into the kinetics of monocyte trafficking from the periphery into the granuloma could be examined by purifying monocyte/macrophage subsets from both peripheral blood and BAL (either through FACS sorting or MACS bead

separation), in control and pulmonary sarcoidosis patients, with and without fibrosis, and assessing their response to gradients of chemokines, with or without receptor blockade. This may address whether sarcoidosis monocytes/macrophages are primed to react in a more pathological manner. A Boyden chamber as part of a standard transwell migration assay could be used for this (320) , and the influence of the chemotactic factors and their receptors highlighted in this study for both activity and pulmonary fibrosis, namely, CCL2 and CCR2, CCXL9 and CXCR3, CCL19 and CCR7, CCL5, CCL14 and CCR1, and CXCL10 and CCL18, can be further understood.

The upregulation of TREM2 in sarcoidosis could potentially be investigated using a monocyte fusion assay (310, 321), establishing the level of expression of TREM2 in sarcoidosis compared to control derived fused monocytes, and then assessing the outcome of blocking the pathway. The role of IDO1 could also be investigated in this way.

Further study should also investigate the notion that what happens in the periphery can reflect what happens in the lung. Measuring the protein levels of CXCL9 and assessing the patient for contemporaneous CTAS or SCAS levels may aid its validation as a marker of disease activity in pulmonary sarcoidosis.

To complement and confirm the functional findings that ADAMDEC1 and MMP12 gene expression is upregulated in the BAL cells of patients with fibrotic pulmonary sarcoidosis, I would like to use Real time PCR (Polymerase Chain Reaction) to

quantify the protein in the BAL fluid from the same cohort of patients, and compare that to healthy controls.

Finally, a longitudinal study of disease phenotype, using gene expression profiling of samples at presentation of the disease, may answer the question of whether patients that develop fibrotic sarcoidosis are pre-disposed from the outset, or whether the persistence of granulomatous inflammation is responsible for the progression to fibrosis. This could then allow targeted treatment in this remarkable and heterogeneic disease entity.

8 Appendices

8.1 Definition of terms for abnormalities on HRCT scoring proforma. All from Fleischner Glossary of Terms apart from asterisked.

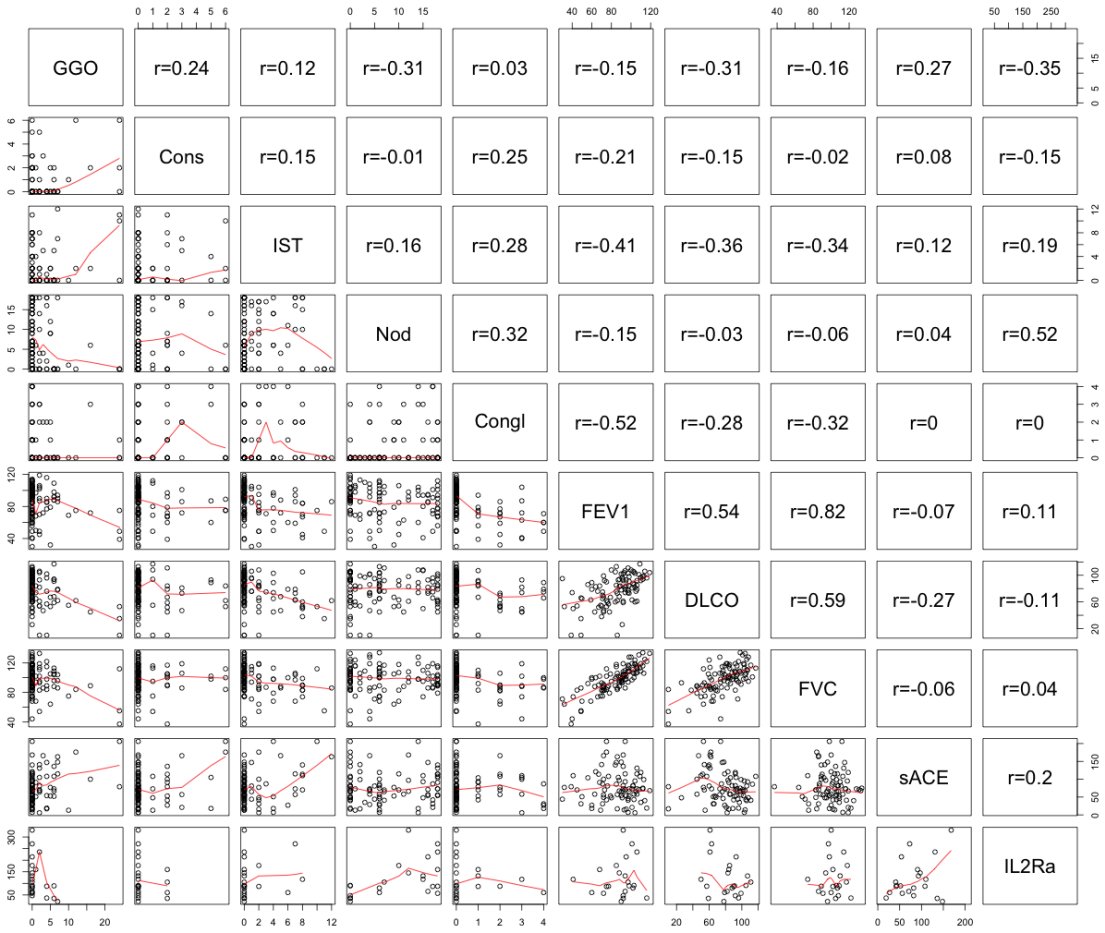
Terms	Definition
Ground glass opacity:	hazy increased lung opacity, with preserved bronchial and vascular margins.
Consolidation:	homogeneous increase in pulmonary parenchymal attenuation which obscures vessels and airway walls.
Interlobular septal thickening:	thickening of the septa between lobules.
Bronchocentric distribution:	disease conspicuously centred on macroscopic bronchovascuar bundles
Tree-in-bud pattern:	centrilobular branching nodular structures that resemble a budding tree
Centrilobular nodules:	nodules which appear separated by several mm from the pleural surfaces, fissures and interlobular septa; they may be soft tissue or ground glass attenuation and are usually ill defined.
Perilobular distribution:	distribution along the structures that border the pulmonary lobules (i.e. interlobular septa, visceral pleura and vessels)
Architectural distortion:	abnormal displacement of bronchi, vessels, fissures or septa and is usually accompanied by fibrosis and volume loss

Honeycombing:	clustered cystic airspaces with well-defined walls, usually 3-10mm in diameter
Intralobular linear opacities:	fine linear opacities within a lobule, when multiple appearing as a reticular pattern
Traction bronchiectasis:	irregular bronchial dilatation due to surrounding fibrosis
Cyst:	a round parenchymal lucency with a well defined interface with adjacent lung
Emphysema:	areas of low attenuation, usually without visible walls
Bronchiectasis:	bronchial dilatation with respect to the accompanying pulmonary artery, lack of tapering of bronchi and bronchi visible within 1cm of the pleural surface.
Linear atelectasis:	a focal area of subsegmental atelectasis with a linear configuration, almost always extending to the pleura
Part solid nodule:	consists of both ground glass and solid components
Mosaic attenuation pattern:	a patchwork of regions of differing attenuation
*Conglomeration:	a large opacity >3cm which often surrounded and encompassed the bronchi and vessels
*Galaxy sign:	a large nodule formed by coalescence of smaller nodules, often with many small surrounding nodules, resembling a galaxy
*Irregular linear opacities:	any linear opacity of irregular thickness of 1-3mm, distinct from interlobular septa, bronchovascular bundles and nodular opacities and without a precise topographic distribution.

8.2 Definition of zones on HRCT scoring proforma

Definition of Zones on HRCT	
Upper Zone	Predominantly above the carina in the cranio-caudal plane
Lower Zone	Predominantly below inferior pulmonary veins in the cranio-caudal plane
Mid Zone	Predominantly between UZ and LZ

8.3 Graphical Representation of Table 3.6



LOWESS lines in red, locally weighted polynomial regression; r calculated by Spearman Rank Sum Test. To read graph, first determine which two parameters you want to compare e.g. IST and DLCO, then draw two lines from the IST and DLCO boxes till they meet at the regression graph or until the lines meet for the r value (r=-0.36).Nod – Nodularity; Congl – Conglomeration; IST – Interlobular septal thickening, Cons – Consolidation.

8.4 GSE accession details for GSEA analysis

Gene set	GEO ID/Platform	Experimental design
M1/M2	GSE35495/Affymetrix	Genome wide analysis of human and mouse macrophages, resting and activated
"	GSE18686/Illumina Human HT-12 v3.0	Macrophages derived from human monocytes were cultures for 7 days then treated with IFNγ, IL4, TNF, LPS and LPS + IFNγ
"	GSE5099/Affymetrix	Monocytes were induced to mature to macrophages with M-CSF. Cells were then activated with IFNγ and LPS or IL-4
"	GSE35449/Illumina Human HT-12 v3.0	Total RNA was isolated from untreated, classically and alternatively activated human macrophages
IPF	GSE10667/Agilent Whole Human Genome Microarry	Lung tissue samples from IPF patients and controls underwent RNA extraction, for cDNA synthesis and expression profiling by array
"	GSE24206/Affymetrix	Lung tissue samples from IPF patients and controls underwent RNA extraction, for cDNA synthesis and expression profiling by array
Th1/Th2	GSE14308/Affymetrix	T helper subsets are profiled for mRNA
"	GSE11924/Affymetrix	Murine splenic T cells were stimulated into differentiation and gene expression analysed
Th1/Th17	GSE14026/ Affymetrix	Different murine T helper subsets are generated and profiled for mRNA
Th2/Th17	GSE14308/Affymetrix	Freshly isolated murine CD4+ T cells are cultured in Th1, Th2 and Th17 conditions and differentiation state verified with DNA microarray analysis of mRNA profiles.

8.5 KEGG_CHEMOKINE_SIGNALING_PATHWAY gene set

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
ADCY1	CCL22	CSK	GNB3	JAK3	PRKACA	TIAM1
ADCY2	CCL23	CX3CL1	GNB4	KRAS	PRKACB	TIAM2
ADCY3	CCL24	CX3CR1	GNB5	LYN	PRKACG	VAV1
ADCY4	CCL25	CXCL1	GNG10	MAP2K1	PRKCB	VAV2
ADCY5	CCL26	CXCL10	GNG11	MAPK1	PRKCD	VAV3
ADCY6	CCL27	CXCL11	GNG12	MAPK3	PRKCZ	WAS
ADCY7	CCL28	CXCL12	GNG13	NCF1	PRKX	WASL
ADCY8	CCL3	CXCL13	GNG2	NFKB1	PTK2	XCL1
ADCY9	CCL3L1	CXCL14	GNG3	NFKBIA	PTK2B	XCL2
ADRBK1	CCL3L3	CXCL16	GNG4	NFKBIB	PXN	XCR1
ADRBK2	CCL4	CXCL2	GNG5	NRAS	RAC1	
AKT1	CCL4L1	CXCL3	GNG7	PAK1	RAC2	
AKT2	CCL4L2	CXCL5	GNG8	PARD3	RAF1	
AKT3	CCL5	CXCL6	GNGT1	PF4	RAP1A	
ARRB1	CCL7	CXCL9	GNGT2	PF4V1	RAP1B	
ARRB2	CCL8	CXCR1	GRB2	PIK3CA	RASGRP2	
BCAR1	CCR1	CXCR2	GRK1	PIK3CB	RELA	
BRAF	CCR10	CXCR3	GRK4	PIK3CD	RHOA	
CCL1	CCR2	CXCR4	GRK5	PIK3CG	ROCK1	
CCL11	CCR3	CXCR5	GRK6	PIK3R1	ROCK2	
CCL13	CCR4	CXCR6	GRK7	PIK3R2	SHC1	
CCL14	CCR5	DOCK2	GSK3A	PIK3R3	SHC2	
CCL15	CCR6	ELMO1	GSK3B	PIK3R5	SHC3	
CCL16	CCR7	FGR	HCK	PLCB1	SHC4	
CCL17	CCR8	FOXO3	HRAS	PLCB2	SOS1	
CCL18	CCR9	GNAI1	IKBKB	PLCB3	SOS2	
CCL19	CDC42	GNAI2	IKBKG	PLCB4	STAT1	
CCL2	CHUK	GNAI3	IL8	PPBP	STAT2	
CCL20	CRK	GNB1	ITK	PPBPL1	STAT3	
CCL21	CRKL	GNB2	JAK2	PREX1	STAT5B	

8.6 IPF v Sarcoidosis Gene signature (downregulated) – Leng et al (279)

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
MSMB	SLC4A4	PFN2	C1orf106	LRRC43	SPA17	RFX2	ROPN1L	DYNLRB2
PLUNC	FA2H	NAT1	GJB5	SPAG16	NME5	AGBL4	EFCAB1	C9orf24
LTF	KIAA2022	DZIP3	GABRP	C1orf92	C14orf45	ERBB4	CAPS	FABP6
KLK12	CLDN1	C14orf142	SERPINB5	SPAG8	LRRC23	SLC4A8	WDR49	CLDN8
SAA4	PIGR	NPHP1	PROM2	CCDC37	IQCG	DIO1	TSGA10	C6
TF	EHF	DZIP1L	ALDH3A1	DLEC1	DNAJA4	TTC21A	DNAH5	ALOX15
CAPN13	TSPAN8	CCDC24	PTPRZ1	CRISP2	SLC22A4	TSNAXIP1	STK33	
HSD17B2	AGR2	DNAJC12	ADH7	SCGB2A1	KCNE1	C10orf92	C3orf15	
OXTR	SLITRK6	C9orf45	SERPINB13	SPATA4	C9orf116	KNDC1	C1orf114	
SPOCK3	FMN2	TJP3	KRT17	C22orf23	PACRG	ANKMY1	C14orf50	
OSTalpha	COL4A6	MAP1A	S100A2	RTDR1	WDR54	C20orf195	ARMC2	
CLDN10	POF1B	TCP11	FAT2	C8orf47	DNAH11	BAIAP3	SPAG17	
GDA	GRHL1	C10orf93	AKR1B10	CXorf41	PZP	KIF19	DNAH7	
THRSP	PPAP2C	GPR172B	KRT5	GSTA3	OSBPL6	RBM24	NEK10	
SCRG1	CYP2J2	CLDN3	KRT15	DNER	IFT172	ALDH1L1	C11orf16	
SCGB3A1	CLIC6	MUC20	GPR87	CXCL6	C20orf12	EFNB3	DNAI1	
SCGB1A1	EPN3	C8orf51	MMP10	TTLL10	NBEA	C6orf154	CCDC19	
SCNN1B	RHBDL2	RPRM	FUT2	PIP	KLHL13	NGEF	C6orf97	
KLK11	IRF6	CHGB	MMP7	CCNA1	FGF14	LRRC18	AK7	
KLK10	CDH26	GRP	CP	ANKRD43	MAK	CNGA3	CCDC65	
ZSCAN4	HSPA4L	C6orf142	SYT8	NXF5	NEK11	C22orf15	C15orf26	
HS6ST2	DSP	AMPD1	KLK13	ZNF214	FBXL13	C9orf98	SLC27A2	
SPDEF	KIF21A	ASB5	FCGBP	PPM1E	DYX1C1	WDR65	SPAG6	
TRPV4	BCAS1	TIMP4	COL17A1	RNF32	DNALI1	C14orf174	CAPSL	
HMGB3	WNK2	CRLF1	CEACAM5	FSIP1	EFHC1	BTG4	C10orf79	
ESR2	CXXC4	OTX1	BNC1	CLGN	CETN2	SLC23A1	CHST9	
WNT4	PRMT8	FOXN4	SOX2	PNLDC1	CAPS2	SPATS1	STOML3	
FAM83F	INHBB	HRK	ST6GALNAC1	HRASLS2	SERPINI2	C6orf118	DNAH9	
NMNAT2	IGFBP2	CYP2A13	EYA1	FAM3D	KIF6	WFDC6	ABCA13	
FAM83A	TNFRSF19	S100P	KCNJ16	TMEM17	FLJ40852	CXorf22	TEKT1	
SIX4	PRSS12	B3GNT3	MUC4	PLEKHB1	ELMOD1	SRD5A2	C1orf87	
SIX1	FZD3	SLN	VTCN1	HHLA2	TMC5	DNAH3	MS4A8B	
SRGAP3	CRISPLD1	GAS8	ATP12A	BBOX1	CDS1	SLC44A4	C20orf85	
EYA2	TTC12	ARVCF	LCN2	KIAA1377	PPP1R14C	MORN3	WDR78	
DST	MAT1A	TFAP2B	KRT23	PKIB	MLF1	TEKT2	AGBL2	
TMEM45A	LRRC56	TRIM29	CLCA2	ESRRG	C8orf34	ZMYND10	CASC1	
BMPR1B	KIF9	TMPRSS4	C20orf114	MYB	GRAMD1C	LRRC46	EFHB	
FHOD3	TSGA14	NTS	TTC16	KATNAL2	ENPP5	SPATA18	ARMC4	
CNTN3	SNCAIP	SUSD4	TEKT3	LRRC34	IL20RA	WDR63	PTPRT	
SERTAD4	TSPAN6	SLC4A11	RIBC2	IQCH	C11orf52	TTC25	CACNG6	

8.7 IPF v Sarcoidosis Gene signature (upregulated) – Leng et al (279)

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
SLC6A4	GCA	GIMAP8	RTN1	SERPINA1	SLIT2	HOXA5	ANKRD29
TMEM100	GPR65	GIMAP4	ITGAL	SCD	GYPE	SPOCK2	FRAS1
FCN3	BACH1	S100A4	SIGLEC9	DOK2	SEMA6A	HPCAL1	GRIA1
PPBP	PTPRC	MFNG	SIGLEC7	TREM1	PTPRB	ACVRL1	SLC1A1
SOSTDC1	LCP2	ITM2A	HCK	C1orf162	PAPSS2	COL4A4	ANXA3
HPGD	GCH1	FLI1	P2RX7	ALOX5AP	KCNMB4	MME	SGEF
SSTR1	EPSTI1	RRAS	PIK3AP1	CPM	DISP1	AFF3	EMP2
CA4	CLEC7A	FXYD6	CD14	ALOX15B	CDO1	CSF3R	C11orf9
SERPINE1	CLEC4A	GIMAP5	P2RY13	LIPG	AGTR1	CD300LG	CLIC5
IL1B	ZNF649	KLRK1	IFNG	FAP	PHLDB2	SCTR	CLIC3
THBS1	NLGN4X	DDX58	EGR2	CCL8	GJA1	RASGRF1	RETN
ICAM1	G0S2	APOL3	TFEC	CXCL9	GNG11	VEPH1	OLR1
HAS2	STX11	LAMB1	SNX10	CXCL11	ITGB6	LGI3	OASL
IL18RAP	SMAD7	CLIC2	BCAT1	LGALS2	ATP11A	F11	STAC
SOD2	FGD5	FAM105A	TLR8	CD86	LPL	CHI3L2	KCNK17
GPR84	VEGFC	WNT2	HK3	P2RY1	ATP8A2	SMAD6	CA2
CLEC4E	CLEC2B	SGIP1	CCR1	SH2D1B	SPARC	SFTPD	TMC7
CLEC4D	SERPINB9	RYR2	SLC6A12	PRF1	LRCH2	SCEL	CLCN5
SLC6A14	NRCAM	LAMA4	CD300LF	NKG7	RAPGEF4	HHIP	DPCR1
TNIP3	COL4A1	GUCY1B3	FBP1	GZMB	LPHN2	LRRK2	BDNF
IL1RN	SPRY4	BTN3A3	LYZ	GNLY	SDPR	LAMP3	ABCC13
TUBB1	DENND3	IFIT3	TLR7	TSPAN7	EMCN	ABCA3	CD5L
NFE2	KDR	BST2	AMICA1	TIE1	ARHGAP29	HSD17B6	APLN
ITGA2B	IL18R1	GPR64	CYBB	GPR146	ROBO4	CACNA2D2	IL1RL1
KLRF1	CLEC1A	PLA1A	ARRDC4	LRRC32	KL	SLC5A9	AGTR2
GRM8	SH2D3C	IFIT2	CYP2C9	THBD	TBX3	GUCY1A2	ZBED2
CLEC1B	ICAM2	CAMTA1	CCR2	TMEM2	NOTCH4	COL4A3	C4BPA
PDCD1LG2	HSPB8	SPHK1	HK2	HYAL2	ARHGEF10	GPR116	PLA2G1B
MCOLN2	SLCO2A1	GPR68	IGSF6	BCL6B	DACH1	DAPK2	FIGF
HSD11B1	ERG	CSTA	FGR	TEAD4	CDH19	MIA2	CLDN18
HSPA6	CD93	CD83	VAMP5	FADS3	RAMP3	EFNB2	AGER
TNFAIP6	ETS1	CD80	SCARF1	PTGIR	RAMP2	ANKRD50	
CD274	BMPR2	ADORA2A	CMKLR1	GPR4	FOXF1	TTN	
TAP2	SIRPD	HMOX1	ACE	PDGFB	KCNK3	ACOXL	
TAP1	BMP2	FCER1G	DICER1	RASIP1	DLC1	ACADL	
IL32	KCNN2	C1QA	ARL4C	NES	AQP1	SEMA6D	
PLEK	CPEB2	IL2RB	ADRB1	EGFL7	RSAD2	LRRN3	
TNFRSF1B	USP25	FSCN1	GZMH	LMCD1	IFI44L	GJA5	
CORO1A	RAP2A	KLRB1	CD244	FGF2	FLJ34503	NPR3	
ITK	RALA	CST7	CSF2	NRGN	STXBP6	HECW2	
TNFSF13B	NETO2	CETP	RGS18	LONRF3	EDNRB	CALCRL	
IL21R	PREX1	SPN	MNDA	S100A3	Gcom1	SEMA3G	
PIK3CG	PTPN12	CCL5	FCN1	ASRGL1	CYP3A5	PCDH12	
FMNL3	PAG1	ITGA5	S100A8	GATA2	SLCO1A2	PRX	
RASGRP1	WARS	GNAI2	AQP9	C20orf46	RGS9	LAMC3	
PLA2G4C	TNS3	ENG	RGS16	RSPO2	C6orf114	ST6GALNAC5	
CD2	CCRL2	CTSW	INHBA	RNF182	CLEC4M	SLC39A8	
CXCR6	HHEX	BATF2	HBEGF	NDRG4	IL7R	LEPREL1	
P2RY14	GIMAP7	SUCNR1	C8orf4	HYAL1	PDGFRA	ETV1	
OBFC2A	GIMAP6	SLC15A3	MSR1	TBX2	INMT	CAV1	

9 References

1. Hutchinson J. Case of livid papillary psoriasis. Illustrations of Clinical Surgery. 1877:42.
2. Besnier E. Lupus pernio de la face. Ann Dermatol Syphiligr (Paris). 1889;10:33-6.
3. Boeck C. Multiple Benign Sarkoid of the Skin (Nor) Norsk Mag Laegevidensk 1899;14:1321-34.
4. Kreibich K. Uber lupus pernio. Arch Dermatol Syph 1904;71:3-12.
5. Kuznitzky EB, A. Boecksches sarkoid mit beteiligung innerer organe. MMW Munch Med Wochenschr. 1915;62:1349-53.
6. Schaumann J. Lymphogranulomatosis benigna in the light of prolonged clinical observations and autopsy findings. Br J Dermatol. 1936;48:399-446.
7. Lofgren S. Primary Pulmonary Sarcoidosis. Acta Med Scand. 1953;145:424-55.
8. Hunninghake GW, Costabel U, Ando M, Baughman R, Cordier JF, du Bois R, et al. ATS/ERS/WASOG statement on sarcoidosis. American Thoracic Society/European Respiratory Society/World Association of Sarcoidosis and other Granulomatous Disorders. Sarcoidosis Vasc Diffuse Lung Dis. 1999 Sep;16(2):149-73. PubMed PMID: 10560120. Epub 1999/11/24. eng.
9. Siltzbach LE, James DG, Neville E, Turiaf J, Battesti JP, Sharma OP, et al. Course and prognosis of sarcoidosis around the world. Am J Med. 1974 Dec;57(6):847-52. PubMed PMID: 4432869.
10. Reich JM. A critical analysis of sarcoidosis incidence assessment. Multidisciplinary respiratory medicine. 2013;8(1):57. PubMed PMID: 24139391. Pubmed Central PMCID: 3765309.
11. Hennessy TW, Ballard DJ, DeRemee RA, Chu CP, Melton LJ, 3rd. The influence of diagnostic access bias on the epidemiology of sarcoidosis: a population-based study in Rochester, Minnesota, 1935-1984. Journal of clinical epidemiology. 1988;41(6):565-70. PubMed PMID: 3385457.
12. Pietinalho A, Hiraga Y, Hosoda Y, Lofroos AB, Yamaguchi M, Selroos O. The frequency of sarcoidosis in Finland and Hokkaido, Japan. A comparative epidemiological study. Sarcoidosis. 1995 Mar;12(1):61-7. PubMed PMID: 7617979. Epub 1995/03/01. eng.
13. Gribbin J, Hubbard RB, Le Jeune I, Smith CJ, West J, Tata LJ. Incidence and mortality of idiopathic pulmonary fibrosis and sarcoidosis in the UK. Thorax. 2006 Nov;61(11):980-5. PubMed PMID: 16844727. Pubmed Central PMCID: 2121155.
14. Rybicki BA, Major M, Popovich J, Jr., Maliarik MJ, Iannuzzi MC. Racial differences in sarcoidosis incidence: a 5-year study in a health maintenance organization. Am J Epidemiol. 1997 Feb 1;145(3):234-41. PubMed PMID: 9012596. Epub 1997/02/01. eng.
15. Baughman RP, Teirstein AS, Judson MA, Rossman MD, Yeager H, Jr., Bresnitz EA, et al. Clinical characteristics of patients in a case control study of sarcoidosis. Am J Respir Crit Care Med. 2001 Nov 15;164(10 Pt 1):1885-9. PubMed PMID: 11734441. Epub 2001/12/06. eng.
16. Byg KE, Milman N, Hansen S. Sarcoidosis in Denmark 1980-1994. A registry-based incidence study comprising 5536 patients. Sarcoidosis Vasc Diffuse Lung Dis. 2003 Mar;20(1):46-52. PubMed PMID: 12737280.

17. Newman LS, Rose CS, Bresnitz EA, Rossman MD, Barnard J, Frederick M, et al. A case control etiologic study of sarcoidosis: environmental and occupational risk factors. Am J Respir Crit Care Med. 2004 Dec 15;170(12):1324-30. PubMed PMID: 15347561. Epub 2004/09/07. eng.
18. Rybicki BA, Iannuzzi MC, Frederick MM, Thompson BW, Rossman MD, Bresnitz EA, et al. Familial aggregation of sarcoidosis. A case-control etiologic study of sarcoidosis (ACCESS). Am J Respir Crit Care Med. 2001 Dec 1;164(11):2085-91. PubMed PMID: 11739139. Epub 2001/12/12. eng.
19. Darlington P, Haugom-Olsen H, von Sivers K, Wahlstrom J, Runold M, Svjatoha V, et al. T-cell phenotypes in bronchoalveolar lavage fluid, blood and lymph nodes in pulmonary sarcoidosis--indication for an airborne antigen as the triggering factor in sarcoidosis. Journal of internal medicine. 2012 Nov;272(5):465-71. PubMed PMID: 22469005.
20. Moller DR. Potential etiologic agents in sarcoidosis. Proc Am Thorac Soc. 2007 Aug 15;4(5):465-8. PubMed PMID: 17684291. Pubmed Central PMCID: 2647598. Epub 2007/08/09. eng.
21. Alavi HA, Moscovic EA. Immunolocalization of cell-wall-deficient forms of Mycobacterium tuberculosis complex in sarcoidosis and in sinus histiocytosis of lymph nodes draining carcinoma. Histology and histopathology. 1996 Jul;11(3):683-94. PubMed PMID: 8839759.
22. Hance AJ. The role of mycobacteria in the pathogenesis of sarcoidosis. Seminars in respiratory infections. 1998 Sep;13(3):197-205. PubMed PMID: 9764950.
23. Gupta D, Agarwal R, Aggarwal AN, Jindal SK. Molecular evidence for the role of mycobacteria in sarcoidosis: a meta-analysis. Eur Respir J. 2007 Sep;30(3):508-16. PubMed PMID: 17537780.
24. Chen ES, Wahlstrom J, Song Z, Willett MH, Wiken M, Yung RC, et al. T cell responses to mycobacterial catalase-peroxidase profile a pathogenic antigen in systemic sarcoidosis. J Immunol. 2008 Dec 15;181(12):8784-96. PubMed PMID: 19050300. Pubmed Central PMCID: 2700300. Epub 2008/12/04. eng.
25. Yamada T, Eishi Y, Ikeda S, Ishige I, Suzuki T, Takemura T, et al. In situ localization of Propionibacterium acnes DNA in lymph nodes from sarcoidosis patients by signal amplification with catalysed reporter deposition. J Pathol. 2002 Dec;198(4):541-7. PubMed PMID: 12434425. Epub 2002/11/16. eng.
26. Negi M, Takemura T, Guzman J, Uchida K, Furukawa A, Suzuki Y, et al. Localization of propionibacterium acnes in granulomas supports a possible etiologic link between sarcoidosis and the bacterium. Mod Pathol. 2012 Sep;25(9):1284-97. PubMed PMID: 22596102. Pubmed Central PMCID: 3434354. Epub 2012/05/19. eng.
27. Homma JY, Abe C, Chosa H, Ueda K, Saegusa J, Nakayama M, et al. Bacteriological investigation on biopsy specimens from patients with sarcoidosis. The Japanese journal of experimental medicine. 1978 Jun;48(3):251-5. PubMed PMID: 713130.
28. Abe C, Iwai K, Mikami R, Hosoda Y. Frequent isolation of Propionibacterium acnes from sarcoidosis lymph nodes. Zentralblatt fur Bakteriologie, Mikrobiologie, und Hygiene Series A, Medical microbiology, infectious diseases, virology, parasitology. 1984 Apr;256(4):541-7. PubMed PMID: 6377763.
29. Ishige I, Eishi Y, Takemura T, Kobayashi I, Nakata K, Tanaka I, et al. Propionibacterium acnes is the most common bacterium commensal in peripheral lung tissue and mediastinal lymph nodes from subjects without sarcoidosis.

Sarcoidosis Vasc Diffuse Lung Dis. 2005 Mar;22(1):33-42. PubMed PMID: 15881278.
30. Sverrild A, Backer V, Kyvik KO, Kaprio J, Milman N, Svendsen CB, et al. Heredity in sarcoidosis: a registry-based twin study. Thorax. 2008 Oct;63(10):894-6. PubMed PMID: 18535119.
31. Iannuzzi MC, Rybicki BA. Genetics of sarcoidosis: candidate genes and genome scans. Proc Am Thorac Soc. 2007 Jan;4(1):108-16. PubMed PMID: 17202299. Pubmed Central PMCID: 2647608. Epub 2007/01/05. eng.
32. Fischer A, Grunewald J, Spagnolo P, Nebel A, Schreiber S, Muller-Quernheim J. Genetics of sarcoidosis. Semin Respir Crit Care Med. 2014 Jun;35(3):296-306. PubMed PMID: 25007082.
33. Martinetti M, Tinelli C, Kolek V, Cuccia M, Salvaneschi L, Pasturenzi L, et al. "The sarcoidosis map": a joint survey of clinical and immunogenetic findings in two European countries. Am J Respir Crit Care Med. 1995 Aug;152(2):557-64. PubMed PMID: 7633707. Epub 1995/08/01. eng.
34. Gardner J, Kennedy HG, Hamblin A, Jones E. HLA associations in sarcoidosis: a study of two ethnic groups. Thorax. 1984 Jan;39(1):19-22. PubMed PMID: 6582657. Pubmed Central PMCID: 459715. Epub 1984/01/01. eng.
35. Simon R, Radmacher MD, Dobbin K. Design of studies using DNA microarrays. Genet Epidemiol. 2002 Jun;23(1):21-36. PubMed PMID: 12112246. Epub 2002/07/12. eng.
36. Dobbin K, Simon R. Sample size determination in microarray experiments for class comparison and prognostic classification. Biostatistics. 2005 Jan;6(1):27-38. PubMed PMID: 15618525. Epub 2004/12/25. eng.
37. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001 Sep 11;98(19):10869-74. PubMed PMID: 11553815. Pubmed Central PMCID: 58566. Epub 2001/09/13. eng.
38. Grutters JC, Sato H, Welsh KI, du Bois RM. The importance of sarcoidosis genotype to lung phenotype. Am J Respir Cell Mol Biol. 2003 Sep;29(3 Suppl):S59-62. PubMed PMID: 14503556.
39. Grunewald J, Eklund A, Olerup O. Human leukocyte antigen class I alleles and the disease course in sarcoidosis patients. Am J Respir Crit Care Med. 2004 Mar 15;169(6):696-702. PubMed PMID: 14656748.
40. Grunewald J, Brynedal B, Darlington P, Nisell M, Cederlund K, Hillert J, et al. Different HLA-DRB1 allele distributions in distinct clinical subgroups of sarcoidosis patients. Respiratory research. 2010 Feb 26;11:25. PubMed PMID: 20187937. Pubmed Central PMCID: 2846896.
41. Berlin M, Fogdell-Hahn A, Olerup O, Eklund A, Grunewald J. HLA-DR predicts the prognosis in Scandinavian patients with pulmonary sarcoidosis. Am J Respir Crit Care Med. 1997 Nov;156(5):1601-5. PubMed PMID: 9372682.
42. Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, et al. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. Nat Genet. 2005 Apr;37(4):357-64. PubMed PMID: 15735647. Epub 2005/03/01. eng.
43. Hofmann S, Franke A, Fischer A, Jacobs G, Nothnagel M, Gaede KI, et al. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. Nat Genet. 2008 Sep;40(9):1103-6. PubMed PMID: 19165924. Epub 2009/01/24. eng.
44. Daniil Z, Mollaki V, Malli F, Koutsokera A, Antoniou KM, Rodopoulou P, et al. Polymorphisms and haplotypes in MyD88 are associated with the development of

sarcoidosis: a candidate-gene association study. Molecular biology reports. 2013 Jul;40(7):4281-6. PubMed PMID: 23666053.
45. Kim HS, Choi D, Lim LL, Allada G, Smith JR, Austin CR, et al. Association of interleukin 23 receptor gene with sarcoidosis. Disease markers. 2011;31(1):17-24. PubMed PMID: 21846945. Pubmed Central PMCID: 3826587.
46. Piotrowski WJ, Gorski P, Pietras T, Fendler W, Szemraj J. The selected genetic polymorphisms of metalloproteinases MMP2, 7, 9 and MMP inhibitor TIMP2 in sarcoidosis. Medical science monitor : international medical journal of experimental and clinical research. 2011 Oct;17(10):CR598-607. PubMed PMID: 21959615. Pubmed Central PMCID: 3539463.
47. Kruit A, Grutters JC, Ruven HJ, van Moorsel CH, Weiskirchen R, Mengsteab S, et al. Transforming growth factor-beta gene polymorphisms in sarcoidosis patients with and without fibrosis. Chest. 2006 Jun;129(6):1584-91. PubMed PMID: 16778279.
48. Pabst S, Franken T, Schonau J, Stier S, Nickenig G, Meyer R, et al. Transforming growth factor-{beta} gene polymorphisms in different phenotypes of sarcoidosis. Eur Respir J. 2011 Jul;38(1):169-75. PubMed PMID: 21148227.
49. Agostini C, Adami F, Semenzato G. New pathogenetic insights into the sarcoid granuloma. Current opinion in rheumatology. 2000 Jan;12(1):71-6. PubMed PMID: 10647958.
50. Zissel G, Prasse A, Muller-Quernheim J. Immunologic response of sarcoidosis. Semin Respir Crit Care Med. 2010 Aug;31(4):390-403. PubMed PMID: 20665389.
51. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. The New England journal of medicine. 2007 Nov 22;357(21):2153-65. PubMed PMID: 18032765. Epub 2007/11/23. eng.
52. Agostini C, Trentin L, Facco M, Sancetta R, Cerutti A, Tassinari C, et al. Role of IL-15, IL-2, and their receptors in the development of T cell alveolitis in pulmonary sarcoidosis. J Immunol. 1996 Jul 15;157(2):910-8. PubMed PMID: 8752945.
53. Gerke AK, Hunninghake G. The immunology of sarcoidosis. Clin Chest Med. 2008 Sep;29(3):379-90, vii. PubMed PMID: 18539233.
54. Katchar K, Eklund A, Grunewald J. Expression of Th1 markers by lung accumulated T cells in pulmonary sarcoidosis. Journal of internal medicine. 2003 Dec;254(6):564-71. PubMed PMID: 14641797.
55. Moller DR, Forman JD, Liu MC, Noble PW, Greenlee BM, Vyas P, et al. Enhanced expression of IL-12 associated with Th1 cytokine profiles in active pulmonary sarcoidosis. J Immunol. 1996 Jun 15;156(12):4952-60. PubMed PMID: 8648147.
56. Shigehara K, Shijubo N, Ohmichi M, Takahashi R, Kon S, Okamura H, et al. IL-12 and IL-18 are increased and stimulate IFN-gamma production in sarcoid lungs. J Immunol. 2001 Jan 01;166(1):642-9. PubMed PMID: 11123348.
57. Hunninghake GW, Glazier AJ, Monick MM, Dinarello CA. Interleukin-1 is a chemotactic factor for human T-lymphocytes. Am Rev Respir Dis. 1987 Jan;135(1):66-71. PubMed PMID: 2948432.
58. Hunninghake GW. Release of interleukin-1 by alveolar macrophages of patients with active pulmonary sarcoidosis. Am Rev Respir Dis. 1984 Apr;129(4):569-72. PubMed PMID: 6608889.
59. Gudmundsson G, Hunninghake GW. Interferon-gamma is necessary for the expression of hypersensitivity pneumonitis. J Clin Invest. 1997 May 15;99(10):2386-90. PubMed PMID: 9153280. Pubmed Central PMCID: 508077.

60. Robinson BW, McLemore TL, Crystal RG. Gamma interferon is spontaneously released by alveolar macrophages and lung T lymphocytes in patients with pulmonary sarcoidosis. J Clin Invest. 1985 May;75(5):1488-95. PubMed PMID: 3923038. Pubmed Central PMCID: 425487. Epub 1985/05/01. eng.
61. Baughman RP, Strohofer SA, Buchsbaum J, Lower EE. Release of tumor necrosis factor by alveolar macrophages of patients with sarcoidosis. J Lab Clin Med. 1990 Jan;115(1):36-42. PubMed PMID: 2299255.
62. Kunkel SL, Lukacs NW, Strieter RM, Chensue SW. Th1 and Th2 responses regulate experimental lung granuloma development. Sarcoidosis Vasc Diffuse Lung Dis. 1996 Sep;13(2):120-8. PubMed PMID: 8893380.
63. Luster AD, Unkeless JC, Ravetch JV. Gamma-interferon transcriptionally regulates an early-response gene containing homology to platelet proteins. Nature. 1985 Jun 20-26;315(6021):672-6. PubMed PMID: 3925348.
64. Agostini C, Cassatella M, Zambello R, Trentin L, Gasperini S, Perin A, et al. Involvement of the IP-10 chemokine in sarcoid granulomatous reactions. J Immunol. 1998 Dec 1;161(11):6413-20. PubMed PMID: 9834133.
65. Palchevskiy V, Hashemi N, Weigt SS, Xue YY, Derhovanessian A, Keane MP, et al. Immune response CC chemokines CCL2 and CCL5 are associated with pulmonary sarcoidosis. Fibrogenesis & tissue repair. 2011;4:10. PubMed PMID: 21463523. Pubmed Central PMCID: 3080805.
66. Cho ML, Yoon BY, Ju JH, Jung YO, Jhun JY, Park MK, et al. Expression of CCR2A, an isoform of MCP-1 receptor, is increased by MCP-1, CD40 ligand and TGF-beta in fibroblast like synoviocytes of patients with RA. Experimental & molecular medicine. 2007 Aug 31;39(4):499-507. PubMed PMID: 17934338.
67. Ho LP, Urban BC, Thickett DR, Davies RJ, McMichael AJ. Deficiency of a subset of T-cells with immunoregulatory properties in sarcoidosis. Lancet. 2005 Mar 19-25;365(9464):1062-72. PubMed PMID: 15781102.
68. Salez F, Gosset P, Copin MC, Lamblin Degros C, Tonnel AB, Wallaert B. Transforming growth factor-beta1 in sarcoidosis. Eur Respir J. 1998 Oct;12(4):913-9. PubMed PMID: 9817168.
69. Hauber HP, Gholami D, Meyer A, Pforte A. Increased interleukin-13 expression in patients with sarcoidosis. Thorax. 2003 Jun;58(6):519-24. PubMed PMID: 12775866. Pubmed Central PMCID: 1746705.
70. Baumer I, Zissel G, Schlaak M, Muller-Quernheim J. Th1/Th2 cell distribution in pulmonary sarcoidosis. Am J Respir Cell Mol Biol. 1997 Feb;16(2):171-7. PubMed PMID: 9032124.
71. Wynn TA, Cheever AW, Jankovic D, Poindexter RW, Caspar P, Lewis FA, et al. An IL-12-based vaccination method for preventing fibrosis induced by schistosome infection. Nature. 1995 Aug 17;376(6541):594-6. PubMed PMID: 7637808.
72. Wynn TA. Fibrotic disease and the T(H)1/T(H)2 paradigm. Nature reviews Immunology. 2004 Aug;4(8):583-94. PubMed PMID: 15286725. Pubmed Central PMCID: 2702150.
73. Emura M, Nagai S, Takeuchi M, Kitaichi M, Izumi T. In vitro production of B cell growth factor and B cell differentiation factor by peripheral blood mononuclear cells and bronchoalveolar lavage T lymphocytes from patients with idiopathic pulmonary fibrosis. Clinical and experimental immunology. 1990 Oct;82(1):133-9. PubMed PMID: 2208788. Pubmed Central PMCID: 1535146.
74. Wallace WA, Ramage EA, Lamb D, Howie SE. A type 2 (Th2-like) pattern of immune response predominates in the pulmonary interstitium of patients with

cryptogenic fibrosing alveolitis (CFA). Clinical and experimental immunology. 1995 Sep;101(3):436-41. PubMed PMID: 7664490. Pubmed Central PMCID: 1553239.
75. Rankin JA, Picarella DE, Geba GP, Temann UA, Prasad B, DiCosmo B, et al. Phenotypic and physiologic characterization of transgenic mice expressing interleukin 4 in the lung: lymphocytic and eosinophilic inflammation without airway hyperreactivity. Proc Natl Acad Sci U S A. 1996 Jul 23;93(15):7821-5. PubMed PMID: 8755560. Pubmed Central PMCID: 38832.
76. Zhu Z, Homer RJ, Wang Z, Chen Q, Geba GP, Wang J, et al. Pulmonary expression of interleukin-13 causes inflammation, mucus hypersecretion, subepithelial fibrosis, physiologic abnormalities, and eotaxin production. J Clin Invest. 1999 Mar;103(6):779-88. PubMed PMID: 10079098. Pubmed Central PMCID: 408149.
77. Tesmer LA, Lundy SK, Sarkar S, Fox DA. Th17 cells in human disease. Immunological reviews. 2008 Jun;223:87-113. PubMed PMID: 18613831. Pubmed Central PMCID: 3299089.
78. Kryczek I, Bruce AT, Gudjonsson JE, Johnston A, Aphale A, Vatan L, et al. Induction of IL-17+ T cell trafficking and development by IFN-gamma: mechanism and pathological relevance in psoriasis. J Immunol. 2008 Oct 01;181(7):4733-41. PubMed PMID: 18802076. Pubmed Central PMCID: 2677162.
79. Facco M, Cabrelle A, Teramo A, Olivieri V, Gnoato M, Teolato S, et al. Sarcoidosis is a Th1/Th17 multisystem disorder. Thorax. Feb;66(2):144-50. PubMed PMID: 21139119. Epub 2010/12/09. eng.
80. Simonian PL, Roark CL, Wehrmann F, Lanham AK, Diaz del Valle F, Born WK, et al. Th17-polarized immune response in a murine model of hypersensitivity pneumonitis and lung fibrosis. J Immunol. 2009 Jan 01;182(1):657-65. PubMed PMID: 19109199. Pubmed Central PMCID: 2766086.
81. Mosser DM, Edwards JP. Exploring the full spectrum of macrophage activation. Nature reviews Immunology. 2008 Dec;8(12):958-69. PubMed PMID: 19029990. Pubmed Central PMCID: 2724991.
82. Pechkovsky DV, Prasse A, Kollert F, Engel KM, Dentler J, Luttmann W, et al. Alternatively activated alveolar macrophages in pulmonary fibrosis-mediator production and intracellular signal transduction. Clinical immunology. 2010 Oct;137(1):89-101. PubMed PMID: 20674506.
83. Shamaei M, Mortaz E, Pourabdollah M, Garssen J, Tabarsi P, Velayati A, et al. Evidence for M2 macrophages in granulomas from pulmonary sarcoidosis: A new aspect of macrophage heterogeneity. Human immunology. 2018 Jan;79(1):63-9. PubMed PMID: 29107084.
84. Judson MA, Thompson BW, Rabin DL, Steimel J, Knattereud GL, Lackland DT, et al. The diagnostic pathway to sarcoidosis. Chest. 2003 Feb;123(2):406-12. PubMed PMID: 12576358.
85. Scadding JG. Prognosis of intrathoracic sarcoidosis in England. A review of 136 cases after five years' observation. Br Med J. 1961 Nov 4;2(5261):1165-72. PubMed PMID: 14497750. Pubmed Central PMCID: 1970202.
86. Baughman RP, Shipley R, Desai S, Drent M, Judson MA, Costabel U, et al. Changes in chest roentgenogram of sarcoidosis patients during a clinical trial of infliximab therapy: comparison of different methods of evaluation. Chest. 2009 Aug;136(2):526-35. PubMed PMID: 19395578.
87. Bradley B, Branley HM, Egan JJ, Greaves MS, Hansell DM, Harrison NK, et al. Interstitial lung disease guideline: the British Thoracic Society in collaboration

with the Thoracic Society of Australia and New Zealand and the Irish Thoracic Society. Thorax. 2008 Sep;63 Suppl 5:v1-58. PubMed PMID: 18757459.
88. Nunes H, Brillet PY, Valeyre D, Brauner MW, Wells AU. Imaging in sarcoidosis. Semin Respir Crit Care Med. 2007 Feb;28(1):102-20. PubMed PMID: 17330195.
89. Braun JJ, Imperiale A, Riehm S, Veillon F. Imaging in sinonasal sarcoidosis: CT, MRI, 67Gallium scintigraphy and 18F-FDG PET/CT features. Journal of neuroradiology Journal de neuroradiologie. 2010 Jul;37(3):172-81. PubMed PMID: 19959235.
90. Teirstein AS, Machac J, Almeida O, Lu P, Padilla ML, Iannuzzi MC. Results of 188 whole-body fluorodeoxyglucose positron emission tomography scans in 137 patients with sarcoidosis. Chest. 2007 Dec;132(6):1949-53. PubMed PMID: 17925421.
91. Gilman MJ, Wang KP. Transbronchial lung biopsy in sarcoidosis. An approach to determine the optimal number of biopsies. Am Rev Respir Dis. 1980 Nov;122(5):721-4. PubMed PMID: 7447156.
92. Garwood S, Judson MA, Silvestri G, Hoda R, Fraig M, Doelken P. Endobronchial ultrasound for the diagnosis of pulmonary sarcoidosis. Chest. 2007 Oct;132(4):1298-304. PubMed PMID: 17890467.
93. Tremblay A, Stather DR, MacEachern P, Khalil M, Field SK. A randomized controlled trial of standard vs endobronchial ultrasonography-guided transbronchial needle aspiration in patients with suspected sarcoidosis. Chest. 2009 Aug;136(2):340-6. PubMed PMID: 19188547.
94. Nagai S, Izumi T. Bronchoalveolar lavage. Still useful in diagnosing sarcoidosis? Clin Chest Med. 1997 Dec;18(4):787-97. PubMed PMID: 9413658.
95. Studdy PR, Bird R, Neville E, James DG. Biochemical findings in sarcoidosis. Journal of clinical pathology. 1980 Jun;33(6):528-33. PubMed PMID: 6249857. Pubmed Central PMCID: 1146136.
96. Studdy PR, Lapworth R, Bird R. Angiotensin-converting enzyme and its clinical significance--a review. Journal of clinical pathology. 1983 Aug;36(8):938-47. PubMed PMID: 6308066. Pubmed Central PMCID: 498427.
97. Sah BP, Goyal S, Iannuzzi MC. Novel pharmacotherapy of sarcoidosis. Pharmacology & therapeutics. 2016 Jan;157:1-9. PubMed PMID: 26478443.
98. Cronstein B. How does methotrexate suppress inflammation? Clinical and experimental rheumatology. 2010 Sep-Oct;28(5 Suppl 61):S21-3. PubMed PMID: 21044428.
99. Goljan-Geremek A, Bednarek M, Franczuk M, Puscinska E, Nowinski A, Czystowska M, et al. Methotrexate as a single agent for treating pulmonary sarcoidosis: a single centre real-life prospective study. Pneumonologia i alergologia polska. 2014;82(6):518-33. PubMed PMID: 25339562.
100. Baughman RP, Winget DB, Lower EE. Methotrexate is steroid sparing in acute sarcoidosis: results of a double blind, randomized trial. Sarcoidosis Vasc Diffuse Lung Dis. 2000 Mar;17(1):60-6. PubMed PMID: 10746262.
101. Pacheco Y, Marechal C, Marechal F, Biot N, Perrin Fayolle M. Azathioprine treatment of chronic pulmonary sarcoidosis. Sarcoidosis. 1985 Sep;2(2):107-13. PubMed PMID: 3031790.
102. Vorselaars ADM, Wuyts WA, Vorselaars VMM, Zanen P, Deneer VHM, Veltkamp M, et al. Methotrexate vs azathioprine in second-line therapy of sarcoidosis. Chest. 2013 Sep;144(3):805-12. PubMed PMID: 23538719.

103. Brill AK, Ott SR, Geiser T. Effect and safety of mycophenolate mofetil in chronic pulmonary sarcoidosis: a retrospective study. Respiration; international review of thoracic diseases. 2013;86(5):376-83. PubMed PMID: 23295253.
104. Baughman RP, Drent M, Kavuru M, Judson MA, Costabel U, du Bois R, et al. Infliximab therapy in patients with chronic sarcoidosis and pulmonary involvement. Am J Respir Crit Care Med. 2006 Oct 01;174(7):795-802. PubMed PMID: 16840744.
105. Rossman MD, Newman LS, Baughman RP, Teirstein A, Weinberger SE, Miller W, Jr., et al. A double-blinded, randomized, placebo-controlled trial of infliximab in subjects with active pulmonary sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis. 2006 Oct;23(3):201-8. PubMed PMID: 18038919.
106. Mana J, Gomez-Vaquero C, Montero A, Salazar A, Marcoval J, Valverde J, et al. Lofgren's syndrome revisited: a study of 186 patients. Am J Med. 1999 Sep;107(3):240-5. PubMed PMID: 10492317.
107. Judson MA, Baughman RP, Thompson BW, Teirstein AS, Terrin ML, Rossman MD, et al. Two year prognosis of sarcoidosis: the ACCESS experience. Sarcoidosis Vasc Diffuse Lung Dis. 2003 Oct;20(3):204-11. PubMed PMID: 14620163. Epub 2003/11/19. eng.
108. Nagai S, Shigematsu M, Hamada K, Izumi T. Clinical courses and prognoses of pulmonary sarcoidosis. Current opinion in pulmonary medicine. 1999 Sep;5(5):293-8. PubMed PMID: 10461533. Epub 1999/08/26. eng.
109. Baughman RP, Judson MA, Teirstein A, Yeager H, Rossman M, Knatterud GL, et al. Presenting characteristics as predictors of duration of treatment in sarcoidosis. QJM : monthly journal of the Association of Physicians. 2006 May;99(5):307-15. PubMed PMID: 16595563.
110. Neville E, Walker AN, James DG. Prognostic factors predicting the outcome of sarcoidosis: an analysis of 818 patients. The Quarterly journal of medicine. 1983 Autumn;52(208):525-33. PubMed PMID: 6657915.
111. Hillerdal G, Nou E, Osterman K, Schmekel B. Sarcoidosis: epidemiology and prognosis. A 15-year European study. Am Rev Respir Dis. 1984 Jul;130(1):29-32. PubMed PMID: 6742607.
112. Judson MA, Boan AD, Lackland DT. The clinical course of sarcoidosis: presentation, diagnosis, and treatment in a large white and black cohort in the United States. Sarcoidosis Vasc Diffuse Lung Dis. 2012 Oct;29(2):119-27. PubMed PMID: 23461074.
113. Reich JM, Johnson RE. Course and prognosis of sarcoidosis in a nonreferral setting. Analysis of 86 patients observed for 10 years. Am J Med. 1985 Jan;78(1):61-7. PubMed PMID: 3966490.
114. Judson MA, Baughman RP, Teirstein AS, Terrin ML, Yeager H, Jr. Defining organ involvement in sarcoidosis: the ACCESS proposed instrument. ACCESS Research Group. A Case Control Etiologic Study of Sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis. 1999 Mar;16(1):75-86. PubMed PMID: 10207945.
115. Drent M, Lower EE, De Vries J. Sarcoidosis-associated fatigue. Eur Respir J. 2012 Jul;40(1):255-63. PubMed PMID: 22441750.
116. Cox CE, Donohue JF, Brown CD, Kataria YP, Judson MA. The Sarcoidosis Health Questionnaire: a new measure of health-related quality of life. Am J Respir Crit Care Med. 2003 Aug 1;168(3):323-9. PubMed PMID: 12738606.
117. https://http://www.optum.com/optum-outcomes/what-we-do/health-surveys/sf-36v2-health-survey.html.

118. Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. Respiratory medicine. 1991 Sep;85 Suppl B:25-31; discussion 3-7. PubMed PMID: 1759018.
119. Michielsen HJ, Drent M, Peros-Golubicic T, De Vries J. Fatigue is associated with quality of life in sarcoidosis patients. Chest. 2006 Oct;130(4):989-94. PubMed PMID: 17035429.
120. Consensus conference: activity of sarcoidosis. Third WASOG meeting, Los Angeles, USA, September 8-11, 1993. Eur Respir J. 1994 Mar;7(3):624-7. PubMed PMID: 8013622. Epub 1994/03/01. eng.
121. Lieberman J. Elevation of serum angiotensin-converting-enzyme (ACE) level in sarcoidosis. Am J Med. 1975 Sep;59(3):365-72. PubMed PMID: 169692. Epub 1975/09/01. eng.
122. Silverstein E, Friedland J, Kitt M, Lyons HA. Increased serum angiotensin converting enzyme activity in sarcoidosis. Israel journal of medical sciences. 1977 Oct;13(10):995-1000. PubMed PMID: 201593.
123. Brice EA, Friedlander W, Bateman ED, Kirsch RE. Serum angiotensin-converting enzyme activity, concentration, and specific activity in granulomatous interstitial lung disease, tuberculosis, and COPD. Chest. 1995 Mar;107(3):706-10. PubMed PMID: 7874941.
124. Kennon B, Connell JM. ACE gene polymorphism and diabetic complications: is there a connection? BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy. 2000 Aug;14(2):73-81. PubMed PMID: 18034559.
125. Smallridge RC, Rogers J, Verma PS. Serum angiotensin-converting enzyme. Alterations in hyperthyroidism, hypothyroidism, and subacute thyroiditis. Jama. 1983 Nov 11;250(18):2489-93. PubMed PMID: 6313976.
126. Lawrence EC, Brousseau KP, Berger MB, Kurman CC, Marcon L, Nelson DL. Elevated concentrations of soluble interleukin-2 receptors in serum samples and bronchoalveolar lavage fluids in active sarcoidosis. Am Rev Respir Dis. 1988 Apr;137(4):759-64. PubMed PMID: 3128148. Epub 1988/04/01. eng.
127. Kobayashi J, Kitamura S. Serum KL-6 for the evaluation of active pneumonitis in pulmonary sarcoidosis. Chest. 1996 May;109(5):1276-82. PubMed PMID: 8625680. Epub 1996/05/01. eng.
128. Miyoshi S, Hamada H, Kadowaki T, Hamaguchi N, Ito R, Irifune K, et al. Comparative evaluation of serum markers in pulmonary sarcoidosis. Chest. 2010 Jun;137(6):1391-7. PubMed PMID: 20081103.
129. Sweiss NJ, Salloum R, Gandhi S, Alegre ML, Sawaqed R, Badaracco M, et al. Significant CD4, CD8, and CD19 lymphopenia in peripheral blood of sarcoidosis patients correlates with severe disease manifestations. PLoS One. 2010;5(2):e9088. PubMed PMID: 20140091. Pubmed Central PMCID: 2816716.
130. Selroos O, Koivunen E. Prognostic significance of lymphopenia in sarcoidosis. Acta Med Scand. 1979;206(4):259-62. PubMed PMID: 506798.
131. Bell DY, Johnson SM, Piantadosi CA. Elevated serum immunoglobulin G levels and bronchoalveolar lymphocytosis as predictors of clinical course in pulmonary sarcoidosis. Ann N Y Acad Sci. 1986;465:672-7. PubMed PMID: 3460402. Epub 1986/01/01. eng.
132. Lawrence EC, Teague RB, Gottlieb MS, Jhingran SG, Lieberman J. Serial changes in markers of disease activity with corticosteroid treatment in sarcoidosis. Am J Med. 1983 May;74(5):747-56. PubMed PMID: 6301276.

133. Ghose T, Landrigan P, Asif A. Localization of immunoglobulin and complement in pulmonary sarcoid granulomas. Chest. 1974 Sep;66(3):264-8. PubMed PMID: 4608492.
134. Boot RG, Verhoek M, de Fost M, Hollak CE, Maas M, Bleijlevens B, et al. Marked elevation of the chemokine CCL18/PARC in Gaucher disease: a novel surrogate marker for assessing therapeutic intervention. Blood. 2004 Jan 1;103(1):33-9. PubMed PMID: 12969956.
135. Bargagli E, Bennett D, Maggiorelli C, Di Sipio P, Margollicci M, Bianchi N, et al. Human chitotriosidase: a sensitive biomarker of sarcoidosis. Journal of clinical immunology. 2013 Jan;33(1):264-70. PubMed PMID: 22878841.
136. Boot RG, Hollak CE, Verhoek M, Alberts C, Jonkers RE, Aerts JM. Plasma chitotriosidase and CCL18 as surrogate markers for granulomatous macrophages in sarcoidosis. Clin Chim Acta. Jan;411(1-2):31-6. PubMed PMID: 19808030. Epub 2009/10/08. eng.
137. Moodley YP, Dorasamy T, Venketasamy S, Naicker V, Lalloo UG. Correlation of CD4:CD8 ratio and tumour necrosis factor (TNF)alpha levels in induced sputum with bronchoalveolar lavage fluid in pulmonary sarcoidosis. Thorax. 2000 Aug;55(8):696-9. PubMed PMID: 10899249. Pubmed Central PMCID: 1745837.
138. Shorr AF, Torrington KG, Parker JM. Serum angiotensin converting enzyme does not correlate with radiographic stage at initial diagnosis of sarcoidosis. Respiratory medicine. 1997 Aug;91(7):399-401. PubMed PMID: 9327039.
139. Ward K, O'Connor C, Odlum C, Fitzgerald MX. Prognostic value of bronchoalveolar lavage in sarcoidosis: the critical influence of disease presentation. Thorax. 1989 Jan;44(1):6-12. PubMed PMID: 2928989. Pubmed Central PMCID: 461656.
140. Baughman RP, Shipley R, Eisentrout CE. Predictive value of gallium scan, angiotensin-converting enzyme level, and bronchoalveolar lavage in two-year follow-up of pulmonary sarcoidosis. Lung. 1987;165(6):371-7. PubMed PMID: 2828772.
141. Klech H, Kohn H, Kummer F, Mostbeck A. Assessment of activity in Sarcoidosis. Sensitivity and specificity of 67Gallium scintigraphy, serum ACE levels, chest roentgenography, and blood lymphocyte subpopulations. Chest. 1982 Dec;82(6):732-8. PubMed PMID: 6291864.
142. Baumann MH, Strange C, Sahn SA. Do chest radiographic findings reflect the clinical course of patients with sarcoidosis during corticosteroid withdrawal? AJR American journal of roentgenology. 1990 Mar;154(3):481-5. PubMed PMID: 2106208.
143. Guidelines for the use of the ILO International Classification of Radiographs of Pneumoconioses, Revised edition 2011. Occupational Safety and Health Series No 22 (Rev 2011). 2011:978-92-2-124542-1(pdf)[ISBN].
144. Muller NL, Mawson JB, Mathieson JR, Abboud R, Ostrow DN, Champion P. Sarcoidosis: correlation of extent of disease at CT with clinical, functional, and radiographic findings. Radiology. 1989 Jun;171(3):613-8. PubMed PMID: 2717730.
145. Line BR, Hunninghake GW, Keogh BA, Jones AE, Johnston GS, Crystal RG. Gallium-67 scanning to stage the alveolitis of sarcoidosis: correlation with clinical studies, pulmonary function studies, and bronchoalveolar lavage. Am Rev Respir Dis. 1981 Apr;123(4 Pt 1):440-6. PubMed PMID: 7224356.
146. Baughman RP, Fernandez M, Bosken CH, Mantil J, Hurtubise P. Comparison of gallium-67 scanning, bronchoalveolar lavage, and serum angiotensin-converting

enzyme levels in pulmonary sarcoidosis. Predicting response to therapy. Am Rev Respir Dis. 1984 May;129(5):676-81. PubMed PMID: 6326626.
147. Oberstein A, von Zitzewitz H, Schweden F, Muller-Quernheim J. Non invasive evaluation of the inflammatory activity in sarcoidosis with high-resolution computed tomography. Sarcoidosis Vasc Diffuse Lung Dis. 1997 Mar;14(1):65-72. PubMed PMID: 9186991. Epub 1997/03/01. eng.
148. Leung AN, Brauner MW, Caillat-Vigneron N, Valeyre D, Grenier P. Sarcoidosis activity: correlation of HRCT findings with those of 67Ga scanning, bronchoalveolar lavage, and serum angiotensin-converting enzyme assay. J Comput Assist Tomogr. 1998 Mar-Apr;22(2):229-34. PubMed PMID: 9530385. Epub 1998/04/08. eng.
149. Nishimura K, Itoh H, Kitaichi M, Nagai S, Izumi T. CT and pathological correlation of pulmonary sarcoidosis. Semin Ultrasound CT MR. 1995 Oct;16(5):361-70. PubMed PMID: 8527169. Epub 1995/10/01. eng.
150. Nishimura K, Itoh H, Kitaichi M, Nagai S, Izumi T. Pulmonary sarcoidosis: correlation of CT and histopathologic findings. Radiology. 1993 Oct;189(1):105-9. PubMed PMID: 8372178. Epub 1993/10/01. eng.
151. Keijsers RG, Grutters JC, van Velzen-Blad H, van den Bosch JM, Oyen WJ, Verzijlbergen FJ. (18)F-FDG PET patterns and BAL cell profiles in pulmonary sarcoidosis. European journal of nuclear medicine and molecular imaging. 2010 Jun;37(6):1181-8. PubMed PMID: 20157709.
152. Adams H, Keijsers RG, Korenromp IH, Grutters JC. FDG PET for gauging of sarcoid disease activity. Semin Respir Crit Care Med. 2014 Jun;35(3):352-61. PubMed PMID: 25007087.
153. Sobic-Saranovic D, Grozdic I, Videnovic-Ivanov J, Vucinic-Mihailovic V, Artiko V, Saranovic D, et al. The utility of 18F-FDG PET/CT for diagnosis and adjustment of therapy in patients with active chronic sarcoidosis. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2012 Oct;53(10):1543-9. PubMed PMID: 22879080.
154. Patterson KC, Hogarth K, Husain AN, Sperling AI, Niewold TB. The clinical and immunologic features of pulmonary fibrosis in sarcoidosis. Translational research : the journal of laboratory and clinical medicine. 2012 Nov;160(5):321-31. PubMed PMID: 22683422. Pubmed Central PMCID: 3910531.
155. Wynn TA, Ramalingam TR. Mechanisms of fibrosis: therapeutic translation for fibrotic disease. Nature medicine. 2012 Jul 06;18(7):1028-40. PubMed PMID: 22772564. Pubmed Central PMCID: 3405917.
156. Wynn TA. Integrating mechanisms of pulmonary fibrosis. The Journal of experimental medicine. 2011 Jul 04;208(7):1339-50. PubMed PMID: 21727191. Pubmed Central PMCID: 3136685.
157. Wynn TA, Barron L. Macrophages: master regulators of inflammation and fibrosis. Seminars in liver disease. 2010 Aug;30(3):245-57. PubMed PMID: 20665377. Pubmed Central PMCID: 2924662.
158. Zuo F, Kaminski N, Eugui E, Allard J, Yakhini Z, Ben-Dor A, et al. Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. Proc Natl Acad Sci U S A. 2002 Apr 30;99(9):6292-7. PubMed PMID: 11983918. Pubmed Central PMCID: 122942.
159. Jiang D, Liang J, Fan J, Yu S, Chen S, Luo Y, et al. Regulation of lung injury and repair by Toll-like receptors and hyaluronan. Nature medicine. 2005 Nov;11(11):1173-9. PubMed PMID: 16244651.

160. Song E, Ouyang N, Horbelt M, Antus B, Wang M, Exton MS. Influence of alternatively and classically activated macrophages on fibrogenic activities of human fibroblasts. Cellular immunology. 2000 Aug 25;204(1):19-28. PubMed PMID: 11006014.
161. Statement on sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and by the ERS Executive Committee, February 1999. Am J Respir Crit Care Med. 1999 Aug;160(2):736-55. PubMed PMID: 10430755.
162. Shigemitsu H, Oblad JM, Sharma OP, Koss MN. Chronic interstitial pneumonitis in end-stage sarcoidosis. Eur Respir J. 2010 Mar;35(3):695-7. PubMed PMID: 20190336.
163. Meltzer EB, Noble PW. Idiopathic pulmonary fibrosis. Orphanet journal of rare diseases. 2008 Mar 26;3:8. PubMed PMID: 18366757. Pubmed Central PMCID: 2330030.
164. Nuyten DS, van de Vijver MJ. Using microarray analysis as a prognostic and predictive tool in oncology: focus on breast cancer and normal tissue toxicity. Seminars in radiation oncology. 2008 Apr;18(2):105-14. PubMed PMID: 18314065.
165. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. 1987 Jan 9;235(4785):177-82. PubMed PMID: 3798106.
166. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science. 1989 May 12;244(4905):707-12. PubMed PMID: 2470152.
167. Selman M, Pardo A, Barrera L, Estrada A, Watson SR, Wilson K, et al. Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity pneumonitis. Am J Respir Crit Care Med. 2006 Jan 15;173(2):188-98. PubMed PMID: 16166619. Pubmed Central PMCID: 2662988.
168. Crouser ED, Culver DA, Knox KS, Julian MW, Shao G, Abraham S, et al. Gene expression profiling identifies MMP-12 and ADAMDEC1 as potential pathogenic mediators of pulmonary sarcoidosis. Am J Respir Crit Care Med. 2009 May 15;179(10):929-38. PubMed PMID: 19218196. Pubmed Central PMCID: 2684019. Epub 2009/02/17. eng.
169. Lockstone HE, Sanderson S, Kulakova N, Baban D, Leonard A, Kok WL, et al. Gene set analysis of lung samples provides insight into pathogenesis of progressive, fibrotic pulmonary sarcoidosis. Am J Respir Crit Care Med. 2010 Jun 15;181(12):1367-75. PubMed PMID: 20194811. Epub 2010/03/03. eng.
170. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J. 2005 Nov;26(5):948-68. PubMed PMID: 16264058.
171. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. Biotechniques. 2003 Feb;34(2):374-8. PubMed PMID: 12613259. Epub 2003/03/05. eng.
172. Computing RFfS. A language and environment for statistical computing. Vienna, Austria. 2013.
173. Valeyre D, Prasse A, Nunes H, Uzunhan Y, Brillet PY, Muller-Quernheim J. Sarcoidosis. Lancet. 2014 Mar 29;383(9923):1155-67. PubMed PMID: 24090799.
174. Abe S, Munakata M, Nishimura M, Tsuneta Y, Terai T, Nakano I, et al. Gallium-67 scintigraphy, bronchoalveolar lavage, and pathologic changes in

patients with pulmonary sarcoidosis. Chest. 1984 May;85(5):650-5. PubMed PMID: 6325097.
175. Caruso C, Candore G, Cigna D, Colucci AT, Modica MA. Biological significance of soluble IL-2 receptor. Mediators of inflammation. 1993;2(1):3-21. PubMed PMID: 18475497. Pubmed Central PMCID: 2365387.
176. Harlander M, Salobir B, Zupancic M, Dolensek M, Bavcar Vodovnik T, Tercelj M. Serial chitotriosidase measurements in sarcoidosis--two to five year follow-up study. Respiratory medicine. 2014 May;108(5):775-82. PubMed PMID: 24594143.
177. Homolka J, Lorenz J, Zuchold HD, Muller-Quernheim J. Evaluation of soluble CD 14 and neopterin as serum parameters of the inflammatory activity of pulmonary sarcoidosis. The Clinical investigator. 1992 Oct;70(10):909-16. PubMed PMID: 1280495.
178. Keir G, Wells AU. Assessing pulmonary disease and response to therapy: which test? Semin Respir Crit Care Med. 2010 Aug;31(4):409-18. PubMed PMID: 20665391.
179. Ober C, Chupp GL. The chitinase and chitinase-like proteins: a review of genetic and functional studies in asthma and immune-mediated diseases. Current opinion in allergy and clinical immunology. 2009 Oct;9(5):401-8. PubMed PMID: 19644363. Pubmed Central PMCID: 3018778.
180. Spagnolo P, Sverzellati N, Wells AU, Hansell DM. Imaging aspects of the diagnosis of sarcoidosis. European radiology. 2014 Apr;24(4):807-16. PubMed PMID: 24399491.
181. Treglia G, Annunziata S, Sobic-Saranovic D, Bertagna F, Caldarella C, Giovanella L. The role of 18F-FDG-PET and PET/CT in patients with sarcoidosis: an updated evidence-based review. Academic radiology. 2014 May;21(5):675-84. PubMed PMID: 24703482.
182. Schurmann M. Angiotensin-converting enzyme gene polymorphisms in patients with pulmonary sarcoidosis: impact on disease severity. American journal of pharmacogenomics : genomics-related research in drug development and clinical practice. 2003;3(4):233-43. PubMed PMID: 12930157.
183. Chen ES, Moller DR. Etiologies of Sarcoidosis. Clinical reviews in allergy & immunology. 2015 Aug;49(1):6-18. PubMed PMID: 25771769.
184. Abehsera M, Valeyre D, Grenier P, Jaillet H, Battesti JP, Brauner MW. Sarcoidosis with pulmonary fibrosis: CT patterns and correlation with pulmonary function. AJR American journal of roentgenology. 2000 Jun;174(6):1751-7. PubMed PMID: 10845518.
185. Akira M, Kozuka T, Inoue Y, Sakatani M. Long-term follow-up CT scan evaluation in patients with pulmonary sarcoidosis. Chest. 2005 Jan;127(1):185-91. PubMed PMID: 15653982.
186. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Muller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. Radiology. 2008 Mar;246(3):697-722. PubMed PMID: 18195376. Epub 2008/01/16. eng.
187. Walsh SL, Hansell DM. High-resolution CT of interstitial lung disease: a continuous evolution. Semin Respir Crit Care Med. 2014 Feb;35(1):129-44. PubMed PMID: 24481766.
188. Drent M, De Vries J, Lenters M, Lamers RJ, Rothkranz-Kos S, Wouters EF, et al. Sarcoidosis: assessment of disease severity using HRCT. European radiology. 2003 Nov;13(11):2462-71. PubMed PMID: 12811502.
189. Remy-Jardin M, Giraud F, Remy J, Wattinne L, Wallaert B, Duhamel A. Pulmonary sarcoidosis: role of CT in the evaluation of disease activity and functional

impairment and in prognosis assessment. Radiology. 1994 Jun;191(3):675-80. PubMed PMID: 8184045. Epub 1994/06/01. eng.
190. Zappala CJ, Desai SR, Copley SJ, Spagnolo P, Sen D, Alam SM, et al. Accuracy of individual variables in the monitoring of long-term change in pulmonary sarcoidosis as judged by serial high-resolution CT scan data. Chest. 2014 Jan;145(1):101-7. PubMed PMID: 24051950.
191. Prasse A, Katic C, Germann M, Buchwald A, Zissel G, Muller-Quernheim J. Phenotyping sarcoidosis from a pulmonary perspective. Am J Respir Crit Care Med. 2008 Feb 1;177(3):330-6. PubMed PMID: 17975200. Epub 2007/11/03. eng.
192. Kataria YP, Holter JF. Immunology of sarcoidosis. Clin Chest Med. 1997 Dec;18(4):719-39. PubMed PMID: 9413655.
193. Asano M, Minagawa T, Ohmichi M, Hiraga Y. Detection of endogenous cytokines in sera or in lymph nodes obtained from patients with sarcoidosis. Clinical and experimental immunology. 1991 Apr;84(1):92-6. PubMed PMID: 1849808. Pubmed Central PMCID: 1535346.
194. Crawshaw A, Kendrick YR, McMichael AJ, Ho LP. Abnormalities in iNKT cells are associated with impaired ability of monocytes to produce IL-10 and suppress T-cell proliferation in sarcoidosis. European journal of immunology. 2014 Jul;44(7):2165-74. PubMed PMID: 24723379. Pubmed Central PMCID: 4746743.
195. Bargagli E, Mazzi A, Rottoli P. Markers of inflammation in sarcoidosis: blood, urine, BAL, sputum, and exhaled gas. Clin Chest Med. 2008 Sep;29(3):445-58, viii. PubMed PMID: 18539237.
196. Eklund A, Blaschke E. Elevated serum neopterin levels in sarcoidosis. Lung. 1986;164(6):325-32. PubMed PMID: 3027463. Epub 1986/01/01. eng.
197. de Boer S, Kolbe J, Wilsher ML. The relationships among dyspnoea, health-related quality of life and psychological factors in sarcoidosis. Respirology. 2014 Oct;19(7):1019-24. PubMed PMID: 25123812.
198. McLoud TC, Epler GR, Gaensler EA, Burke GW, Carrington CB. A radiographic classification for sarcoidosis: physiologic correlation. Investigative radiology. 1982 Mar-Apr;17(2):129-38. PubMed PMID: 7076445.
199. Office IL. Guidleines for the use of ILO international classification of radiographs of pneumoconioses. International Labour Office occupational safety and health series no 22. 1980;Geneva: International Labour Office.
200. Benamore R, Kendrick YR, Repapi E, Helm E, Cole SL, Taylor S, et al. CTAS: a CT score to quantify disease activity in pulmonary sarcoidosis. Thorax. 2016 Jul 27. PubMed PMID: 27701118.
201. Salaffi F, Peroni M, Ferraccioli GF. Discriminating ability of composite indices for measuring disease activity in rheumatoid arthritis: a comparison of the Chronic Arthritis Systemic Index, Disease Activity Score and Thompson's articular index. Rheumatology. 2000 Jan;39(1):90-6. PubMed PMID: 10662880.
202. de Groot K GW, Herlyn K, Reinhold-Keller E. Development and validation of a disease extent index for Wegener's granulomatosis. Clin Nephrol. 2001;55(1):31-8.
203. Castrejon I, Tani C, Jolly M, Huang A, Mosca M. Indices to assess patients with systemic lupus erythematosus in clinical trials, long-term observational studies, and clinical care. Clinical and experimental rheumatology. 2014 Nov-Dec;32 Suppl 85(5):85-95. PubMed PMID: 25365095.
204. van der Heijde DM, van't Hof MA, van Riel PL, van Leeuwen MA, van Rijswijk MH, van de Putte LB. Validity of single variables and composite indices for measuring disease activity in rheumatoid arthritis. Annals of the rheumatic diseases.

1992 Feb;51(2):177-81. PubMed PMID: 1550400. Pubmed Central PMCID: 1005654.
205. van der Heijde DM, van't Hof MA, van Riel PL, Van de Putte L. Disease activity score. Annals of the rheumatic diseases. 1992 Jan;51(1):140. PubMed PMID: 1540024.
206. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum. 1993 Jun;36(6):729-40. PubMed PMID: 8507213.
207. Luqmani RA, Bacon PA, Moots RJ, Janssen BA, Pall A, Emery P, et al. Birmingham Vasculitis Activity Score (BVAS) in systemic necrotizing vasculitis. QJM : monthly journal of the Association of Physicians. 1994 Nov;87(11):671-8. PubMed PMID: 7820541.
208. Liang MH, Socher SA, Larson MG, Schur PH. Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. Arthritis Rheum. 1989 Sep;32(9):1107-18. PubMed PMID: 2775320.
209. Bombardier C, Gladman DD, Urowitz MB, Caron D, Chang CH. Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. Arthritis Rheum. 1992 Jun;35(6):630-40. PubMed PMID: 1599520.
210. Staton GW, Jr., Gilman MJ, Pine JR, Fajman WA, Check IJ. Comparison of clinical parameters, bronchoalveolar lavage, gallium-67 lung uptake, and serum angiotensin converting enzyme in assessing the activity of sarcoidosis. Sarcoidosis. 1986 Mar;3(1):10-8. PubMed PMID: 3033784.
211. Rosenbach M, Yeung H, Chu EY, Kim EJ, Payne AS, Takeshita J, et al. Reliability and convergent validity of the cutaneous sarcoidosis activity and morphology instrument for assessing cutaneous sarcoidosis. JAMA dermatology. 2013 May;149(5):550-6. PubMed PMID: 23677081. Pubmed Central PMCID: 3868362.
212. Baughman RP, Judson MA, Teirstein A, Lower EE, Lo K, Schlenker-Herceg R, et al. Chronic facial sarcoidosis including lupus pernio: clinical description and proposed scoring systems. American journal of clinical dermatology. 2008;9(3):155-61. PubMed PMID: 18429644.
213. Hanefeld M SE, Gouni-Berthold I, Melichar M, Vesela I, Johnson D, Miao S, Sullivan T, Jaen J, Schall T, Bekker P, and the CCX140-B Diabetes Study Group. Orally-Administered Chemokine Receptor CCR2 Antagonist CCX140-B in Type 2 Diabetes: A Pilot Double-Blind, Randomized Clinical Trial. J Diabetes Metab. 2012;3(9):225.
214. Levin AM, Iannuzzi MC, Montgomery CG, Trudeau S, Datta I, Adrianto I, et al. Admixture fine-mapping in African Americans implicates XAF1 as a possible sarcoidosis risk gene. PLoS One. 2014;9(3):e92646. PubMed PMID: 24663488. Pubmed Central PMCID: 3963923.
215. Dubrey S, Shah S, Hardman T, Sharma R. Sarcoidosis: the links between epidemiology and aetiology. Postgraduate medical journal. 2014 Oct;90(1068):582-9. PubMed PMID: 25230946.
216. Hofmann S, Fischer A, Nothnagel M, Jacobs G, Schmid B, Wittig M, et al. Genome-wide association analysis reveals 12q13.3-q14.1 as new risk locus for sarcoidosis. Eur Respir J. 2013 Apr;41(4):888-900. PubMed PMID: 22936702.

217. Kieszko R, Krawczyk P, Chocholska S, Dmoszynska A, Milanowski J. TNF-alpha and TNF-beta gene polymorphisms in Polish patients with sarcoidosis. Connection with the susceptibility and prognosis. Sarcoidosis Vasc Diffuse Lung Dis. 2010 Jul;27(2):131-7. PubMed PMID: 21319595.
218. Rossman MD, Thompson B, Frederick M, Maliarik M, Iannuzzi MC, Rybicki BA, et al. HLA-DRB1*1101: a significant risk factor for sarcoidosis in blacks and whites. American journal of human genetics. 2003 Oct;73(4):720-35. PubMed PMID: 14508706. Pubmed Central PMCID: 1180597.
219. Sato H, Woodhead FA, Ahmad T, Grutters JC, Spagnolo P, van den Bosch JM, et al. Sarcoidosis HLA class II genotyping distinguishes differences of clinical phenotype across ethnic groups. Human molecular genetics. 2010 Oct 15;19(20):4100-11. PubMed PMID: 20685690. Pubmed Central PMCID: 4351376.
220. Brewerton DA, Cockburn C, James DC, James DG, Neville E. HLA antigens in sarcoidosis. Clinical and experimental immunology. 1977 Feb;27(2):227-9. PubMed PMID: 849654. Pubmed Central PMCID: 1540783.
221. Smith MJ, Turton CW, Mitchell DN, Turner-Warwick M, Morris LM, Lawler SD. Association of HLA B8 with spontaneous resolution in sarcoidosis. Thorax. 1981 Apr;36(4):296-8. PubMed PMID: 7281099. Pubmed Central PMCID: 471495.
222. Rossman MD, Thompson B, Frederick M, Iannuzzi MC, Rybicki BA, Pander JP, et al. HLA and environmental interactions in sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis. 2008 Dec;25(2):125-32. PubMed PMID: 19382531.
223. Rutherford RM, Brutsche MH, Kearns M, Bourke M, Stevens F, Gilmartin JJ. HLA-DR2 predicts susceptibility and disease chronicity in Irish sarcoidosis patients. Sarcoidosis Vasc Diffuse Lung Dis. 2004 Oct;21(3):191-8. PubMed PMID: 15554075.
224. Naruse TK, Matsuzawa Y, Ota M, Katsuyama Y, Matsumori A, Hara M, et al. HLA-DQB1*0601 is primarily associated with the susceptibility to cardiac sarcoidosis. Tissue antigens. 2000 Jul;56(1):52-7. PubMed PMID: 10958356.
225. Voorter CE, Drent M, Hoitsma E, Faber KG, van den Berg-Loonen EM. Association of HLA DQB1 0602 in sarcoidosis patients with small fiber neuropathy. Sarcoidosis Vasc Diffuse Lung Dis. 2005 Jun;22(2):129-32. PubMed PMID: 16053028.
226. Voorter CE, Drent M, van den Berg-Loonen EM. Severe pulmonary sarcoidosis is strongly associated with the haplotype HLA-DQB1*0602-DRB1*150101. Human immunology. 2005 Jul;66(7):826-35. PubMed PMID: 16112030.
227. Seitzer U, Swider C, Stuber F, Suchnicki K, Lange A, Richter E, et al. Tumour necrosis factor alpha promoter gene polymorphism in sarcoidosis. Cytokine. 1997 Oct;9(10):787-90. PubMed PMID: 9344512.
228. Zhou T, Zhang W, Sweiss NJ, Chen ES, Moller DR, Knox KS, et al. Peripheral blood gene expression as a novel genomic biomarker in complicated sarcoidosis. PLoS One.7(9):e44818. PubMed PMID: 22984568. Pubmed Central PMCID: 3440319. Epub 2012/09/18. eng.
229. Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. The Journal of experimental medicine. 2003 Mar 17;197(6):711-23. PubMed PMID: 12642603. Pubmed Central PMCID: 2193846.
230. Illei GG, Tackey E, Lapteva L, Lipsky PE. Biomarkers in systemic lupus erythematosus: II. Markers of disease activity. Arthritis Rheum. 2004 Jul;50(7):2048-65. PubMed PMID: 15248202.

231. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50. PubMed PMID: 16199517. Pubmed Central PMCID: 1239896. Epub 2005/10/04. eng.
232. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25-9. PubMed PMID: 10802651. Pubmed Central PMCID: 3037419.
233. Townsend A, Trowsdale J. The transporters associated with antigen presentation. Seminars in cell biology. 1993 Feb;4(1):53-61. PubMed PMID: 8453065.
234. Su R, Nguyen ML, Agarwal MR, Kirby C, Nguyen CP, Ramstein J, et al. Interferon-inducible chemokines reflect severity and progression in sarcoidosis. Respiratory research. 2013;14:121. PubMed PMID: 24199653. Pubmed Central PMCID: 4176097.
235. Spagnolo P, Renzoni EA, Wells AU, Sato H, Grutters JC, Sestini P, et al. C-C chemokine receptor 2 and sarcoidosis: association with Lofgren's syndrome. Am J Respir Crit Care Med. 2003 Nov 15;168(10):1162-6. PubMed PMID: 12882757.
236. Kuziel WA, Morgan SJ, Dawson TC, Griffin S, Smithies O, Ley K, et al. Severe reduction in leukocyte adhesion and monocyte extravasation in mice deficient in CC chemokine receptor 2. Proc Natl Acad Sci U S A. 1997 Oct 28;94(22):12053-8. PubMed PMID: 9342361. Pubmed Central PMCID: 23699.
237. Qian BZ, Li J, Zhang H, Kitamura T, Zhang J, Campion LR, et al. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. Nature. 2011 Jun 08;475(7355):222-5. PubMed PMID: 21654748. Pubmed Central PMCID: 3208506.
238. Negus RP, Stamp GW, Relf MG, Burke F, Malik ST, Bernasconi S, et al. The detection and localization of monocyte chemoattractant protein-1 (MCP-1) in human ovarian cancer. J Clin Invest. 1995 May;95(5):2391-6. PubMed PMID: 7738202. Pubmed Central PMCID: 295866.
239. Shi C, Pamer EG. Monocyte recruitment during infection and inflammation. Nature reviews Immunology. 2011 Oct 10;11(11):762-74. PubMed PMID: 21984070. Pubmed Central PMCID: 3947780.
240. Haringman JJ, Smeets TJ, Reinders-Blankert P, Tak PP. Chemokine and chemokine receptor expression in paired peripheral blood mononuclear cells and synovial tissue of patients with rheumatoid arthritis, osteoarthritis, and reactive arthritis. Annals of the rheumatic diseases. 2006 Mar;65(3):294-300. PubMed PMID: 16107514. Pubmed Central PMCID: 1798063.
241. Valentonyte R, Hampe J, Croucher PJ, Muller-Quernheim J, Schwinger E, Schreiber S, et al. Study of C-C chemokine receptor 2 alleles in sarcoidosis, with emphasis on family-based analysis. Am J Respir Crit Care Med. 2005 May 15;171(10):1136-41. PubMed PMID: 15750046.
242. Hizawa N, Yamaguchi E, Furuya K, Jinushi E, Ito A, Kawakami Y. The role of the C-C chemokine receptor 2 gene polymorphism V64I (CCR2-64I) in sarcoidosis in a Japanese population. Am J Respir Crit Care Med. 1999 Jun;159(6):2021-3. PubMed PMID: 10351956.
243. Petrek M, Drabek J, Kolek V, Zlamal J, Welsh KI, Bunce M, et al. CC chemokine receptor gene polymorphisms in Czech patients with pulmonary

sarcoidosis. Am J Respir Crit Care Med. 2000 Sep;162(3 Pt 1):1000-3. PubMed PMID: 10988120.
244. Peters W, Scott HM, Chambers HF, Flynn JL, Charo IF, Ernst JD. Chemokine receptor 2 serves an early and essential role in resistance to Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2001 Jul 3;98(14):7958-63. PubMed PMID: 11438742. Pubmed Central PMCID: 35450.
245. Feria M D-GF. The CCR2 receptor as a therapeutic target. Expert Opin Ther Patents. 2006;16(1):49-57.
246. Leuschner F, Dutta P, Gorbatov R, Novobrantseva TI, Donahoe JS, Courties G, et al. Therapeutic siRNA silencing in inflammatory monocytes in mice. Nature biotechnology. 2011 Nov;29(11):1005-10. PubMed PMID: 21983520. Pubmed Central PMCID: 3212614.
247. Oshima T, Sonoda KH, Tsutsumi-Miyahara C, Qiao H, Hisatomi T, Nakao S, et al. Analysis of corneal inflammation induced by cauterisation in CCR2 and MCP-1 knockout mice. The British journal of ophthalmology. 2006 Feb;90(2):218-22. PubMed PMID: 16424537. Pubmed Central PMCID: 1860158.
248. Kang YS, Cha JJ, Hyun YY, Cha DR. Novel C-C chemokine receptor 2 antagonists in metabolic disease: a review of recent developments. Expert opinion on investigational drugs. 2011 Jun;20(6):745-56. PubMed PMID: 21466412.
249. Charo IF, Ransohoff RM. The many roles of chemokines and chemokine receptors in inflammation. The New England journal of medicine. 2006 Feb 9;354(6):610-21. PubMed PMID: 16467548.
250. Carter PH. Progress in the discovery of CC chemokine receptor 2 antagonists, 2009 - 2012. Expert opinion on therapeutic patents. 2013 May;23(5):549-68. PubMed PMID: 23428142.
251. Seitzer U, Haas H, Gerdes J. A human in vitro granuloma model for the investigation of multinucleated giant cell and granuloma formation. Histology and histopathology. 2001 Apr;16(2):645-53. PubMed PMID: 11332720.
252. Sahashi K, Ina Y, Takada K, Sato T, Yamamoto M, Morishita M. Significance of interleukin 6 in patients with sarcoidosis. Chest. 1994 Jul;106(1):156-60. PubMed PMID: 8020265.
253. Agostini C, Trentin L, Perin A, Facco M, Siviero M, Piazza F, et al. Regulation of alveolar macrophage-T cell interactions during Th1-type sarcoid inflammatory process. The American journal of physiology. 1999 Aug;277(2 Pt 1):L240-50. PubMed PMID: 10444517.
254. Muro S, Taha R, Tsicopoulos A, Olivenstein R, Tonnel AB, Christodoulopoulos P, et al. Expression of IL-15 in inflammatory pulmonary diseases. J Allergy Clin Immunol. 2001 Dec;108(6):970-5. PubMed PMID: 11742275.
255. Cyster JG. Chemokines and cell migration in secondary lymphoid organs. Science. 1999 Dec 10;286(5447):2098-102. PubMed PMID: 10617422.
256. Krzysiek R, Lefevre EA, Bernard J, Foussat A, Galanaud P, Louache F, et al. Regulation of CCR6 chemokine receptor expression and responsiveness to macrophage inflammatory protein-3alpha/CCL20 in human B cells. Blood. 2000 Oct 1;96(7):2338-45. PubMed PMID: 11001880.
257. Dieu MC, Vanbervliet B, Vicari A, Bridon JM, Oldham E, Ait-Yahia S, et al. Selective recruitment of immature and mature dendritic cells by distinct chemokines expressed in different anatomic sites. The Journal of experimental medicine. 1998 Jul 20;188(2):373-86. PubMed PMID: 9670049. Pubmed Central PMCID: 2212459.

258. Campbell JJ, Hedrick J, Zlotnik A, Siani MA, Thompson DA, Butcher EC. Chemokines and the arrest of lymphocytes rolling under flow conditions. Science. 1998 Jan 16;279(5349):381-4. PubMed PMID: 9430588.
259. Facco M, Baesso I, Miorin M, Bortoli M, Cabrelle A, Boscaro E, et al. Expression and role of CCR6/CCL20 chemokine axis in pulmonary sarcoidosis. J Leukoc Biol. 2007 Oct;82(4):946-55. PubMed PMID: 17615381.
260. Paradowska-Gorycka A, Jurkowska M. Structure, expression pattern and biological activity of molecular complex TREM-2/DAP12. Human immunology. 2013 Jun;74(6):730-7. PubMed PMID: 23459077.
261. Pelham CJ, Agrawal DK. Emerging roles for triggering receptor expressed on myeloid cells receptor family signaling in inflammatory diseases. Expert review of clinical immunology. 2014 Feb;10(2):243-56. PubMed PMID: 24325404.
262. Turnbull IR, Gilfillan S, Cella M, Aoshi T, Miller M, Piccio L, et al. Cutting edge: TREM-2 attenuates macrophage activation. J Immunol. 2006 Sep 15;177(6):3520-4. PubMed PMID: 16951310.
263. Takahashi K, Rochford CD, Neumann H. Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2. The Journal of experimental medicine. 2005 Feb 21;201(4):647-57. PubMed PMID: 15728241. Pubmed Central PMCID: 2213053.
264. N'Diaye EN, Branda CS, Branda SS, Nevarez L, Colonna M, Lowell C, et al. TREM-2 (triggering receptor expressed on myeloid cells 2) is a phagocytic receptor for bacteria. The Journal of cell biology. 2009 Jan 26;184(2):215-23. PubMed PMID: 19171755. Pubmed Central PMCID: 2654305.
265. Suchankova M, Bucova M, Tibenska E, Tedlova E, Demian J, Majer I, et al. Triggering receptor expressed on myeloid cells-1 and 2 in bronchoalveolar lavage fluid in pulmonary sarcoidosis. Respirology. 2013 Apr;18(3):455-62. PubMed PMID: 23231660.
266. Schutyser E, Richmond A, Van Damme J. Involvement of CC chemokine ligand 18 (CCL18) in normal and pathological processes. J Leukoc Biol. 2005 Jul;78(1):14-26. PubMed PMID: 15784687. Pubmed Central PMCID: 2665283.
267. Schraufstatter IU, Zhao M, Khaldoyanidi SK, Discipio RG. The chemokine CCL18 causes maturation of cultured monocytes to macrophages in the M2 spectrum. Immunology. 2012 Apr;135(4):287-98. PubMed PMID: 22117697. Epub 2011/11/29. eng.
268. Schutyser E, Struyf S, Wuyts A, Put W, Geboes K, Grillet B, et al. Selective induction of CCL18/PARC by staphylococcal enterotoxins in mononuclear cells and enhanced levels in septic and rheumatoid arthritis. European journal of immunology. 2001 Dec;31(12):3755-62. PubMed PMID: 11745396.
269. Xanthou G, Polihronis M, Tzioufas AG, Paikos S, Sideras P, Moutsopoulos HM. "Lymphoid" chemokine messenger RNA expression by epithelial cells in the chronic inflammatory lesion of the salivary glands of Sjogren's syndrome patients: possible participation in lymphoid structure formation. Arthritis Rheum. 2001 Feb;44(2):408-18. PubMed PMID: 11229473.
270. Pardo A, Smith KM, Abrams J, Coffman R, Bustos M, McClanahan TK, et al. CCL18/DC-CK-1/PARC up-regulation in hypersensitivity pneumonitis. J Leukoc Biol. 2001 Oct;70(4):610-6. PubMed PMID: 11590198. Epub 2001/10/09. eng.
271. Cai M, Bonella F, He X, Sixt SU, Sarria R, Guzman J, et al. CCL18 in serum, BAL fluid and alveolar macrophage culture supernatant in interstitial lung diseases. Respiratory medicine. 2013 Sep;107(9):1444-52. PubMed PMID: 23831213.

272. Sabat R, Grutz G, Warszawska K, Kirsch S, Witte E, Wolk K, et al. Biology of interleukin-10. Cytokine & growth factor reviews. 2010 Oct;21(5):331-44. PubMed PMID: 21115385.
273. Harlander M, Salobir B, Zupancic M, Tercelj M. Bronchoalveolar lavage chitotriosidase activity as a biomarker of sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis. 2015;32(4):313-7. PubMed PMID: 26847098.
274. Geyer AI, Kraus T, Roberts M, Wisnivesky J, Eber CD, Hiensch R, et al. Plasma level of interferon gamma induced protein 10 is a marker of sarcoidosis disease activity. Cytokine. 2013 Oct;64(1):152-7. PubMed PMID: 23899720.
275. Criado E, Sanchez M, Ramirez J, Arguis P, de Caralt TM, Perea RJ, et al. Pulmonary sarcoidosis: typical and atypical manifestations at high-resolution CT with pathologic correlation. Radiographics : a review publication of the Radiological Society of North America, Inc. 2010 Oct;30(6):1567-86. PubMed PMID: 21071376.
276. Antoniou KM, Tzouvelekis A, Alexandrakis MG, Tsiligianni I, Tzanakis N, Sfiridaki K, et al. Upregulation of Th1 cytokine profile (IL-12, IL-18) in bronchoalveolar lavage fluid in patients with pulmonary sarcoidosis. Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research. 2006 Jun;26(6):400-5. PubMed PMID: 16734560.
277. Prasse A, Pechkovsky DV, Toews GB, Jungraithmayr W, Kollert F, Goldmann T, et al. A vicious circle of alveolar macrophages and fibroblasts perpetuates pulmonary fibrosis via CCL18. Am J Respir Crit Care Med. 2006 Apr 1;173(7):781-92. PubMed PMID: 16415274. Epub 2006/01/18. eng.
278. Reynolds HY. Lung inflammation and fibrosis: an alveolar macrophage-centered perspective from the 1970s to 1980s. Am J Respir Crit Care Med. 2005 Jan 15;171(2):98-102. PubMed PMID: 15557133.
279. Leng D, Huan C, Xie T, Liang J, Wang J, Dai H, et al. Meta-analysis of genetic programs between idiopathic pulmonary fibrosis and sarcoidosis. PLoS One. 2013;8(8):e71059. PubMed PMID: 23967151. Pubmed Central PMCID: 3743918.
280. Konishi K, Gibson KF, Lindell KO, Richards TJ, Zhang Y, Dhir R, et al. Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2009 Jul 15;180(2):167-75. PubMed PMID: 19363140. Pubmed Central PMCID: 2714820.
281. Meltzer EB, Barry WT, D'Amico TA, Davis RD, Lin SS, Onaitis MW, et al. Bayesian probit regression model for the diagnosis of pulmonary fibrosis: proof-of-principle. BMC medical genomics. 2011 Oct 05;4:70. PubMed PMID: 21974901. Pubmed Central PMCID: 3199230.
282. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic acids research. 2016 Jan 04;44(D1):D457-62. PubMed PMID: 26476454. Pubmed Central PMCID: 4702792.
283. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000 Jan 01;28(1):27-30. PubMed PMID: 10592173. Pubmed Central PMCID: 102409.
284. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research. 2017 Jan 04;45(D1):D353-D61. PubMed PMID: 27899662. Pubmed Central PMCID: 5210567.

285. Yoshida R, Imai T, Hieshima K, Kusuda J, Baba M, Kitaura M, et al. Molecular cloning of a novel human CC chemokine EBI1-ligand chemokine that is a specific functional ligand for EBI1, CCR7. The Journal of biological chemistry. 1997 May 23;272(21):13803-9. PubMed PMID: 9153236.
286. Lo JC, Chin RK, Lee Y, Kang HS, Wang Y, Weinstock JV, et al. Differential regulation of CCL21 in lymphoid/nonlymphoid tissues for effectively attracting T cells to peripheral tissues. J Clin Invest. 2003 Nov;112(10):1495-505. PubMed PMID: 14617751. Pubmed Central PMCID: 259130.
287. Rangel-Moreno J, Moyron-Quiroz JE, Hartson L, Kusser K, Randall TD. Pulmonary expression of CXC chemokine ligand 13, CC chemokine ligand 19, and CC chemokine ligand 21 is essential for local immunity to influenza. Proc Natl Acad Sci U S A. 2007 Jun 19;104(25):10577-82. PubMed PMID: 17563386. Pubmed Central PMCID: 1965555.
288. Astrup E, Ranheim T, Damas JK, Davi G, Santilli F, Jensenius M, et al. Increased expression of the homeostatic chemokines CCL19 and CCL21 in clinical and experimental Rickettsia conorii infection. BMC infectious diseases. 2014 Feb 09;14:70. PubMed PMID: 24507453. Pubmed Central PMCID: 3923395.
289. Kim CH, Pelus LM, White JR, Applebaum E, Johanson K, Broxmeyer HE. CK beta-11/macrophage inflammatory protein-3 beta/EBI1-ligand chemokine is an efficacious chemoattractant for T and B cells. J Immunol. 1998 Mar 01;160(5):2418-24. PubMed PMID: 9498785.
290. Kim CH, Pelus LM, White JR, Broxmeyer HE. Macrophage-inflammatory protein-3 beta/EBI1-ligand chemokine/CK beta-11, a CC chemokine, is a chemoattractant with a specificity for macrophage progenitors among myeloid progenitor cells. J Immunol. 1998 Sep 01;161(5):2580-5. PubMed PMID: 9725259.
291. Kim CH, Pelus LM, Appelbaum E, Johanson K, Anzai N, Broxmeyer HE. CCR7 ligands, SLC/6Ckine/Exodus2/TCA4 and CKbeta-11/MIP-3beta/ELC, are chemoattractants for CD56(+)CD16(-) NK cells and late stage lymphoid progenitors. Cellular immunology. 1999 May 01;193(2):226-35. PubMed PMID: 10222066.
292. Gibejova A, Mrazek F, Subrtova D, Sekerova V, Szotkowska J, Kolek V, et al. Expression of macrophage inflammatory protein-3 beta/CCL19 in pulmonary sarcoidosis. Am J Respir Crit Care Med. 2003 Jun 15;167(12):1695-703. PubMed PMID: 12626344.
293. Mosser DM. The many faces of macrophage activation. J Leukoc Biol. 2003 Feb;73(2):209-12. PubMed PMID: 12554797.
294. Hesse M, Modolell M, La Flamme AC, Schito M, Fuentes JM, Cheever AW, et al. Differential regulation of nitric oxide synthase-2 and arginase-1 by type 1/type 2 cytokines in vivo: granulomatous pathology is shaped by the pattern of L-arginine metabolism. J Immunol. 2001 Dec 01;167(11):6533-44. PubMed PMID: 11714822.
295. https://http://www.ncbi.nlm.nih.gov/gene/4321 [24/1/17].
296. Matute-Bello G, Wurfel MM, Lee JS, Park DR, Frevert CW, Madtes DK, et al. Essential role of MMP-12 in Fas-induced lung fibrosis. Am J Respir Cell Mol Biol. 2007 Aug;37(2):210-21. PubMed PMID: 17446527. Pubmed Central PMCID: 1976544.
297. Bauer Y, White ES, de Bernard S, Cornelisse P, Leconte I, Morganti A, et al. MMP-7 is a predictive biomarker of disease progression in patients with idiopathic pulmonary fibrosis. ERJ open research. 2017 Jan;3(1). PubMed PMID: 28435843. Pubmed Central PMCID: 5395293 openres.ersjournals.com.
298. Uchida T, Shirasawa M, Ware LB, Kojima K, Hata Y, Makita K, et al. Receptor for advanced glycation end-products is a marker of type I cell injury in

acute lung injury. Am J Respir Crit Care Med. 2006 May 01;173(9):1008-15. PubMed PMID: 16456142. Pubmed Central PMCID: 2662912.
299. Teirstein AT, Morgenthau AS. "End-stage" pulmonary fibrosis in sarcoidosis. The Mount Sinai journal of medicine, New York. 2009 Feb;76(1):30-6. PubMed PMID: 19170216.
300. Moller DR. Pulmonary fibrosis of sarcoidosis. New approaches, old ideas. Am J Respir Cell Mol Biol. 2003 Sep;29(3 Suppl):S37-41. PubMed PMID: 14503552.
301. Barbarin V, Xing Z, Delos M, Lison D, Huaux F. Pulmonary overexpression of IL-10 augments lung fibrosis and Th2 responses induced by silica particles. Am J Physiol Lung Cell Mol Physiol. 2005 May;288(5):L841-8. PubMed PMID: 15608148.
302. Ostadkarampour M, Eklund A, Moller D, Glader P, Olgart Hoglund C, Linden A, et al. Higher levels of interleukin IL-17 and antigen-specific IL-17 responses in pulmonary sarcoidosis patients with Lofgren's syndrome. Clinical and experimental immunology. 2014 Nov;178(2):342-52. PubMed PMID: 24962673. Pubmed Central PMCID: 4233383.
303. Busuttil A, Weigt SS, Keane MP, Xue YY, Palchevskiy V, Burdick MD, et al. CXCR3 ligands are augmented during the pathogenesis of pulmonary sarcoidosis. Eur Respir J. 2009 Sep;34(3):676-86. PubMed PMID: 19386686.
304. Kyriakides TR, Foster MJ, Keeney GE, Tsai A, Giachelli CM, Clark-Lewis I, et al. The CC chemokine ligand, CCL2/MCP1, participates in macrophage fusion and foreign body giant cell formation. Am J Pathol. 2004 Dec;165(6):2157-66. PubMed PMID: 15579457. Pubmed Central PMCID: 1618731.
305. Bouchon A, Hernandez-Munain C, Cella M, Colonna M. A DAP12-mediated pathway regulates expression of CC chemokine receptor 7 and maturation of human dendritic cells. The Journal of experimental medicine. 2001 Oct 15;194(8):1111-22. PubMed PMID: 11602640. Pubmed Central PMCID: 2193511.
306. Daws MR, Lanier LL, Seaman WE, Ryan JC. Cloning and characterization of a novel mouse myeloid DAP12-associated receptor family. European journal of immunology. 2001 Mar;31(3):783-91. PubMed PMID: 11241283.
307. Cella M, Buonsanti C, Strader C, Kondo T, Salmaggi A, Colonna M. Impaired differentiation of osteoclasts in TREM-2-deficient individuals. The Journal of experimental medicine. 2003 Aug 18;198(4):645-51. PubMed PMID: 12913093. Pubmed Central PMCID: 2194167.
308. Chen LC, Laskin JD, Gordon MK, Laskin DL. Regulation of TREM expression in hepatic macrophages and endothelial cells during acute endotoxemia. Experimental and molecular pathology. 2008 Apr;84(2):145-55. PubMed PMID: 18222421. Pubmed Central PMCID: 2752215.
309. Seno H, Miyoshi H, Brown SL, Geske MJ, Colonna M, Stappenbeck TS. Efficient colonic mucosal wound repair requires Trem2 signaling. Proc Natl Acad Sci U S A. 2009 Jan 06;106(1):256-61. PubMed PMID: 19109436. Pubmed Central PMCID: 2629230.
310. Helming L, Tomasello E, Kyriakides TR, Martinez FO, Takai T, Gordon S, et al. Essential role of DAP12 signaling in macrophage programming into a fusion-competent state. Science signaling. 2008;1(43):ra11. PubMed PMID: 18957693. Pubmed Central PMCID: 2673960.
311. Pelham CJ, Pandya AN, Agrawal DK. Triggering receptor expressed on myeloid cells receptor family modulators: a patent review. Expert opinion on therapeutic patents. 2014 Dec;24(12):1383-95. PubMed PMID: 25363248. Pubmed Central PMCID: 4596234.

312. Aoki N, Zganiacz A, Margetts P, Xing Z. Differential regulation of DAP12 and molecules associated with DAP12 during host responses to mycobacterial infection. Infection and immunity. 2004 May;72(5):2477-83. PubMed PMID: 15102754. Pubmed Central PMCID: 387866.
313. Lucas M, Daniel L, Tomasello E, Guia S, Horschowski N, Aoki N, et al. Massive inflammatory syndrome and lymphocytic immunodeficiency in KARAP/DAP12-transgenic mice. European journal of immunology. 2002 Sep;32(9):2653-63. PubMed PMID: 12207350.
314. Helming L, Gordon S. Molecular mediators of macrophage fusion. Trends in cell biology. 2009 Oct;19(10):514-22. PubMed PMID: 19733078.
315. Yamashita M, Iwama N, Date F, Shibata N, Miki H, Yamauchi K, et al. Macrophages participate in lymphangiogenesis in idiopathic diffuse alveolar damage through CCL19-CCR7 signal. Human pathology. 2009 Nov;40(11):1553-63. PubMed PMID: 19540558.
316. Choi ES, Pierce EM, Jakubzick C, Carpenter KJ, Kunkel SL, Evanoff H, et al. Focal interstitial CC chemokine receptor 7 (CCR7) expression in idiopathic interstitial pneumonia. Journal of clinical pathology. 2006 Jan;59(1):28-39. PubMed PMID: 16394278. Pubmed Central PMCID: 1860265.
317. Wollin L, Wex E, Pautsch A, Schnapp G, Hostettler KE, Stowasser S, et al. Mode of action of nintedanib in the treatment of idiopathic pulmonary fibrosis. Eur Respir J. 2015 May;45(5):1434-45. PubMed PMID: 25745043. Pubmed Central PMCID: 4416110.
318. Lee BS, Margolin SB, Nowak RA. Pirfenidone: a novel pharmacological agent that inhibits leiomyoma cell proliferation and collagen production. The Journal of clinical endocrinology and metabolism. 1998 Jan;83(1):219-23. PubMed PMID: 9435445.
319. Zhang S, Shiels IA, Ambler JS, Taylor SM. Pirfenidone reduces fibronectin synthesis by cultured human retinal pigment epithelial cells. Australian and New Zealand journal of ophthalmology. 1998 May;26 Suppl 1:S74-6. PubMed PMID: 9685030.
320. Kramer N, Walzl A, Unger C, Rosner M, Krupitza G, Hengstschlager M, et al. In vitro cell migration and invasion assays. Mutation research. 2013 Jan-Mar;752(1):10-24. PubMed PMID: 22940039.
321. Helming L, Gordon S. Macrophage fusion induced by IL-4 alternative activation is a multistage process involving multiple target molecules. European journal of immunology. 2007 Jan;37(1):33-42. PubMed PMID: 17154265.