Using Synovial Fluid Biomarkers to Define a Phenotype of Osteoarthritis in the Hip

A dissertation submitted to the University of Oxford in partial fulfilment of the requirements for the degree of

Doctor of Philosophy



John Alexander James Broomfield

Lincoln College

Michaelmas Term 2020

Acknowledgements

To my fiancé Grace, my mother Jennifer and my father Alex: I cannot thank you enough for your unwavering support and patience with me throughout this work. This is for you.

Thank you to the following people for your help and guidance with this work:

UCB Biopharma for generously supporting and funding this work.

My supervisors; Philippa Hulley, Andrew Judge and Sion Glyn-Jones.

My examiners; Linda Troeberg and Simon Donell.

Antony Palmer

Scott Fernquest

Hannah Wilson

Chethan Jayadev

Tamer Malak

Geraint Thomas

Benjamin Kendrick

Dani Prieto-Alhambra

Anastasios Chanalaris

Gary Collins

Stephen Gerry

Peter Taylor

Geoffrey Johnston

Chris Chamberlain

Abstract

Using Synovial Fluid Biomarkers to Define a Phenotype of Osteoarthritis in the Hip

Doctor of Philosophy in Musculoskeletal Sciences

JAJ Broomfield, Lincoln College, Michaelmas 2020

Osteoarthritis (OA) is a chronic, multi-factorial, progressive joint disease. The understanding of OA has evolved away from a wear and tear phenomenon to a complex disease of the joint taken as a whole. A method of stratifying disease, and monitoring progression, is necessary for clinical decision making and for use as end points in future research studies to develop disease modifying OA drugs (DMOADs). The creation of such a tool requires detailed understanding of the disease biopathogenesis. Synovial fluid (SF) is the most likely target to generate the insight required due to its proximity to the intra-articular structures, whose changes are fundamental to the symptomatology of the disease. No work to date has identified any useful single biomarkers of OA and for progress to be made, a consideration of a number of proteins as a panel of markers, whose interrelations describe a signature of different disease states, is required. The emergence of the slow off-rate modified aptamer scan (SOMAscan), a broad spectrum proteomic platform, presents the opportunity to analyse 1310 proteins on a single sample of fluid that measures less than 1 ml. This thesis aims to examine the utility of the SOMAscan in OA biomarker discovery by validation against traditional bench assays, use multivariate statistical analysis methods to identify candidate markers of OA disease states, and contextualise their best clinical timing in the patient pathway by examining the treatment pathways of patients from initial presentation to final treatment.

Synovial fluid samples from hip and knee joints were collected from patients with different stages of clinically classified disease including knee injury, early and late OA cohorts, and hip early and late OA cohorts with the early hip OA group requiring the development of an entirely new collection method. The performance of the SOMAscan was validated using ELISA and MSD assays that were performed on a curated list of proteins and compared to the same markers in the SOMAscan showing robust comparison. Multivariate regression analysis showed separation of the knee injury and late OA cohorts using PLS-DA, but the early group were not separable from the injury group. Markers identified by LASSO that were strongly associated with late disease included elevated TSG-6, TIMP-1 and R-spondin 4. These

markers were also identified in the PLS-DA model generating potential targets for use in future OA studies.

Hip SF analysis showed separation between the early and late OA groups on PLS-DA. LASSO analysis revealed 6 markers: TSG-6, DDR-2, Ephrin-A5, S100A7, MICB, IDS all of which were present in the top 50 marker model produced using PLS-DA. This combined analysis approach has shown differences between the hip groups, differences between hip and knee OA biology and suggested novel targets for future OA studies.

An epidemiological study investigating the patient pathways, from presentation in primary care through to eventual arthroplasty, was conducted to identify where a diagnostic test would be best placed in the patient journey. Results showed marked variation in the rates of treatments offered to patients during their disease, with wide variation in analgesia, physiotherapy and intermediate surgical interventions. Most notably it showed that there was a continued increase in the use of opioid analgesia overall, particularly in the most deprived patients, indicating a need to investigate ways to harmonise this patient experience such that patients can be identified reliably at earlier stages of disease.

In summary, the works in this thesis have validated the SOMAscan. A new system of biomarker discovery in OA, for use in SF analysis. The SOMAscan analysis of SF has shown a variety of differently expressed markers and pathways including well described OA processes in both knee and hip. It has suggested that early knee OA appears biologically closer to joint injury than to late knee OA, and that hip and knee OA appear to be biologically distinct processes with common thematic pathways as well as differences. This work has collected a new and unique SF cohort of human early hip OA SF to define a biological phenotype of late OA that discriminates between the groups studied. This work has further shown that the patient journey from presentation to definitive treatment is hugely varied, both in duration and treatments delivered, demonstrating that ongoing work is needed not only to identify biomarkers of early OA but to identify means of accessing patients earlier in their clinical disease states. Together these critical elements of disease description and patient identification require optimisation such that future OA biomarkers may be utilised as disease diagnosis, stratification and outcome tools in DMOAD studies.

Contents

Chapter 1 Introduction 1

1.1 Osteoarthritis 1

1.1.1 Definition 2

1.1.2 Classification 3

1.1.3 Epidemiology 4

1.1.4 Aetiology 5

1.1.5 Diagnosis 5

1.1.5.1 Clinical Features 6

1.1.5.2 Radiography 7

1.1.5.3 Magnetic Resonance Imaging 10

1.1.6 Staging 11

1.1.7 Management in the Community 12

1.1.8 Timing of Referral to Secondary Care 13

1.1.9 Variations in Access to Care 14

1.2 Anatomy 15

1.2.1 The Hip 15

1.2.2 The Knee 17

1.3 The Synovial Joint 18

1.4 Synovium and Synovial Fluid 19

1.5 Cartilage 20

1.5.1 Cartilage Structure 21

1.5.2 Proteoglycans 24

1.5.3 Collagens 24

1.5.4 Cartilage Degradation and OA 25

1.6 Biomarkers 28

1.6.1 Definition 30

1.6.2 Description of OA Biomarkers 31

1.7 Immunoassay & Multiplex Approaches to Biomarker Analysis 33
1.7.1 Enzyme-linked Immunosorbent Assay 34
1.7.2 Meso Scale Discovery 35
1.7.3 SOMAscan 36
1.8 Bioinformatic Analysis Approaches 38
1.8.1 Multivariate Statistics in Biomarker Discovery 39
1.8.2 Principal Component Analysis 40
1.8.3 Partial Least Squares – Discriminant Analysis 40
1.8.4 Least Absolute Shrinkage & Selection Operator 41
1.9 Summary 41
1.9.1 Literature Gaps 43
1.9.2 Scope of Thesis 44
1.9.3 Thesis Aims & Structure 45
Chapter 2 Methods 47
2.1 Introduction 47
2.2 Synovial Fluid Acquisition 47
2.2.1 Ethical Approval 48
2.2.1.1 Synovial Fluid 48
2.2.1.2 CPRD Data 48
2.2.2 Knee Sampling Technique 48
2.2.3 Knee Group Descriptions 49
2.2.4 Development of a Novel Technique for Sampling of Hip Synovial Fluid 49
2.2.5 Hip Group Descriptions 51
2.2.6 Sample Treatment 53
2.3 Validating SOMAscan for Synovial Fluid 54
2.3.1 PlexOA 55
2.3.2 Biomarker Comparison 56
2.3.3 Immunoassays 57
2.4 Patients, Materials & Methods 59

2.4.1 Validation of SOMAscan 60
2.5 Statistical Analysis 67
2.6 Patient Risk Assessment 70
Chapter 3 The Biological Patterns of Synovial Fluid in Knee OA 71
3.1 Introduction 71
3.2 Aim 72
3.3 Groups 72
3.4 PLS-DA 73
3.4.1 Knee Early OA, Late OA and Injury Three Group Comparison Using the Whole SOMAscan 73
3.4.2 Knee Late OA and Injury Group Comparison Using the Whole SOMAscan 78
3.4.3 Knee Late OA and Injury Group Comparison the Most Descriptive 50 Markers From Whole SOMAscan Comparison 81
3.5 LASSO 91
3.5.1 PLS-DA Model Using 3 LASSO markers 95
3.6 PCA 96
3.6.1 Knee Early OA, Late OA and Injury Three Group Comparison Using PCA of the Whole SOMAscan 97
3.6.2 Knee Late OA Group Individual Comparison to Early OA and Injury Groups Using the Whole SOMAscan 100
3.7 Network Analysis 103
3.8 Discussion 107
Chapter 4 The Biological Patterns of Synovial Fluid in Hip OA 117
4.1 Introduction 117
4.2 Aim 117
4.3 Groups 117

4.4 PLS-DA 118
4.4.1 Hip Early OA and Late OA Group Comparison Using the Whole SOMAscan 118
4.4.2 Hip Early OA and Late OA Group Comparison Using the Most Descriptive 50 Markers From the Whole SOMAscan Comparison 121
4.5 LASSO 131
4.5.1 PLS-DA model of 6 LASSO markers 134
4.6 PCA 134
4.6.1 Hip Early OA and Late OA Group Comparison Using the Whole SOMAscan 135
4.7 Network Analysis 138
4.8 Discussion 141
Chapter 5 Variations in Treatment Pathways to Access Care 148
5.1 Introduction 148
5.2 Aims 149
5.3 Proposed pathways 149
5.4 Materials and methods 151
5.4.1 Data source 151
5.4.2 Case identification 151
5.4.3 Analyses 152
5.5 Results 153
5.5.1 Demographics 153
5.5.2 Analgesic Pathway 153
5.5.3 Physiotherapy Pathway 158
5.5.4 Intermediate Surgery Pathway 159
5.5.5 Inequalities in access to care 161
5.6 Discussion 168
Chapter 6 Thesis Discussion 173

6.1 Chapter 2 – Validation of SF for SOMAscan 174

6.2 Chapter 3 – Knee SF SOMAscan Analysis 175

6.3 Chapter 4 – Hip SF SOMAscan Analysis 178

6.4 Chapter 5 – Variations in Treatment Pathways 180

6.5 Clinical Relevance 181

Chapter 7 Future Directions 185

Bibliography 188

Appendix A 242

A.1 – Sample OMB Patient Consent Form 243

A.2 – Sample FAIT Consent Form 244

Appendix B 245

B.1 – SOMAscan Assay 1310 Content 245

Appendix C 256

C.1 - Sample R code for LASSO analysis 256

Appendix D 262

D.1 - Sample OPCS-4 code data dictionary for THR 262

Figures

Figure 1-1 - Plain antero-posterior radiograph of a left hip joint demonstrating normal appearances with no evidence of OA (Image courtesy of Sion Glyn-Jones) 9

Figure 1-2 - Plain antero-posterior radiograph showing an osteoarthritic hip demonstrating joint space narrowing, femoral and acetabular marginal osteophytes, subchondral sclerosis and subchondral cysts (Image courtesy of Sion Glyn-Jones).....9

Figure 1-3 - Image taken during arthroscopy of the hip showing the normal appearances of healthy femoral head hyaline cartilage (Image courtesy of Sion Glyn-Jones).....21

Figure 2-1 – Heatmap showing variable regression coefficient plot for 8 markers from the PlexOA study. 56

Figure 2-2 - Distributions of COMP and PIIANP ELISA assay results by group. 61

Figure 2-3 - Comparison of distributions of individual markers between ELISA & MSD assays and SOMAscan by group. 62

Figure 2-4 - Comparison of distributions of individual markers between MSD assays and SOMAscan by group. 63

Figure 2-5 - Correlation of five SOMAlogic assays performed on remaining synovial fluid from 23 of 26 knee patients. 65

Figure 2-6 - Correlation of five SOMAlogic assays performed on remaining synovial fluid from 20 of 22 hip patients. 66

Figure 3-1 - Score scatter plot for PLS-DA model of knee early OA, injury and late OA group comparison using the whole SOMA panel of 1310 markers. 74
Figure 3-2 - Loading scatter plot for PLS-DA model of knee early OA, injury and late OA group comparison using the whole SOMA panel of 1310 markers. 75
Figure 3-3 – VIP plot showing the top 50 markers from the PLS-DA model of knee early OA, injury and late OA group comparison using the whole SOMA panel of 1310 markers. 77
Figure 3-4 - Score scatter plot for PLS-DA model of knee injury vs. late OA group comparison using the whole SOMA panel of 1310 markers. 78
Figure 3-5 - Loading scatter plot for PLS-DA model of knee injury vs. late OA group comparison using the whole SOMA panel of 1310 markers.....79
Figure 3-6 – VIP plot showing the top 50 markers from the PLS-DA model of knee injury vs. late OA group comparison using the whole SOMA panel of 1310 markers. 3 component model (R^2 = 0.940 Q^2 = 0.555).....80
Figure 3-7 - Score scatter plot for knee injury vs. late group comparison using top 50 marker panel. 3 component model (R^2 = 0.984 Q^2 = 0.921).....82
Figure 3-8 - Loading scatter plot for PLS-DA model of knee injury vs. late OA group comparison using top 50 marker panel. 3 component model (R^2 = 0.984 Q^2 = 0.921).....82
Figure 3-9 - VIP plot for PLS-DA model of knee injury vs. late OA group comparison using top 50 marker panel. 3 component model (R^2 = 0.984 Q^2 = 0.921) 83

Figure 3-10 - Graphs showing variance between groups for metalloproteinase inhibitors: TIMP 1, 2 & 3. Significance tested using ANOVA. P-values shown for those markers that met significance. 88
Figure 3-11 - Graphs showing the variance between groups for selected matrix metalloproteinases: MMP-1, 2, 3, 9 & 13. Significance tested using ANOVA. No knee group differences met significance. 89
Figure 3-12 - Graphs showing the variance between groups for selected ADAM and ADAMTS proteins: ADAM 9 & 12 and ADAMTS-1,4,5 13 & 15. Significance tested using ANOVA. No knee group differences met significance 90
Figure 3-13 - Graphs showing the variance between groups for selected proteins that were elevated in the early OA and injury groups: Aggrecan, PARK7 and Osteocalcin. Significance tested using ANOVA. P-values shown for those markers that met significance. 91
Figure 3-14 - Graphs showing the variance between groups for markers identified by LASSO analysis: TIMP-1, TSG-6 and RSPO-4 94
Figure 3-15 - Loading scatter plot for PLS-DA model of knee injury vs. late group comparison using a panel of the 3 LASSO markers alone showing that all three markers load to the late OA group. 1 component model (R^2 = 0.651 Q^2 = 0.592) .95
Figure 3-16 – Observation score plot of PCA of quantile normalised SF proteins from comparison of knee early OA, injury and late OA groups using the whole SOMAscan panel of 1310 proteins. 98
Figure 3-17 – Heatmap and hierarchical clustering of PCA of quantile normalised SF proteins from comparison of knee early OA, injury and late OA groups using the whole SOMAscan panel of 1310 proteins. 99

Figure 3-18 - Observation score plot of PCA of quantile normalised SF proteins from comparison of knee early OA and late OA groups using the whole SOMAscan panel of 1310 proteins.....101
Figure 3-19 - Observation score plot of PCA of quantile normalised SF proteins from comparison of knee injury and late OA groups using the whole SOMAscan panel of 1310 proteins.....102
Figure 3-20 - Network of protein - protein interactions based on the PCA top loading markers from the first 5 principal components of PCA of quantile normalised SF proteins from comparison of knee injury and late OA groups using the whole SOMAscan panel of 1310 proteins. 105
Figure 3-21 - Network of protein - protein interactions based on the top 100 markers of the PLS-DA comparison of knee injury and late OA groups using the whole SOMAscan panel of 1310 proteins.....106
Figure 4-1 - Score scatter plot for hip early vs. late OA group comparison using whole SOMA panel of 1310 markers showing clear separation of the groups. 3 component model (R^2 = 0.940 Q^2 = 0.669) 119
Figure 4-2 - Loading scatter plot for hip early vs. late group comparison using whole SOMA panel of 1310 markers.....119
Figure 4-3 – VIP plot showing the top 50 markers from the hip early vs. late group comparison using the whole SOMA panel of 1310 markers. 3 component model (R^2 = 0.940 Q^2 = 0.669) 120
Figure 4-4 - Score scatter plot for hip early vs. late group comparison using the top 50 markers from the SOMA panel showing improved separation and within group clustering. 3 component model (R^2 = 0.948 Q^2 = 0.852).....122

Figure 4-5 - Loading scatter plot for hip early vs. late group comparison using the top 50 markers from the SOMA panel. 3 component model (R^2 = 0.948 Q^2 = 0.852) 122

Figure 4-6 - VIP plot for hip early vs. late group comparison using the top 50 markers from the SOMA panel. 3 component model (R^2 = 0.948 Q^2 = 0.852).....123

Figure 4-7 - Graphs showing variance between groups for metalloproteinase inhibitors: TIMP 1, 2 & 3. Significance tested using 2-tailed t-test. P-values shown for those markers that met significance. 127

Figure 4-8 - Graphs showing the variance between groups for selected matrix metalloproteinases: MMP-1, 2, 3, 9 & 13. Significance tested using 2-tailed t-test. P-values shown for those markers that met significance.....128

Figure 4-9 - Graphs showing the variance between groups for selected ADAM and ADAMTS proteins: ADAM-9 & 12 and ADAMTS-1,4,5 13 & 15. Significance tested using 2-tailed t-test. P-values shown for those markers that met significance....129

Figure 4-10 - Graphs showing the variance between groups for selected proteins that were elevated in the early hip OA group: Glypican 3, BMP14, WIF-1, JAK2, SOD3 and CATC. Significance tested using 2-tailed t-test. P-values shown for those markers that met significance. 130

Figure 4-11 - Graphs showing the variance between groups for markers identified by LASSO analysis: TSG-6, DDR-2, S100A7, MICB, EFNA5 and IDS. Significance tested using 2-tailed t-test. P-values shown for those markers that met significance. 133

Figure 4-12 - Loading scatter plot for hip early vs. late group comparison using LASSO marker panel. 1 component model (R^2 = 0.845 Q^2 = 0.821).....134

Figure 4-13 - Observation score plot of PCA of quantile normalised SF proteins from comparison of hip early OA and late OA groups.....136
Figure 4-14 - Heatmap and hierarchical clustering of PCA of quantile normalised SF proteins from comparison of hip early OA and late OA groups.....137
Figure 4-15 – Focused network map of protein-protein interactions based solely on hip LASSO markers. 139
Figure 4-16 – Interactome network of protein - protein interactions. 140
Figure 5-1 - Variation in analgesic pathway over time showing increase in numbers of patients on pathways that include full opiate analgesia. 155
Figure 5-2 - Change in rates of prescription by class of medication over time....156
Figure 5-3 - Cumulative interval (months) between last analgesic prescription and THR 157
Figure 5-4 - Cumulative interval (years) from first analgesic prescription to THR by pathway 157
Figure 5-5 - Trend in rate of referral for physiotherapy prior to THR over time 158
Figure 5-6 – Cumulative interval between first and last referral to physiotherapy and THR 159
Figure 5-7 - Trend in referral for steroid injection over time.....160
Figure 5-8 - Interval between first and last intra-articular steroid injection and THR 161
Figure 5-9 - Trend in rates of full opioid prescription over time by WHO BMI group 162
Figure 5-10 - Trend in rates of full opioid prescription over time by IMD class .162

Figure 5-11 - Trend in rates of full opioid prescription highlighting the most and least deprived patient groups over time by IMD class 163
Figure 5-12 - Trend in stratified age of patients referred for intra-articular steroid injection over time.....164
Figure 5-13 - Trend in BMI of patients undergoing THR over time.....165
Figure 5-14 - Percentage of patients referred for physiotherapy by WHO BMI groups.....166
Figure 5-15 - Percentage of patients referred for physiotherapy by IMD class ...166
Figure 5-16 - Percentage of patients referred for steroid injection by WHO BMI group.....167
Figure 5-17 - Percentage of patients referred for physiotherapy by IMD class ...167

Tables

Table 2-1 – Summary of descriptions of the groups used for the biological experiments in this thesis.....53

Table 2-2 - Description and manufacturer details for all ELISA & MSD bench assays performed during the SOMAscan validation study.....58

Table 2-3 - Demographic data for age and sex of all patients in the study cohorts 60

Table 3-1 – Knee cohort individual group participant demographics.....73

Table 3-2 - Table showing top 50 markers from the injury vs late OA group comparison panel (Fig. 3-6) with UniProt ID, group that the marker loads towards and brief description, ** indicates LASSO marker.....85

Table 4-1 - Hip cohort individual group participant demographics.....118

Table 4-2 - Table showing top 50 markers from the early vs late hip OA group comparison panel (Fig. 4-3) with UniProt ID and description, ** indicates LASSO marker.....124

Table 5-1 – Interval (years) between first prescription and eventual surgery by pathway (Mean and SD) 158

Abbreviations

AAOS	American Academy of Orthopaedic Surgeons
ACR	American College of Rheumatology
ADAMTS	A Disintegrin and Metalloproteinase with Thrombospondin Motifs
ANOVA	Analysis Of Variance
BCL2L1	B Cell Lymphoma-2-Like Protein 1
BMI	Body Mass Index
CHST6	Carbohydrate Sulfotransferase 6
COMP	Cartilage Oligomeric Protein
COX	Cyclo-oxegenase
CPRD	Clinical Practice Research Datalink
CTX	C-terminal Telopeptide of Collagen
DAF	Decay Accelerating Factor
DDH	Developmental Dysplasia of the Hip
DDR	Discoidin Domain Receptor
dGEMRIC	delayed Gadolinium-Enhanced MRI of Cartilage
DMOADs	Disease Modifying Osteoarthritis Drugs
ECL	Electrochemiluminescence
ELISA	Enzyme Linked Immunosorbent Assay
EULAR	European League Against Rheumatism

FAI	Femoroacetabular Impingement
FCN2	Ficolin-2
FGF 12	Fibroblast Growth Factor 12
GAG	Glycosaminoglycan
HA	Hyaluronic Acid
HAse	Hyaluronidase
HS6ST1	Heparan Sulfate 6-O-Sulfotransferase 1
ICD-10	International Statistical Classification of Diseases and Related Health Problems – Tenth Revision
IDS	Iduronate 2 Sulfatase
IL	Interleukin
IMD	Index of Multiple Deprivation
ISAC	Independent Scientific Advisory Committee
K&L	Kellgren & Lawrence
LASSO	Least Absolute Shrinkage and Selection Operator
MAPK	Mitogen-Activated Protein Kinase
MCP-1	Monocyte Chemoattractant Protein
MICB	MHC Class I Polypetide Related Sequence B
MIP	Macrophage Inhibitory Protein
MMP	Matrix Metalloproteinase
MRI	Magnetic Resonance Imaging
MSD	Meso Scale Discovery

NCK1	Non-Catalytic Region of Tyrosine Kinase Adaptor Protein 1
NICE	National Institute for health and Care Excellence
NSAID	Non-Steroidal Anti-Inflammatory Drug
OA	Osteoarthritis
OARSI	Osteoarthritis Research Society International
P13K-Akt	Phosphoinositide 3-kinase – Protein Kinase B
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PLS-DA	Partial Least Squares – Discriminant Analysis
RA	Rheumatoid Arthritis
RAGE	Receptor for Advanced Glycation End products
RANTES	Regulated on Activation, Normal T cell Expressed and Secreted
RSPO-4	R Spondin 4
SF	Synovial Fluid
SLPI	Secretory Leukocyte Proteinase Inhibitor
SOMA	Slow Off-rate Modified Aptamer
TGF-β1	Transforming Growth Factor β1
THR	Total Hip Replacement
TKR	Total Knee Replacement
TIMP	Tissue Inhibitor of Metalloproteinases
TSG-6	Tumour necrosis factor Stimulated Gene 6
WHO	World Health Organization

Chapter 1 Introduction

1.1 Osteoarthritis

Osteoarthritis (OA) is a chronic multi-factorial, progressive joint disease that is a major cause of disability with symptoms typically including pain, limitation of function and joint swelling [1]. Historically OA was considered simply as a disease of joint overuse, the so called "wear and tear" arthritis. Whilst OA is, without question, a disease with a large mechanical component in the activation of key underlying biological processes [2] it is apparent that the disease can be manifested in the absence of clinical symptoms or signs and progress at variable rates [3]. OA should be considered as a disease of the entire joint and is characterised by irreversible loss of articular cartilage with bone remodelling and inflammation of the synovium [4].

The financial impact of OA, both in terms of direct and indirect costs, is difficult to measure accurately. However, in some developed countries such costs have been estimated to be in the region of up to 2.5% of GDP [5] and according to the National Institute for health and Care Excellence (NICE) the total annual cost of OA to the UK economy alone is estimated at 1% of GDP, which currently totals approximately £1.5 trillion.

OA is a global problem, which continues to rise, with cases worldwide steadily increasing year on year. Indeed data shows that rates of OA rose by 31.4% between

2007 – 2017 [6] and OA is projected to affect 29.5% of the global population over 45 years of age by 2032 [7]. The most commonly affected joints in the body are knee, hand and hip with the most commonly reported age of onset between 50 and 75 years of age [8].

1.1.1 Definition

The definition of OA has been the subject of continued debate with ongoing controversy over a unifying definition [9]. Over time, key stakeholders including governing bodies, societies and organizations have attributed different weight to a wide variety of elements ranging from presence of symptoms such as pain or functional limitation through to empirical measurable characteristics such as radiographic changes and patient reported outcome scores to name a few. Various definitions for OA have been proposed over recent years by NICE [10], the centres for disease control and prevention (CDC) [11] and the World Health Organization (WHO) in collaboration with the American Academy of Orthopaedic Surgeons (AAOS) [12] among others, however in 2015 the Osteoarthritis Research Society International (OARSI) proposed the following comprehensive definition [13] with the goal of standardising the nomenclature of OA:

> *"Osteoarthritis is a disorder involving movable joints characterized by cell stress and extracellular matrix degradation initiated by micro- and macro-injury that activates maladaptive repair responses including pro-inflammatory pathways of innate immunity. The disease manifests first as a molecular derangement (abnormal joint tissue metabolism) followed by anatomic, and/or physiologic*

> *derangements (characterized by cartilage degradation, bone remodeling, osteophyte formation, joint inflammation and loss of normal joint function), that can culminate in illness."*

The NICE system is simple, but clinically focussed, due to its aim for use in the treatment of OA, whilst the CDC system uses case definitions that are common to all types of arthritis and not specific to OA. The WHO/AAOS system is comprehensive but does not involve consideration of the inflammatory and immune components seen in OA. Accordingly, the OARSI system is preferred because it provides the only unified description including all the cellular processes that precede the signs and symptoms of disease.

1.1.2 Classification

The clinical classification of OA has historically been challenging. The WHO produces a standardised document termed the International Statistical Classification of Diseases and Related Health Problems, the tenth revision of which is currently in use globally and is commonly referred to as ICD-10. The production of this document allows international alignment and consistency when recording and observing diseases. The thirteenth chapter of this document concerns diseases of the musculoskeletal system and connective tissues, and includes OA. Subsections within chapter 13 of this document consider a classification of OA in respect of the location of the joint, the side of the occurrence of the disease, and the underlying pathology of the disease. Whilst this system is valuable for tracking disease in populations, and for coding purposes, it is also limited in that it does

not have a clinical application beyond documentation purposes. Furthermore, this system of classification of OA does not consider the underlying biological pathways, which may, of course, be relevant.

Beyond systems such as the ICD-10, the American College of Rheumatology (ACR) supported the development of criteria for the classification of OA of the hip [14] and knee [15] which have become widely utilised. The ACR criteria for the classification of both hip and knee OA focuses upon various iterations of clinical, laboratory and radiographic criteria. In both cases the results of the combination of all relevant parameters generated the set of criteria with highest combined sensitivity and specificity at 91% and 89% respectively for the hip and 91% and 86% respectively for the knee. In both cases sensitivity could be marginally increased at the expense of specificity by systems excluding radiographic measures.

1.1.3 Epidemiology

The prevalence and incidence of radiographic, clinical, and symptomatic OA of the hip and knee has been frequently discussed by commentators and variously reported in a growing body of literature on the topic [8,16–19]. All estimates agree that there is an extremely high burden of disease within the population. The most significant risk factors for development of OA are advancing age [18] and obesity [2] with women more frequently affected than men [20]. OA is a mechanical disease and so it follows that abnormalities in joint biomechanics and morphology also contribute to development of the disease [21]. Notwithstanding this, it does

not follow that all people with identifiable abnormalities of joint morphology will necessarily progress to symptomatic OA [22]. Significant differences exist between individuals due to complex interactions of biological, psychological and behavioural factors that generate the experience of pain [23]. Sleep disturbance has been linked with central sensitisation in OA pain [24] whilst lower socioeconomic status has been shown to be an independent predictor of pain in OA with no association with functional scores [25].

1.1.4 Aetiology

By tradition, the aetiology of OA has generally been considered to be either primary, or alternatively idiopathic, or secondary, most commonly as a result of trauma, but also associated with many factors. However, over recent years a growing body of literature has considered and developed an understanding of the causation of OA that focuses upon the interaction of both genetic and environmental components and their influence on risk factors for development and progression of OA [26,27]. Thereby suggesting that the aetiology of OA is, in fact, far more complex than was previously thought.

1.1.5 Diagnosis

The diagnosis of OA, as in any medical encounter, should begin with taking a detailed clinical history of the patient. Classically patients will describe a gradual onset of worsening pain in the affected joint which is usually associated with disruption of activities of daily living and recreational activities. As the disease worsens and conditions deteriorate, the symptoms progress to include pain at rest,

disturbance of sleep as a result of pain and morning stiffness typically lasting approximately thirty minutes before mobility improves. This debilitating loss of function has a wider psychosocial impact on the patient and their relationships, which may be both personal and professional. The biopsychosocial model of illness first proposed by Engel [28] is clearly applicable to the case of OA where such loss of function and impact on wider activities may be closely linked with feelings of reduced quality of life [29]. Pain is the core symptom of OA but is perceived differently by each individual and can be variable in its effects on activity and function [30]. Accordingly the biopsychosocial impact is considered a central component of the European League Against Rheumatism (EULAR) recommendations for pain management in OA [31].

1.1.5.1 Clinical Features

A clinical examination of the affected joint should aim to confirm the historical findings and exclude alternative diagnoses. The examiner may observe joint swelling due to effusion, deformity, abnormality of limb alignment or an abnormal (usually antalgic) gait. There may be tenderness to palpation, crepitus and restriction of range of motion. Indeed, using the ACR criteria it may be possible to make a diagnosis of OA without the use of laboratory testing or imaging at all with sensitivity of 86% and specificity of 75% [14]. At present plain radiography is the primary investigation of choice with laboratory tests being of little additional value in clinical practice, save for the exclusion of infection. Accordingly, there is a pressing need for better tests with the capability to detect this disease earlier and

monitor its progression once detected. Such tests would provide insights into the disease state that may facilitate patient stratification, and inform clinical decision making about appropriate therapies.

1.1.5.2 Radiography

OA is most commonly investigated using plain radiography, otherwise known as X-Ray examination. It is freely available in almost all hospitals, quick to perform and low cost. As such it remains the primary investigation of choice for any clinical presentation where the differential diagnosis includes OA [32]. Radiography is capable of reliably identifying the pathognomonic features of OA: joint space narrowing, marginal osteophytes, subchondral cysts and subchondral sclerosis [9]. These are the features most commonly utilised clinically although osteophytes alone are included in the ACR clinical and radiographic criteria for diagnosis of OA [14]. This modality not only forms an integral part of the clinical investigation and diagnosis of OA but is also key in the grading and classification of the disease both for treatment planning and as an outcome assessment tool. There have been many proposed radiographic grading systems used in OA of diarthrodial joints including OARSI [33] and Ahlback [34]. Perhaps the most commonly used, however, is the method described by Kellgren and Lawrence in 1957 [35]. The method denotes that the observer views the radiographs and considers the presence or absence of features of OA. Based on this evaluation the observer may then place the radiographic appearances into one of five grades:

- None 0
- Doubtful 1
- Minimal 2
- Moderate 3
- Severe 4

The critical threshold is taken at grade 2 where below this grade there is no convincing evidence of OA and grades 2 and above have at least minimal evidence of OA. Although this system has become widely used due to its simplicity, it is also frequently inconsistently applied [36]. As a result, more recently it has been proposed that the measurement of joint space width alone can be more reliable than grading systems for the radiographic diagnosis of OA of the hip [37] and knee [38]. Examples of normal (Figure 1-1) and OA (Figure 1-2) hip radiographs are shown below.



Figure 1-1 - Plain antero-posterior radiograph of a left hip joint demonstrating normal appearances with no evidence of OA (Image courtesy of Sion Glyn-Jones)



Figure 1-2 - Plain antero-posterior radiograph showing an osteoarthritic hip demonstrating joint space narrowing, femoral and acetabular marginal osteophytes, subchondral sclerosis and subchondral cysts (Image courtesy of Sion Glyn-Jones)

1.1.5.3 Magnetic Resonance Imaging

Plain radiography generates an image that is a two-dimensional representation of a three-dimensional object. This fact highlights the obvious limitations of identifying pathology within the joint. Radiography is also incapable of providing any insight into potential pathology of any joint tissues other than bone such as early cartilage injury [39]. Accordingly, radiographs are only capable of detecting bony pathology late in the disease chronology. In order to shed any light on early disease, whether in future clinical practice or research studies, an alternative modality is required. Images obtained from Magnetic Resonance Imaging (MRI) give three dimensional information about the joint and are, therefore, the cross-sectional imaging modality of choice for investigating articular cartilage morphology and early OA changes [40,41]. In spite of this there remains a critical need for a method of identifying the disease at its earliest stage and tracking its chronological development.

This problem is most prominent in younger patients since they pose a particular treatment challenge as a result of limited surgical options, particularly in the hip. They may often undergo a protracted course of conservative treatment prior to presentation by which time their disease state has often progressed significantly. Another challenge for the clinician is that significant pain can result from a joint where minimal cartilage damage is seen and, conversely, some patients with appearances of end stage disease are relatively painless [42]. The presence of pain alone cannot be relied upon to inform treatment decisions. As a result, patients

may therefore endure years of pain before the damage is sufficient for detection by current imaging modalities.

Physiological MRI has emerged with the prospect of being capable of detecting the very earliest evidence of OA by assessing glycosaminoglycan content of cartilage such as delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) [43]. However, this method of imaging would be challenging to implement in clinical practice due to the time taken to produce the scan following administration of the contrast agent which itself can, in a small number of cases, prove nephrotoxic [44]. Other sequences may have better clinical utility such as T2 mapping, T2* mapping, T1Rho and diffusion weighted scanning [45–48]. T2 mapping is the best studied of these sequences so far with growing use in clinical research [49]. Potential future MRI sequences also include chemical exchange saturation transfer and sodium scanning. These sequences however require scanners that utilise ultra-high field strength magnets that are currently limited to only a small number of research centres. These methods have value as tools in research studies to quantify proteoglycan content in cartilage but the universal application of these methods in the clinical setting is unlikely in the foreseeable future [50,51].

1.1.6 Staging

As discussed above at paragraph 1.1.2, OA has been classified by the ACR and others using imaging and clinical features which has led to a robust agreement of the features of end stage disease. Such classification is clearly of importance.

However, for the purposes of research it is necessary to consider what is meant by the term "early OA" because, as a term, it is in isolation somewhat nebulous.

In truth, patients in the early phase of OA may be rather different by appreciable measures [52] and, although challenging, the defining signs of early OA should include structural changes in the cartilage identified by arthroscopy or radiology [53]. The chondral injury that leads to OA is commonly associated with anterior cruciate ligament (ACL) rupture and meniscal injury in the knee [54–57]. Similarly in the hip, a chondral injury that commonly leads to OA has been shown to be caused by femoroacetabular impingement syndrome (FAI) [58–61].

1.1.7 Management in the Community

The current NICE guidance (CG 177 Osteoarthritis: care & management) [10] advises a diagnosis of OA may be made clinically without investigations in patients aged over 45 with activity related joint pain and morning stiffness lasting less than 30 minutes. Patients should then be offered advice about weight loss if necessary and be provided with appropriate sources of information about activity and exercise. Initial non-pharmacological management may include local muscle strengthening and improving general aerobic stiffness. Before consideration of surgery a holistic approach should be taken to understand the effects of the disease on the individual's life. Treatments prior to medication may include transcutaneous electrical nerve stimulation and walking aids such as sticks, crutches, frames etc. Next, pharmacological treatments should be considered to include the prescription of paracetamol and/or topical non-steroidal anti-

inflammatory drugs (NSAIDs). The failure of symptoms to resolve at this stage should be followed by an escalation of treatment to oral NSAIDs. Finally, following all other pharmacological options, should symptoms not improve, an escalation to consideration of opiate analgesia may then be considered. However, it is important to consider that pain and mental health have a complex interaction. The wider psychosocial context of the individual and its impact upon their symptomatology must be taken into account.

Following these conservative and medical measures, if symptoms persist then practitioners should discuss referral for a consultation with an orthopaedic surgeon where, if appropriate, further treatment options will be discussed that may include intra-articular injections as a diagnostic aid and eventual total, or partial, joint arthroplasty.

1.1.8 Timing of Referral to Secondary Care

Patients present to medical care at differing points in their disease chronology which, as a result, influences the type of treatment, and the duration of the course of the treatment they are offered. In addition, there remain questions in respect of the influence that patient related factors may exert on their outcomes. In particular, the regional variations in availability of, and access to, services as well as the relative impact of comorbidities on outcomes in musculoskeletal diseases may be influential factors. Since published NICE guidance continues to recommend NSAIDs among the first line therapies for OA prescription of these drugs

represents a reliable marker of patients making their initial, and on-going, presentations in the community with joint related pathology.

It is clear that OA is under-diagnosed and under-treated and yet the severity of OA disability is associated with a significant increase in all-cause mortality even after controlling for multiple confounders [62]. However, OA outcomes cannot improve until a reliable method of diagnosis emerges, but it will be of little use clinically without understanding how patients progress from disease onset to medical care. This requires integration of biological and population studies to inform when, in the chronology of disease, a test would be best used.

1.1.9 Variations in Access to Care

Over 160,000 people with severe hip or knee pain caused by osteoarthritis undergo joint replacement surgery annually in the UK [63]. However, very little is known about how they access surgery. It would appear that their journey begins by consulting their general medical practitioner after the onset of symptoms but the onward journey to definitive care varies widely. Despite a common healthcare service, across the UK there is significant variation in the treatment options offered and models of care delivered [64]. Patients may be unable to access some treatments simply based on where they live. National guidelines have been published about which treatments the NHS should offer to patients presenting with symptoms of OA, and that surgery should only be used if non-surgical treatments have failed [10]. However, it remains unclear which pathway is best for patients, or most cost effective for the NHS. It is unclear whether some patient

groups are less likely to be offered surgery based on intrinsic characteristics such as age, gender, geographical location, socioeconomic status or comorbid conditions.

1.2 Anatomy

In order to attempt to better understand the disease processes in OA that underpin the clinical pathology and begin to consider avenues of investigation, it is first necessary to consider the anatomical features of the joints in question. This thesis focuses on the hip and, to a lesser extent, the knee. These two joints have many similarities but also some key differences.

1.2.1 The Hip

The *articulatio coxae*, or hip joint as it is more commonly known, is configured as a ball and socket articulation between the ball shaped head of the femur and the cup shaped acetabulum of the pelvis. Both surfaces are covered in hyaline cartilage with the exception of the central area of the head of the femur at which the ligamentum teres femoris runs from the fovea capitis femoris of the femoral head in two bands to the acetabular notch, where it becomes continuous with the transverse acetabular ligament. There as an extensive synovium within the capsule of the hip joint extending from beyond the articular cartilage of the head of the femur, lining the capsule to beyond the extra-articular aspect of the acetabular labrum.

The cotyloid ligament, more commonly known as the acetabular labrum, is a triangular structure composed of fibrocartilage which serves to increase the depth and surface area of the acetabulum [65]. Another key role of the labrum is to create and maintain a "seal" around the articulating surfaces to permit synovial lubrication of the joint and prevent articular cartilage degeneration [66]. In addition, cadaveric studies have demonstrated increased cartilage strain and evidence of chondral damage in the absence of a functional acetabular labrum [67,68]. Joint movement has been shown to influence the intra-articular fluid pressure due to joint position causing alterations in the shape and volume of the joint meaning the joint pressure is, at times, subatmospheric [69]. Factors predisposing to labral tears include (but are not limited to) trauma, femoro-acetabular impingement (FAI), and developmental dysplasia of the hip (DDH). Whilst the true incidence of labral tears in the general population is not known, published evidence shows that labral tears may be present in 22-55% of patients presenting with hip pain [68,70]. The improved understanding, in recent years, of the role of hip joint morphology on function has seen FAI emerge as a leading cause of labral tears [71]. Due to the surface continuity at the chondro-labral junction labral tears may co-exist with anterosuperior acetabular cartilage loss termed a 'carpet lesion' representing cartilage delamination [72]. In addition, as a result of such direct and indirect processes, FAI is recognised as a primary cause of hip OA [73] along with numerous patient defined risk factors including previous history of trauma, family history of OA, increasing age, female sex,

occupation, and obesity [74]. Whether a labral tear should be considered an injury which leads to secondary OA, as in the case of ACL rupture or meniscal tear in the knee [75], or as a sign of progression due to morphology-induced wear and tear, more akin to the degenerative meniscal tears seen in early idiopathic knee OA, is an important issue to debate. Effective clinical trials rely on robust stratification of patient groups and need a mechanistic understanding, rooted in tissue biology, of the appropriate therapeutic directions to take [76].

1.2.2 The Knee

The knee is a synovial joint of the lower limb comprising an articulation between three bones; the tibia, femur, and patella. It is considered to comprise three compartments. The articulation of the condyles of the femur and the tibial plateaus are considered separately and divided into the medial and lateral compartments whilst the articulation between the femur and patella are termed the patella-femoral compartment. Together these three compartments make the knee the largest joint in the body.

Mechanically speaking there is no stability conferred by the osseous components of the knee joint and so, in contrast to the hip, there are further structures required to regulate and stabilise the joint in the locomotion and weight bearing of the gait cycle. There are four main ligaments of the knee; the two pairs of cruciate and collateral ligaments. The anterior and posterior cruciate ligaments are named by the position of their tibial attachments and confer stability in the antero-posterior

plane. The medial or tibial and lateral or fibular collateral ligaments, named after their distal attachment, confer stability in the medio-lateral plane.

In contrast to the hip, where there is a large surface area of femoral head and acetabulum in contact throughout the range of motion, the convex femoral condyles and the tibial plateaus do not have a large surface area in contact at any point during the range of motion. The medial tibial plateau is concave whilst the lateral tibial plateau is not, meaning there is necessarily an extra structure contained within both the medial and lateral tibio-femoral compartments. These structures are termed the menisci which are C-shaped pieces of fibrocartilage whose function is to better match the congruity between the joint surfaces and distribute the load more evenly throughout the range of motion.

1.3 The Synovial Joint

Classically the term synovial joint is synonymous with the term diarthrodial joint or diarthrosis. This type of joint represents the majority of the joints in the human body which, under normal circumstances, permits a large range of stable, pain free motion. The key features of the synovial joint are of opposing surfaces covered by articular cartilage, bones connected by stabilizing ligaments, and a lining of synovium. They may also be divided by a disc or meniscus. Whilst this clearly reflects the macroscopic structure of the knee, it should be noted that all these key features are present in the hip joint. The hip is also classified as a ball and socket

joint, or enarthrosis, and it can therefore be considered to be a multi-axial diarthrosis [77].

1.4 Synovium and Synovial Fluid

The capsule of synovial joints is lined with synovium which serves to seal the synovial joint space and regulate the constitution of synovial fluid (SF) [78]. This tissue is comprised of two layers; the outer is largely acellular, thicker and more fibrous layer called the subintima and has a high proportion of type 1 collagen, whilst the inner layer, called the intima, is significantly thinner [79]. In healthy synovium the cellular component of the intima consists of type A synoviocytes that have features in keeping with macrophages, and type B synoviocytes that display features more consistent with those of fibroblasts [80]. It is the type B synoviocytes that secrete SF. Joint effusions are common in both inflammatory arthropathies and degenerative joint disease despite the different modes of pathogenesis. In the early phase of rheumatoid arthritis (RA) fibroblasts have been shown to transition from an immunosuppressive function to the acquisition of a pro-inflammatory phenotype [81]. Despite the degenerative nature of OA it has been shown that low grade inflammation plays a pivotal role in the pathogenesis of OA [82]. Indeed, recent evidence has shown that there are, in fact, distinct subsets of fibroblasts responsible for mediating inflammation, or tissue damage, in arthritis with immune effector fibroblasts located in the synovial sub-lining and destructive fibroblasts being restricted to the synovial lining layer [83]. Using a

similar single-cell transcriptomic analysis technique it has further been shown that there are potentially overlapping pathogenic mechanisms in RA and OA [84].

SF is a complex viscous fluid present in the healthy diarthrodial joint in small volumes that covers the articular cartilage. The roles of SF include lubrication of the joint to permit smooth motion, and to serve as a vector for cartilage nutrition, since cartilage is accepted as being essentially avascular save for the area immediate contact with the subchondral bone at the chondro-osseous junction. The continual turnover of cartilage also means there is ongoing presence of low concentrations of cartilage breakdown products in the SF. The major protein component of SF is hyaluronic acid (HA) secreted, along with lubricin, by the type B synoviocytes. HA is a complex high molecular weight glycosaminoglycan (GAG) which contributes significantly to the unusual visco-elastic properties of SF [85]. Lubricin, or PRG4, is a proteoglycan of smaller molecular size, which was thought not to play a significant role in the properties of SF. However studies have suggested it can directly modulate the visco-elastic properties of SF [86]. It should be noted that all tissues of the joint release substances into the SF, including microvesicles, exosomes, matrix and signaling molecules, and that this changes in response to injury, damage and disease [87].

1.5 Cartilage

Articular cartilage is a specialised connective tissue that covers the articulating surfaces of diarthrodial joints in layers of approximately 2mm thickness.

Specifically termed "hyaline" cartilage meaning glass-like as a result of its smooth, translucent appearance, it serves to absorb shock and distribute load across its surface during motion to provide a low friction bearing surface (Figure 1-3). It is avascular, aneural and alymphatic.

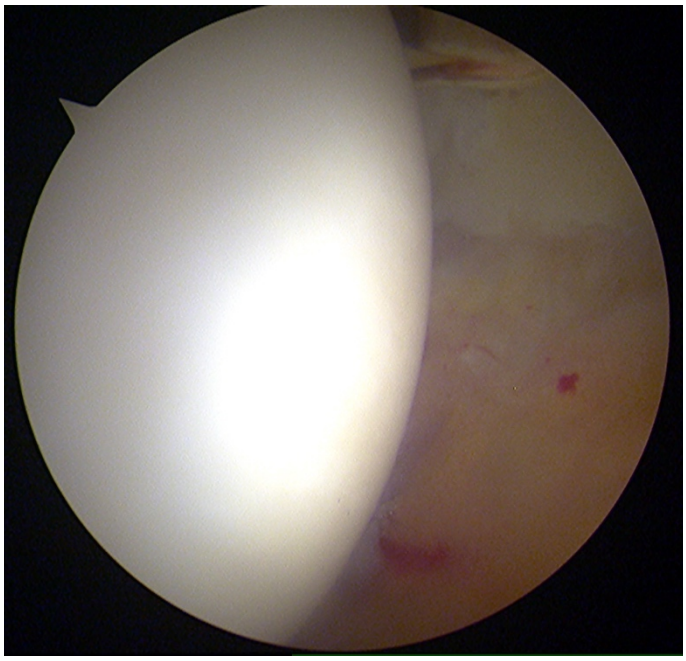


Figure 1-3 - Image taken during arthroscopy of the hip showing the normal appearances of healthy femoral head hyaline cartilage (Image courtesy of Sion Glyn-Jones)

1.5.1 Cartilage Structure

The majority of articular cartilage is composed of extracellular matrix (ECM) which contributes to approximately 95% of its volume. The ECM comprises water, collagen fibres (predominantly type 2), proteoglycans and glycosaminoglycans.

The cellular component of cartilage is the chondrocyte. A small number of cells are distributed throughout the ECM comprising only approximately 5% of the cartilage volume. The primary role of the chondrocyte is cartilage maintenance and

homeostasis. As a result of their isolation within the ECM, and the avascular conditions, the source of nutrition for the chondrocyte is via diffusion from the SF. Chondrocytes have very limited rates of cellular turnover.

Chondrocytes may be considered as having three distinct subtypes relating to their location within the ECM and reflecting the differences in their local environment [88]. They are named with respect to their proximity to the cartilage / bone interface with the zone furthest away termed the superficial zone. Next is the middle zone with the zone closest termed the deep zone.

There are discrete zonal differences in the structure of articular cartilage in terms of ECM and chondrocyte appearance and function:

1. **Superficial Zone**: this contributes approximately 20% of the overall depth of articular cartilage. The chondrocytes in this zone are flattened and secrete lubricin into the joint. Alternatively termed the tangential zone due to the orientation of the collagen fibres being parallel to the joint. This permits smooth gliding movement over its surface thereby resisting shear forces.
2. **Middle Zone**: this contributes approximately 50% of the overall depth of articular cartilage where the chondrocytes have a larger, more rounded appearance. Alternatively termed the "transitional zone" due to the collagen fibres here running more obliquely. This allows for the

transition from resistance to shear forces superficially towards resistance to compressive forces in the deep zone.

3. **Deep Zone**: this contributes approximately 30% of the overall depth of articular cartilage. The chondrocytes and collagen fibres here are larger again and are orientated perpendicular to the joint surface, reflecting the proximity to the calcified cartilage. The collagen fibres cross into the deep zone form the calcified cartilage providing the resistance to compressive forces. Alternatively termed the "radial zone" in reference to the orientation of the fibres.

Concentration of collagen is highest superficially, whilst proteoglycan concentration is highest in the deep zone; a reflection of the apparent difference in role of the different layers. Deep to these three layers is the zone of calcified cartilage that anchors the cartilage to the underlying bone.

The ECM displays cartographical variations reflecting the relative proximity to the chondrocytes. The peri-cellular region is the area immediately adjacent to the chondrocyte where proteoglycan predominates and plays a role in signal transduction during load bearing [89]. The territorial matrix is beyond the peri-cellular matrix containing a network of finer collagen fibres. The largest component of the ECM is the interterritorial matrix comprising larger collagen fibres that contribute to the biomechanical properties of articular cartilage [90]. Overwhelmingly the predominant constituent of ECM is water comprising approximately 75% and is most abundant superficially facilitating diffusion of

nutrients to and from the synovial fluid. The remaining portion is mostly comprised of collagen (mostly type II) and proteoglycan of which aggrecan is the most abundant.

1.5.2 Proteoglycans

Proteoglycans are large molecules that consist of a central protein core bound to multiple glycosaminoglycan side chains that contain a negatively charged subgroup, usually a sulfate group [91] most commonly chondroitin sulfate or keratan sulfate. This negative charge creates an osmotic pressure that maintains the hydration state of articular cartilage. The most abundant proteoglycan is aggrecan which is capable of interacting with hyaluronan. In fact, numerous aggrecan molecules may interact with a single hyaluronan molecule to create a large proteoglycan aggregate that is incarcerated within the ECM. The high water concentration and large aggregates fixed within the ECM contribute to the generation of a firm structure capable of resisting deformation under high compressive loads [92].

1.5.3 Collagens

Collagens constitute over half of the dry weight of cartilage. The most abundant collagen in articular cartilage is type II [93]. However, numerous further subtypes have been identified in the literature including types I, III, IV, VI, IX, X, XI, XII, XIII, XIV [94,95]. Type II collagen is comprised of three α1 polypeptide chains that are arranged as a triple helix. The resultant fibrils orientate into polymers that form the framework of articular cartilage [93]. Type II cartilage predominates but

heteropolymers including other minor collagens may occur, most commonly with type IX collagen [96].

1.5.4 Cartilage Degradation and OA

OA and its pathogenesis is polyfactorial. It may occur in a single joint, or many, and it may present with dramatic symptoms, or be silent. At its core though, it is characterised by degradation and loss of articular cartilage. That said, it should not be forgotten that the joint must be considered as an organ and that failure of any component will have a wide impact on the joint as a whole. This process is mediated by a complex interaction of chemokines, cytokines and matrix degrading enzymes, under genetic and epigenetic control [97].

At present the identification of OA occurs late in the disease process, meaning management options are limited to reactive approaches that may perhaps be most aptly described as palliation rather than treatment. Radiographic features of OA are late signs but patients frequently only have mild, or even no symptoms or signs early in their disease. During surgery to the joint, whether open or arthroscopic, the condition of the articular cartilage can be classified under direct vision. Perhaps the most common such classification system is that originally described in the assessment of chondromalacia patellae by Outerbridge in 1961 which describes 4 macroscopic grades [98]:

- 1 – Softening and swelling of the cartilage
- 2 – Fragmentation and fissuring of an area half an inch or less in diameter

- 3 – Fragmentation and fissuring of an area more than half an inch diameter
- 4 – Erosion of cartilage down to the bone

This system remains in common use but has customarily been extended to classify normal cartilage as grade 0 and include a modest update to a metric measurement of area.

Chondrocytes regulate the bio-architecture of cartilage by responding to chemical and mechanical disturbances [99]. This activity is continuous and steady response to normal load and activity because the chondrocyte has a very low rate of cellular turnover. The half life of the aggrecan core protein is variably reported up to as long as 24 years [100]. However, after disruption of this steady state the chondrocytes proliferate and cluster producing higher quantities of both matrix proteins and matrix degrading enzymes, including matrix metalloproteinases (MMP) and a disintegrin and metalloproteinase with thrombospondin-like motifs (ADAMTS) [101].

The MMP and ADAMTS enzymes can be broadly considered as a group of zinc and calcium dependent proteinases that facilitate tissue remodelling in both healthy and diseased states [102]. They have established roles in the degradation of cartilage ECM as well as that of other connective tissues [103,104]. In particular ADAMTS-4 and 5 are potent aggrecanases that have been repeatedly detected in cartilage of human OA subjects, hence they have been alternatively named aggrecanase 1 and 2 respectively [104–106]. Also associated with cartilage

breakdown are ADAMTS-7 and 12 whilst ADAMTS-2, 3 and 14 have shown cartilage anabolic activity [107]. In addition MMP-1 and 13 have been shown to be key type II collagenases [108,109] and have been the focus of much investigation in OA as possible therapeutic targets. The tissue inhibitors of metalloproteinases (TIMP) are a family of four proteins which inhibit the activity of MMPs, ADAMs and ADAMTSs [110–112]. Through this action they play a role in regulation of ECM turnover by direct inhibition of proteolysis or by restriction of ECM deposition via regulation of inflammation [113].

The role of immune mediators, including chemokines and cytokines, has been the subject of much investigation. Both chemokines and cytokines are produced by chondrocytes and synoviocytes and go on to deliver local actions within the joint [114]. Cytokines may be pro-inflammatory such as interleukin-1β (IL-1β), tumour necrosis factor alpha (TNF-α) and IL-6 [115,116] or anti-inflammatory such as IL-1 receptor antagonist (IL-1RA), IL-4, IL-10 and IL-13 [117,118]. These effects may be direct or driven through activation of other mediators to cause downstream effects. Chemokines are a subfamily of cytokines that function as a chemoattractant directing immune cell traffic to sites of injury or infection [119]. They are subclassified into four groups based on the position of cysteine residues and are interchangeably named by acronyms describing their action or by subfamily motif descriptor (C, CC, CXC or CX3C) followed by a number indicating the order of discovery. The letter L may be added to indicate a ligand or R to indicate a receptor [120]. Among the most studied chemokines in OA are monocyte chemoattractant

protein 1 (MCP-1) also named CCL2, macrophage inhibitory protein 1α (MIP-1α) also named CCL3, macrophage inhibitory protein 1β (MIP-1β) also named CCL4 and regulated on activation, normal T cell expressed, and secreted (RANTES) also named CCL5 [121–124]. Also implicated in the signalling pathway of OA are the growth factors: transforming growth factor β (TGF-β) which mediates osteochondral crosstalk [125], insulin-like growth factor 1 (IGF-1) which plays a role in inducing matrix synthesis [126], and the fibroblast growth factor family which are known to activate chondrogenesis although under certain conditions can display negative effects on matrix deposition [127,128].

1.6 Biomarkers

In the search for information about the biological processes underpinning OA pathogenesis we must go to the source of the disease and examine the biology of the affected sites directly. Peripheral serum sampling and urine collection has, traditionally, been the mainstay of the OA biomarker discovery field. However, neither serum nor urine have been able to yield any "game-changing" step with no biomarkers so far sufficiently validated for clinical use [129]. Furthermore, there are significant challenges in achieving a deep understanding of the changes and causation of the underlying disease processes due to the potential for confounding when the chain of custody of the output is interrupted by systemic filtering and distant pathology [130]. Ageing also presents a challenge since both the cellular and matrix components of cartilage decrease with age negatively affecting the

mechano-protective mechanisms over time [131]. In particular the key to changing the patient level outcomes of OA, by generating new interventions and delaying or preventing joint replacement surgery, will never be possible if we cannot identify patients in early disease states and stratify their risk of progression [132]. Only then can we begin attempts to modulate the pathophysiology with a genuine prospect of influencing disease progression. To this end SF is advantageous for measuring OA biomarkers, since they are present in higher concentrations than in serum or urine [133], given that it is in direct physical contact with articular structures and this limits dilution in other biological fluids [134,135]. Although SF represents the most accurate assessment of the intra-articular biology, and accordingly, offers the best measure of joint pathophysiology, it is not present in the volumes necessary for the analysis of more than a small number of markers using traditional bench assay platforms such as enzyme-linked immunosorbent assays (ELISA) and meso scale discovery (MSD). Therefore, when there is such scarcity of samples with only very small volumes available, it is necessary to gain as much information as is possible from each sample. Research has begun to characterise the proteome of SF using a variety of approaches including 2D Gel Electrophoresis [136], ELISA and histological staining [137] and mass spectroscopy [138,139]. These approaches have broadened the awareness of the components of SF and mass-spectrometry has shown greater promise with broad proteomics in the search for definable differences between stages of OA [140]. However, further

progress requires comparison of samples from tightly defined groups of patients at differential stages of disease.

1.6.1 Definition

The term biomarker has evolved over recent years. Classically the consideration centred upon the identification of a singular measure that could be quantified in isolation to determine, in an almost binary fashion, the presence or absence of a disease state. Proteins, such as prostate specific antigen or rheumatoid factor, are common examples in clinical practice. Initial definitions involved only cellular, biochemical or molecular measures, such as the National Institutes of Health (NIH) biomarkers working group definition of the term biomarker in 2001:

> *"a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to therapeutic intervention"* [141].

Unfortunately, in many disease areas, this is far too limited a view and disorders from a wide range of systems and aetiologies including Alzheimer's disease [142], liver cirrhosis [143] and metastatic renal cell carcinoma [144], to name just three, have required broader approaches in the search for biomarkers to aid disease diagnosis, progression, classification and prognostication. In the case of OA the term now encompasses alterations in symptomatology and imaging techniques [145]. A new classification system proposed in 2006 was termed "BIPED", an acronym denoting the five categories of markers:

- ***B***urden of disease
- ***I***nvestigative
- ***P***rognostic
- ***E***fficacy of intervention
- ***D***iagnostic.

This classification system presented a broad and detailed system to outline the different potential roles of biochemical biomarkers in OA, as well as the nature and quality of the evidence for them [146]. The BIPED groups are not mutually exclusive meaning a marker can be considered to have multiple roles and also provides for the possibility of a method to combine more than one marker in isolation. As a result, the system has helped shape the presentation and interpretation of new and emerging biomarkers in OA since its inception. However, there has yet to be a major step change in the OA biomarker space.

1.6.2 Description of OA Biomarkers

There is a desperate need for methods to identify early OA and to find clinically useful markers of disease progression. Such a requirement is most pressing in younger patients where limited surgical options lead to protracted courses of conservative treatment prior to presentation, during which time they may progress significantly. Pain is unreliable for informing treatment since pain can result from minimal cartilage damage, whilst conversely, end stage disease can be

relatively painless [126]. Patients can, as a result, experience years of pain before OA is clinically detected since there are currently no effective interventions available for early OA. The odds are stacked against their discovery, and development, by the current inability to find and monitor early disease, coupled with a fundamental lack of understanding in respect of the underlying processes in the biology of early OA pathogenesis. This inevitably limits clinicians' ability to grade the disease, and robustly separate groups of patients for use in trials. Our group has defined a range of biomarkers that when combined together present a fingerprint of end stage OA [147]. However, this needs to be validated using larger cohorts and does not address fingerprints that could be used to define early stages of disease and injury.

To identify therapeutic biological targets the first consideration is to identify the best tissue to investigate. The traditional approach has been urine or serum due to the ready access, ease of sampling, sample volumes and familiarity with handling and preparation. The chondrocyte represents only around 1% of the total volume of articular cartilage [99]. The remainder is composed of a balanced equilibrium of collagen fibrils and hydrated proteoglycans such as aggrecan [148]. Accordingly, the most commonly studied markers to date are cartilage breakdown products; serum cartilage oligomeric protein (sCOMP) and urinary C-terminal crosslinking telopeptide of collagen type II (uCTX-II). Studies have shown elevated concentration of uCTX-II is associated with increased risk of radiographic hip OA [149] and in the knee, as sCOMP increased, the risk of cartilage loss on MRI also

increased [150]. Despite such associations no current biomarkers are sufficiently discriminating to aid OA diagnosis and prognosis [151]. Biomarkers that better reflect the disease process are required.

Investigations of individual proteins as markers of OA have proved unreliable and to date there have been no concerted efforts to consider multivariate methods. However, in other disease areas there have been successful examples, such as the Ottawa ankle rules, that used multiple clinical parameters for assessment of injuries of the foot and ankle [152], and the Framingham Heart Study cardiovascular risk profiles, which use a variety of measurable risk factors to predict disease [153,154] to name just two. This may be, in part, linked to the poor reporting of development and validation of multivariable prediction models in medicine [155] resulting from the inherent challenges in the application of multivariate methods. This concept has resulted in recent exciting progress in the search for biomarkers in Alzheimer's disease [156], type II diabetes [157], ischaemic heart disease [158] and Duchenne muscular dystrophy [159] where the novel SOMAscan assay, in concert with detailed multivariate analysis, has helped better define disease groups.

1.7 Immunoassay & Multiplex Approaches to Biomarker Analysis

The studies within this thesis focused on the use of the commercial SOMAscan assay for the first time on SF samples. Assay validation was done by comparison with ELISA and MSD assays, for well accepted SF components in OA.

1.7.1 Enzyme-linked Immunosorbent Assay

ELISA is the harnessing of an enzyme reaction which relies on the specificity of an antibody binding to a specific epitope of an antigen. This specificity allows for detection of even very small amounts of a substance of interest and also for a wide range of targets in a wide variety of media. This reaction takes place in the wells of a plastic assay plate and, typically, the resultant solution, following the binding event, will be associated with a colour change. The colour of the fluid in the wells of the plate can then be read to measure the absorbance values which are compared to a standard curve produced from serial dilutions of a known concentration of the target. The assay may be performed as a qualitative analysis to test for the simple presence of a target where the reading simply identifies the presence of the target above a lower limit of detection. Alternatively, the assay may be performed as a quantitative analysis where the colour produced in the test samples is directly proportional to the concentration. The absorbance values can be read off the standard curve to generate concentrations for the tested samples.

There are four common types of ELISA. In a direct ELISA the detection antibody is directly conjugated with the enzyme generating the colour change. In an indirect ELISA a two stage process is required. First the antigen antibody reaction occurs as in the direct method. However, this does not produce the colour change, meaning a second conjugated antibody must be introduced to then bind the primary antibody/antigen complex for concentration measurement. Sandwich ELISAs use two detection antibodies to bind different epitopes of the same antigen.

These reactions may each be either direct or indirect to produce the eventual colour change for measurement. Finally, in competitive ELISAs the sample antigen competes for limited numbers of antibody binding sites. This can be of use for small molecules with limited epitope availability.

1.7.2 Meso Scale Discovery

MSD uses a binding event type and assay technique similar to that seen in the ELISA method. The measurement is, however, not performed using a colorimetric reading, instead MSD uses electrochemiluminescence (ECL). The plates are specific to MSD and to each assay kit, with the base of the wells housing an electrode and "spot" to which the capture antibody is attached. When the spot is electrified during the plate read, it triggers an oxidation reaction of the "SULFO-TAG" labels in the presence of a "read buffer solution" that produces light that is received by a camera. The relative intensity of light produced in each well generates values for the target substance concentration. The MSD system minimises background noise because the stimulation and read are not coupled, and multiple excitation cycles allows signal amplification to increase the dynamic range decreasing the need for sample dilutions. A significant advantage of the MSD platform is the multiplex offering. This allows for up to ten spots per well of a 96 well plate without requiring increased sample volume. This is of great value when sample volume is at a premium, allowing for wider analysis of samples.

1.7.3 SOMAscan

The most advanced multiplex proteomic platform currently available is the Slow Off-Rate Modified Aptamer (SOMA) assay termed "SOMAscan". It uses DNA-based protein affinity reagents or *aptamers* that are highly sensitive and reproducible, permitting measurement of 1310 analytes from a single fluid sample of 150µl [160] on the version of the platform used for the work in this thesis. More recent expansion of the platform has developed to 5000 analytes on a single sample [161]. An aptamer is a short single strand of nucleic acid, between 20 and 100 base pairs in length [162]. SOMAscan aptamers are the latest evolution of the systematic evolution of ligands by exponential enrichment (SELEX) procedure for amplification of selected RNA's as DNA ligands for specified target proteins first described by Tuerk et al. in 1990 [163] and similarly also described simultaneously by another group and reported in Ellington et al [164]. Rather than requiring direct antibody/antigen binding, as in assays such as those described above, this procedure utilises aptamers, nucleic acid based oligonucleotides that bind with high affinity to proteins and peptides [165]. In the case of the SOMAscan these aptamers, or SOMAmers, are DNA based and are nucleotides chemically modified with protein-like functional groups [166]. The system has minimal risk of cross reactivity due, in part, to the specificity, meaning that only a single SOMAmer per analyte is required, but also to the "slow off rate". SOMAmers are selected based on the duration of their binding and dissociation time, where the strongest interactions bind for the longest time, thereby maximising the number of analytes

per sample and minimising the risk of dissociation prior to analysis [167,168]. The system has been automated and optimised over time. Its highly multiplexed nature allows multiple assays on a single sample improving efficiency and productivity [169] whilst also exceeding the breadth of other proteomic profiling platforms including mass-spectrometry for biomarker discovery [170]. Although mass spectrometry can provide a complete proteome approach, it is highly specialised and labour intensive requiring multiple sample pre-processing and cleaning steps [171]. When compared with SOMAscan, the sensitivity of MS is limited by the high dynamic range of protein concentrations which precludes the direct detection of low abundance biomarkers [172]. The SOMAscan has also been shown to give stable readings between fresh and archived samples demonstrating its utility over longer term cohort analyses [173]. Quantification of protein concentrations in test samples is calculated by measurement of the strong "cognate" SOMAmer-protein complexes whose bindings exhibit the slow off-rate whilst the "incognate" bindings are weaker and, as such, they dissociate more rapidly. The SOMAmers comprise a 5′ biotin with a photocleavable group and fluorescent tag and the binding occurs on streptavidin coated beads via a biotin-streptavidin reaction after incubation with the sample [174] and after a series of partitioning and wash steps, a nucleic acid signal is generated that is quantified using custom DNA microarrays and read in relative fluorescence units [160].

The SOMAscan has emerged as a tool with the power to accelerate biomarker discovery across a range of diseases and subject media. It has yielded biomarker

discoveries in samples of serum, plasma, cerebrospinal fluid, synovial fluid and purified exosomes from patients with Duchenne muscular dystrophy (serum) [159], inflammatory arthritis (serum) [162], heart failure (serum) [161], multiple sclerosis (plasma) [175], Alzheimer's disease (plasma and cerebrospinal fluid) [156], ACL rupture (synovial fluid) [176], prostate carcinoma (exosomes) [177] and chronic rhinosinusitis (exosomes) [178]. Its capability to investigate over 1000 proteins in a single small sample of fluid or material presents a new opportunity to derive new understanding about diseases which present greater challenges with minimal sample material. This opportunity, however, comes at great cost. The SOMAlogic platform, whilst able to perform such broad-spectrum analysis, is far more expensive than traditional bench assays, and is priced in excess of £1000 per sample. As such, a large scale, longitudinal, study of a complex disease using an untested matrix would not be possible as a first foray into this new space. It is first necessary to conduct exploratory experiments to ascertain the validity of the tool for SF and its utility for OA biomarker discovery.

1.8 Bioinformatic Analysis Approaches

Biomarker discovery is limited by the accessibility of data, whether in the form of tissue, fluids, imaging or patient evaluation data. Even in common diseases such as OA there are innumerable variations of patient factors creating subtle variations in an individual's disease compared to another. Indeed, the ideal point to study disease prevention may be considered to be in the presymptomatic state. It is, of

course, impossible to fully investigate all the factors of all patients at all time points and therefore all clinical research studies require a representative population to examine. It follows that methods to extract the most information from datasets are required and combining data from observations of multiple different variables in the same cohort are necessary. Where tissue and fluid samples are in short supply, such as SF, multiplex analysis methods require multivariate statistical approaches to gain the greatest insight. Multivariate methods are able to account for the noise within multicollinear data whilst simultaneously analysing all the variables in the dataset [179–181].

1.8.1 Multivariate Statistics in Biomarker Discovery

In the studies discussed in the following chapters of this thesis, for the first time, sufficient fluid was obtained to enable large-scale, single platform proteomics on hip SF. The SOMAscan assay results comprise values in relative fluorescence units (RFU) for each of the 1310 proteins for each of the patients. The challenge with such data is to reduce a high number of variables that are related in unknown ways to a number that continue to demonstrate the variation, using a comparatively small number of observations. The statistical methods that will be utilised in this thesis to reliably explore this highly dimensional data include: unsupervised principal component analysis (PCA) for hierarchical clustering and discovery of top loading markers for use in network pathway analysis, supervised partial least squares – discriminant analysis (PLS-DA) for group discrimination modelling and discovery of top loading markers, and, finally, the strict penalisation method, least absolute

shrinkage and selection operator (LASSO). This was used to account for the small number of observations, as compared with the number of variables, with the aim of identification of only the very strongest markers that describe the group differences.

1.8.2 Principal Component Analysis

PCA is an unsupervised analysis technique commonly used with the goal of reducing the dimensionality of a given dataset whilst preserving the variation within it. This is achieved by identification of linear directions of maximal variance of the data being examined. The direction of the first principal component represents the most descriptive combination of the original variables generating the most significant variation. The second is then in the direction of the next most significant variation that is not correlated with the first component [182–185]. PCA can be valuable in describing multivariate data but must be used with caution, as the user may inadvertently introduce significant systematical errors and bias, if physically meaningful components are neglected in the reconstruction. This may happen easily if their variance lies below the noise variance, so that their information is spread out over many principal components [186]. The limitation of PCA is the inability to discriminate between groups pre-defined by the user; new variables identified are defined by the dataset at hand, rather than a priori.

1.8.3 Partial Least Squares – Discriminant Analysis

PLS-DA is a supervised multivariate technique popularised due to its capacity to handle strongly collinear and noisy data with numerous X-variables (observations

in the data) while also simultaneously modelling several response Y-variables (prediction of membership of a specified class or group) by identification of a linear subspace of the explanatory variables [187,188]. However, despite its benefits in graphical exploration of highly dimensional data, one limitation is that where the number of variables greatly exceeds the number of observations, there is a risk of erroneously accepting chance associations due to model overfitting [189–191].

1.8.4 Least Absolute Shrinkage & Selection Operator

LASSO is a supervised method of penalised regression that is of particular value where the data being analysed has a low ratio of observations per variable and is thereby at risk of overfitting [192]. This penalisation is achieved by shrinkage of the regression coefficients of the variables closer to zero according to predicted risk, and, if the predicted risk is too high for any variable the coefficient, may be shrunk to zero thereby excluding the variable altogether [193]. This penalisation attribute improves model performance in multicollinear data and may be considered an alternative to PLS-based techniques [194]. Due to the graphical flexibility of PLS-DA and the penalisation in LASSO, both techniques can be used in concert to reliably explore wide datasets with low numbers of observations and high collinearity, such as broad spectrum proteomic data [155,195].

1.9 Summary

OA is a multi-factorial progressive disease that affects the entire joint. It continues to be a leading cause of pain, limitation of function, and, as a result, disability for

a great many people. Existing tools for the early detection, classification, and monitoring of the disease, are limited resulting in a series of management and treatment options that are limited to late stage, reactive measures that would at best be characterised as palliative. There remains a desperate need for disease modifying treatments in addition to the lifestyle modification measures including weight loss, diet, exercise, and smoking cessation. However, these cannot be developed in the absence of methods of early diagnosis of disease at a stage when such therapies could be of use [196]. In order to gain greater insights into the biological processes that underpin the clinical presentation it is necessary to go to the source using SF to examine the markers of disease activity present. Without the means to define different biological states of disease that can be translated into clinically useful endpoints, the hunt for disease modifying osteoarthritis drugs may be fruitless [197].

The most common sites of OA are the knee, hip and hand [198]. The small joints of the hand do not offer viable access to SF. Therefore, the hip and knee represent excellent targets to access SF for analysis. The commonest surgical management for hip and knee OA is arthroplasty with over 160,000 patients undergoing hip and knee arthroplasty surgery annually in England and Wales alone. The two-year cost of hip and knee replacement have been reported as £11,987 and £12,578 respectively when primary and outpatient care costs are added to the procedure costs [199]. This is an extremely costly treatment which would save significant sums of money if a suitable early therapy was available. Current early

interventions termed "cartilage repair techniques" such as the variety of chondrocyte implantation and osteochondral autologous transplantation procedures are expensive, challenging and have limited evidence of efficacy and safety [200].

The hunt for biomarkers in OA reflects the complexity of the pathobiology of the disease. It is unlikely that the identification of a single marker, or even a small panel of biomarkers, of disease will sufficiently diagnose or describe the disease state[201]. This requires an acceptance of combining multiple markers which, in turn, requires the use of multivariate statistical approaches for analysis and investigation. Ultimately, OA is a clinical problem experienced by patients and if a biomarker, or combination of biomarkers, were available it would not be sufficient to simply show it can work on curated samples in the experimental setting. It will be necessary to identify how to access patients at the correct timing in order to maximise its utility in the clinical setting. Currently there is wide variation in patient presentation to primary care, subsequent treatment and timing of eventual referral to secondary care for those patients that ultimately receive surgical intervention in the form of arthroplasty.

1.9.1 Literature Gaps

There is currently a paucity of literature describing the biology of early OA in the hip. It is commonly thought that OA is a single disease and that as a result the findings from one joint can be extrapolated to cover any other joint whether in human or animal studies. In part, at least where the hip is concerned, this may

relate to the challenges in accessing joint fluid and tissue samples. Hip arthrocentesis cannot be performed in the clinic room or at the bedside, and hip arthroscopy is a less common procedure when compared with the knee, making for fewer opportunities to access samples.

At present there is little evidence examining the variations in the patient experience in the community prior to accessing definitive orthopaedic care. Musculoskeletal presentations to primary care are common. However, a diagnosis of OA in a younger patient is typically not considered during the early phase of their symptoms, despite evidence to show significant effects on quality of life and psychosocial health, particularly in the younger, working age population [202–204].

1.9.2 Scope of Thesis

This thesis is primarily a discovery study, aimed at investigating the biological appearances of OA of the hip and knee using SF harvested from groups of patients with differing clinical phenotypes of disease to enable investigation into the corresponding differences in their biological phenotype. The hip and knee are two of the commonest joints affected by OA and also two of the largest synovial joints in the body that can be accessed for collection of SF. The work contained herein will:

- Validate the SOMAscan for use on SF against ELISA and MSD assays for known late stage OA markers

- Develop a novel method of collection of SF in the hip to collect a new cohort of early OA hip samples
- Analyse SF using the SOMAscan from patients with clinically diagnosed early OA, late OA (end stage) and knee injury cohorts
- Analyse SF using the SOMAscan from patients with clinically diagnosed early OA and late OA (end stage) hip cohorts
- Use clinical practice research datalink (CPRD) data to investigate the routes of access to care taken by patients ultimately receiving arthroplasty surgery to identify differences, inequalities and barriers to identifying patients at a suitably early stage that any tool could be used.

Statistical approaches also required adjustment for this highly dimensional dataset with relatively few samples. We compared classical PCA and hierarchical clustering with multivariate analysis using the PLS-DA approach piloted by our group [147] and LASSO. Standard bioinformatics was performed on the biological processes and pathways differentially expressed by patient groups to provide a biological sense check.

1.9.3 Thesis Aims & Structure

The aims and structure of this thesis will be as follows:

Chapter 2 describes the range of methods used in this thesis including a description of the novel method created for harvesting SF from the hip. It further

describes the patient groups, the samples acquired and the validation of SOMAscan for use in SF.

Chapter 3 aims to discover SF biomarker patterns of different clinical and structural disease states of knee OA groups using multivariate analyses of SOMAscan data.

Chapter 4 aims to discover SF biomarker patterns of different clinical and structural disease states of hip OA groups using multivariate analyses of SOMAscan data.

Chapter 5 aims to contextualise the future use of a potential diagnostic biological tool by use of CPRD data to determine the treatment pathway of patients that undergo joint replacement. Pathways are described relating to the order and timings of receiving analgesia, physiotherapy and intermediate surgical interventions. Variations in the pathway taken are considered, according to patient factors including age, sex, social class and BMI.

Chapters 6 and 7 comprise an overall discussion of the thesis and conclusions with suggestions for future directions of work.

Chapter 2 Methods

2.1 Introduction

This chapter provides an outline of the methods used in the studies contained within this thesis, discussing the samples used, their acquisition and validation, as well as descriptions of the patient groups, and the relevant ethical approvals.

2.2 Synovial Fluid Acquisition

SF can be aspirated from the knee with relative ease and, frequently, in large volumes. However, the hip is an entirely different prospect. Whilst it is not rare to encounter a large effusion at the time of total hip replacement (THR) this is far less common than in total knee replacement (TKR) and is often considered to be associated with inflammatory arthropathy or a rapidly destructive osteoarthritis [205]. Furthermore, with respect to the injured knee, whether the tissue injured is articular cartilage, ligament, meniscus, bone, synovium or tendon, a clinical effusion will be generated. In the hip, however, not only is it clinically challenging to identify the fluid due to the deep position of the joint, even if one was confident of the presence of a large effusion, sampling would require ultrasound or fluoroscopic image guidance and the use of a larger needle to access the fluid contained within the joint. Such a method of acquisition is therefore much less tolerated by patients.

2.2.1 Ethical Approval

The studies contained within this thesis all received appropriate ethical approvals with all study participants giving written, informed consent to be included. All human tissue used for the enclosed studies was sampled, handled and stored in accordance with the Human Tissue Act 2004.

2.2.1.1 Synovial Fluid

Samples of knee and hip SF were collected from a total of 48 patients for analysis. All samples were collected under the umbrella ethical approval via the Oxford Biobank (HTA License 12217 and Oxford REC C 09/H0606/11) or from patients enrolled in the femoro-acetabular impingement trial (FAIT) cohort (FAIT study was approved by Health Research Authority, National Research Ethics Services Committee South Central – Berkshire [REC reference: 13/SC/0154]). All patients gave informed consent for sampling on the day of surgery.

2.2.1.2 CPRD Data

Scientific approval for the epidemiological analysis of CPRD data for this study was granted by the CPRD Independent Scientific Advisory Committee (ISAC), protocol number 17_127R.

2.2.2 Knee Sampling Technique

The knee joint is superficial and lends itself to aspiration of joint fluid with ease. Furthermore, it is a capacious joint with a large capsule capable of holding large volumes of fluid generated during pathology. Fluid was aspirated under aseptic

conditions prior to the operation via superolateral portal approach with 21 gauge needle and 20ml syringe.

2.2.3 Knee Group Descriptions

Synovial fluid samples from the knee were acquired, by a collaborator, from 26 patients. The patients were admitted to the Nuffield Orthopaedic Centre for knee arthroscopy or arthroplasty. Three distinct groups of patients were identified:

1. Early OA (8 patients) – knee arthroscopy patients with no evidence of OA i.e. no loss of joint space width on radiographs, Kellgren & Lawrence score <2, Outerbridge classification at surgery <2.

2. Late OA (11 patients) – medial unicompartmental arthroplasty patients with radiographic diagnosis of late stage OA (matched to Jayadev study 2020 [147]).

3. Injury (7 patients) – patients undergoing arthroscopic surgery for Anterior Cruciate Ligament (ACL) injuries or meniscal debridement (matched to Jayadev study 2020 [147]).

2.2.4 Development of a Novel Technique for Sampling of Hip Synovial Fluid

The hip joint is unlike all other large joints in that it is a deep structure with no easily definable external landmarks that permit reliable, blind aspiration. As a result, the inherent challenges of accessing the hip joint mean that, typically, aspiration of fluid from the hip joint has been performed under image intensifier guidance in the operating theatre or fluoroscopy suite. Notwithstanding this, it is

often challenging and commonly results in a "dry tap", with no fluid being acquired. In order to acquire fluid from both end stage and early osteoarthritis cohorts a different approach was required. End stage patients were patients undergoing total hip arthroplasty and, therefore, the most reliable point for aspiration was during the operation. All THRs in this cohort were performed using the posterior approach. After dividing the piriformis tendon, and immediately prior to the capsulotomy, the hip aspiration was performed, and SF harvested. The early cohort presented a unique challenge in that all procedures are performed arthroscopically and, therefore, no direct vision method such as that used for the end stage group is possible. Furthermore, arthroscopy requires continual irrigation of fluid into the joint using sterile saline which, if introduced into the joint, dilutes the SF making biological analysis unmeasurable. It was therefore necessary to develop a novel, bespoke, approach to acquiring hip SF.

This bespoke method was successfully used to harvest samples of SF from a cohort of patients undergoing hip arthroscopy for FAI. The technique required the initial hip arthroscopy portal to be made using the conventional technique under image intensifier control. After the arthroscope is placed within the joint, the fluid is not connected as per protocol. A dry arthroscopy is then performed to identify if an effusion is present. If SF is identified then an unfiltered 22-gauge catheter, such as might be commonly used as a paravertebral catheter, is advanced into the joint via the second "working" portal and can then be used to aspirate the fluid. Catheters with embedded filters cannot be used in this method because the filter is too fine

and will not permit aspiration of SF due to its high viscosity. No samples were harvested with heavy, macroscopic, blood contamination.

2.2.5 Hip Group Descriptions

Hip SF samples were prospectively harvested from 22 patients undergoing hip arthroscopy or arthroplasty. Two groups were identified:

1. Early OA (10 patients) – a clinical diagnosis of FAI, confirmed on MRI. Sampling was performed, after considerable, and lengthy, optimisation, using a 22 gauge paraspinal catheter under direct vision via the second placed arthroscopy portal immediately prior to joint infiltration to ensure no saline contamination.

2. Late OA (12 patients) – patients undergoing hip arthroplasty with a plain radiographic diagnosis of late stage OA. Sampling was performed by needle aspiration intra-operatively after the surgical approach prior to capsulotomy.

The early group were patients with clinically and radiologically proven symptomatic FAI syndrome undergoing hip arthroscopy. Patients were diagnosed using morphological and physiological MRI sequences to define the shape of the femoral head having presented clinically with symptoms of hip impingement. If diagnosed with FAI and plain imaging revealed no evidence of degeneration, as defined by KL grade <2, they were considered suitable for inclusion. This group of patients represents a unique and extremely rare cohort whose pathology

includes FAI with evidence of acetabular carpet lesions constituting pre-radiographic changes with symptoms often presenting at K&L grade 0. All patients included in the study exhibited co-existing acetabular labral tears; an extremely common finding at arthroscopy due to the surface continuity between the acetabular labrum and the acetabular articular cartilage.

The end stage group were patients with clinically and radiologically proven hip osteoarthritis undergoing THR. Radiological evidence of osteoarthritis was defined using the Kellgren and Lawrence scale and only patients with K&L grading of three or four were included. Clinical evidence of OA was defined as patients who presented with ongoing hip pain, limitation of function and restriction of range of motion as well as having failed conservative treatment measures.

Only patients who met all criteria were approached for sample acquisition. Patients were not included who had any history of systemic inflammatory disease or any current ongoing comorbidities including diabetes, hormone disturbance, atopy, blood borne viruses or malignancy due to the potential influence on biomarker patterns.

Table 2-1 – Summary of descriptions of the groups used for the biological experiments in this thesis. Hip samples collected prospectively. Knee samples acquired from an established cohort in a collaborators group.

Study Group	Description
Knee Early OA	Knee arthroscopy patients with no evidence of OA ie: no loss of joint space width on radiographs, Kellgren & Lawrence score <2, Outerbridge classification at surgery <2
Knee Late OA	Medial unicompartmental arthroplasty patients with radiographic diagnosis of late stage OA
Knee Injury	Patients undergoing arthroscopic surgery for Anterior Cruciate Ligament (ACL) injuries or meniscal debridement
Hip Early OA	Diagnosis of femoro-acetabular impingement (FAI) made clinically and confirmed on MRI
Hip Late OA	Patients undergoing hip arthroplasty with a plain radiographic diagnosis of late stage OA

2.2.6 Sample Treatment

Upon collection, all hip samples were taken directly for processing by one researcher (John Broomfield). The samples were centrifuged at 2500 g for 15 minutes to separate cellular debris. Samples of the supernatant were then aliquoted into 1.5 ml cryovials. The samples were then snap frozen in liquid nitrogen and stored at -80 °C. Upon thawing, to room temperature, the samples were vortexed for 5 seconds and treated with hyaluronidase (HAse) following the protocol established by our lab [206]. HAse treatment was performed with a 4 mg/ml solution of bovine testicular hyaluronidase using type I-S, lyophilised

powder, 618.4 units/mg from Sigma-Aldrich. The samples were diluted 1:1 with the resultant solution and immediately vortexed for 5 s. This resulted, after dilution, in a 2 fold sample dilution with overall 2 mg/ml HAse. After treatment, these final samples were incubated for 60 minutes on a shaker at approx. 300 RPM. All hip SF samples were processed in parallel as a single batch with the same reconstituted HAse whilst knee samples were treated in series at the point of acquisition immediately after harvesting. Knee SF samples were treated with HAse, using the above protocol, on date of collection after centrifugation. They were subsequently aliquoted into cryovials, snap frozen in liquid nitrogen and stored at -80 °C.

2.3 Validating SOMAscan for Synovial Fluid

Prior to the work contained within this thesis the SOMAscan platform has been used primarily in serum studies and remained largely untested with SF. One study has been published, since the analysis described in this thesis, which analysed SF from an ACL cohort although the report did not describe any work to validate the the platform for the SF matrix [176]. Whilst the fluid properties of serum supernatant are well known and its use established, the effects of a new matrix cannot be predicted, nor can the performance of a fluid with different viscoelastic properties. Therefore SF, in and of itself, poses a technical challenge as it is a viscous ultrafiltrate of plasma that lubricates articulating joint motion [207]. When performing laboratory assays HAse treatment of SF samples reduces sample

viscosity permitting better precision during multiplex assays using Luminex while conferring no detriment on MSD [206]. SF sampling, when compared with serum or urine, yields low volumes which necessitates multiplexed platforms capable of analysing a broad range of targets using only small fluid volumes. The SOMAscan panel has the benefit of avoiding the numerous cleaning steps required in mass spectrometry to reduce highly abundant proteins. However, SOMAscan was untested for the SF matrix and, therefore, its suitability for the analysis of SF needed to be proven.

2.3.1 PlexOA

The name *PlexOA* was created as a portmanteau of Multiplex and OA and represents the product of earlier work in our group identified by Jayadev et al 2020 [147]. *PlexOA* is a synovial "fingerprint" of 20 markers (further refined to 8) expressed differentially between endstage OA, Injury and Inflammatory arthritis states. Figure 2-1 shows a heatmap representation of the 8 marker panel with most markers loading the model away from injury towards either the inflammatory or end stage OA groups. To validate the sensitivity of SOMAscan for discriminating between disease states a subset of the markers available on the SOMAscan were taken to compare the pattern across disease groups against PlexOA. All 8 markers from the refined PlexOA fingerprint 8-Plex were performed as lab assays and compared with SOMAscan on the same samples. It should be noted that all of these markers discriminated between end stage OA and knee injury, so although the inflammatory comparison is irrelevant to the current study the 8 markers are

relevant in the comparison of injury and end stage knee OA. Of the 8 markers, six were present in the SOMAscan: tissue inhibitor of metalloproteinases 1 (TIMP-1), transforming growth factor beta 1 (TGF-β1), interleukin-6 (IL-6), A disintegrin and metalloproteinase with thrombospondin motifs-4 (ADAMTS-4), interferon gamma induced protein 10 (IP-10) and monocyte chemotactic protein 1 (MCP-1). The remaining two markers were cartilage oligomeric protein (COMP) and N-propeptide of collagen IIA (PIIANP) which were not present on the SOMAscan but were tested on all three knee group samples.

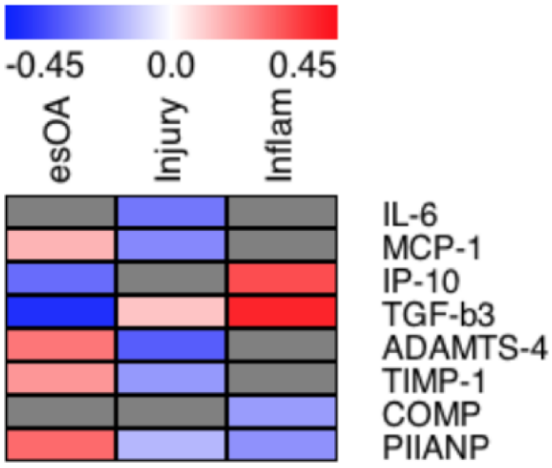


Figure 2-1 – Heatmap showing variable regression coefficient plot for 8 markers from the PlexOA study.PLS-DA model using 8 marker panel comprising IL-6, MCP-1, IP-10, TGF-β3, ADAMTS-4, TIMP-1, COMP and PIIANP for comparison of end stage OA, injury and inflammatory knee groups. (2 component model $R^2 = 0.694$, $Q^2 = 0.673$). Grey boxes denote non-significant regression coefficients where CI includes zero [147].

2.3.2 Biomarker Comparison

Aliquots of HAse treated SF samples prepared using the optimised protocol detailed above were separated for SOMAscan and manual ELISA and MSD

laboratory assays. SOMAscan samples were shipped to Somalogic, Boulder, CO, for commercial multiplex assay. Insufficient SF was available for two knee injury patients and one early hip patient (both groups with very low SF volume). Samples from these patients were only shipped for SOMAscan. Samples passed assay performance and sample quality criteria and were assayed using their standard three dilution series taking into account that HAse treated samples are diluted by 50% during the treatment process. The dilution series uses 20%, 1% and 0.005% dilutions to account for the dynamic range seen between the extremes of high and low abundance proteins within the sample medium [208].

2.3.3 Immunoassays

Laboratory immunoassays performed were competitive ELISA (PIIANP), sandwich ELISA (COMP, ADAMTS-4) and MSD (TIMP-1, MCP-1, IP-10, TGF-β1, IL-6). Manufacturer details for all assays are shown in Table 2-2.

Table 2-2 - Description and manufacturer details for all ELISA & MSD bench assays performed during the SOMAscan validation study

Assay Target	Assay Details	Catalog Number
PIIANP	EMD Millipore Human PIIANP ELISA Kit	EZPIIANP-53K
COMP	Biovendor Human COMP ELISA	RD194080200
ADAMTS-4	Cusabio Human ADAMTS-4 ELISA Kit	CSB-E11848h
TIMP-1	MSD Human TIMP-1 Assay Kit	K151JFC-1
MCP-1 & IP-10	MSD Chemokine Panel 1 (human) V-PLEX	K15047D
TGF-β1	MSD Human TGF-β1 Assay Kit	L451IUA-1
IL-6	MSD Proinflammatory Panel 1 (human) V-PLEX	K15049D

All ELISA and MSD assays were performed by a single researcher (John Broomfield) in accordance with manufacturer instructions and returned standard curve performance as expected except TGF-β1. The first attempt at TGF-β1 returned no standard curve data and repeat performance returned extremely poor data with no reading for undiluted samples. This assay was the most complex of the panel requiring a staged pre-acidification phase and overnight incubation. As a result, this presented numerous stages for investigation of possible future protocol optimisation and as such it was decided to abandon the assay in order to mitigate the risk of further waste of SF.

2.4 Patients, Materials & Methods

The knee SF samples were collected by a collaborator. Immediately after collection the samples were kept chilled prior to initial processing for a period no longer than 30 minutes. Samples was primarily diluted at ratio 1:1 with phosphate buffered saline (PBS) and incubated with 1U/ml bovine testicular hyaluronidase (H 3506, Type 1-S, Sigma-Aldrich, Poole, UK). The samples were subsequently agitated for 60 minutes at room temperature. Samples then underwent centrifugation at 200x g for 10 minutes at room temperature. The supernatant was then transferred to a clean tube and centrifuged at 1000x g for a further 10 minutes at room temperature. The resultant supernatant was then aliquoted and stored at -80 °C until assay. All samples were stored for between 28 and 32 months between sampling and assay. No samples had been defrosted between initial processing and assay.

The hip SF samples were prepared using the method described above in paragraph 2.2.6 as a single batch prior to shipping. Aliquots were shipped on dry ice to SOMALogic (Boulder Colorado, USA) for SOMAscan analysis. The demographic data including age and sex of all patients included in the study are shown in Table 2-3 below.

Table 2-3 - Demographic data for age and sex of all patients in the study cohorts

Knee Early OA		Knee Late OA		Knee Injury		Hip Early OA		Hip Late OA	
Sex	**Age**	**Sex**	**Age**	**Sex**	**Age**	**Sex**	**Age**	**Sex**	**Age**
F	39	M	73	M	61	F	32	M	62
F	38	M	64	M	38	F	46	M	42
F	66	M	53	M	20	M	46	F	60
F	50	F	78	F	43	F	33	F	59
M	73	M	69	M	46	M	49	M	67
M	57	M	51	F	28	F	33	M	61
F	61	M	71	M	55	F	58	F	28
F	50	M	78			F	39	F	50
		M	85			M	53	M	42
		F	70			F	54	M	74
		M	67					F	57
								F	83

2.4.1 Validation of SOMAscan

Standard curves for the laboratory immunoassays performed well with the majority of samples falling within the detection range of the curve producing valid interpolated results for unknowns using ELISA (COMP, PIIANP, ADAMTS-4) and MSD (TIMP-1, MCP-1, IP-10, IL-6). All markers were chosen from the PlexOA streamlined marker model as having contributed to group discrimination [147]. In that study there was an inflammatory cohort which is not featured in the analyses in this work. All sample handling was performed to the same protocol. Neither COMP nor PIIANP were proteins available on the SOMAscan 1310 protein panel and are therefore displayed as isolated ELISA profiles (Figure 2-2).

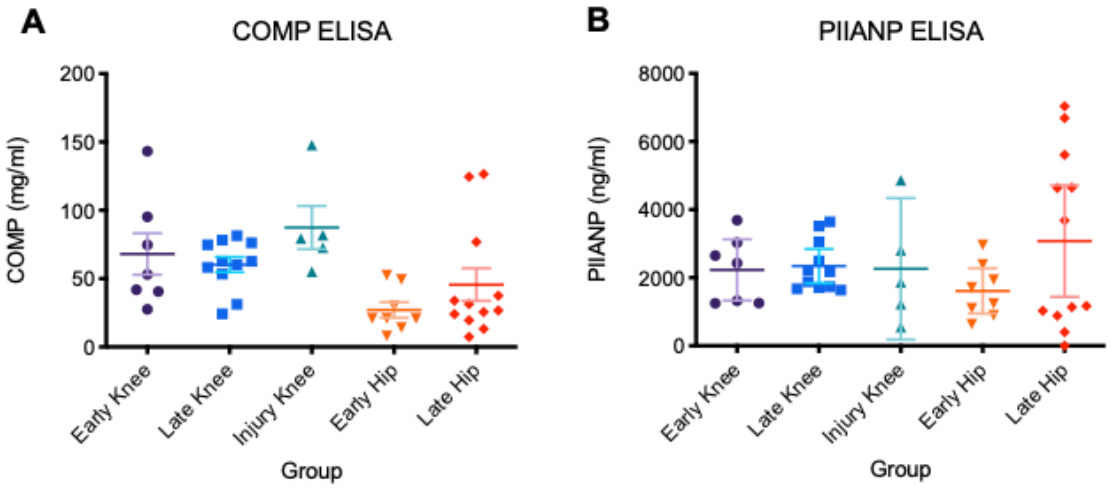


Figure 2-2 - Distributions of COMP and PIIANP ELISA assay results by group. A - COMP ELISA (mg/ml), B - PIIANP ELISA (ng/ml)

Comparison of the relative distributions of the results of ADAMTS-4, IL-6, IP-10, TIMP-1 and MCP-1 demonstrated reproducible patterns for the SOMAscan as compared with the ELISA and MSD assays suggesting reliable output data from the SOMAscan was achieved with SF (Figure 2-3, Figure 2-4). In particular, TIMP-1, MCP-1, IP-10, IL-6 and ADAMTS-4 all appeared higher in the late knee OA group, in agreement with Jayadev et al. 2020 [147], and also higher in the late hip OA group.

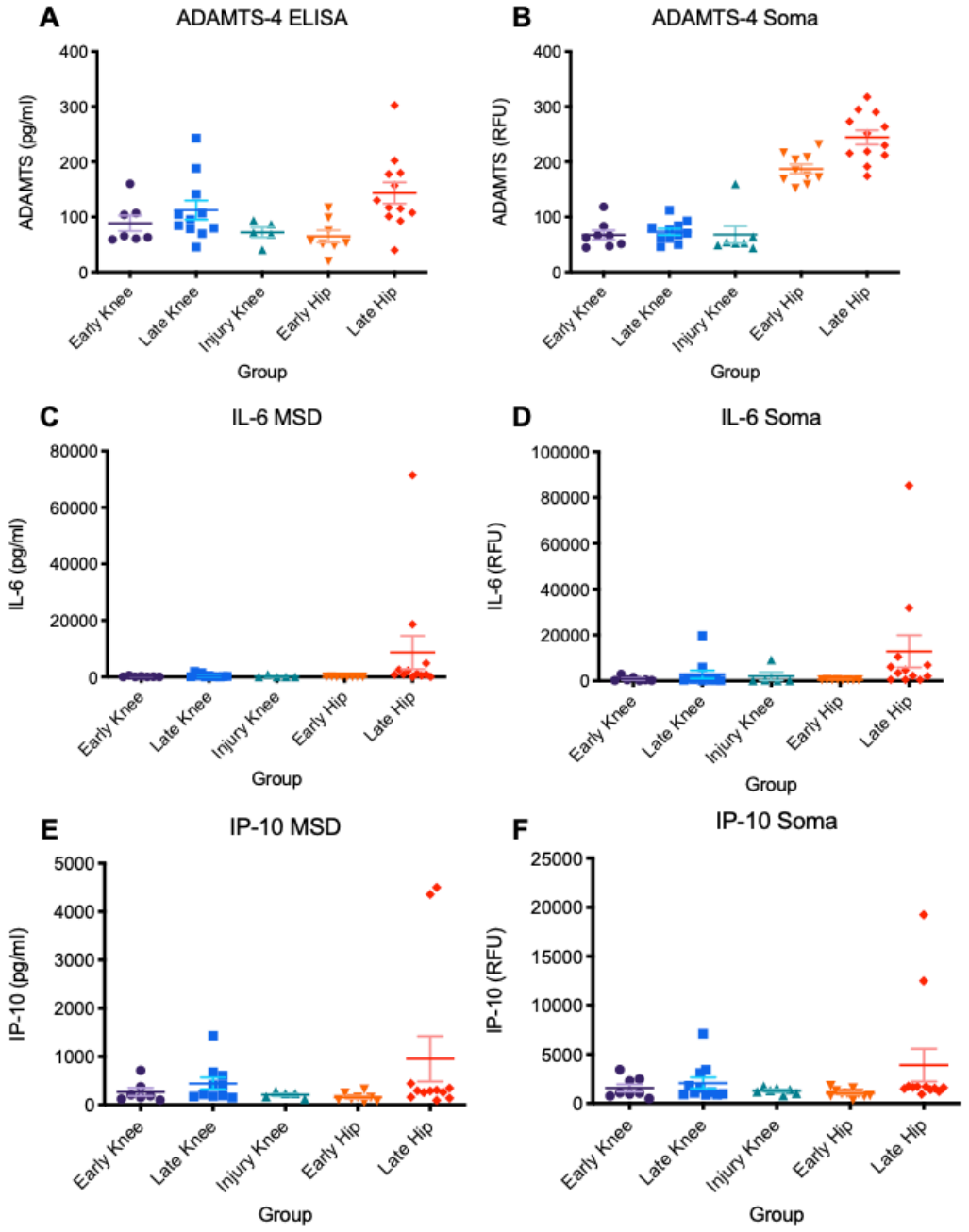


Figure 2-3 - Comparison of distributions of individual markers between ELISA & MSD assays and SOMAscan by group. A - ADAMTS-4 ELISA (pg/ml), B - ADAMTS-4 SOMAscan (RFU), C - IL-6 MSD (pg/ml), D - IL-6 SOMAscan (RFU), E - IP-10 MSD (pg/ml), F - IP-10 SOMAscan (RFU).

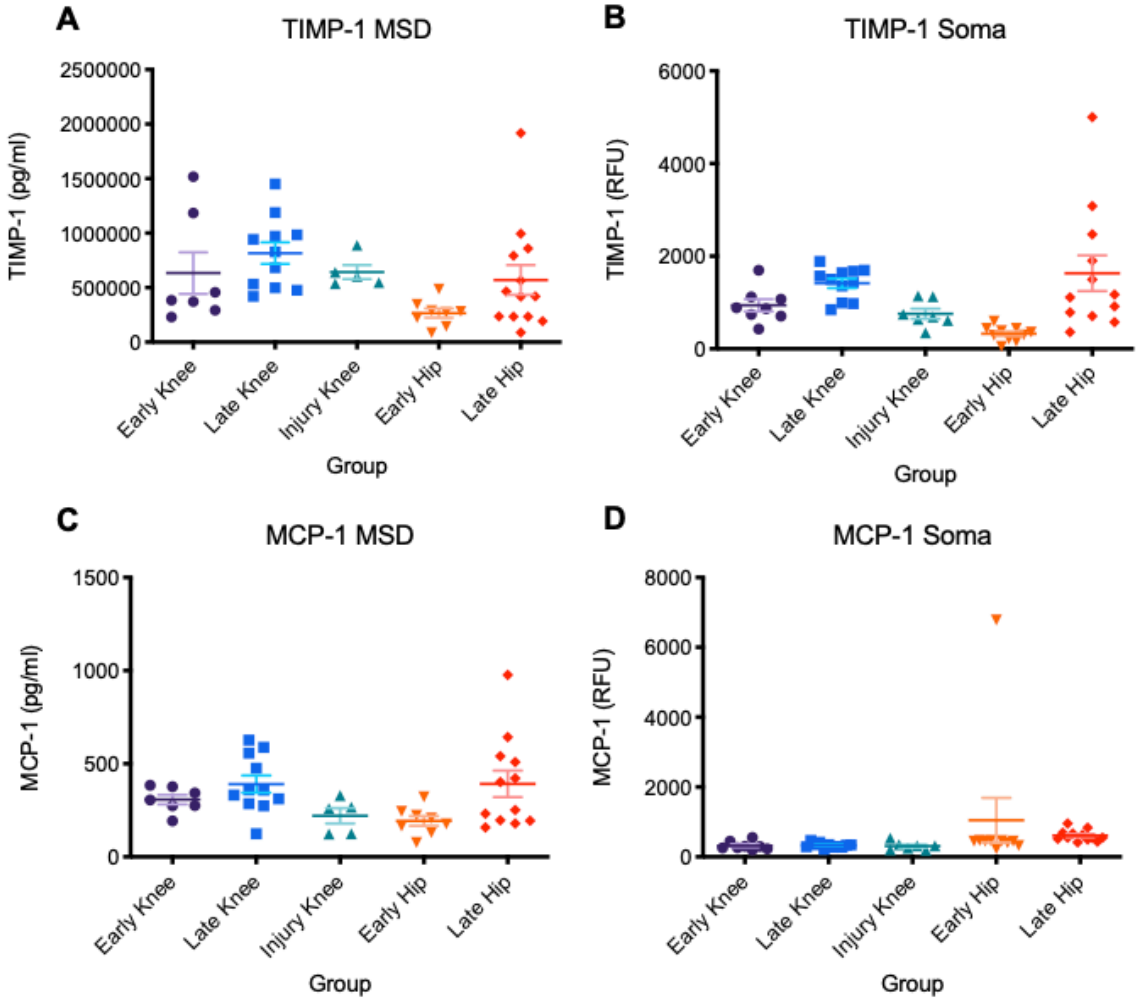


Figure 2-4 - Comparison of distributions of individual markers between MSD assays and SOMAscan by group.A – TIMP-1 MSD (pg/ml), B – TIMP-1 SOMAscan (RFU), C – MCP-1 MSD (pg/ml), D – MCP-1 SOMAscan (RFU)

The MSD assays correlated well against the SOMAscan data for IL-6, IP-10, TIMP-1 and MCP-1, however, the ADAMTS-4 ELISA did not correlate well in either knee or hip although the hip correlation did trend towards significance (p=0.0618) as shown in Figure 2-5 and Figure 2-6. It is not possible to determine whether the

ELISA assay and the SOMAlogic aptamer are measuring the same protein or epitope (Figure 2-5).

It is interesting to note that the markers with the highest dynamic range, IL-6 and IP-10, demonstrated the strongest correlations ($p<0.0001$). In both cases the correlation was also performed with exclusion of the outlying values and remained significant. Data is presented for convenience on the same graph but it should be noted that hip and knee SF were collected and processed in different batches. There may be a batch effect that affects scaling, such that hip analytes consistently show a higher dynamic range. However, the observation that those markers showing a wider range closely correlated between the different assay platforms indicates the possibility that this may be attributable to as yet unidentified, underlying sub groups of disease.

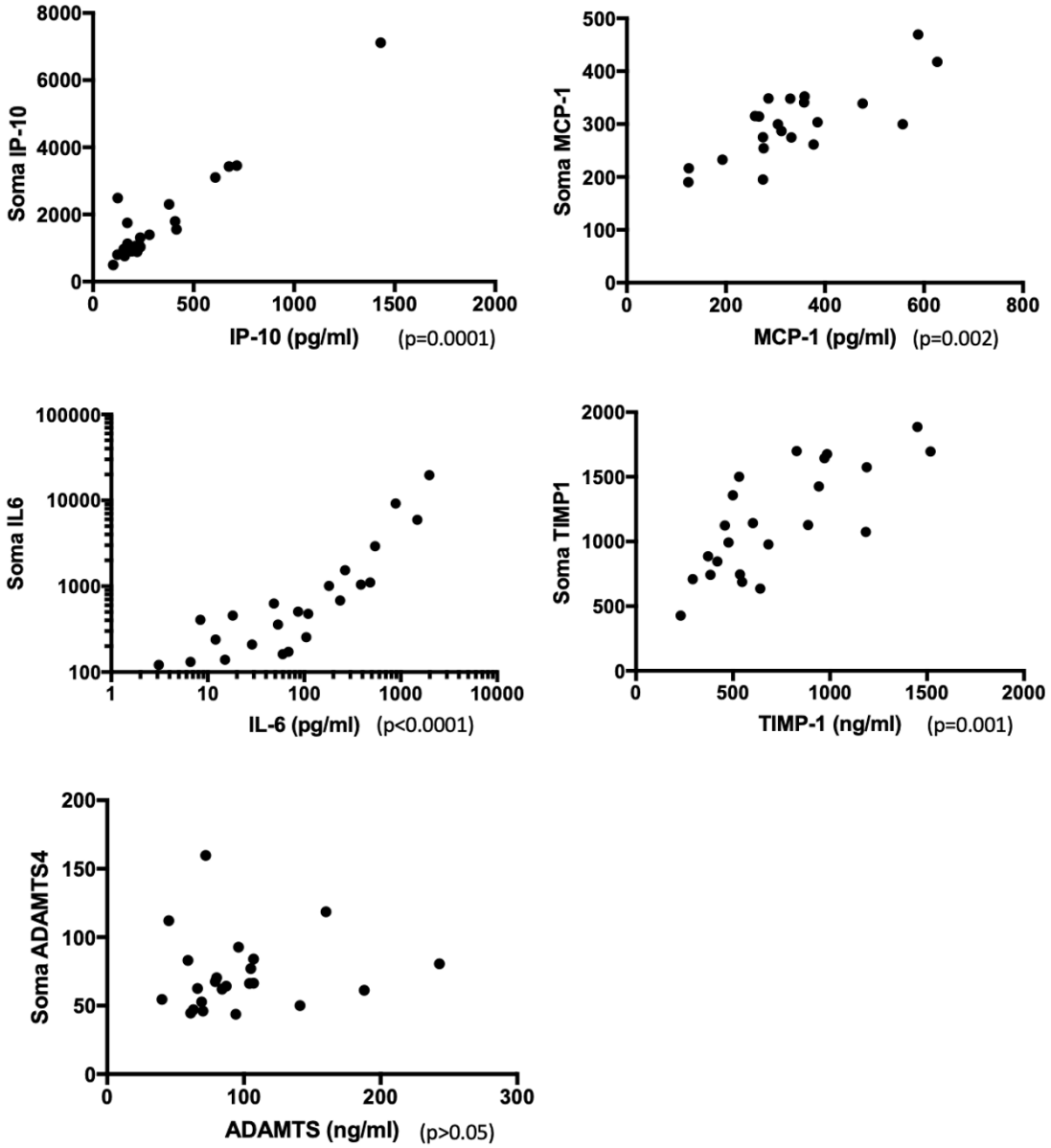


Figure 2-5 - Correlation of five SOMAlogic assays performed on remaining synovial fluid from 23 of 26 knee patients.IP10, MCP-1, TIMP-1 and IL6 were quantified by MSD assays and compared with SOMAscan data from the same samples. ADAMTS-4 was quantified by SOMAscan and ELISA. All markers demonstrate good linear correlation except ADAMTS-4 which does not correlate.

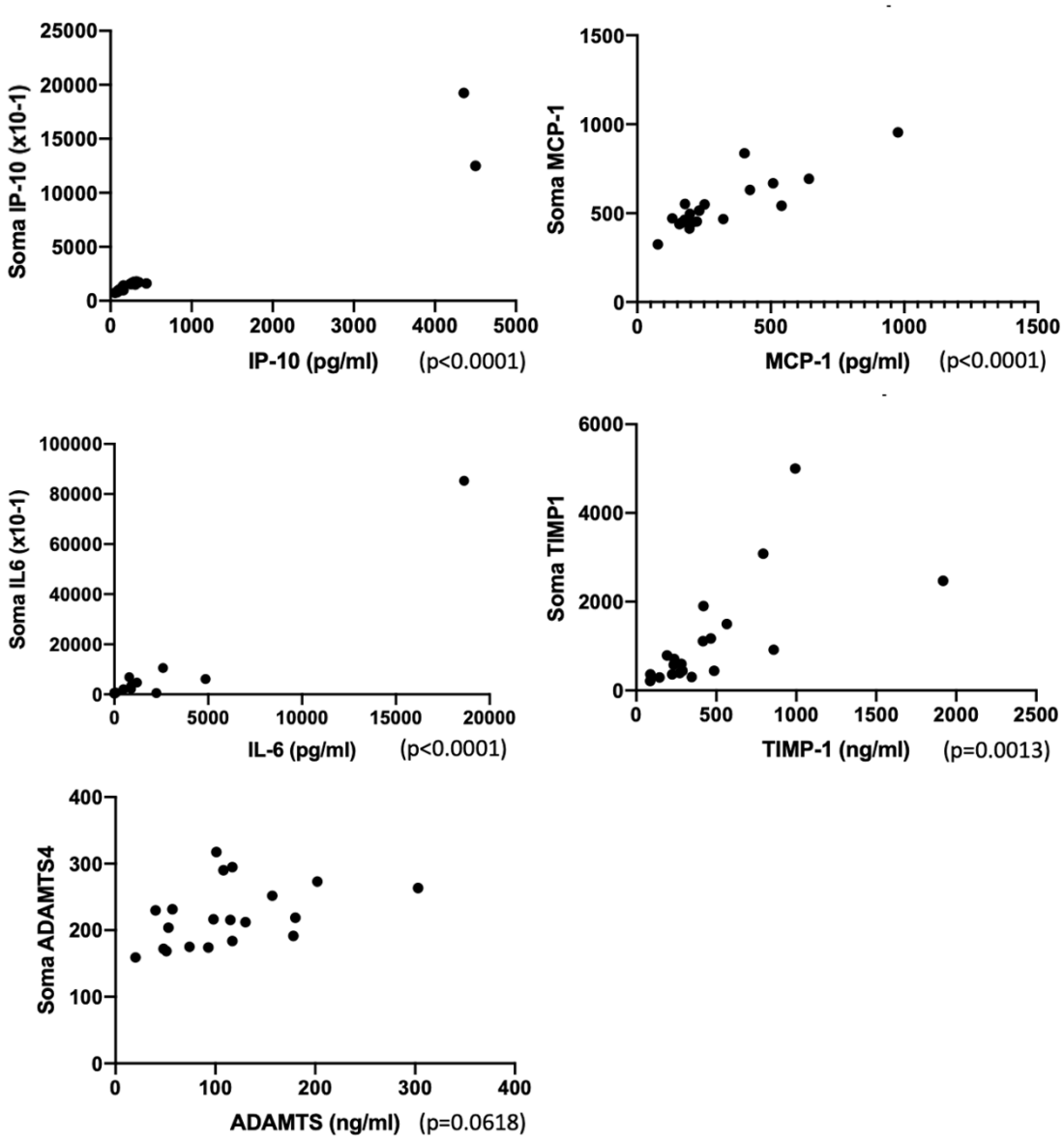


Figure 2-6 - Correlation of five SOMAlogic assays performed on remaining synovial fluid from 20 of 22 hip patients.IP10, MCP-1, TIMP-1 and IL6 were quantified by MSD assays and compared with SOMAscan data from the same samples. ADAMTS-4 was quantified by SOMAscan and ELISA. All markers demonstrate good linear correlation except ADAMTS-4 which does not correlate.

2.5 Statistical Analysis

For comparison of the SOMAscan against laboratory assays the pearson correlation was performed due to the data exhibiting a non-normal distribution. For between group comparisons t-test was used for 2 group comparisons and analysis of variance (ANOVA) was used for all comparisons involving greater than 2 groups. Analysis performed using Prism version 7.0 (GraphPad, San Diego, California, USA)

For analysis of the SOMAscan data the use of conventional statistical approaches would be unsuitable due to the wide dataset with over 1300 variables. Therefore, the multivariate methods PCA, PLS-DA and LASSO were used.

PCA was used as an unsupervised analysis to investigate patient clustering due to marker variation with graphical bioinformatic support from Dr. A. Chanalaris.

PLS-DA was used as a supervised analysis to investigate for discrimination between groups where differences are based on defined, pre-assigned classes. All SOMAscan data was in relative fluorescence units and there was no missing data. The data was initially prepared by performing a log transformation of all variables to account for skewing as a result of differential abundance of markers. The dataset was then centred and scaled to create a baseline of equal weight for all markers in the subsequent analysis [187,209]. Predictive models were then generated using SOMAscan markers as continuous X-variables and the group classification as categorical Y-variables. Model performance was assessed using R^2 and Q^2 values

where R^2 is a measure of the quality of fit and Q^2 is a measure of the quality of the predictivity of the model for predicting class membership. It is well documented that PLS-DA is at risk of model over fitting. It is accepted that external validation is the gold standard for model optimisation but that where strict internal validation is performed internal validation can be acceptable [188,210]. However, in this study, this was not possible as a result of the smaller group sizes. Therefore, a seven round internal cross validation was employed to generate the Q^2 value with Q^2 of greater than 0.5 accepted. Graphical representations of the model performance were generated. Score scatter plots describe the separation of the groups showing the relative locations of individual observations. Variable loading bi-plots describe the relative contribution of each marker to describing the differences between the groups where the theoretical location of the centre of each group is marked and the markers closest to them contribute the most to the model. Variable influence on the projection (VIP) plots show the influence of each X-variable (marker) has on the Y-variables (observations) where the sum of squares of all markers in the model is equal to the total number of markers. As a result, the average VIP score is 1 and accordingly, a score of greater than 1 indicates an above average influence on the model.

LASSO was used as a further supervised analysis by specifying the class of observations a priori. Its use further mitigates the risk of overfitting of the PLS-DA and allows confirmation of the most discriminatory markers from the PLS-DA model because it is ideally suited to data containing a small number of moderate

sized effects [192]. As with PLS-DA all markers are log transformed, centred and scaled prior to the analysis and a 10-fold multiple imputation method applied to generate the internal cross validation. The penalisation is defined such that covariates in the analysis undergo continuous shrinkage that can produce coefficients that are exactly zero where very little, or no, value is observed in the variable. Variables with a regression coefficient of zero are then dropped from the analysis altogether [195]. The resultant short list of markers generated by the penalised LASSO represents the markers whose maximal variation best describes the differences between the groups within the dataset.

Statistical analysis was undertaken with a variety of software packages. PCA and LASSO were performed using R version 3.3.3 (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/). PLS-DA was performed using SIMCA-P version 13.0 (Umetrics, Sweden). All multivariate analyses using these methods were reviewed and validated by a senior statistician (Prof. Gary Collins, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, Oxford).

Epidemiological analysis of CPRD was performed using STATA MP version 14.2 (Stata Corp., College Station, TX). Graphing was performed using Prism version 7.0 (GraphPad, San Diego, California, USA), Discovery Workbench 4.0 desktop

edition (Mesoscale Discovery, Rockville, MD) and Microsoft Excel for Mac version 16.41 (Microsoft Corporation, Redmond, WA).

2.6 Patient Risk Assessment

The studies contributing to this thesis did not require direct administration of any treatments or medications to any patients. SF samples were taken at the time of operation at the discretion of the operating surgeon. All patients had given informed consent for the collection of samples either as part of the Oxford Musculoskeletal Biobank (OMB) or the femoroacetabular impingement trial (FAIT). Examples of the participant consent forms are shown in the accompanying appendices A.1 and A.2 respectively.

Chapter 3 The Biological Patterns of Synovial Fluid in Knee OA

3.1 Introduction

In the search for biomarkers of OA the prospect of a single diagnostic marker has yielded no fruit. The focus has to shift to the recognition that there are too many actors to expect the influence of a single marker to rise above all others. The disease may be isolated to a single joint, but more commonly exists in multiple joints at different states of disease progression. Furthermore, within each joint there a number of possible tissues of origin that could be the driver of pathogenesis. Before there may be a prospect of developing a treatment to address underlying pathology in OA there will need to be deeper understanding of the common initial biological changes associated with the disease. Furthermore, this means understanding the relative changes in associated markers of the different states of OA will be required to derive a fingerprint.

The diagnostic methods and biological understanding of OA fall woefully behind that of other chronic progressive diseases such as RA. The OA space is desperate for a more informative description of the biological phenotype of the disease. Advances in broad scope proteomics platforms, such as the SOMAscan, have generated the opportunity to attempt to develop this fingerprint by virtue of being capable of analysing 1310 proteins in a single sample. Traditional assay platforms have not permitted the analysis of such significant numbers of proteins due to the

amount of sample required to perform each assay and the limited volume of sample available. The full list of 1310 proteins contained in the assay is shown in Appendix B.1. Traditional methods also are limited by cross reactivity in the assay as well as the simple need for different treatment environments for individual analytes limiting the number of proteins that can be simultaneously analysed. This chapter discusses how the SOMAscan assay was used for the first time on SF from the knees of patients with different stages of OA.

3.2 Aim

The aim of this study was to identify biomarker fingerprints of different clinical and structural disease states of knee OA groups using multivariate analyses of SOMAscan data. Specifically, we sought to identify a separate signature for the early knee OA group, since multiple groups, including ours, have found biomarker differences between knee injury and late OA.

3.3 Groups

The knee SF samples were obtained from three groups of patients (as described in chapter 2 above and Table 3-1 as follows:

1. Early OA (8 patients)
2. Late OA (11 patients)
3. Injury (7 patients)

Table 3-1 – Knee cohort individual group participant demographics

Group	Early OA	Late OA	Injury
Participants	**8**	**11**	**7**
Mean Age (yrs)	54.2	69.0	41.6
Female	6	2	2
Male	2	9	5
K&L Grade	<2	3/4	<2

3.4 PLS-DA

Traditional analysis of data begins with graphical representation by plotting the results of analysis of one variable at a time. Multiple variables can be plotted on a single graph but more than a relatively low number of variables becomes challenging for the reader to interpret. This type of representation also permits no inference of linkage of results between variables besides the visual similarity. Accordingly, multivariate techniques were developed for use in situations where there are numerous variables where changes in the "input" variable are responsible for, or describe, changes in the "output" variables. It has been developed specifically for use in data which is highly dimensional and where the number of variables is significantly greater than the number of observations.

3.4.1 Knee Early OA, Late OA and Injury Three Group Comparison Using the Whole SOMAscan

Taking the whole SOMAscan panel of 1310 markers and building the model with all three groups of knee patients demonstrates the overlap of each of the groups and that the model is unable to clearly separate the groups. The score scatter plot displays the observations within the model and how well they are described by

the variables (Figure 3-1). Observations that are similar lie together. The confidence ellipse represents the area of the model space lying within the significance level specified by default at 0.05. The R^2 gives a measure of "goodness of fit" and the Q^2 gives a measure of "goodness of prediction", with values at and above R^2 of 0.6 and Q^2 of 0.5 respectively taken to be acceptable. The Q^2 of 0.185 gives no confidence that this is a predictive model. The loading scatter plot displays the variables within the model and their contribution to it (Figure 3-2). The variables with the largest values dominate the projection and variables that lie near to each other are closely correlated in their relative contribution to the overall model projection. The centre of each of the groups of observations as described by the model are shown on the loading scatter plot.

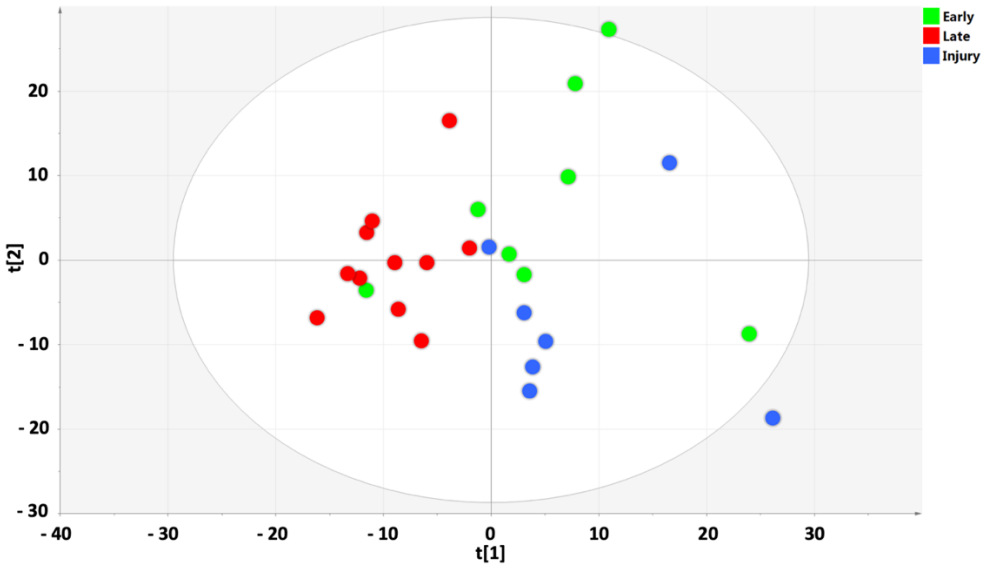


Figure 3-1 - Score scatter plot for PLS-DA model of knee early OA, injury and late OA group comparison using the whole SOMA panel of 1310 markers.The patients comprising the late OA group (indicated in red) are more closely clustered together when compared to the early OA and injury groups (indicated in green and blue respectively) which are more spread out and significantly overlap each other. 3 component model (R^2 = 0.672 Q^2 = 0.185).

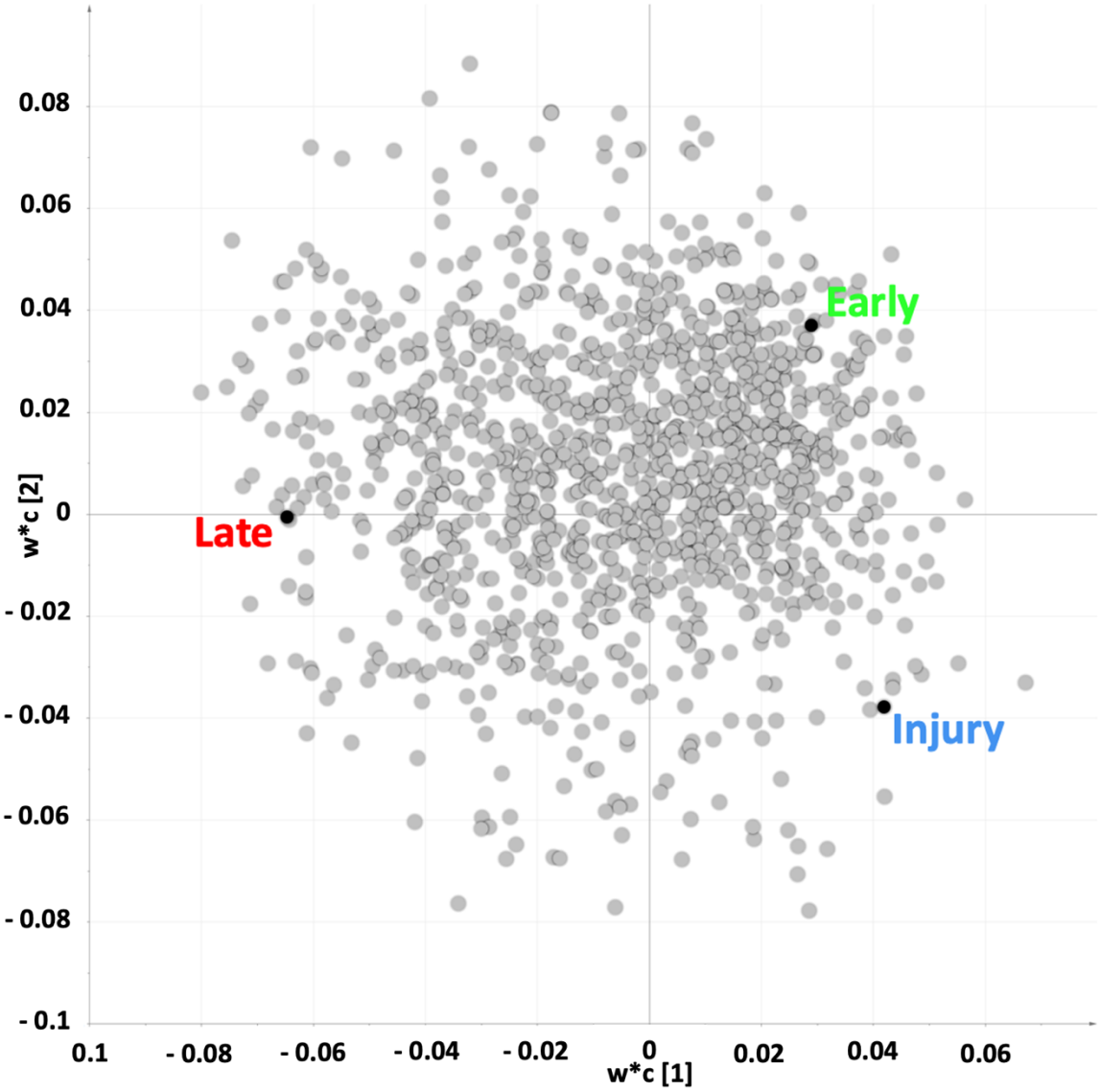


Figure 3-2 - Loading scatter plot for PLS-DA model of knee early OA, injury and late OA group comparison using the whole SOMA panel of 1310 markers. Grey points represent relative contribution of each marker to the model. The position of the group centroids are indicated showing that the majority of markers cluster towards the centre of the model space indicating a weaker comparison with significant noise. 3 component model (R^2 = 0.672 Q^2 = 0.185)

As can be seen from this model the groups cannot be reliably separated by the whole panel of markers taken as an entire model, although late OA samples tend to cluster. However, it is also clear from the model performance characteristics and the huge number of markers close to the centre of the plot that the vast majority of

the markers in the model offer no descriptive value to the model and are instead introducing noise.

The VIP plot demonstrates the relative contribution of each marker to the model. It is constructed such that the mean VIP score for all variables is 1. The degree to which a variable scores above or below 1 indicates the magnitude of its effect. Customarily a threshold of 0.8 is taken above which a marker is considered to significantly contribute to the model and a score below 0.5 indicates no comparative value for that variable.

The VIP for this model (Figure 3-3) demonstrated the most informative markers carried a significant weight despite the noise within the model. The top 50 are shown here for simplicity with all top 50 markers ranked above 1.5 and therefore giving significant weight to the model.

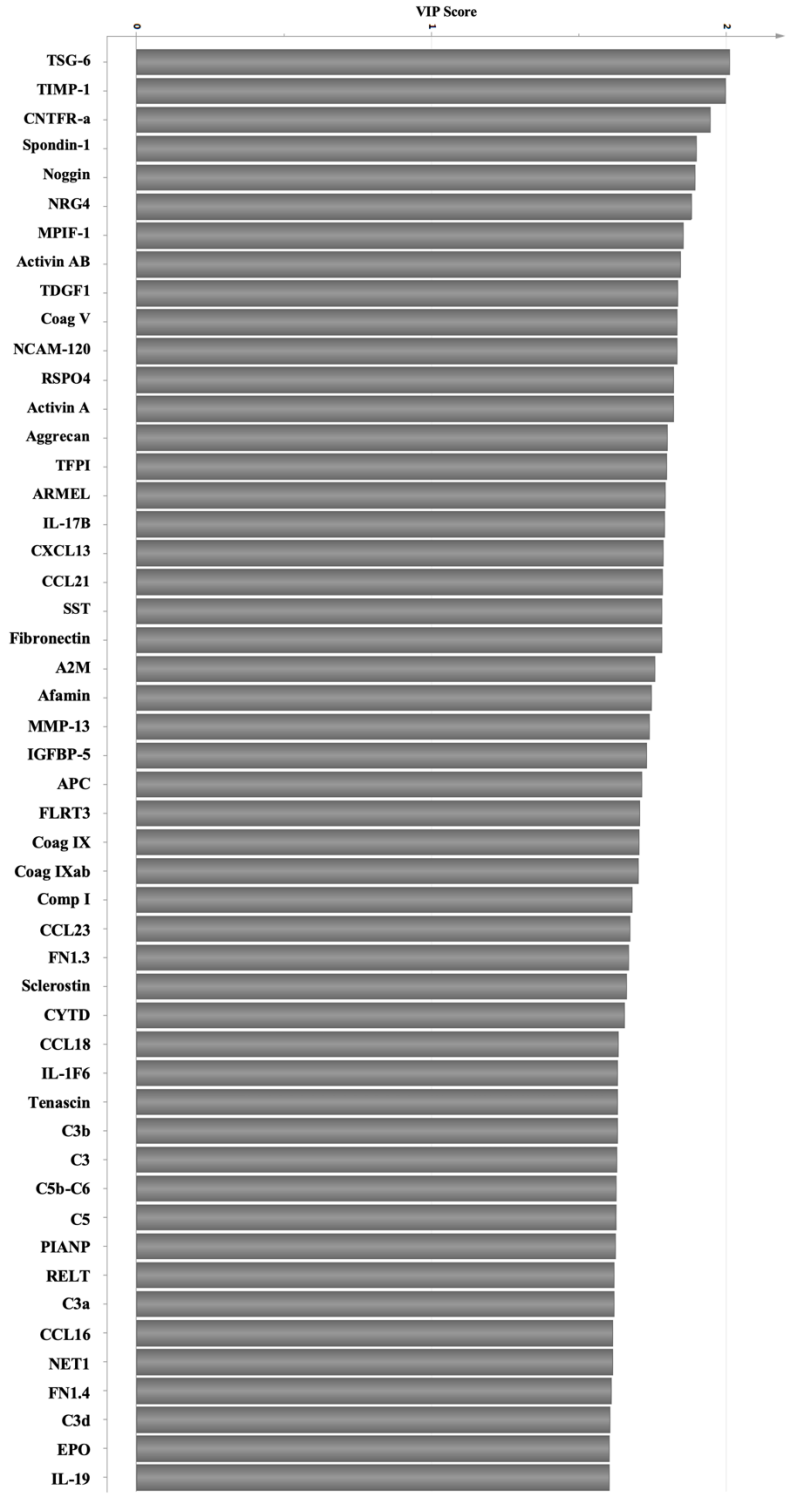


Figure 3-3 – VIP plot showing the top 50 markers from the PLS-DA model of knee early OA, injury and late OA group comparison using the whole SOMA panel of 1310 markers.3 component model (R^2 = 0.672 Q^2 = 0.185). All markers shown rank above 1.5 in this model, where 0.8 is customarily taken as a cut off for significant contribution and 2.0 is indicated by the right hand line on the graph area. See appendix B.1 for details on markers displayed.

3.4.2 Knee Late OA and Injury Group Comparison Using the Whole SOMAscan

In view of the lack of separation seen in the model including all three groups and the relative positions of the groups it appeared that the injury and late OA groups were well separated. As such when the early group is removed from the model the result improved to show clearer separation of the groups (Figure 3-4) with a greater number of markers loading towards the late OA group (Figure 3-5) and improved performance characteristics of $R^2 = 0.940$ and $Q^2 = 0.555$.

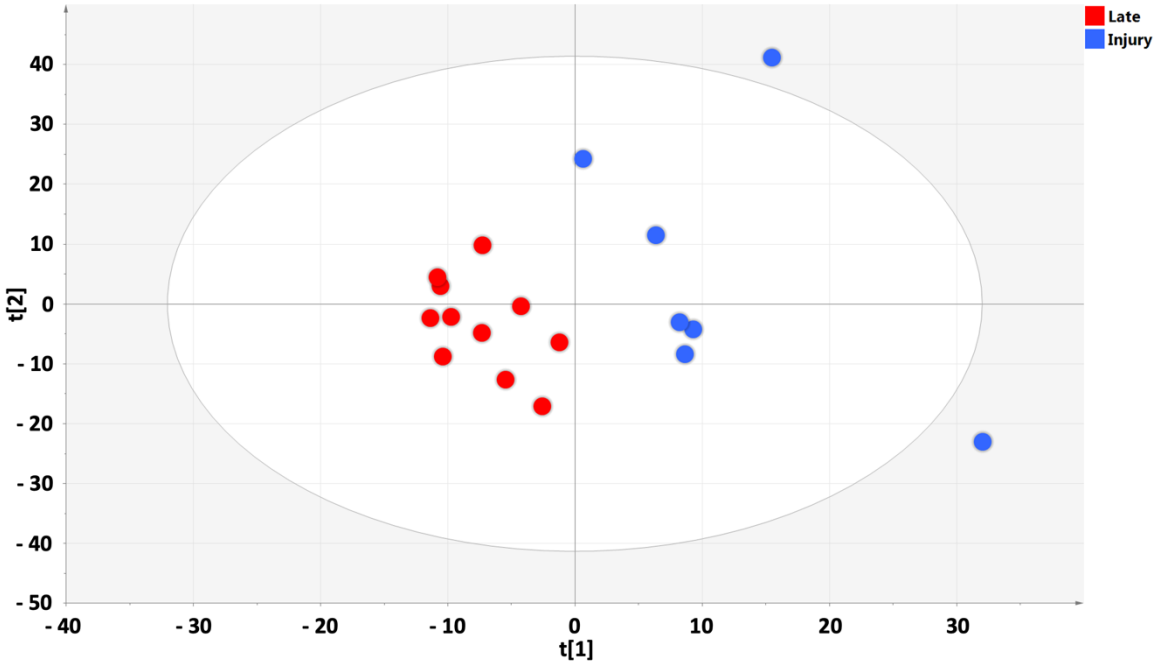


Figure 3-4 - Score scatter plot for PLS-DA model of knee injury vs. late OA group comparison using the whole SOMA panel of 1310 markers. Late OA group (indicated in red) are very closely clustered compared to the injury group (indicated in blue). 3 component model ($R^2 = 0.940$ $Q^2 = 0.555$)

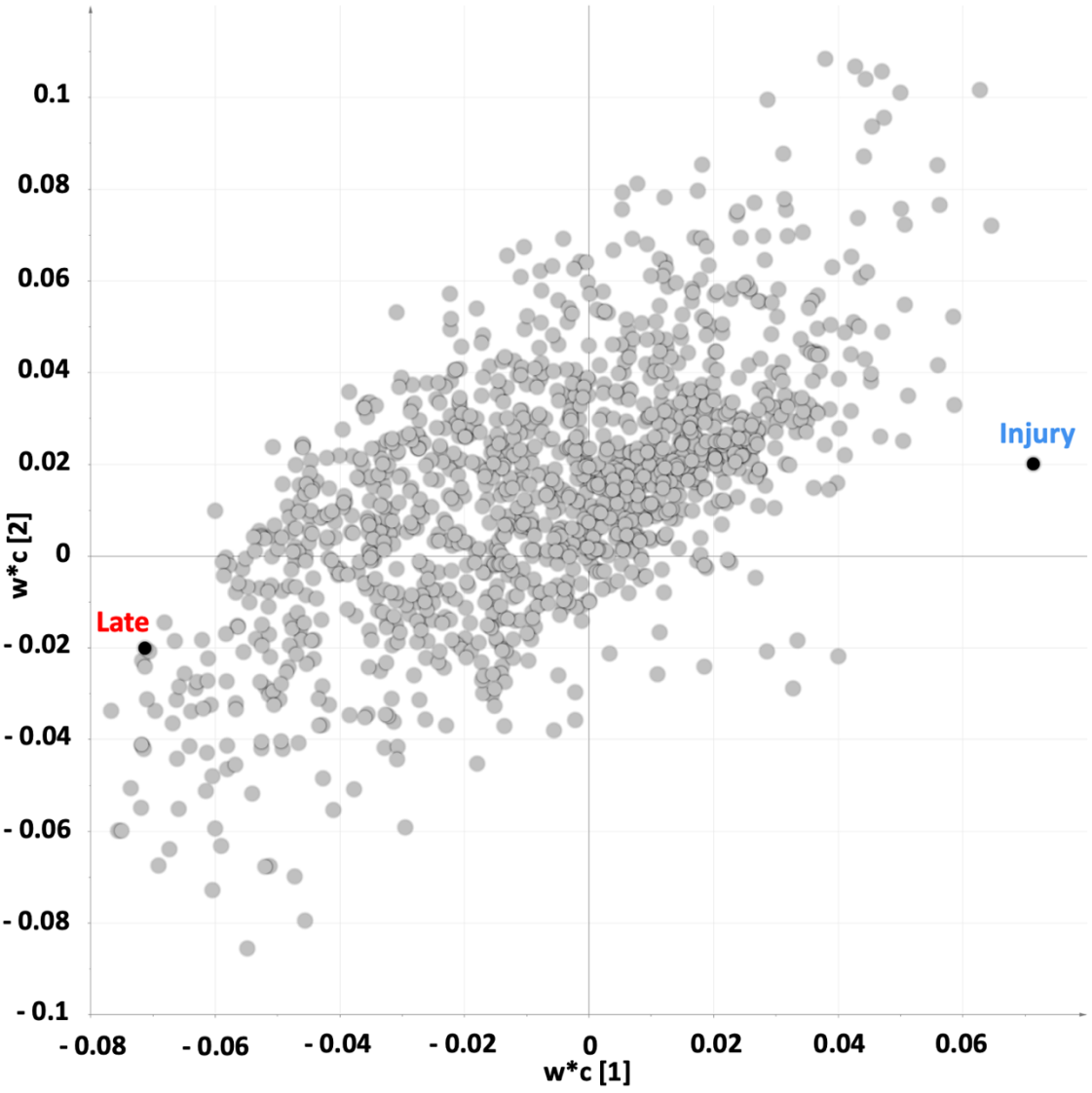


Figure 3-5 - Loading scatter plot for PLS-DA model of knee injury vs. late OA group comparison using the whole SOMA panel of 1310 markers. Grey points represent relative contribution of each marker to the model. The position of the group centroids are indicated showing that the majority of markers cluster towards the centre of the model space. A significantly higher marker clustering is seen around the late group centroid than the early group centroid. 3 component model (R^2 = 0.940 Q^2 = 0.555)

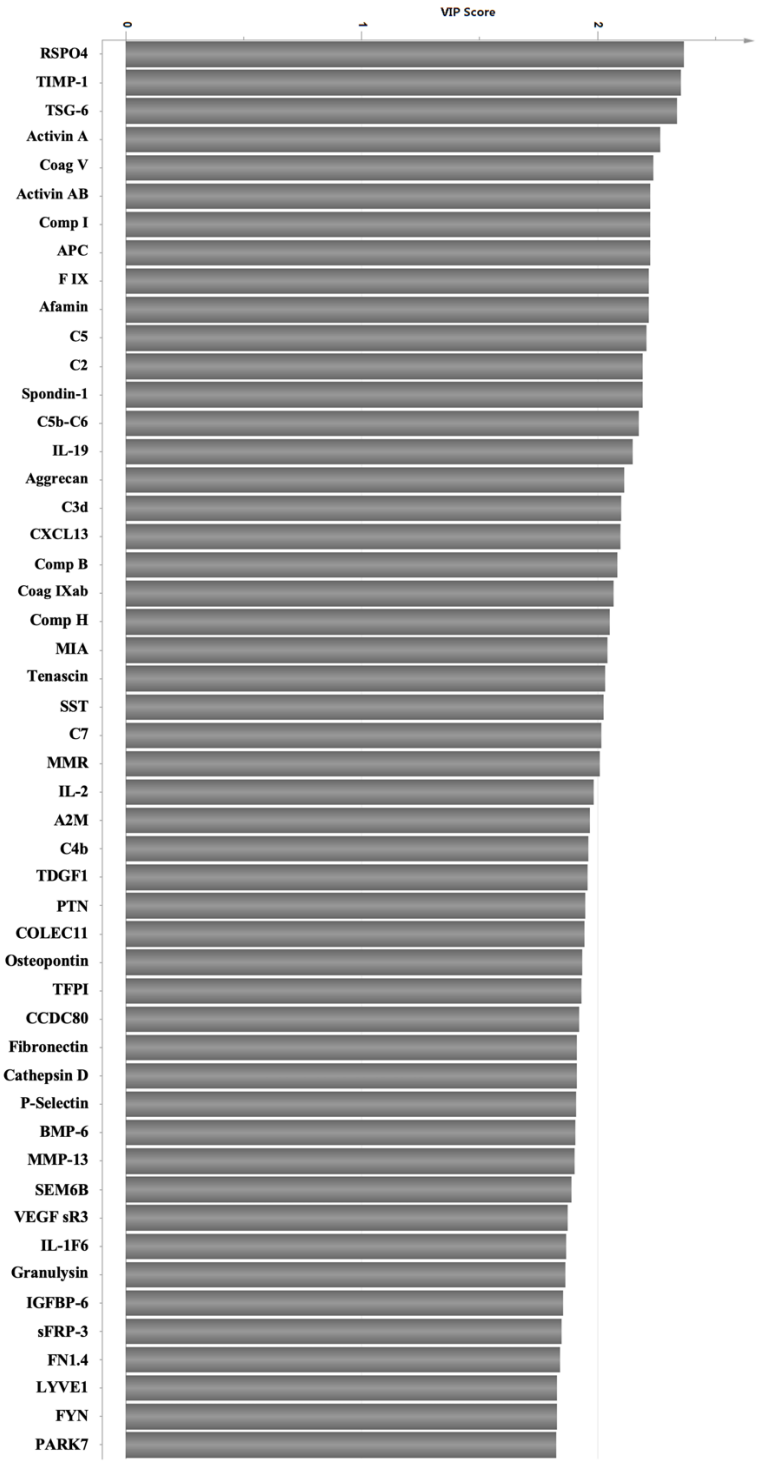


Figure 3-6 – VIP plot showing the top 50 markers from the PLS-DA model of knee injury vs. late OA group comparison using the whole SOMA panel of 1310 markers. 3 component model (R^2 = 0.940 Q^2 = 0.555)All markers shown rank above 1.75 in this model, where 0.8 is customarily taken as a cut off for significant contribution and 2.0 is indicated by the right hand line on the graph area. See table 3-2 and appendix B.1 for details on markers displayed.

Similar to the results of the three group comparison the vast majority of the markers in the model add little to the model and are introducing noise whilst the VIP plot demonstrates that the most important markers score extremely highly indicating that the lowest markers are only creating noise (Figure 3-6). It is evident that more markers cluster around the late OA centroid, with low density of markers near the injury group centroid demonstrating that this model is predicted by strong late stage markers.

3.4.3 Knee Late OA and Injury Group Comparison the Most Descriptive 50 Markers From Whole SOMAscan Comparison

Continuing this iterative approach to exploration of the marker profile, next a comparison of the injury and late groups was subsequently performed using only the top 50 markers from the previous model. In this model, the noise from non-contributing markers is removed and allows a greater exploration of the individual proteins and their relative contribution to the discrimination of the groups. This resulted in a further improved model with $R^2 = 0.984$ and $Q^2 = 0.921$ showing very close grouping of the late OA group and excellent separation from the injury group.

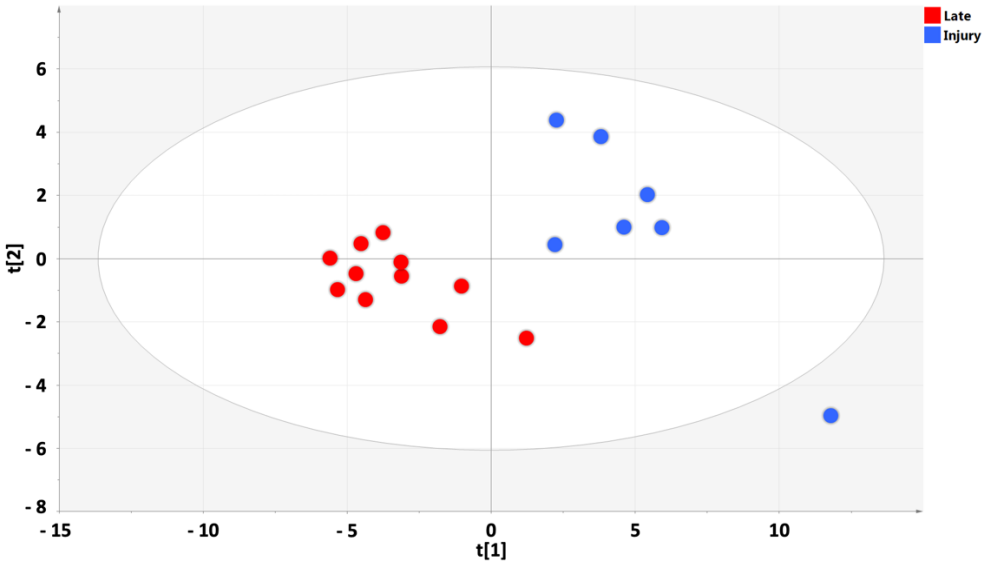


Figure 3-7 - Score scatter plot for knee injury vs. late group comparison using top 50 marker panel. 3 component model (R^2 = 0.984 Q^2 = 0.921)

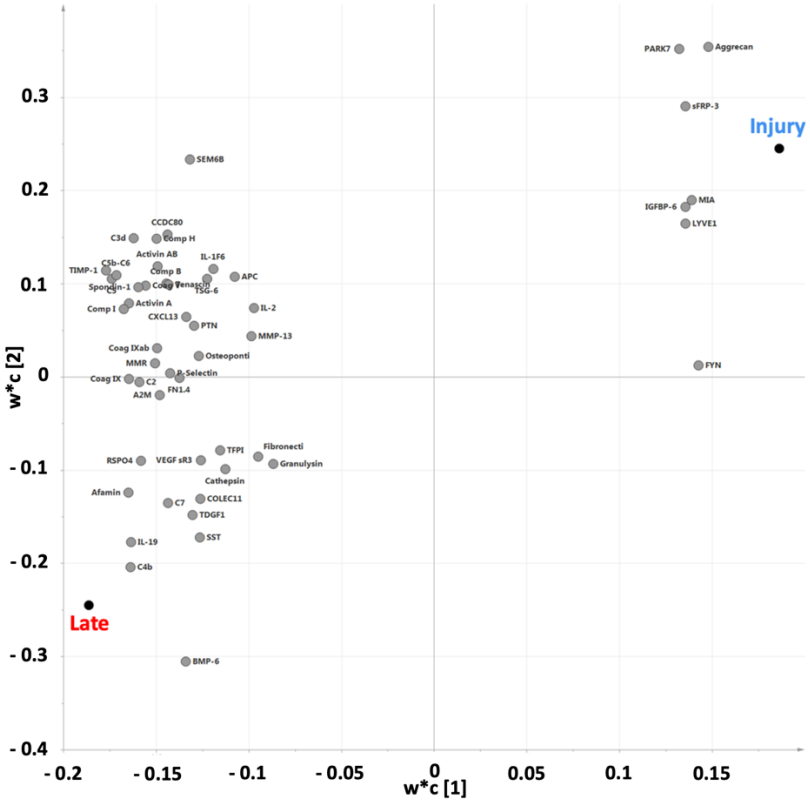


Figure 3-8 - Loading scatter plot for PLS-DA model of knee injury vs. late OA group comparison using top 50 marker panel. 3 component model (R^2 = 0.984 Q^2 = 0.921)

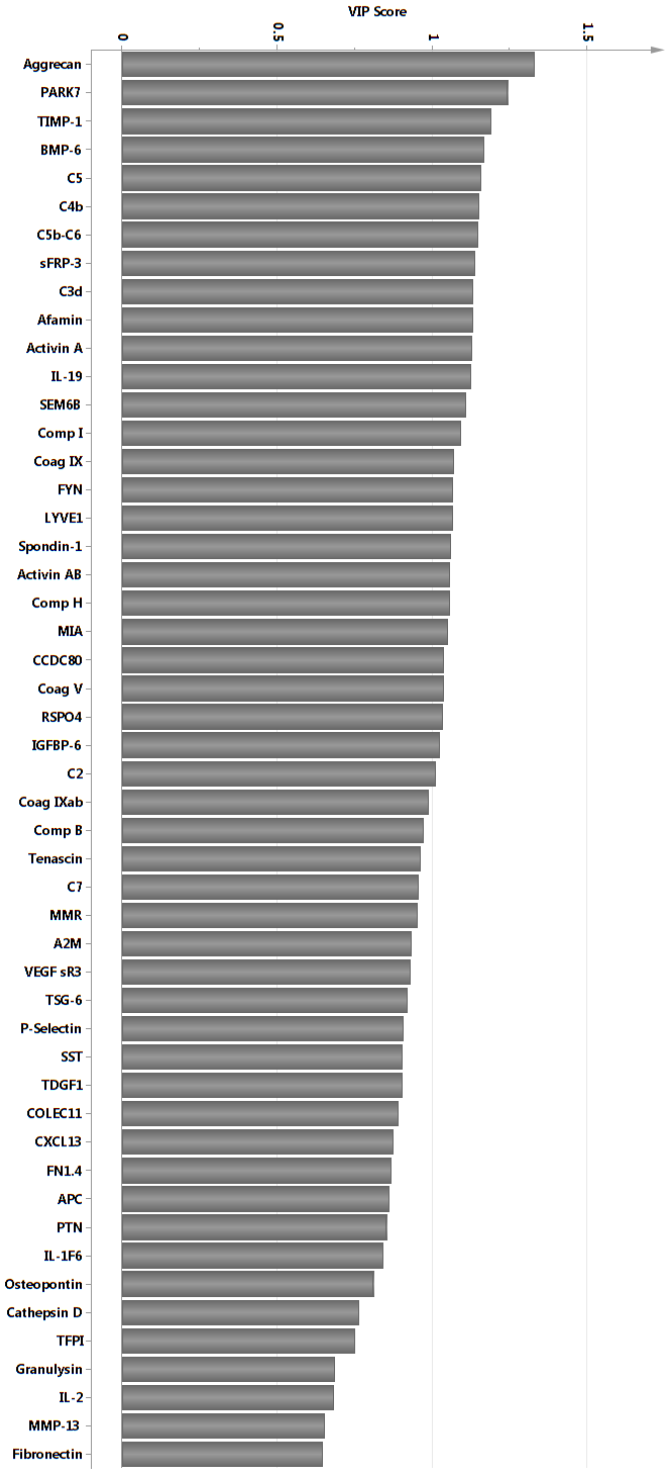


Figure 3-9 - VIP plot for PLS-DA model of knee injury vs. late OA group comparison using top 50 marker panel. 3 component model (R^2 = 0.984 Q^2 = 0.921)44 of the 50 markers shown rank above 0.8 in this model, where 0.8 is customarily taken as a cut off for significant contribution and 1.5 is indicated by the right hand line on the graph area. See table 3-2 and appendix B.1 for details on markers displayed.

The removal of less informative markers from the model demonstrates a significant improvement in the model performance with an allied improvement in separation (Figure 3-7). The loading scatter plot (Figure 3-8) demonstrates which markers describe the observations in each group. The majority of the markers load the model towards the late OA group with fewer loading towards injury. The seven markers loading towards injury are aggrecan, MIA, IGFBP-6, sFRP-3, LYVE1, FYN and PARK7 ranked 16, 22, 45, 46, 48, 49 and 50 respectively in the VIP rank order as shown below in Figure 3-9 above and Table 3-2 below.

Table 3-2 - Table showing top 50 markers from the injury vs late OA group comparison panel (Fig. 3-6) with UniProt ID, group that the marker loads towards and brief description, ** indicates LASSO marker.

Full Name	Marker ID	UniProt ID	Group	Description
R-Spondin-4 **	RSPO4	Q2I0M5	Late	Potentiate Wnt signaling by interaction with leucine-rich G protein receptors [211]
Metalloproteinase inhibitor 1 **	TIMP-1	P01033	Late	Metalloproteinase inhibitor broadly active against matrix metalloproteinases [212]
Tumour necrosis factor-inducible gene 6 protein **	TSG-6	P98066	Late	Glycoprotein up regulated in OA progression and acute knee injury [213]
Inhibin beta A chain	Activin A	P08476	Late	TGF-β cytokine family member with activities in many systems [214]
Coagulation Factor V	F V	P12259	Late	Central factor in coagulation for activation of prothrombin to thrombin
Inhibin beta A chain / B chain heterodimer	Activin AB	P08476 P09529	Late	TGF-β family heterodimer with diverse activity in numerous systems [214]
Complement Factor I	Factor I	P05156	Late	Serine protease that regulates immunity by control of all complement pathways [215]
Activated Protein C	APC	P04070	Late	Regulator of anticoagulation & inflammation by inhibition of Factor Va & VIIIa [216]
Coagulation Factor IX	F IX	P00740	Late	Vitamin K dependent coagulation factor in the intrinsic pathway
Afamin	Afamin	P43652	Late	Induces signaling by making vitamin E and Wnt ligands soluble [217]
C5	C5	P01031	Late	Local inflammatory mediator and component of the membrane attack complex
C2	C2	P06681	Late	Classical complement pathway factor. Combination with C4b activates C3 & C5
Spondin-1	Spondin-1 F-spondin	Q9HCB6	Late	An ECM protein that directs cell migration during development [218]
Complement C5b / C6 complex	C5b/6	P01031 P13671	Late	Complex of C5b & C6 which forms the basis of the C5-9 membrane attack complex
Interleukin 19	IL-19	Q9UHD0	Late	Pro-inflammatory cytokine, activates JAK/STAT pathway, stimulates TNF & IL-6 [219]
Aggrecan core protein	Aggrecan	P16112	Injury	The major PG in human articular cartilage required for normal function
Complement C3d Fragment	C3d	P01024	Late	The central factor of both classical and alternate complement cascades
CXC motif chemokine 13	BLC CXCL13	O43927	Late	Chemokine with key role in inflammation and immunity as B cell chemoattractant [220]

Complement Factor B	Factor B	P00751	Late	Component of the alternate pathway, generates the C3/C5 convertase [221]
Coagulation Factor IXab	F IXab	P00740	Late	Vitamin K dependent coagulation factor in the intrinsic pathway
Complement Factor H	Factor H	P08603	Late	Glycoprotein inhibitor of complement activation and amplification [222]
Melanoma-derived growth regulatory protein	MIA	Q16674	Injury	Expressed and secreted by melanoma cells, interacts with integrins alpha & beta [223]
Tenascin	Tenascin	P24821	Late	ECM guidance protein for synaptic plasticity and neuronal regeneration [224]
Somatostatin-28	SST	P61278	Late	Inhibitor of secretion of pituitary hormones: GH, TSH, PRL
C7	C7	P10643	Late	Component of complement cascade that forms part of the membrane attack complex
Macrophage Mannose Receptor	MMR	P22897	Late	Suggested to mediate allergic sensitisation to multiple glycoallergens [225]
Interleukin 2	IL 2	P60568	Late	Central cytokine in immunity, stimulates B cells, NK cells, monocytes.
a2-Macroglobulin	A2M	P01023	Late	A promiscuous proteinase inhibitor which functions by entrapment [226]
C4b	C4b	P0C0L4	Late	Classical complement pathway factor. Combination with C2a activates C3 & C5
Teratocarcinoma-derived growth factor 1	TDGF1 CRIPTO	P13385	Late	A cell membrane protein forming activin complexes in nodal signaling [227]
Pleiotrophin	PTN	P21246	Late	Growth factor that binds with GAG and has a role in tissue regeneration [228]
Collectin-11	COLEC11	Q9BWP8	Late	A lectin which binds glycoproteins and has a role in innate immunity [229]
Osteopontin	Osteopontin	P10451	Late	Cytokine that binds to hydroxyapatite and integrins alpha and beta [230]
Tissue factor pathway inhibitor	TFPI	P10646	Late	Inhibits thrombin generation via factor X, VIIa and tissue factor inhibition [231]
Coiled-coil domain-containing protein 80	CCDC80	Q76M96	Late	Widely expressed, repressed by insulin and TNF and acts via Wnt signaling [232]
Fibronectin	FN1	P02751	Late	Ubiquitous ECM glycoprotein that binds various targets including collagens [233]
Cathepsin D	Cathepsin D	P07339	Late	An intracellular acid protease cleaved and activated by ADAM-30 [234]

P-Selectin	P-selectin SELP	P16109	Late	Cell adhesion receptor which mediates tethering of leukocytes [235]
Bone morphogenetic protein 6	BMP-6	P22004	Late	TGF-β family member that induces cartilage and bone formation [236]
Collagenase 3	MMP-13	P45452	Late	ECM degrading enzyme with highest activity to type II collagen [237]
Semaphorin-6B	SEM6B	Q9H3T3	Late	Extracellular signaling protein regulating nervous system development [238]
Vascular endothelial growth factor receptor 3	VEGF sR3	P35916	Late	Cell surface TKR for VEGFC & D, regulates endothelial cells and angiogenesis [239]
Interleukin-36 alpha	IL36A IL-1F6	Q9UHA7	Late	Pro-inflammatory cytokine that activates NF-κB and MAPK signaling [240]
Granulysin	Granulysin	P22749	Late	An immune alarmin that indirectly increases the immune response [241]
Insulin-like growth factor-binding protein 6	IGFBP-6	P24592	Injury	Binds its cell surface receptor to influence MAPK mediated cell migration [242]
Secreted frizzled-related protein 3	sFRP-3	Q92765	Injury	Regulates cell growth and differentiation by direct Wnt interaction [243]
Fibronectin Fragment 4	FN1.4	P02751	Late	Fibronectin fragment that binds various targets including collagens [233]
Lymphatic vessel endothelial hyaluronic acid receptor 1	LYVE1	Q9Y5Y7	Injury	A lymph specific, membrane glycoprotein HA receptor for transport or uptake [244]
Tyrosine-protein kinase Fyn	FYN	P06241	Injury	Activates β-catenin signaling in chondrocytes accelerating OA development [245]
Protein deglycase DJ-1	PARK7	Q99497	Injury	Protects against oxidative stress and cell death via NF-κB signaling [246]

Plots showing the actual variance of selected groups of proteins between the study groups are shown below. Metalloproteinase inhibitors (Figure 3-10), matrix metalloproteinases (Figure 3-11), ADAMs and ADAMTS (Figure 3-12) and selected markers that loaded to early and injury groups (Figure 3-13) are shown demonstrating the relative marker distributions. Significance was tested with ANOVA and p-values are shown for those knee group markers where significance was identified ($p<0.05$).

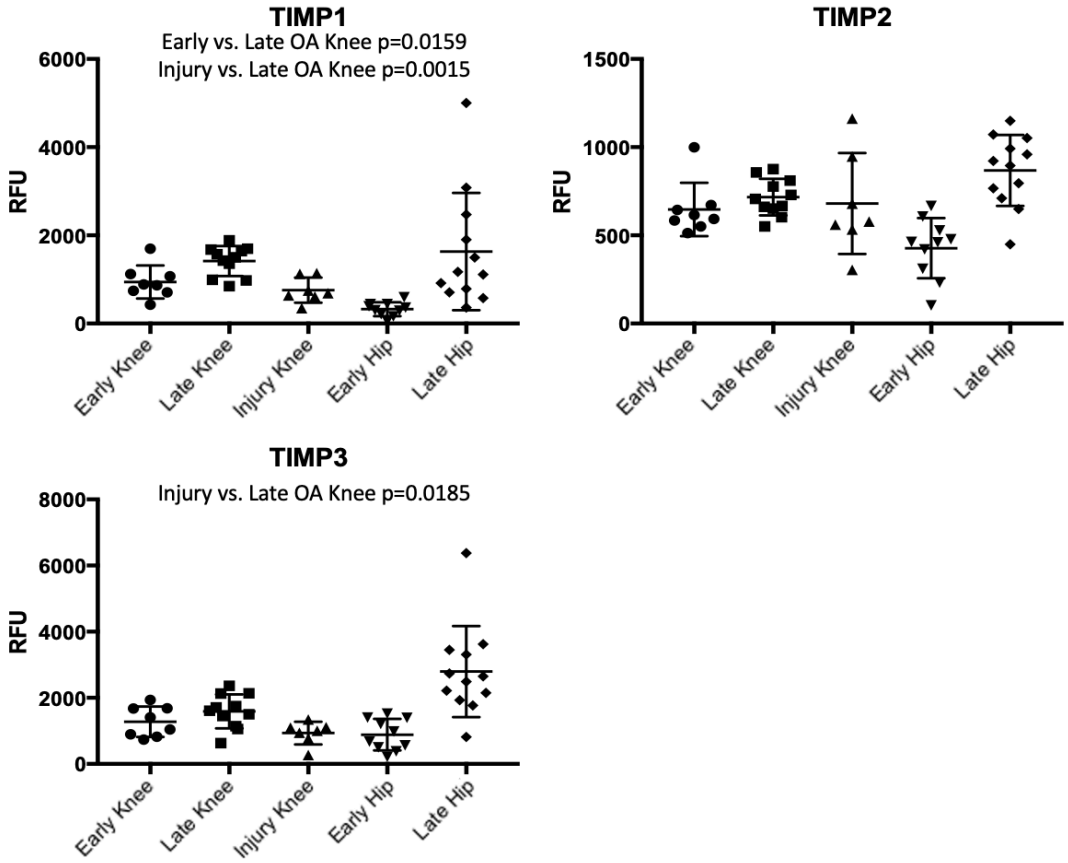


Figure 3-10 - Graphs showing variance between groups for metalloproteinase inhibitors: TIMP 1, 2 & 3. Significance tested using ANOVA. P-values shown for those markers that met significance.

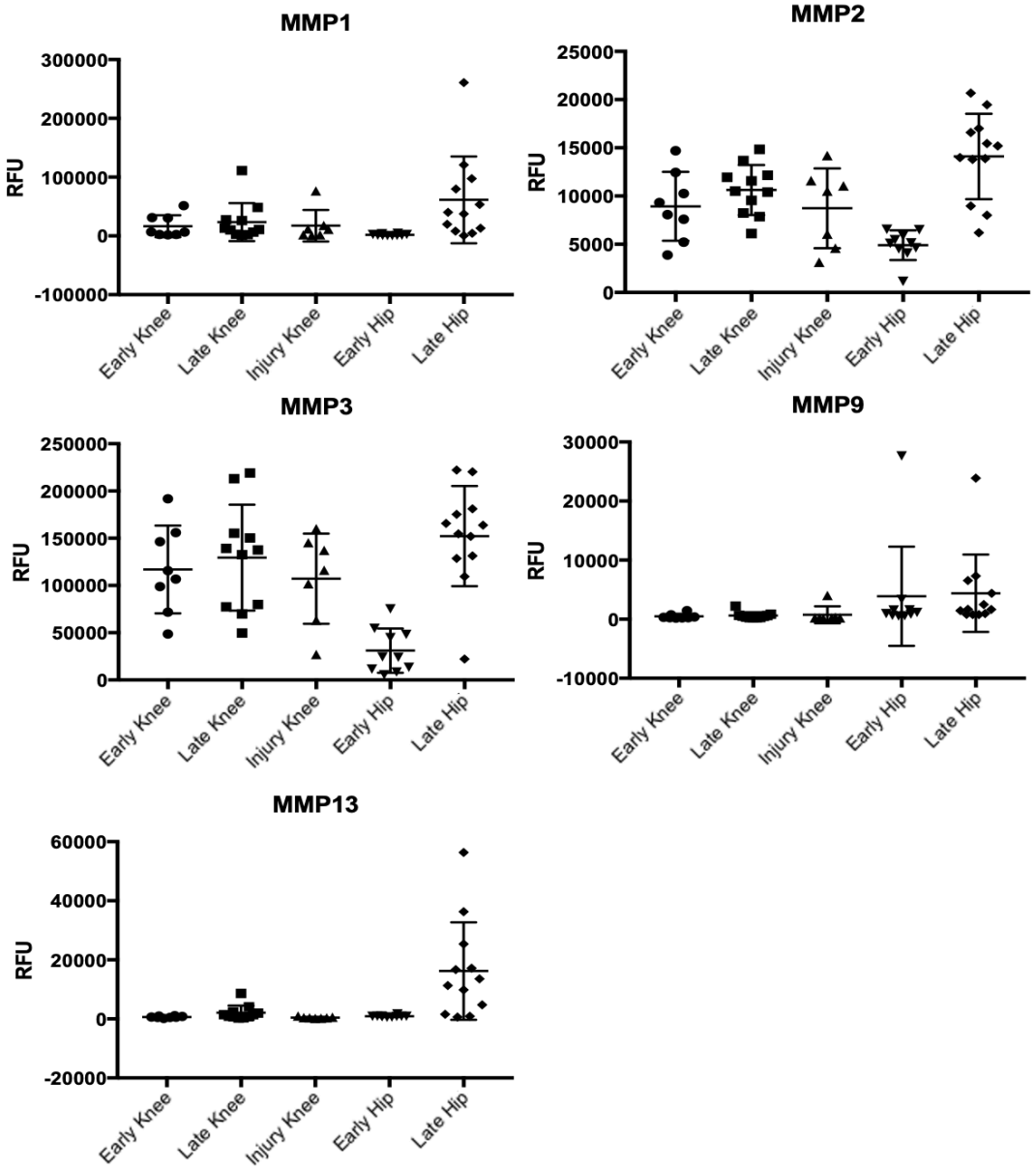


Figure 3-11 - Graphs showing the variance between groups for selected matrix metalloproteinases: MMP-1, 2, 3, 9 & 13. Significance tested using ANOVA. No knee group differences met significance.

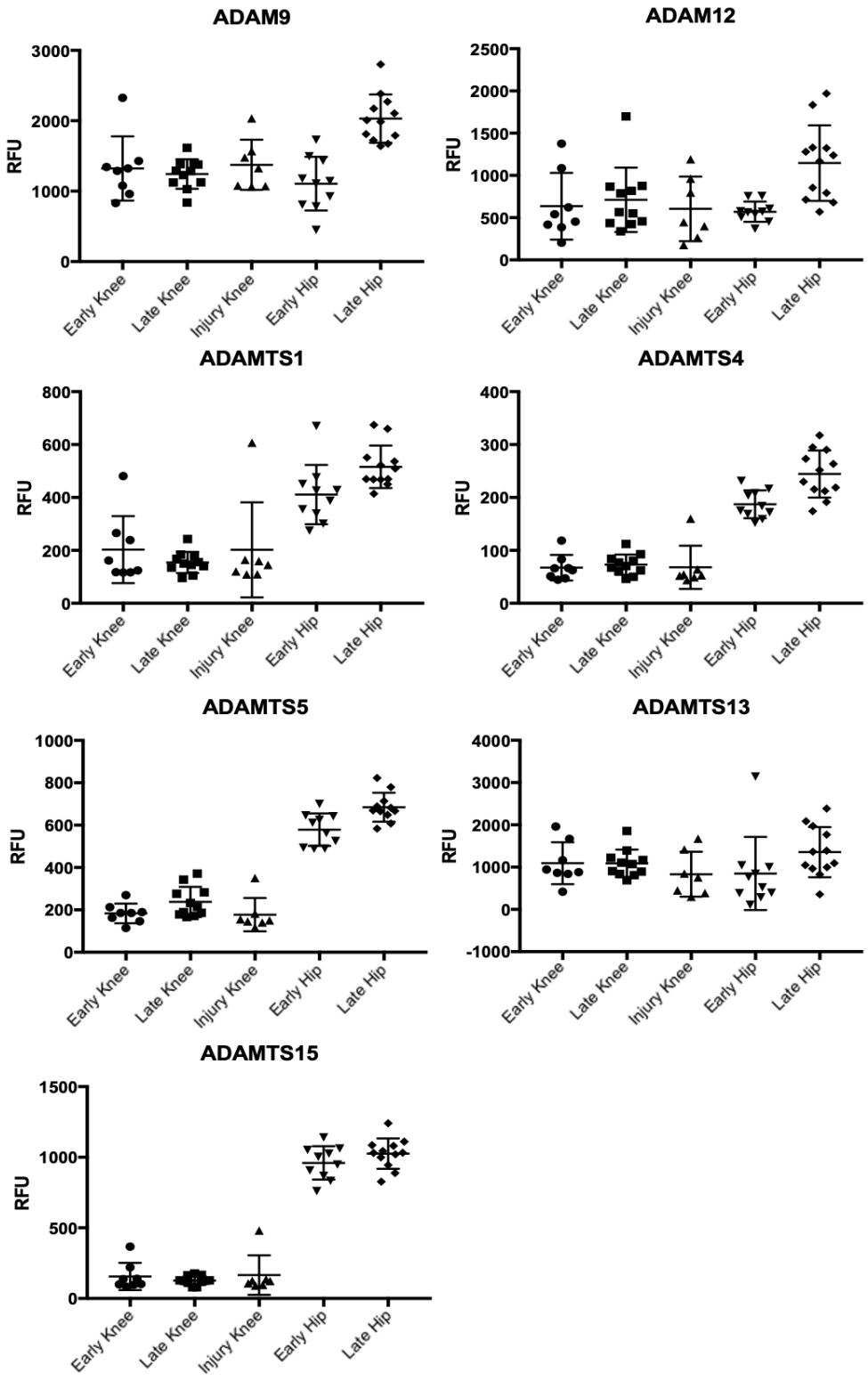


Figure 3-12 - Graphs showing the variance between groups for selected ADAM and ADAMTS proteins: ADAM 9 & 12 and ADAMTS-1,4,5 13 & 15. Significance tested using ANOVA. No knee group differences met significance

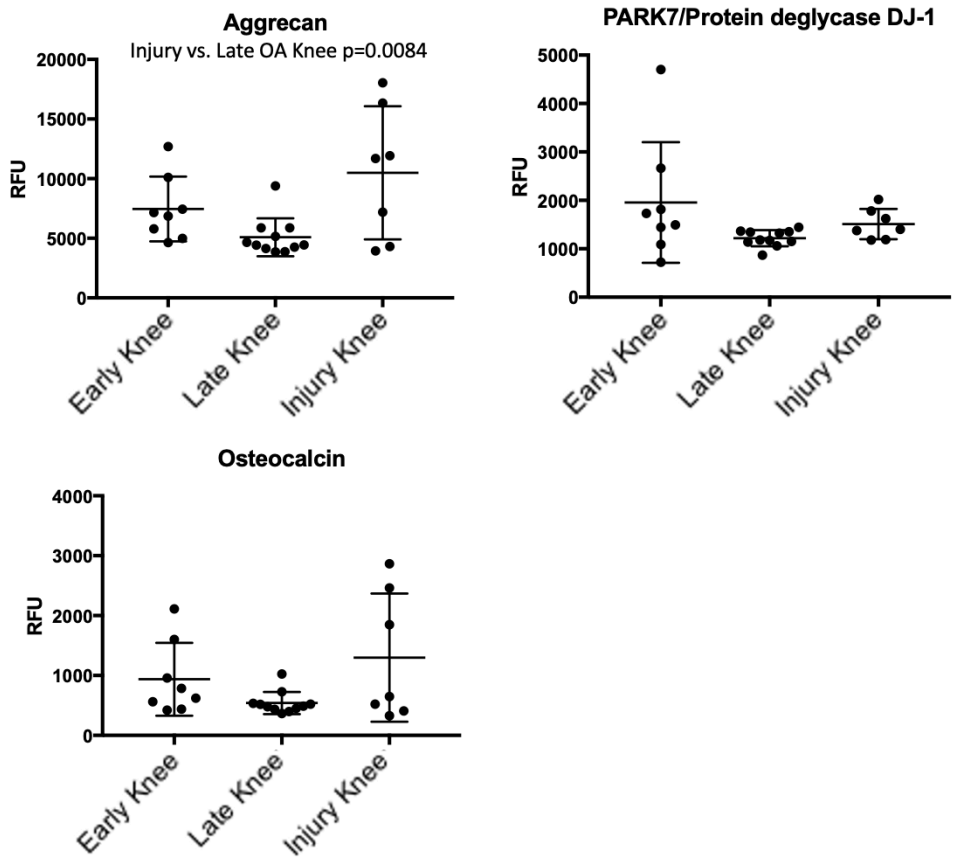


Figure 3-13 - Graphs showing the variance between groups for selected proteins that were elevated in the early OA and injury groups: Aggrecan, PARK7 and Osteocalcin. Significance tested using ANOVA. P-values shown for those markers that met significance.

3.5 LASSO

Refinement of the PLS-DA models of the analysis of the knee groups has shown that the late and injury groups are biologically different, although the early OA group fails to fully separate from either injury or late OA. As previously discussed, it is possible that some markers that appear to contribute significantly to the model may be linked by virtue of being biologically associated and therefore, in essence, may not actually contribute different information to the model. As such, in order to improve the model performance in the case of multi-collinear data, LASSO

analysis of the whole data set for late and injury groups and for late and early groups was performed using R, version 3.3.3 (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/). Data were scaled, log transformed and centred prior to LASSO analysis which was performed using the "glmnet" package (https://cloud.r-project.org/package=glmnet). Internal model performance was checked by serial exclusion and re-introduction of each observation to determine if the output markers were able to appropriately predict the group with no misclassification events. Sample code for LASSO analysis is shown in Appendix C.1.

LASSO identified no markers describing differences between late and early OA samples. This is supported by the PLS-DA analysis which showed only 7 of the top 50 markers loading towards injury with aggrecan (ranked 16 of 50) the highest ranked injury marker. It is expected that LASSO would select only highly ranked markers due to its highly penalistic nature.

The markers identified from LASSO analysis as most specifically describing the differences between observations in injury and late groups were:

- Tumour necrosis factor stimulated gene 6 (TSG-6)
- Tissue inhibitor of metalloproteinase 1 (TIMP-1)
- R spondin 4 (RSPO-4)

These markers were highly ranked in the whole SOMAscan 3 group analysis VIP ranking with TSG-6, TIMP-1 and RSPO-4 1, 2 and 12 respectively indicating good agreement in preferred markers. Plots showing the actual variance in the relative marker distributions between the groups are shown in Figure 3-14. Significance was tested with ANOVA and p-values are shown for markers where significance was identified ($p<0.05$).

Comparing all 3 groups simultaneously using LASSO was considered however it has been suggested that where group sample sizes differ, and where sample size and variable number are significantly different there may be a requirement to consider a priori model tuning. However where there is expectation of sub group differences it is challenging to know if such measures would be beneficial [247]. As such it was considered the results of this analysis would be challenging to interpret and therefore not performed.

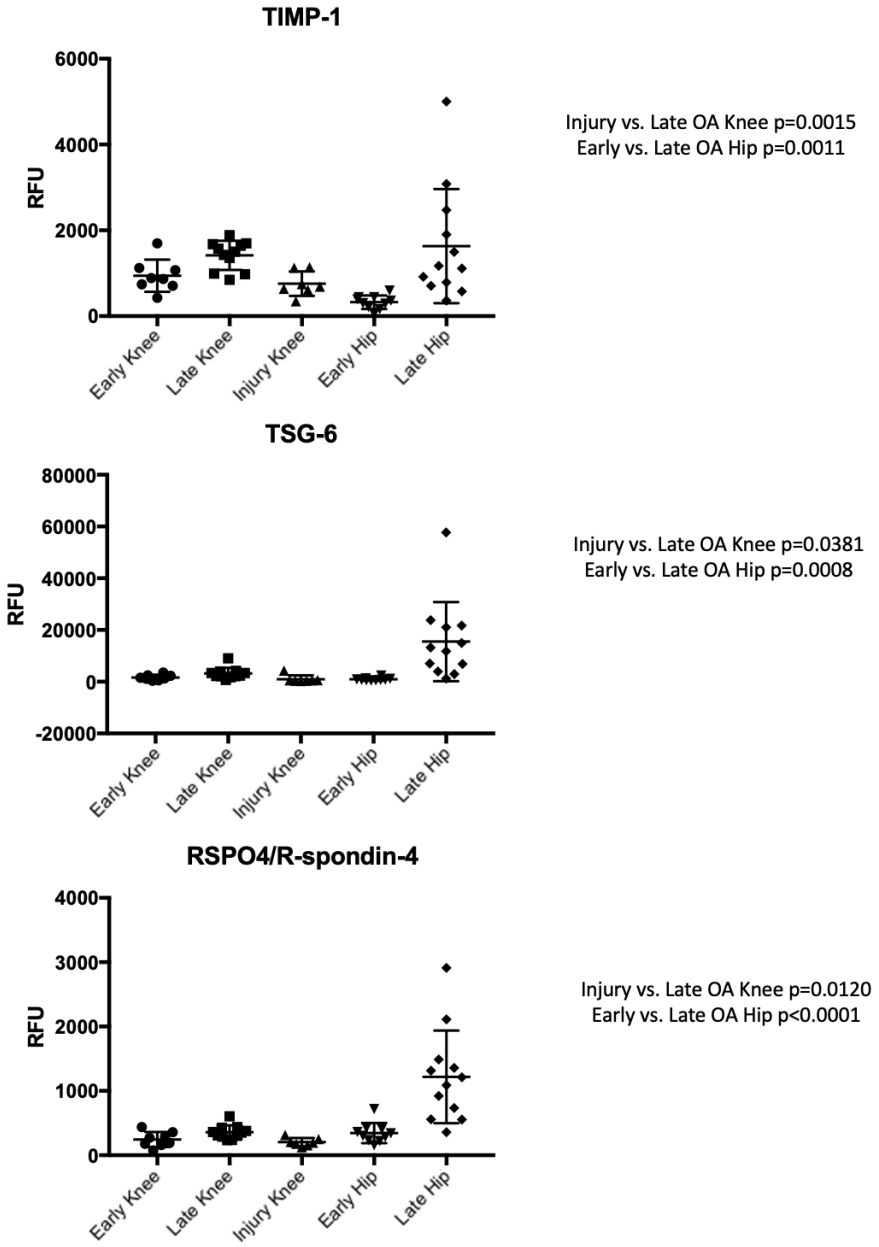


Figure 3-14 - Graphs showing the variance between groups for markers identified by LASSO analysis: TIMP-1, TSG-6 and RSPO-4 Significance tests for knee groups using ANOVA for knee and hip groups.

3.5.1 PLS-DA Model Using 3 LASSO markers

To further assess the selectivity of the markers identified by LASSO analysis a PLS-DA model was performed using only these markers. This demonstrated that the injury and late OA groups could be confidently discriminated on the basis of these three markers alone, as shown in the loading scatter plot below, with acceptable $R^2 = 0.651$ and $Q^2 = 0.592$ (Figure 3-15).

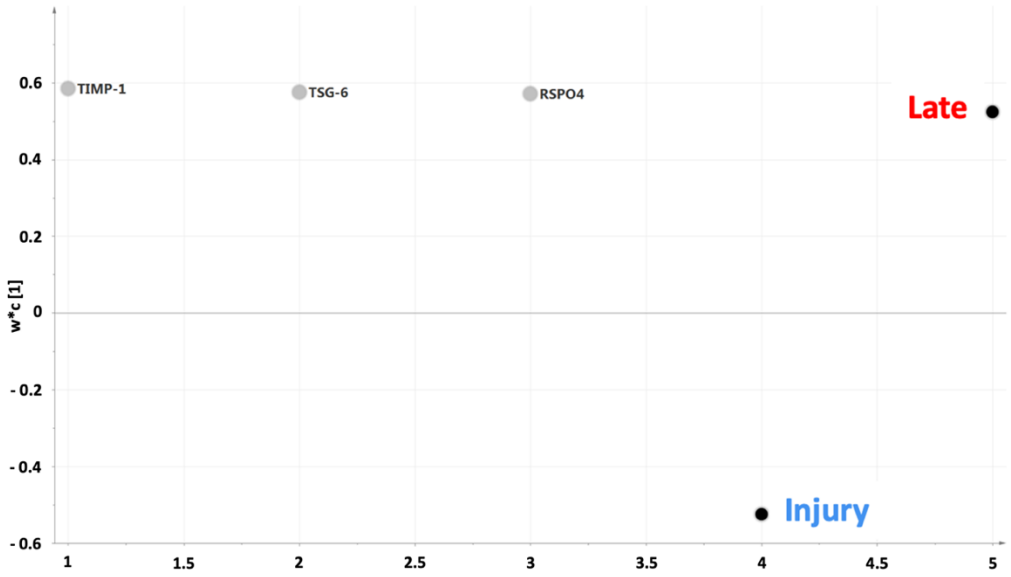


Figure 3-15 - Loading scatter plot for PLS-DA model of knee injury vs. late group comparison using a panel of the 3 LASSO markers alone showing that all three markers load to the late OA group. 1 component model ($R^2 = 0.651$ $Q^2 = 0.592$)

3.6 PCA

Principal component analysis was performed using the base R package with visualisations created using the "pca3d" package (https://CRAN.R-project.org/package=pca3d). Factor analysis was performed for extraction of the top loadings on PCA using the FactoMineR package [248] (https://cran.r-project.org/package=FactoMineR). The top 20 variables, which accounted for up to 70% of the cumulative variance of each principal component, were selected. A further five loading variables were selected which contributed to component 2 since this had the eigenvalue most highly correlated with the late OA samples. The variables contributing the highest loading from the factor analysis were selected. The normalised protein values were transformed to fold changes by subtraction of the values from each observation of a particular protein from the average of the protein values of either the early OA or injury samples in their respective comparisons against the late OA group. The subtracted values were then transformed to linear space using the antilog of base 2. The fold changes were clustered for both protein values and samples using canberra distance and complete linkage. The results were visualised using the gplots package (https://cran.r-project.org/package=gplots) and RColorBrewer package (https://cran.r-project.org/package=RColorBrewer).

3.6.1 Knee Early OA, Late OA and Injury Three Group Comparison Using PCA of the Whole SOMAscan

PCA comparing all three knee groups supports the earlier findings detailed above that a comparison of all three knee groups together cannot detect sufficient differences between the samples of the injury and early OA groups. The PCA analysis demonstrates that the early and injury groups cluster very closely with the two group centroids overlapping whilst there is an apparent difference between these two groups and the late OA group (Figure 3-16). The first two components can explain 56.53% of the variance in the data and the leading proteins in each component are BCL2L1 (B cell lymphoma 2 like protein 1) and HS6ST1 (Heparan Sulfate 6-O-Sulfotransferase 1) for components 1 and 2 respectively. Heirarchical clustering (Figure 3-17) shows the markers that separate the late OA samples, and the overlap between early OA and injury samples. TSG-6 and TIMP-1 (both high in PLS-DA & LASSO) were identified as late group markers and aggrecan was again a strong marker for early and injury groups.

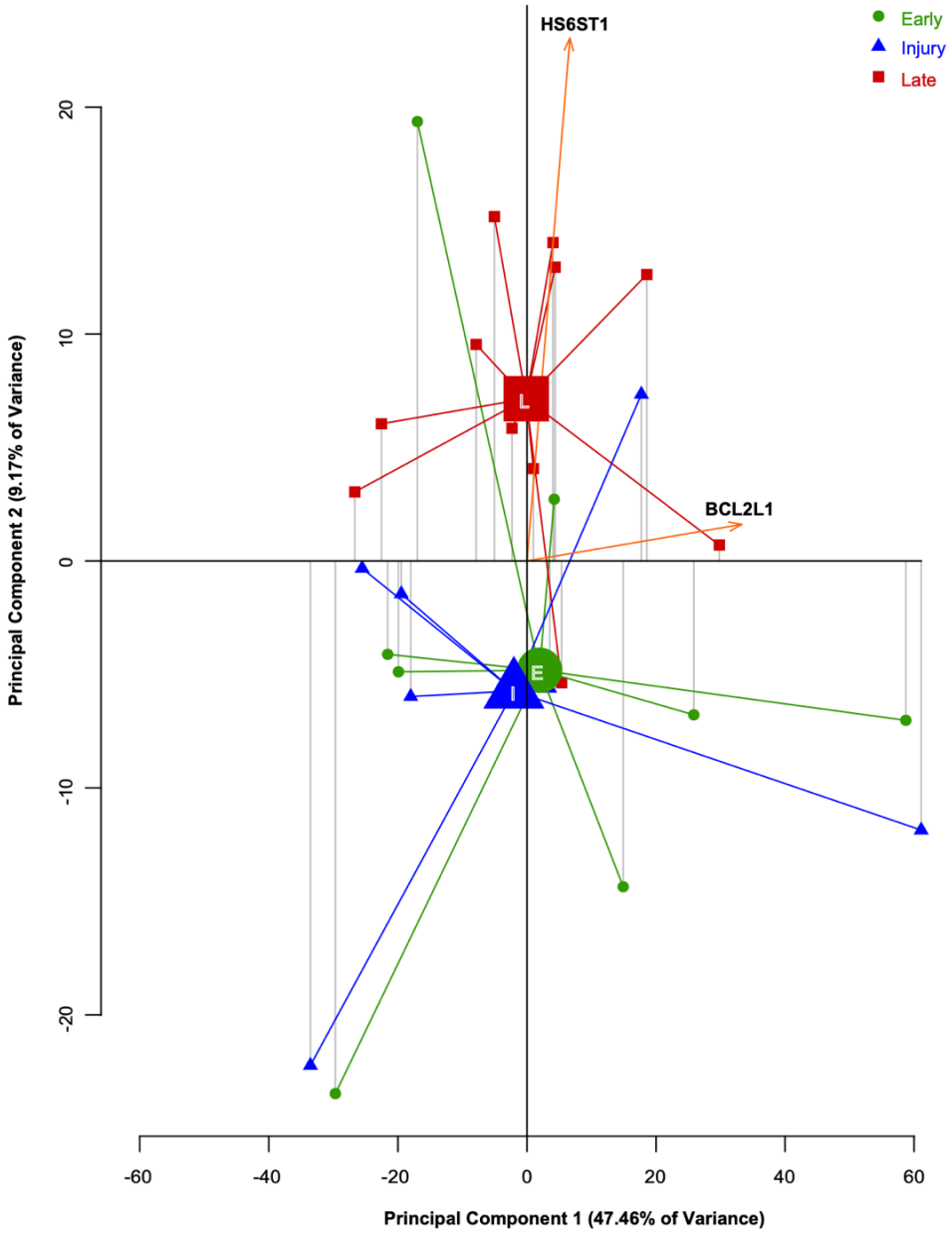


Figure 3-16 – Observation score plot of PCA of quantile normalised SF proteins from comparison of knee early OA, injury and late OA groups using the whole SOMAscan panel of 1310 proteins.Plot shows group centroids (large coloured shapes to correspond with groups). Orange arrows show the first two orthogonal variables (BCL2L1 first loading marker for PC 1 and HS6ST1 first loading marker for PC 2). The first two components explain 56.53% of the overall variance.

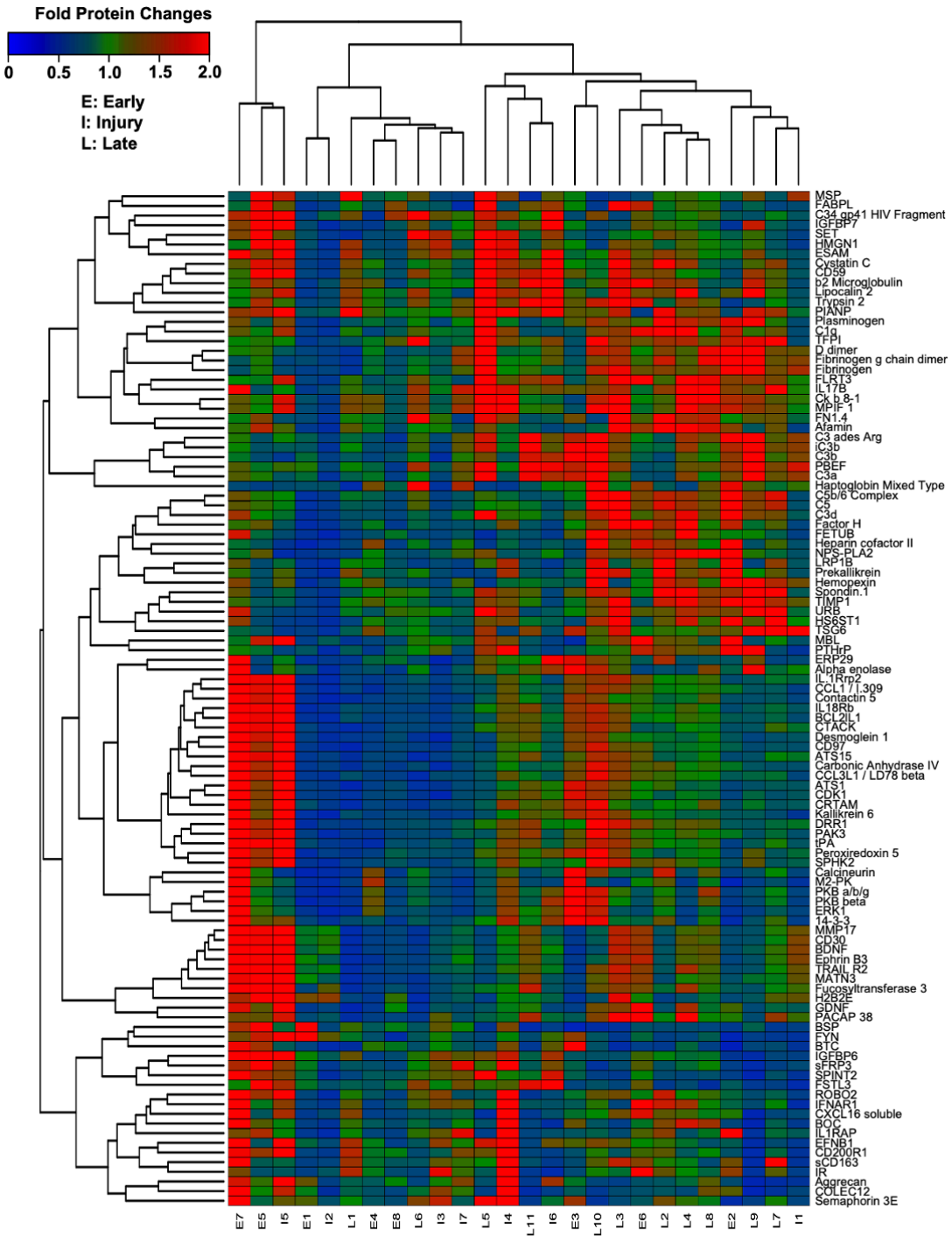


Figure 3-17 – Heatmap and hierarchical clustering of PCA of quantile normalised SF proteins from comparison of knee early OA, injury and late OA groups using the whole SOMAscan panel of 1310 proteins.Heatmap and hierarchical clustering based on first 5 principal components showing samples cluster into two main groups with the right group composed predominantly of late group samples. Courtesy of Dr. A. Chanalaris.

3.6.2 Knee Late OA Group Individual Comparison to Early OA and Injury Groups Using the Whole SOMAscan

Separate individual comparisons of the knee injury and early OA groups to the late OA group reflected the three group comparison preserving the difference between the injury and early OA groups as compared to the late OA group. In both cases there were evident differences however, for the comparison of early and late OA there was more sample crossover towards the other group than was seen in the injury to late OA comparison. Graphical representations of the analyses are shown below (Figure 3-18 and Figure 3-19).

For comparison of the early and late OA groups the first two components can explain 55.06% of the variance in the data and the strongest loading proteins in components 1 and 2 are FGF12 (fibroblast growth factor 12) and SLPI (secretory leukocyte proteinase inhibitor) respectively.

For comparison of the injury and late OA groups the first two components can explain 56.52% of the variance in the data and the strongest loading proteins in components 1 and 2 are FCN2 (ficolin-2) and DAF (decay accelerating factor) respectively.

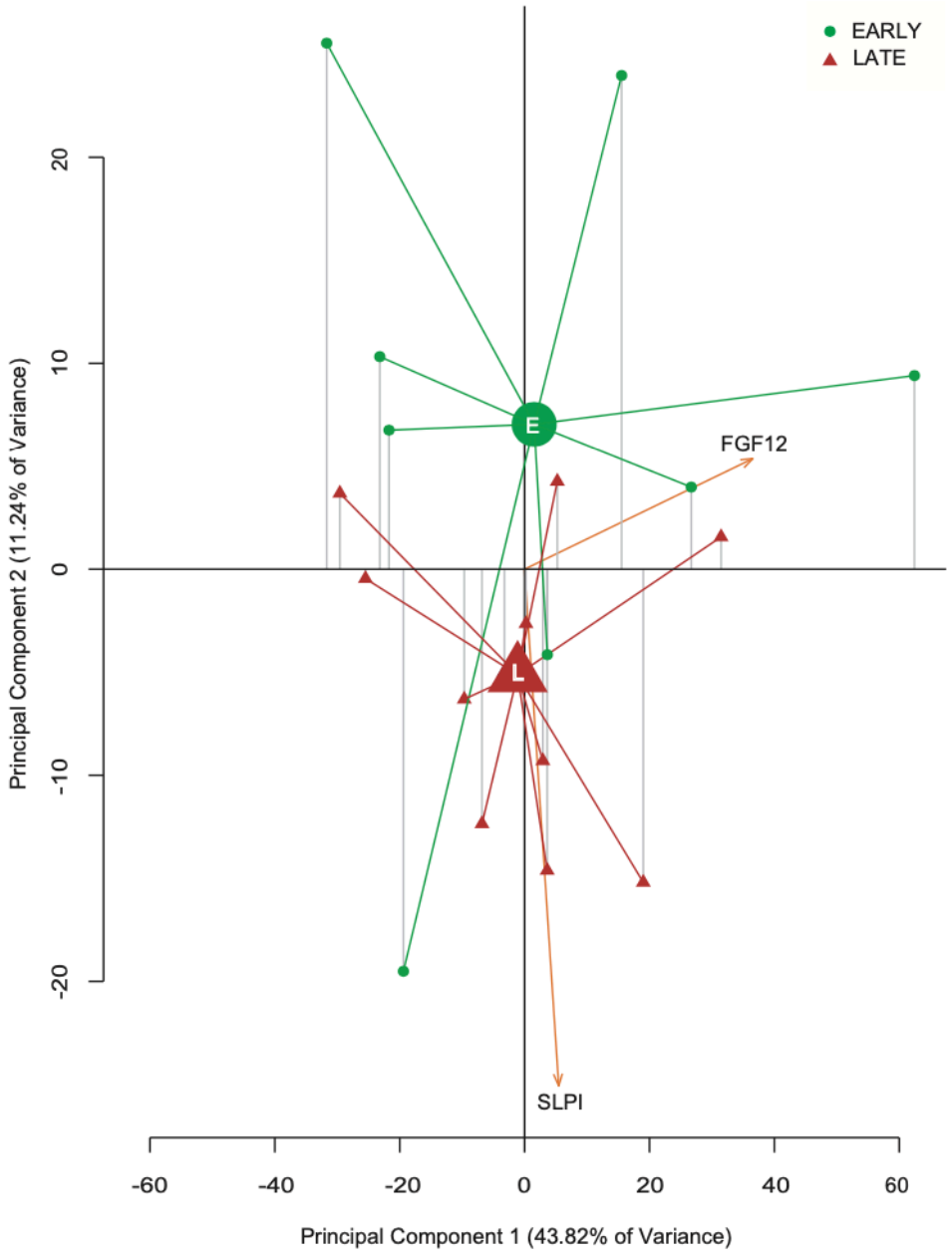


Figure 3-18 - Observation score plot of PCA of quantile normalised SF proteins from comparison of knee early OA and late OA groups using the whole SOMAscan panel of 1310 proteins.Plot shows group centroids (large coloured shapes to correspond with groups). Orange arrows show the first two orthogonal variables (FGF12 first loading marker for PC 1 and SLP1 first loading marker for PC 2). The first two components explain 55.06% of the overall variance. Courtesy of Dr. A. Chanalaris.

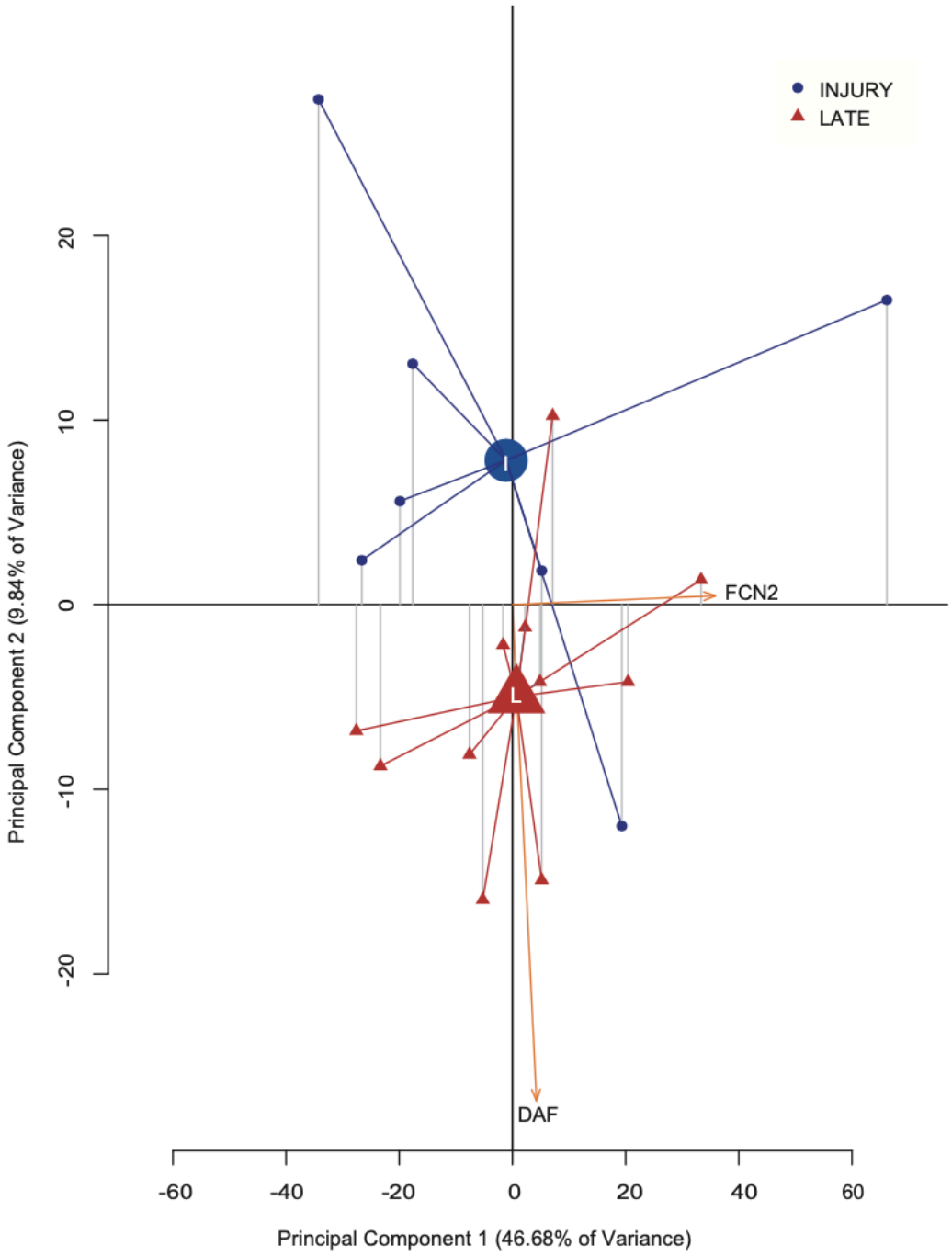


Figure 3-19 - Observation score plot of PCA of quantile normalised SF proteins from comparison of knee injury and late OA groups using the whole SOMAscan panel of 1310 proteins.Plot shows group centroids (large coloured shapes to correspond with groups). Orange arrows show the first two orthogonal variables (FCN2 first loading marker for PC 1 and DAF first loading marker for PC 2). The first two components explain 56.52% of the overall variance. Courtesy of Dr. A. Chanalaris.

3.7 Network Analysis

Network analysis of the proteins identified by the top five components from the comparison of early OA, late OA and injury group PCA analysis (Figure 3-20), and the top 100 proteins identified from the PLS-DA analysis (Figure 3-21), performed using the STRING v11 database [249], demonstrates the human protein – protein interactions based only on text mining and experimental evidence. Support and graphical representations provided by Dr. A. Chanalaris. Only interactions with a probability $p > 0.4$ are shown. Unconnected nodes have been removed. The node colours represent selective gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) mappings.

Vesicles were the most highly enriched network seen for the markers in the PCA analysis and this was also one of the most highly enriched networks in the PLS-DA analysis with 39 and 42 markers respectively. This is in keeping with literature evidence demonstrating the presence of transmembrane and intracellular exosomes and vesicles in OA SF [250–252]. The most highly enriched pathway in the network analysis of the markers from the PLS-DA analysis was response to stress with 59 markers which may indicate a variety of mechanisms that have been shown to be implicated in OA with some agents emerging as potential therapeutic targets [253–255]. Immune response and complement – coagulation cascade pathways were the next most common in both analyses with numerous chemokines, cytokines and complement factors and activators implicated which again is strongly supported by literature exploring the emergence of the role of

innate and adaptive immunity pathways in OA pathogenesis [229,256]. Further pathways identified including apoptosis, osteoclast differentiation and phosphoinositide 3-kinase – protein kinase B (P13K-Akt) signalling pathways in the PCA network analysis and ECM organisation and response to wounding in the PLS-DA marker networks.

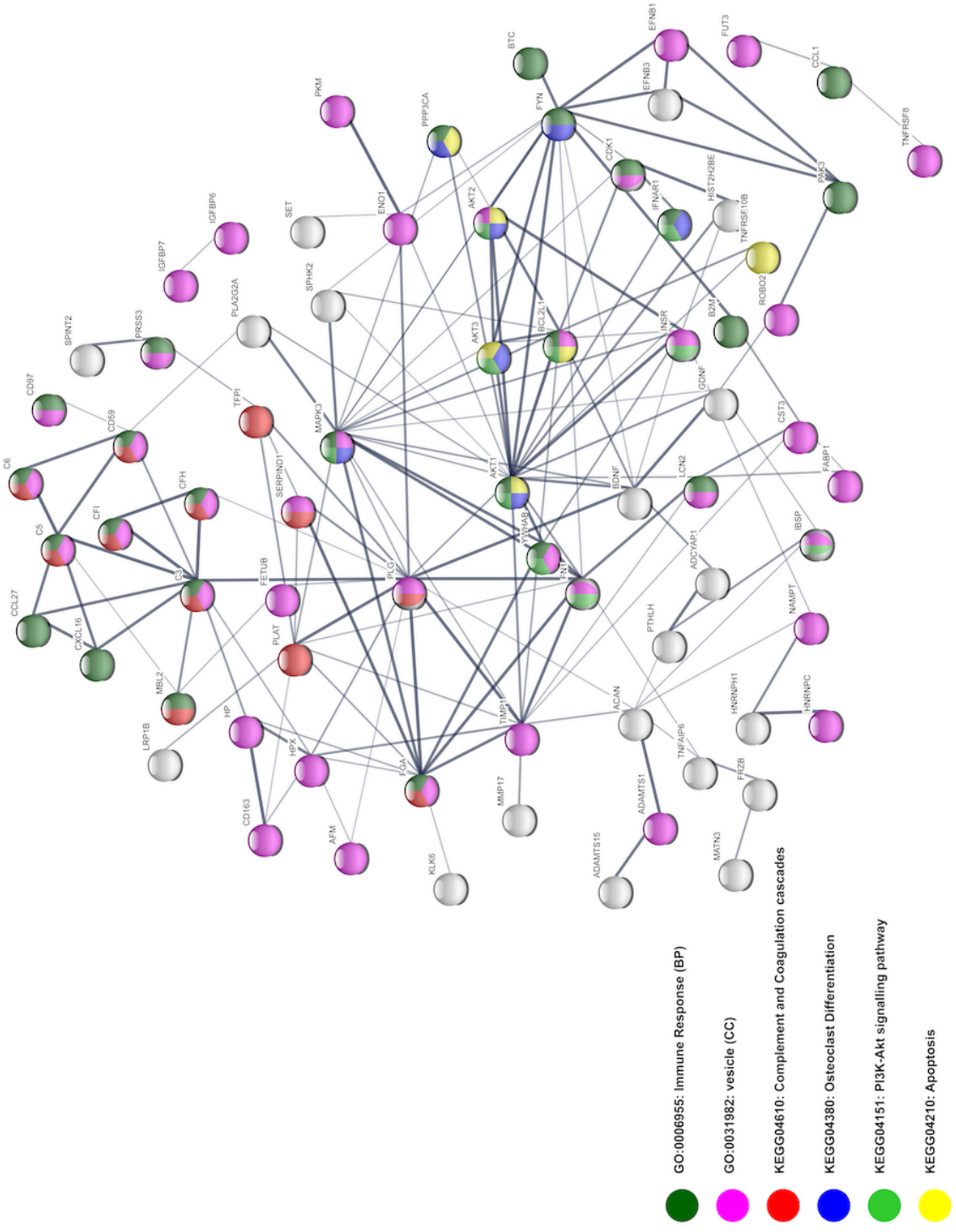


Figure 3-20 - Network of protein - protein interactions based on the PCA top loading markers from the first 5 principal components of PCA of quantile normalised SF proteins from comparison of knee injury and late OA groups using the whole SOMAscan panel of 1310 proteins. Node colour meanings are indicated in the legend. The thickness of the node connection is proportional to the confidence of the interaction. Courtesy of Dr. A. Chanalaris.

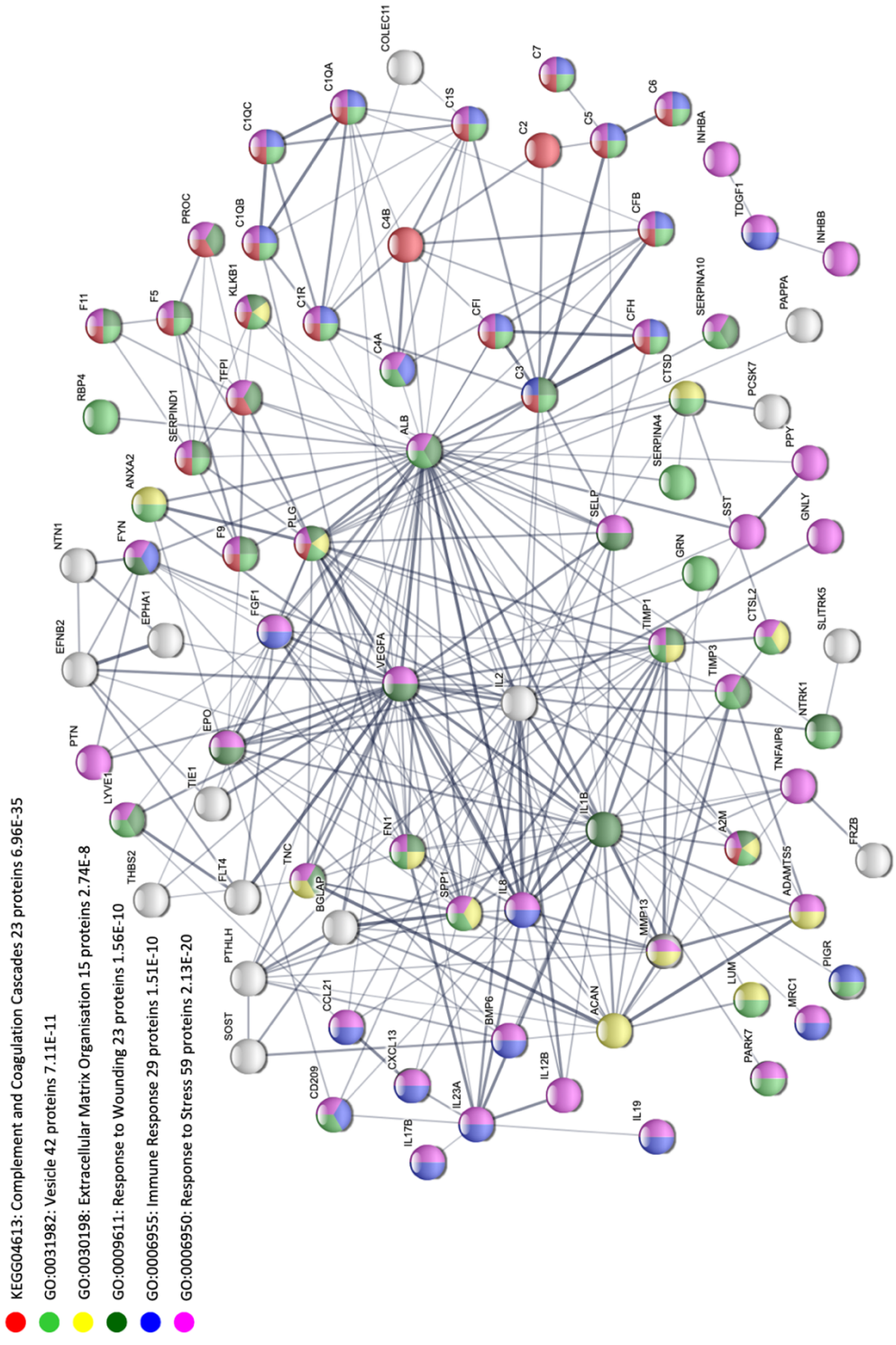


Figure 3-21 - Network of protein - protein interactions based on the top 100 markers of the PLS-DA comparison of knee injury and late OA groups using the whole SOMAscan panel of 1310 proteins Node colour meanings are indicated in the legend. The thickness of the node connection is proportional to the confidence of the interaction. Courtesy of Dr. A. Chanalaris.

3.8 Discussion

This study aimed to identify biomarker fingerprints of different clinical and structural disease states of knee OA groups using multivariate analyses of SOMAscan data. Specifically, we sought to identify a separate signature for the early knee OA group. The results, in this analysis, have shown that the early OA group appear to be biologically similar to the injury group and were unable to be distinguished, as a distinct group from either the injury or late group, whilst the injury and late OA groups did show separation from each other. This finding is very interesting since the injuries selected in this group of ACL rupture and meniscal tear, have been shown to clinically predict progression to OA irrespective of surgical intervention [75,257] suggesting the possibility that the biological phenotype of early OA may begin to develop only shortly after the index injury. This is in contrast to evidence showing that radiographic appearances of OA zmay take longer to appear with 50% of patients undergoing meniscectomy developing radiographic OA between 15 and 22 years later [258]. When the early and late OA groups were compared, they did not produce a PLS-DA model that could discriminate between the groups and LASSO analysis identified no markers that sufficiently described any group differences. The PCA hierarchical clustering plot showed that the injury, and late OA groups, separated best. The early OA group patients clustered with both the injury, and late OA, groups although, on balance, more patients clustered towards the injury group. These analyses have highlighted

some candidate markers of interest in the description of these clinical groups and their biology must be considered.

TSG-6 is a glycoprotein that is coded by the TNFα-induced protein 6 gene induced by TNF and IL-1 [259]. It is significantly up regulated in OA progression and acute knee injury [260,261] and has been shown to be present in the SF of patients with OA and associated with disease progression [262], whilst not being isolated from healthy joints [263]. It binds hyaluronan, aggrecan, heparan sulfate, chondroitin sulfate and multiple chemokines [264–267]. TSG-6 disrupts matrix production in OA cartilage [213] and its activity in SF has been shown to be associated with both incidence and progression of OA [268,269].

TIMP-1 is a soluble, secreted metalloproteinase inhibitor which acts by competitively binding the active site of metalloproteinases and has an established role in metalloproteinase independent cell signaling [212]. It has a broad activity against numerous non-membrane bound metalloproteinases [270]. TIMP-1 has been isolated in the SF of human patients and has been shown to be higher in patients diagnosed with OA than in controls [271] as well as being higher in those with osteoarthritic effusions compared to those without [272].

The R spondins are part of a superfamily of proteins that were first described in 1971 when they were isolated from platelets after thrombin stimulation and named as they contained thrombospondin type 1 (TSR-1) repeat proteins [273]. The R spondins are a sub family of four proteins that were discovered much more

recently with R Spondin 4 (RSPO-4) first isolated in 2006 [211]. All four R spondins strongly potentiate Wnt signaling by a mechanism suggested to be related to binding of leucine-rich repeat containing G protein coupled receptors (LGR). This clears ubiquitin ligases that normally ubiquitinate Wnt receptors for degradation [211,274]. Whilst RSPO-4 has not yet been implicated in OA pathogenesis others of the family have been shown to be mechanistically involved. RSPO-1 has been shown to play a role in prevention of osteoclast development while promoting osteoblast differentiation, and hence bone formation [275]. RSPO-1 can also be identified in the serum of rheumatoid arthritis patients and correlated with disease activity and structural progression [276]. Decreased RSPO-2 in OA osteoblasts are associated with reduced Wnt/β-catenin signaling and abnormal mineralisation [277]. RSPO-3 has been shown to bind syndecan 4; a transmembrane protein that carries three to five heparan sulfate and chondroitin sulfate chains which interact with fibroblast growth factors, TGF-β, and other ECM molecules [278,279]. Although RSPO-4 is a relatively newly identified protein and has not been directly linked with OA in the literature to date, the mechanistic links within its wider family show a potential link via the LGR4 receptor. RSPO-4 is totally dependent upon LGR4 to potentiate Wnt/β-catenin signaling [279] and the expression of LGR4 has been shown to be reduced in a rat model of traumatic OA [280].

BCL2L1 belongs to the BCL2 family of proteins and has been variously linked with apoptosis in a number of disease areas including OA [281–283]. BCL2L1 protein is exosomal and acts by blockage of caspases by blockade of the voltage dependent

anion channel found in the mitochondrial outer membrane [284]. It has been linked to exosome mediated protection against apoptosis in malignancy [285]. Its expression has been shown to promote two principal isoforms: BCLXL which is shown to be anti-apoptotic in OA [286,287], and BCLXS which is shown to be pro-apoptotic [288].

HS6ST1 is a sulfotransferase, responsible for the O-sulfation of N-acetyl glucosamine at the 6 position, that modulates the structure of heparan sulfate, a component of ECM proteoglycans. Its full role is unclear although it has been implicated in the angiogenesis of ovarian cancer via a heparan sulfate dependent pathway of cytokine modulation [289,290]. It has been shown that RNA and protein levels of HS6ST1 are elevated in late OA compared to non-OA human knee cartilage and that there is an increased amount of 6-O-S sulfated HS proteoglycans in late OA knee cartilage [291].

The FGF family of proteins has a wide range of activities on numerous different cell types [292]. FGF12 is a heparin binding growth factor with receptor independent uptake by cells [293]. Interestingly, the extracellular effects of FGF12 remain unclear since it is unable to activate any FGF receptors. However it has been shown to be expressed in proliferating and hypertrophic chondrocytes during embryonic endochondral ossification [294] and has been identified as a component of a biomarker for prediction of response to infliximab treatment of RA [295].

SLPI is a protease inhibitor that is synthesised by cells local to the inflammatory process in response to the inflammatory cytokines [296]. It is known to inhibit numerous cartilage proteases and is upregulated by superficial chondrocytes in OA [297] where it is thought to suppress degradation of type II collagen [298]. It has been shown to be elevated compared to healthy levels in pre-surgery patients with OA awaiting THR [299].

The first ficolins were identified as TGF-β1-binding proteins with subsequent identification of a human form in the liver, hence FCN2 is also termed L-ficolin [300,301]. More recently they have shown to have a central role in the lectin pathway, one of the complement activation pathways, acting as a pattern recognition molecule [302,303].

DAF is alternatively identified as CD55 and serves to regulate complement activation by accelerating decay of C3 and thereby suggested to have a role in ameliorating autoimmune diseases such as glomerulonephritides [304]. It has been shown to have a role in protection against immune-complex mediated arthritis [305] although it should be noted that the same group had previously reported that deletion of CD55 might have a slight protective effect in the pathogenesis of RA [306] which, in their later study, they postulate may have been due to failing to set a suitable threshold for disease detection.

SF represents the most biologically active fluid for exploratory biomarker study. The fact that it is only available in low volumes makes wide scale multiplex assay

techniques an advantageous tool in the analysis. This is the first experiment to compare the proteomic differences in knee SF between different disease states in OA using the SOMAscan. There is currently only one report in the literature of synovial fluid analysis using the SOMAscan, a longitudinal investigation of six patients following ACL rupture with samples taken at 6 and 14 days post injury [176]. This study was published after the work described in this chapter was undertaken. Before this, no study had investigated the use of the SOMAlogic platform for any SF analysis and their study presents no evidence of optimisation of the matrix for bioassay since the SOMAlogic platform has previously been primarily used for plasma and serum studies [156,162,307,308]. Due to the very early sample harvest after injury in their study samples were highly likely to be heavily blood stained since rapid, large haemarthrosis is a classic clinical feature of ACL rupture at presentation although this is not discussed in their report. The patients were recruited from the placebo arm of a randomised trial and all received intra-articular infiltration of saline after first sample aspiration and it is unclear how this influenced the marker concentrations seen in the samples taken at the second time point. Their analysis of the earlier 1129 SOMAscan identified a candidate panel of 30 markers which included some thematic similarities to those identified in this study with ADAM12, fibronectin, IL-6 subunit alpha and IL-17 receptor A indicating matrix degradation and inflammatory pathways. Their network analysis showed enrichment of cytokine-cytokine receptor and osteoclast

differentiation pathways which again, share adaptive inflammatory and OA progression themes similar the findings discussed in this chapter.

In the studies presented in this thesis hyaluronidase treated SF samples were prepared using a previously published optimisation protocol for direct comparative ELISA and MSD assays [206]. Sample and matrix validation (demonstrated in chapter 2) showed favourable comparison overall demonstrating reliability of the protocol and the medium for analysis of SOMAlogic aptamers.

In this exploratory study the large number of target proteins available for analysis utilising the SOMAscan of 1310 markers provides a strong basis for further biomarker exploration and discovery. The relatively low patient numbers used in this study, due to the cost of the SOMAscan, means that a level of caution is required in the interpretation of the data output. However, the statistical analysis tools of PCA, PLS-DA and LASSO are created to robustly analyse datasets with broad variable to observation ratio and highly collinear data whilst limiting overfitting to provide biological insights with positive implications for future studies.

The platform was untested for this matrix and, as such, comparison of biologically and clinically different groups was critical to understanding the underlying processes. The nature and chronicity of the key biological factors in the progression from early to end-stage OA remain unclear. The groups chosen in this study allowed investigation of the biology of early and end stage disease states as well

as the relative constitution of injury and its appearance compared with those of early and late disease.

The findings in this chapter are in keeping with the literature, with respect to the biological factors identified as discriminating between patient groups of known disease states. The most descriptive markers identified described the changes seen in the late OA group with enrichment of vesicle and response to wounding pathways, as well as complement, and coagulation, cascade pathways. Aggrecan was seen to be the best descriptor of the injury group in this study, although it was a weak loader that did not meet LASSO analysis threshold but is implicated in ECM organisation and response to stress pathways.

The findings prove the utility of SF as a matrix for study on the SOMAscan platform. Broad, multiplex proteomics represent valuable tools in the biomarker discovery process facilitating detailed pathway analysis, rather than relying on the traditional single marker study axiom. The elucidated markers are all biologically logical OA candidates, with literature evidence, supporting the value of this approach.

This study is limited by the relatively low number of samples under investigation. However, this limitation is itself a direct result of the high assay costs associated with the SOMAscan. It should also be remembered that prior to this work the SOMAscan had not been formally validated for examination of SF making this a vital and necessary technical pilot study and proof of concept exercise. It would be

unfeasible to commit the scale of financial resources for a project sufficiently larger than this to generate results of discernibly greater power on an untested system. There are also inherent group differences, most notably the mean age difference between early / injury and late groups. This is unavoidable due to the nature of the disease and its timing. Even if age matched groups were possible it is likely that older injury cohorts and younger degeneration cohorts may also represent different biological disease subtypes and, as a result, true matched groups in the OA space are, in all likelihood, unrealistic aspirations. It is also recognised that the injury cohort examined in this chapter were sampled at varying time points after initial injury. This is a result of the clinical healthcare construct and the variation in timing of patient presentation. In an ideal world all patients would present immediately to the emergency department and undergo standardised subsequent care. However, individual differences in personal, social and clinical factors allied to external healthcare system pressure make this impossible. This issue is less relevant in the late group since the nature of reaching a common disease end point has the effect of unifying patients in their symptomatology.

This study has suggested that biological differences exist between these clinically defined injury and late groups. Interestingly, these differences became clear on exclusion of the early group with the analysis demonstrating that, in this study, the early and injury groups lay very close to each other in the model space. This further suggests that there may be biological similarities between the injury and early OA groups in the dataset under investigation. This finding warrants further

investigation to determine whether, with larger cohorts, the common underlying pathologies that may indicate a pathway junction at which potentially treatable injury phenotypes transition into early OA phenotypes requiring different therapeutic approaches.

Chapter 4 The Biological Patterns of Synovial Fluid in Hip OA

4.1 Introduction

Whilst SF from the knee is relatively easily accessible at the bedside in good volumes, unfortunately the hip is less forthcoming. The hip is a deep joint and cannot be reproducibly aspirated "blind" by the bed or in the clinic room. The hip joint also does not generate SF in the same volumes as the knee joint. As such, to analyse the hip SF samples the wide spectrum proteomic analysis of low volumes of scarce material is hugely valuable and critical to gaining an insight into the biological description of the different stages of hip OA.

4.2 Aim

The aim of this study was to identify biomarker fingerprints of different clinical and structural disease states of hip OA groups using multivariate analyses of SOMAscan data.

4.3 Groups

Hip SF samples were prospectively harvested from 22 patients at the time of THR or hip arthroscopy. Two groups were identified (described further in chapter 2 and Table 4-1) as follows:

1. Early OA (10 patients)

2. Late OA (12 patients)

Table 4-1 - Hip cohort individual group participant demographics

Group	Early OA	Late OA
Participants	**10**	**12**
Mean Age (yrs)	44.3	57.1
Female	7	6
Male	3	6
K&L Grade	<2	3/4

4.4 PLS-DA

4.4.1 Hip Early OA and Late OA Group Comparison Using the Whole SOMAscan

Initially a comparison of the two groups using the whole panel of 1310 SOMAscan markers was performed to explore the differences between the two groups. This first look at the hip SF showed good separation of the groups as evidenced in the score scatter plot (Figure 4-1) with good grouping of the late group observations and good model characteristics. The loading scatter plot demonstrates that there are candidate markers describing the two groups but the vast majority of markers in the panel are located far from the two groups in the model space indicating limited contribution to the discrimination seen (Figure 4-2).

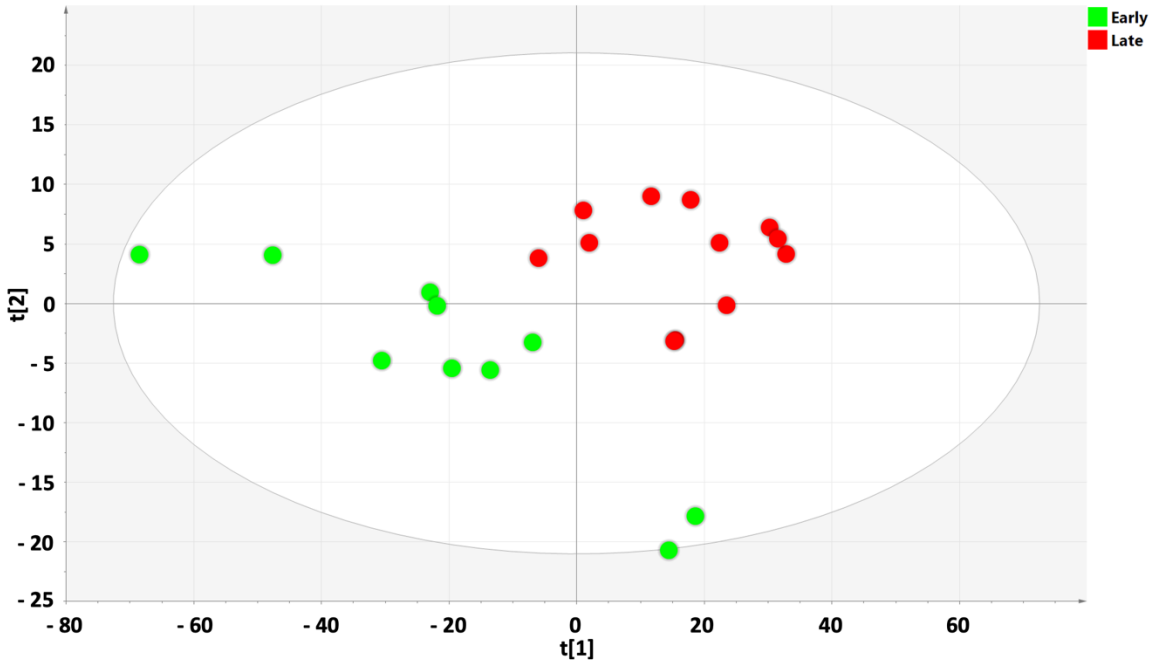


Figure 4-1 - Score scatter plot for hip early vs. late OA group comparison using whole SOMA panel of 1310 markers showing clear separation of the groups. 3 component model (R^2 = 0.940 Q^2 = 0.669)

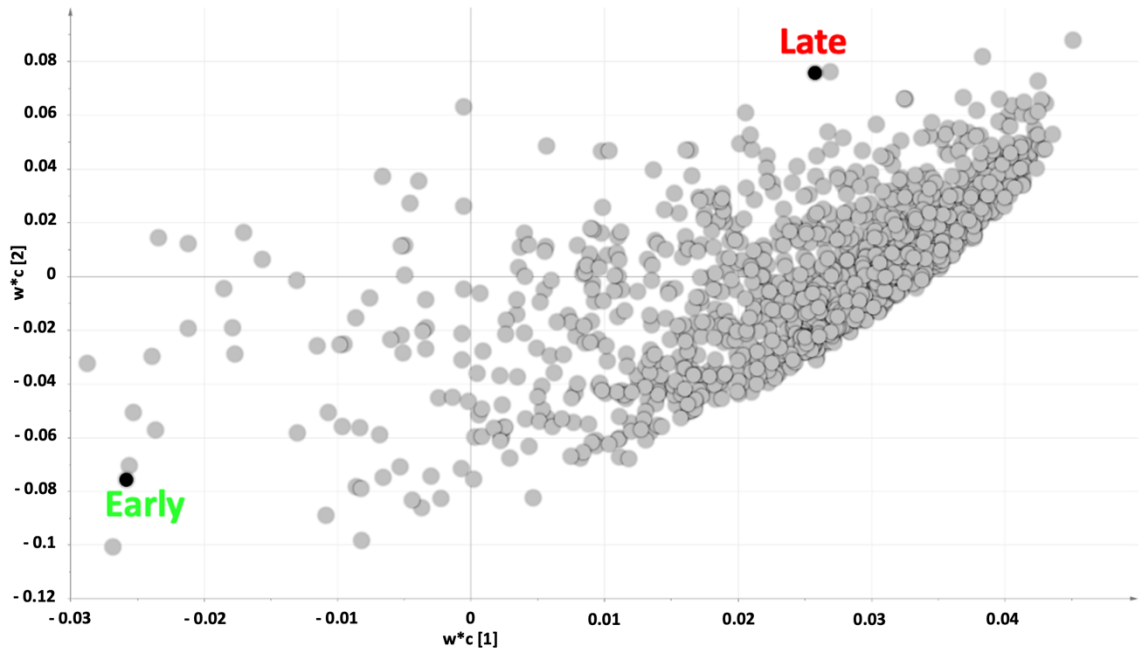


Figure 4-2 - Loading scatter plot for hip early vs. late group comparison using whole SOMA panel of 1310 markers.Grey points represent relative contribution of each marker to the model. The position of the group centroids are indicated showing that a high number of markers cluster towards the centre of the model space indicating a weaker comparison with significant noise. A significantly higher number of markers are seen loading towards the late group than the early group. 3 component model (R^2 = 0.940 Q^2 = 0.669)

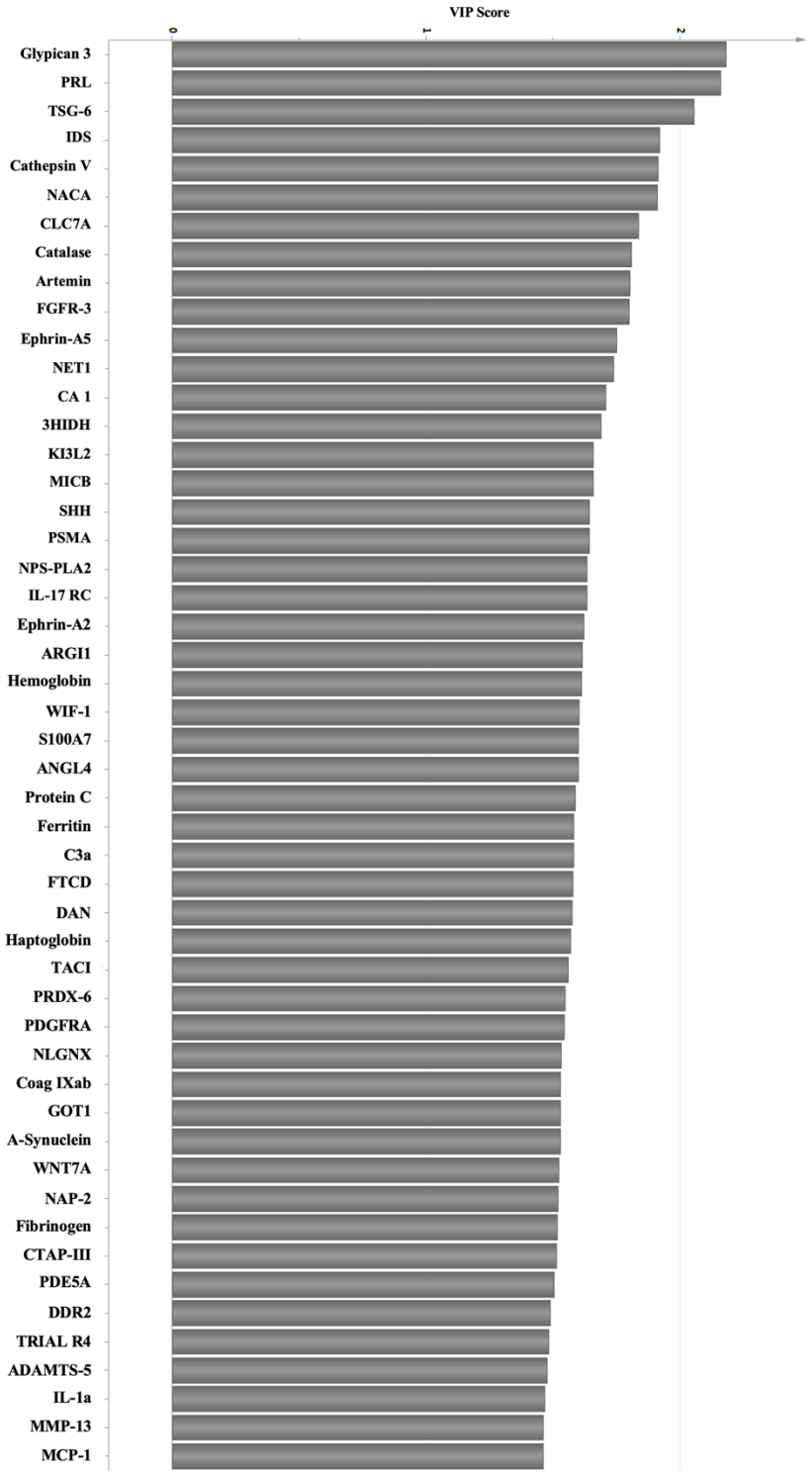


Figure 4-3 – VIP plot showing the top 50 markers from the hip early vs. late group comparison using the whole SOMA panel of 1310 markers. 3 component model (R^2 = 0.940 Q^2 = 0.669)All markers shown rank above 1.5 in this model, where 0.8 is customarily taken as a cut off for significant contribution and 2.0 is indicated by the right hand line on the graph area. See appendix B.1 for details on markers displayed.

The VIP plot again demonstrates the relative weight exerted by the top markers in the model and therefore that the lower weighted markers in effect detract from the model performance by introducing noise. Figure 4-3 shows the VIP plot for the model showing all of the top 50 markers score above 1.5 indicating the most significant contribution to the model of the 1310 markers.

4.4.2 Hip Early OA and Late OA Group Comparison Using the Most Descriptive 50 Markers From the Whole SOMAscan Comparison

Taking these top 50 markers from the model above to create a PLS-DA model comparing the early and late hip groups again demonstrated improved performance characteristics whilst showing markers loading towards the individual groups. There was tight clustering of the late group samples around the group centroids seen on the score scatter plot (Figure 4-4) with the loading scatter plot indicating those markers very close to the centre of the late group having the greatest contribution to the discrimination (Figure 4-5). The VIP plot demonstrates that when taken as a separate model only the top 25 markers have significant discriminative function within the model with a score above 0.8 (Figure 4-6). Table 4-2 shows the top 50 markers and their background information.

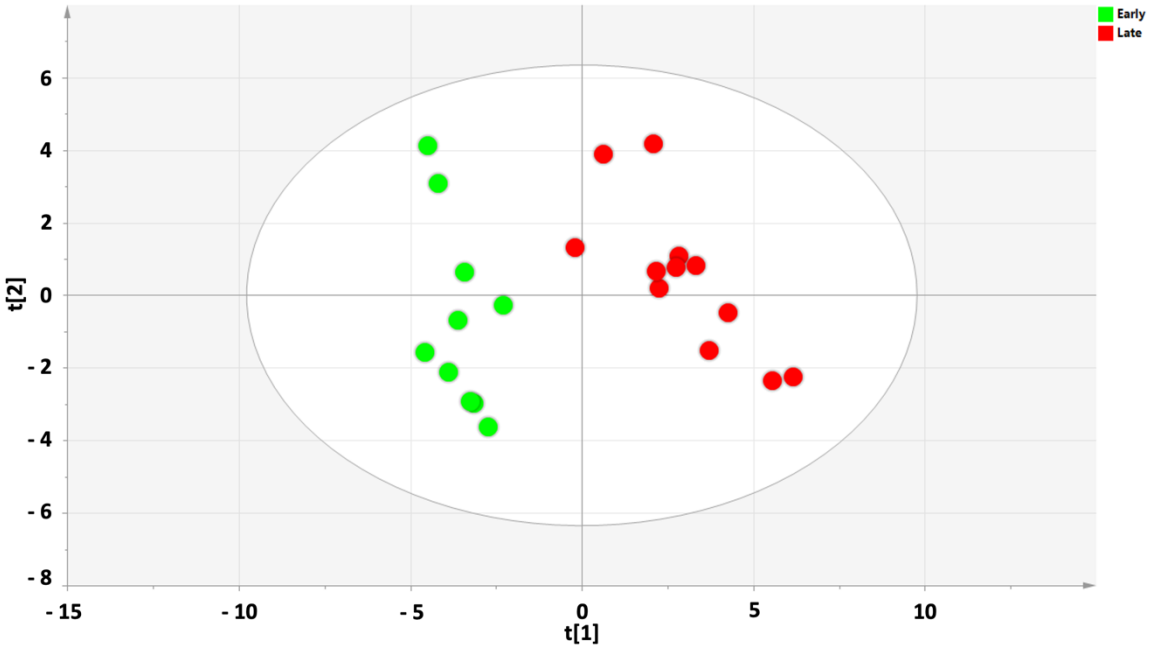


Figure 4-4 - Score scatter plot for hip early vs. late group comparison using the top 50 markers from the SOMA panel showing improved separation and within group clustering. 3 component model (R^2 = 0.948 Q^2 = 0.852)

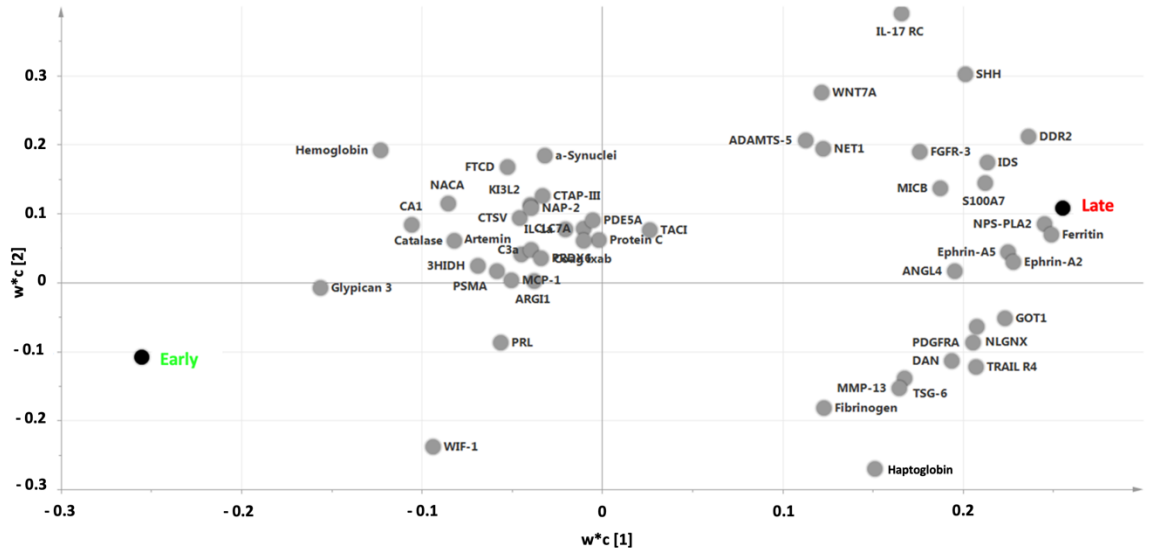


Figure 4-5 - Loading scatter plot for hip early vs. late group comparison using the top 50 markers from the SOMA panel. 3 component model (R^2 = 0.948 Q^2 = 0.852)

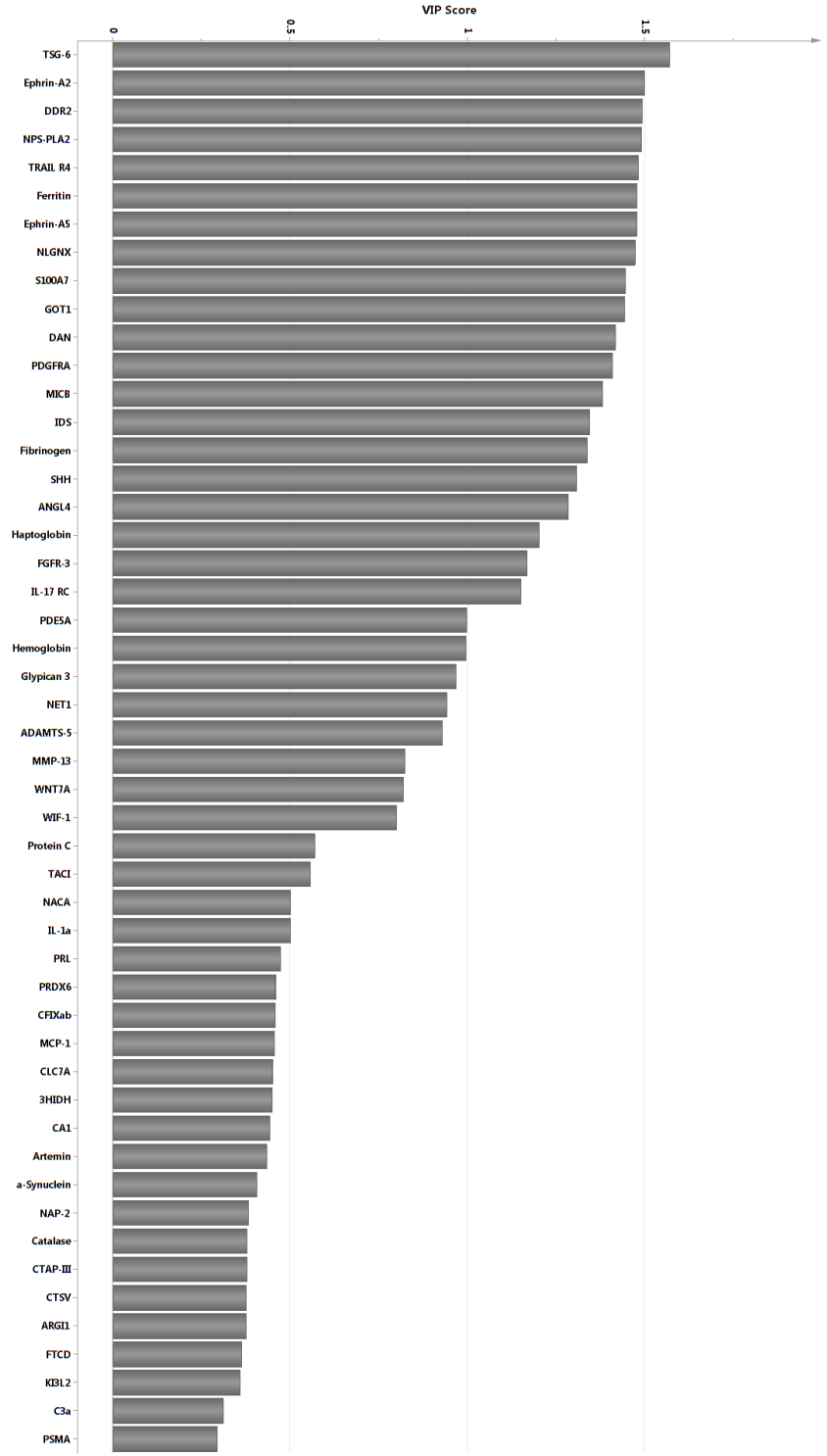


Figure 4-6 - VIP plot for hip early vs. late group comparison using the top 50 markers from the SOMA panel. 3 component model (R^2 = 0.948 Q^2 = 0.852)28 of the 50 markers shown rank above 0.8 in this model, where 0.8 is customarily taken as a cut off for significant contribution and 1.5 is indicated by the right hand line on the graph area. See appendix B.1 for details on markers displayed.

Table 4-2 - Table showing top 50 markers from the early vs late hip OA group comparison panel (Fig. 4-3) with UniProt ID and description, ** indicates LASSO marker.

Full Name	Marker ID	UniProt ID	Loading	Description
Glypican-3	Glypican-3	P51654	Early	A cell surface PG bearing HS, regulates Wnt signaling [309]
Prolactin	PRL	P01236	Early	Pituitary hormone promotes lactation and immune regulation [310]
**TSG-6 ** **	TSG-6	P98066	Late	Glycoprotein up regulated in OA progression and acute knee injury [213]
**Iduronate 2-sulfatase ** **	IDS	P22304	Late	Lysosomal enzyme. Plays role in GAG degradation including HS [311]
Cathepsin V	CTSV	O60911	Early	A cysteine protease with collagenolytic activity [312]
Nascent polypeptide-associated complex subunit alpha	NACA	Q13765	Early	Transcriptional co-activator, potentiates AP-1 activity in osteoblasts [313]
C-type lectin domain family 7 member A	CLC7A	Q9BXN2	Early	A lectin pattern recognition receptor with TLR2 mediated inflammatory activity [314]
Catalase	Catalase	P04040	Early	Ubiquitous enzyme required for neutralization of hydrogen peroxide [315]
Artemin	Artemin	Q5T4W7	Early	A neurotrophic factor that induces axon migration and neuron proliferation [316]
Fibroblast growth factor receptor 3	FGFR-3	P22607	Late	Cell surface receptor with a role in chondrocyte differentiation [317]
**Ephrin-A5 ** **	EFNA5	P52803	Late	Membrane bound signaling ligand influencing cell adhesion [318]
Netrin-1	NET1	O95631	Late	A guidance cue involved in axon projection towards the spinal cord [319]
Carbonic Anhydrase 1	CA1	P00915	Early	A metalloenzyme that catalyses the CO_2 / HCO_3 reversible reaction [320]
3-hydroxyisobutyrate dehydrogenase, mitochondrial	3HIDH	P31937	Early	A protein expressed in neural cells and involved in L-valine metabolism [321]
Killer cell immunoglobulin-like receptor 3DL2	KI3L2	P43630	Early	Receptor on NK cells & T cells which binds MHC class I molecules [322]
**MHC class I polypeptide-related sequence B ** **	MICB	Q29980	Late	A self antigen with activity in the susceptibility and pathogenesis of RA [323]
Sonic Hedgehog Protein	SHH	Q15465	Late	Hedgehog family member that directs cell fate and maintains homeostasis [324]

Glutamate carboxypeptidase 2	PSMA	Q04609	Early	Membrane bound metallo-enzyme, catalyzes neuro-transmitter hydrolysis [325]
Phospholipase A2, membrane associated	NPS-PLA2	P14555	Late	A phospholipase active in inflammation and as a ligand for integrins [326]
Interleukin-17 receptor C	IL-17 RC	Q8NAC3	Late	The receptor for IL-17A & IL-17F, key cytokines in the immune response [327]
Ephrin-A2	EFNA2	O43921	Late	Membrane bound signaling ligand influencing cell adhesion [318]
Arginase-1	ARGI1	P05089	Early	Enzyme expressed by PMN granulocytes active in chronic inflammation [328]
Haemoglobin	Hb	P69905 P68871	Early	Alpha and beta subunits of the oxygen transport molecule
Wnt inhibitory factor 1	WIF-1	Q9Y5W5	Early	Inhibitor of factors involved in cartilage development [329]
Protein S100-A7 **	S100-A7	P31151	Late	Calcium binding protein in wound healing and immune modulation [330]
Angiopoietin-related protein 4	ANGL4	Q9BY76	Late	Mediator of lipoprotein lipase in the regulation of lipid metabolism [331]
Vitamin K-dependent protein C	Protein C	P04070	Early	Serine protease that inactivates coagulation factors Va and VIIIa [332]
Ferritin	Ferritin	P02794 P02792	Late	Heavy and light chains of the iron storage and delivery molecule
C3a anaphylatoxin	C3a	P01024	Early	The central factor of both classical and alternate complement cascades
Formimidoyltransferase-cyclodeaminase	FTCD	O95954	Early	Folate dependent enzyme required for catabolism of histidine [333]
Neuroblastoma suppressor of tumorigenicity 1	DAN NBL-1	P41271	Late	Prominent antagonist of extracellular proteins including BMPs [334]
Haptoglobin	Haptoglobin	P00738	Late	A hepatic protein that binds free haemoglobin for iron recycling
Tumor necrosis factor receptor superfamily member 13B	TACI TNFRSF13B	O14836	Late	Binds TNFSF13 ligands to mediate activation of NF-κB and AP-1 [335]
Peroxiredoxin-6	PRDX6	P30041	Early	Hydrogen peroxidase with a role in protection against oxidative stress [336]
Platelet-derived growth factor receptor alpha	PDGFRA	P16234	Late	Cell surface receptor for PDGF in cell survival and chemotaxis [337]
Neuroligin-4, X-linked	NLGNX	Q8N0W4	Late	Involved in synapse activity, neuronal plasticity and development [338]

Coagulation Factor IXab	F IXab	P00740	Early	Vitamin K dependent coagulation factor in the intrinsic pathway
Aspartate aminotransferase, cytoplasmic	GOT1	P17174	Late	Regulates levels of the excitatory neurotransmitter; glutamate [339]
Alpha-synuclein	a-synuclein	P37840	Early	Role in Parkinson's disease, suggested role in release of neurotransmitters [340]
Wnt-7a	WNT7A	O00755	Late	Ligand for frizzled receptors in Wnt signaling pathway [341]
Neutrophil-activating peptide 2	NAP-2	P02775	Early	A ligand for CXCR1 & 2 and chemoattractant for neutrophils [342]
Fibrinogen	Fibrinogen	P02671 P02675 P02679	Late	Alpha, beta and gamma chains of the primary component of blood clots
Connective tissue-activating peptide III	CTAP-III	P02775	Early	Chemokine with immuno-regulatory activity as a precursor to NAP-2 [343]
cGMP-specific 3',5'-cyclic phosphodiesterase	PDE5A	O76074	Early	Phosphodiesterase of cGMP with a role in signal transduction [344]
Discoidin domain receptor 2 **	DDR2	Q16832	Late	Tyrosine kinase receptor that is activated by type II collagen [345]
Tumor necrosis factor receptor superfamily member 10D	TRAIL R4 TNFRSF10D	Q9UBN6	Late	Receptor for TRAIL which protects TRAIL mediated apoptosis [346]
A disintegrin and metalloproteinase with thrombospondin motifs5	ADAMTS-5	Q9UNA0	Late	One of the central aggrecanases in cartilage ECM degradation [347]
Interleukin-1 alpha	IL-1a	P01583	Early	Endogenous pyrogen in the acute inflammatory response
Collagenase 3	MMP-13	P45452	Late	ECM degrading enzyme with highest activity to type II collagen [237]
C-C motif chemokine 2	MCP-1 CCL2	P13500	Early	Chemoattractant involved in inflammatory processes of arthritis [348]

Plots showing the actual variance of selected groups of proteins between the study groups are shown below. Metalloproteinase inhibitors (Figure 4-7), matrix metalloproteinases (Figure 4-8), ADAMs and ADAMTSs (Figure 4-9) and selected markers that loaded to the hip early OA group (Figure 4-10) are shown demonstrating the relative marker distributions. Significance was tested with ANOVA and is shown for the hip groups. P-values are shown for those markers where significance was identified ($p<0.05$).

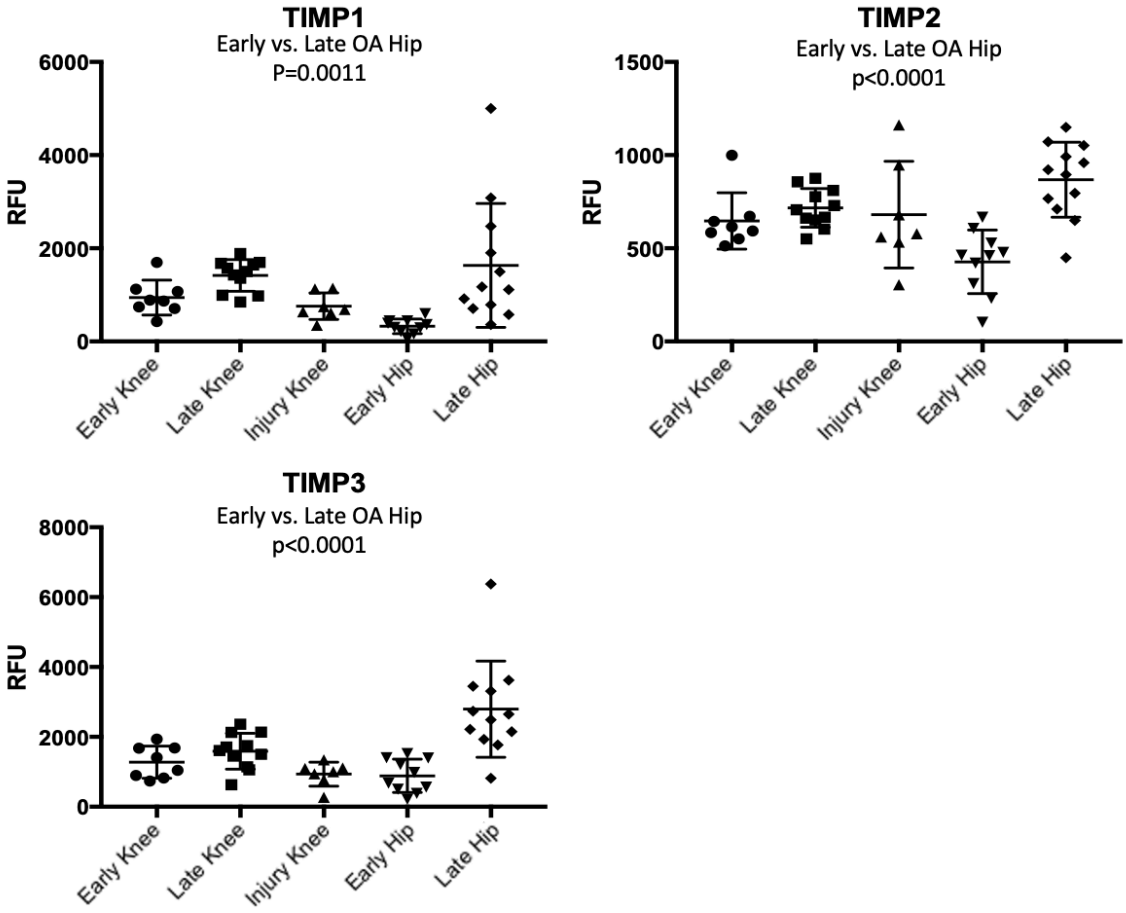


Figure 4-7 - Graphs showing variance between groups for metalloproteinase inhibitors: TIMP 1, 2 & 3. Significance tested using 2-tailed t-test. P-values shown for those markers that met significance.

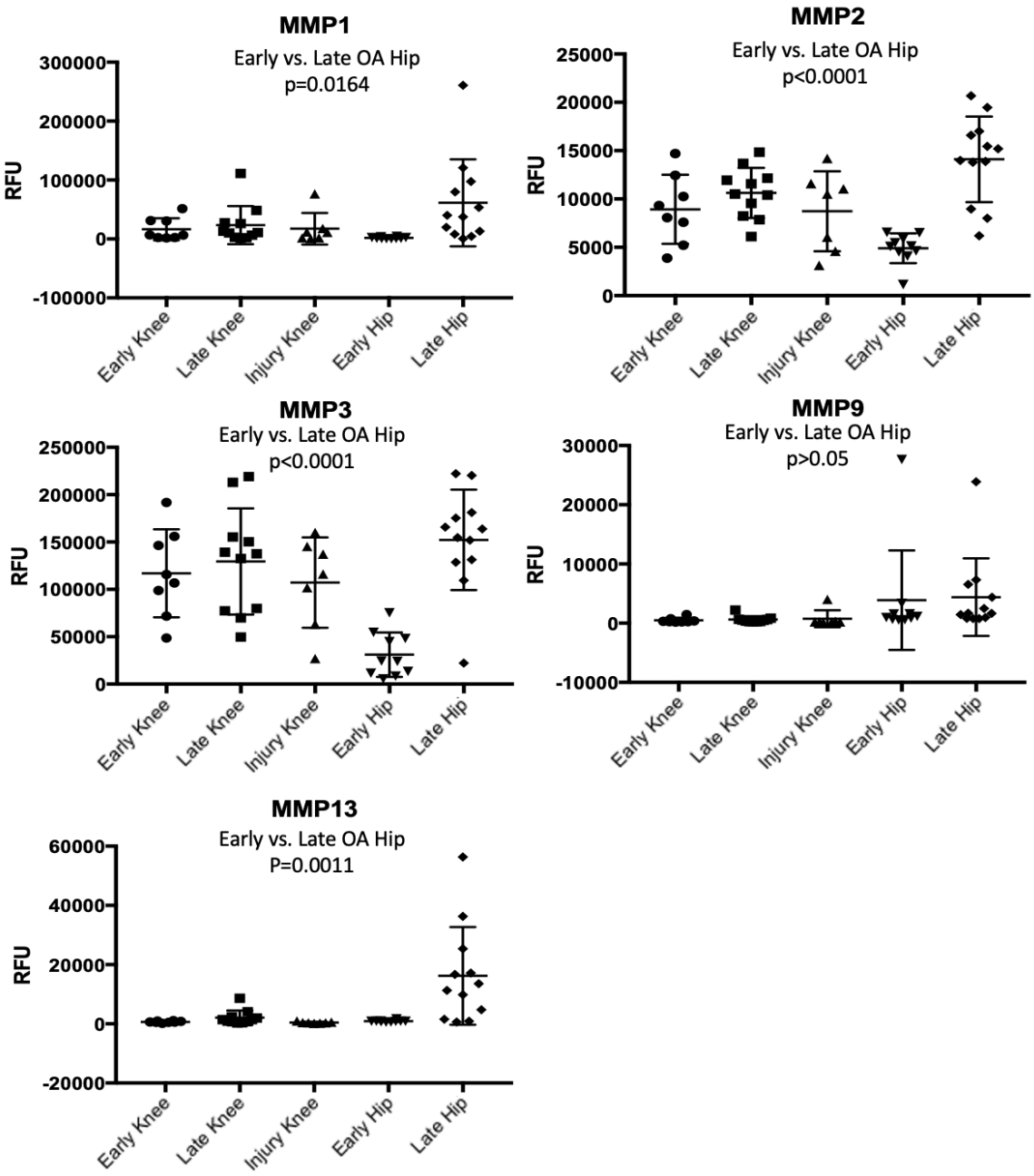


Figure 4-8 - Graphs showing the variance between groups for selected matrix metalloproteinases: MMP-1, 2, 3, 9 & 13. Significance tested using 2-tailed t-test. P-values shown for those markers that met significance.

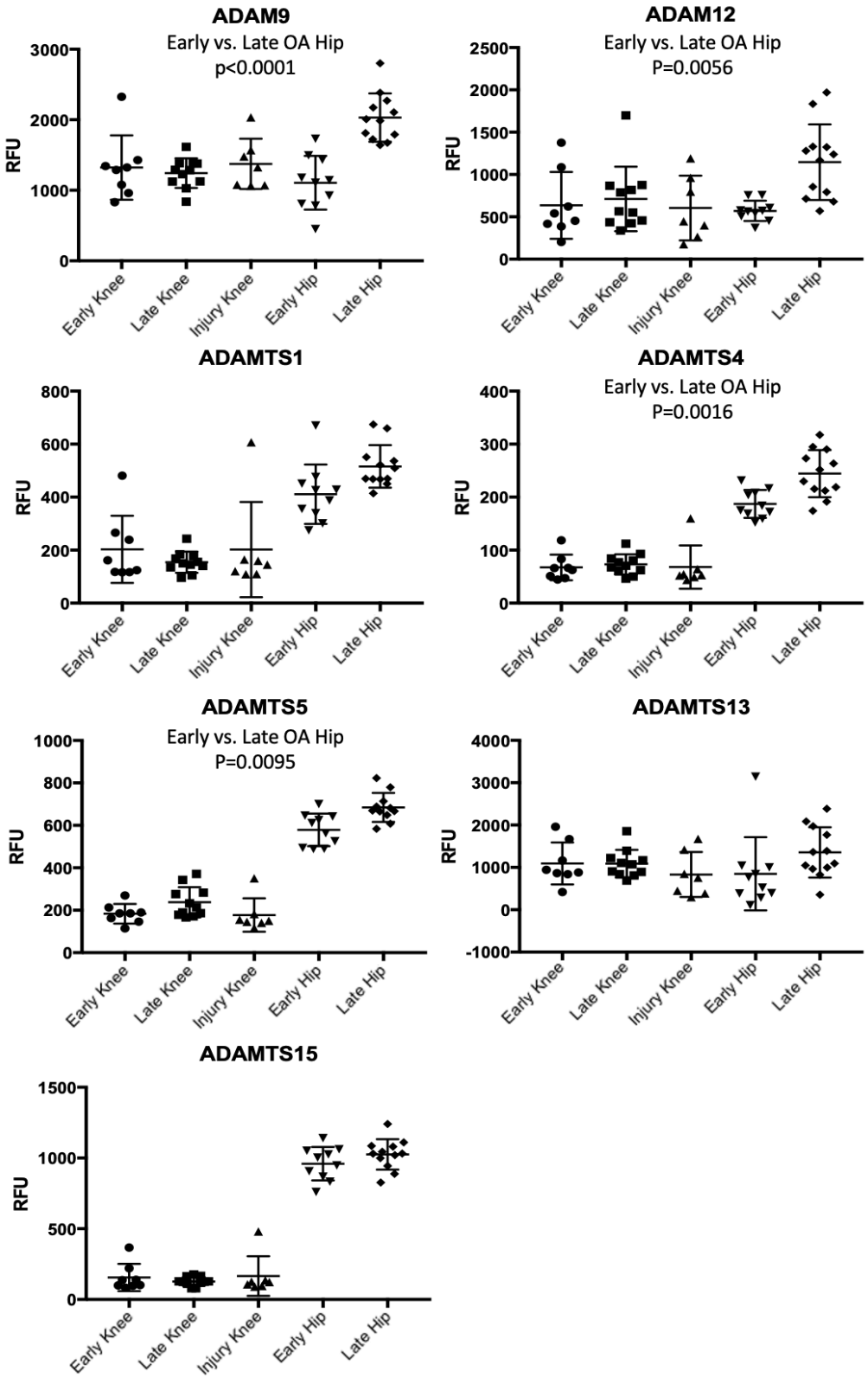


Figure 4-9 - Graphs showing the variance between groups for selected ADAM and ADAMTS proteins: ADAM-9 & 12 and ADAMTS-1,4,5 13 & 15. Significance tested using 2-tailed t-test. P-values shown for those markers that met significance.

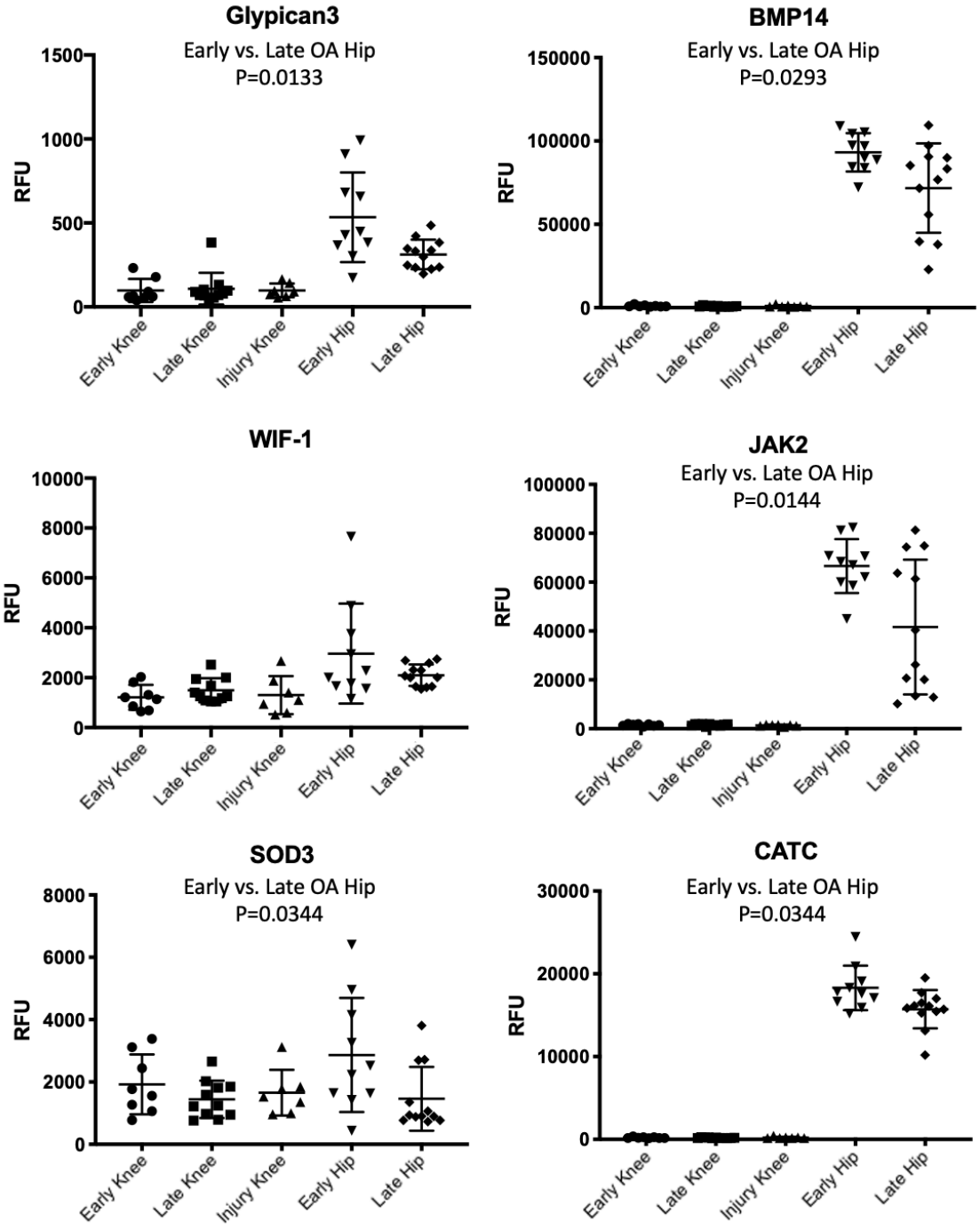


Figure 4-10 - Graphs showing the variance between groups for selected proteins that were elevated in the early hip OA group: Glypican 3, BMP14, WIF-1, JAK2, SOD3 and CATC. Significance tested using 2-tailed t-test. P-values shown for those markers that met significance.

4.5 LASSO

Refinement of the PLS-DA models of the analysis of the hip groups has shown that the early and late hip groups appear biologically different. As with the knee models, in order to improve the model performance in the case of multi-collinear data, LASSO analysis of the whole data set was performed using R, version 3.3.3 (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/). Data were scaled, log transformed and centred prior to LASSO analysis which was performed using the "glmnet" package. Internal model performance was checked by serially exclusion and re-introduction of each observation to determine if the output markers were able to appropriately predict the group with no misclassification events. The markers identified from LASSO analysis as most specifically describing the differences between observations in each group were:

- Tumour necrosis factor stimulated gene 6 (TSG-6)
- Iduronate 2 sulfatase (IDS)
- Ephrin-A5 (EFNA5)
- MHC class I polypeptide related sequence B (MICB)
- S100 calcium binding protein A7 (S100A7)
- Discoidin Domain Receptor 2 (DDR-2)

These markers were highly ranked in the whole 1310 marker SOMAscan analysis VIP ranking with TSG-6, IDS, EFNA5, MICB, S100A7, and DDR-2 ranked 3, 4, 11, 16, 25 and 45 respectively indicating good agreement in preferred markers. Plots showing the actual variance between the groups are shown in Figure 4-11 showing the relative marker distributions. Significance was tested with 2-tailed t-test for the hip groups. The p-values are shown on the plots where markers met significance ($p<0.05$).

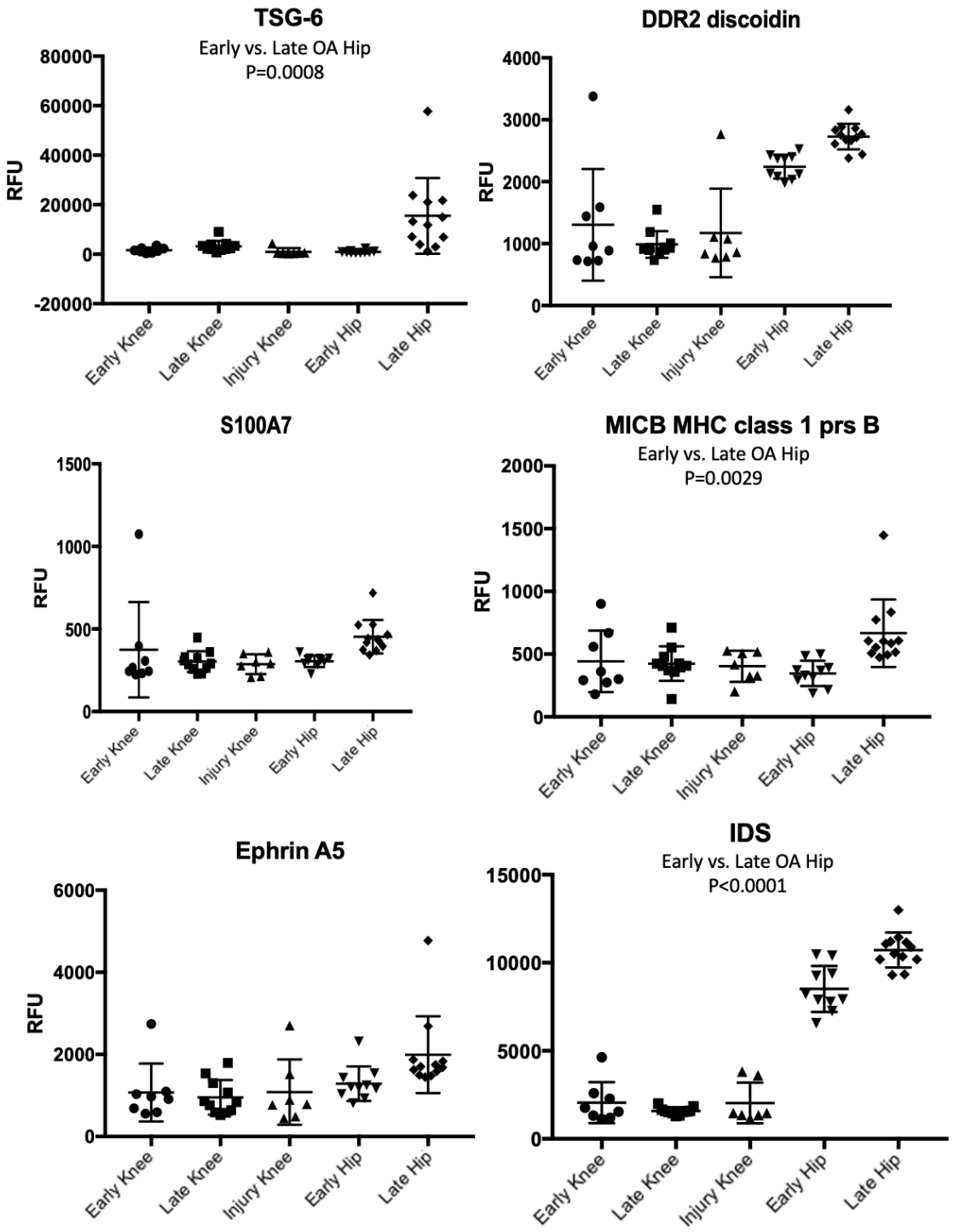


Figure 4-11 - Graphs showing the variance between groups for markers identified by LASSO analysis: TSG-6, DDR-2, S100A7, MICB, EFNA5 and IDS. Significance tested using 2-tailed t-test. P-values shown for those markers that met significance.

4.5.1 PLS-DA model of 6 LASSO markers

To further assess the selectivity of the markers identified by LASSO analysis of the hip SF data from the SOMAscan a PLS-DA model was performed using only these markers. This demonstrated that the groups could be confidently discriminated on the basis of these six markers as shown in the loading scatter plot (Figure 4-12) with excellent model performance characteristics of $R^2 = 0.845$ $Q^2 = 0.821$.

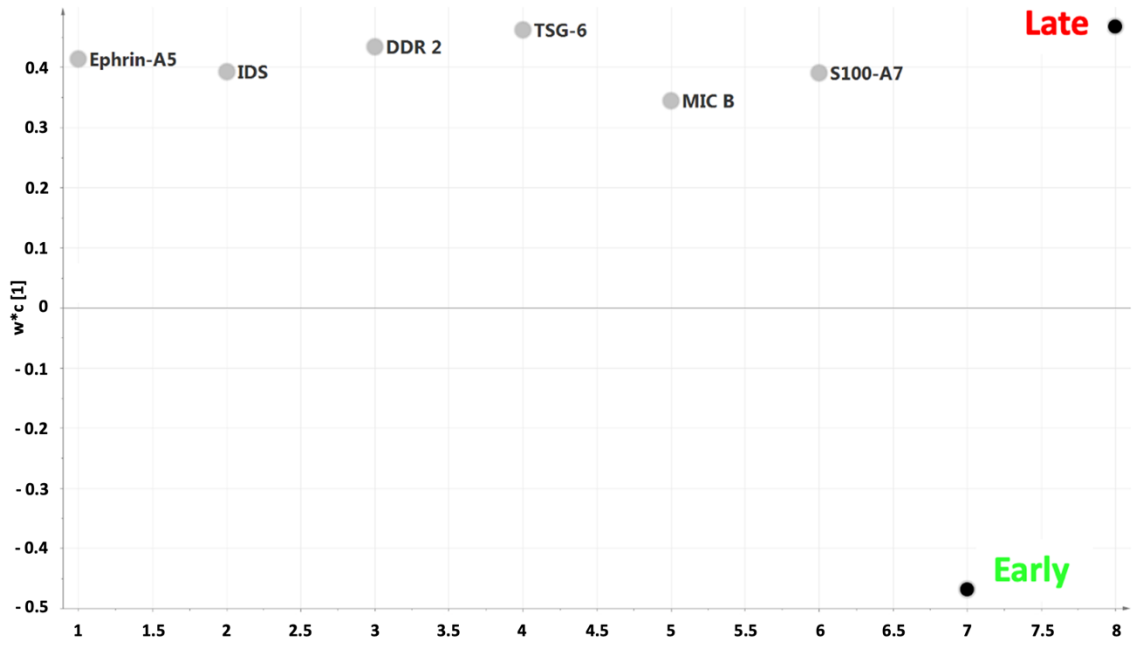


Figure 4-12 - Loading scatter plot for hip early vs. late group comparison using LASSO marker panel. 1 component model ($R^2 = 0.845$ $Q^2 = 0.821$)

4.6 PCA

Principal component analysis was performed using the base R package with visualisations created using the "pca3d" package (https://CRAN.R-project.org/package=pca3d). Factor analysis was performed for extraction of the top loadings on PCA using the FactoMineR package. Using the same method as

that employed in chapter 3 the top 20 variables, which accounted for the majority of variance of each principal component, were selected with further loading variables selected which contributed to component 2. The variables contributing the highest loading from the factor analysis were selected. The normalised protein values were transformed to fold changes by subtraction of the values from each observation of a particular protein from the average of the protein values of either the early OA or injury samples in their respective comparisons against the late OA group. The subtracted values were then transformed to linear space using the antilog of base 2. The fold changes were clustered for both protein values and samples using canberra distance and complete linkage. The results were visualised using the gplots package (https://cran.r-project.org/package=gplots) and RColorBrewer package (https://cran.r-project.org/package=RColorBrewer).

4.6.1 Hip Early OA and Late OA Group Comparison Using the Whole SOMAscan

PCA comparing hip early and late OA groups supports the above findings with clear separation of the groups with localisation of the samples around the centroids (Figure 4-13). The leading proteins in each component are cytoplasmic protein NCK1 (non-catalytic region of tyrosine kinase adaptor protein 1) and CHST6 (Carbohydrate Sulfotransferase 6) for components 1 and 2 respectively. When hierarchical clustering was displayed as a heatmap (Figure 4-14) it is evident that the majority of markers loading for late hip OA are highly expressed compared to early hip OA which is characterised by lower levels of the same markers with early

markers lying to the left of the plot area and late to the right. This observation is in agreement with the PLS-DA analysis and is reflected by the LASSO markers, all showing the same late group preponderance.

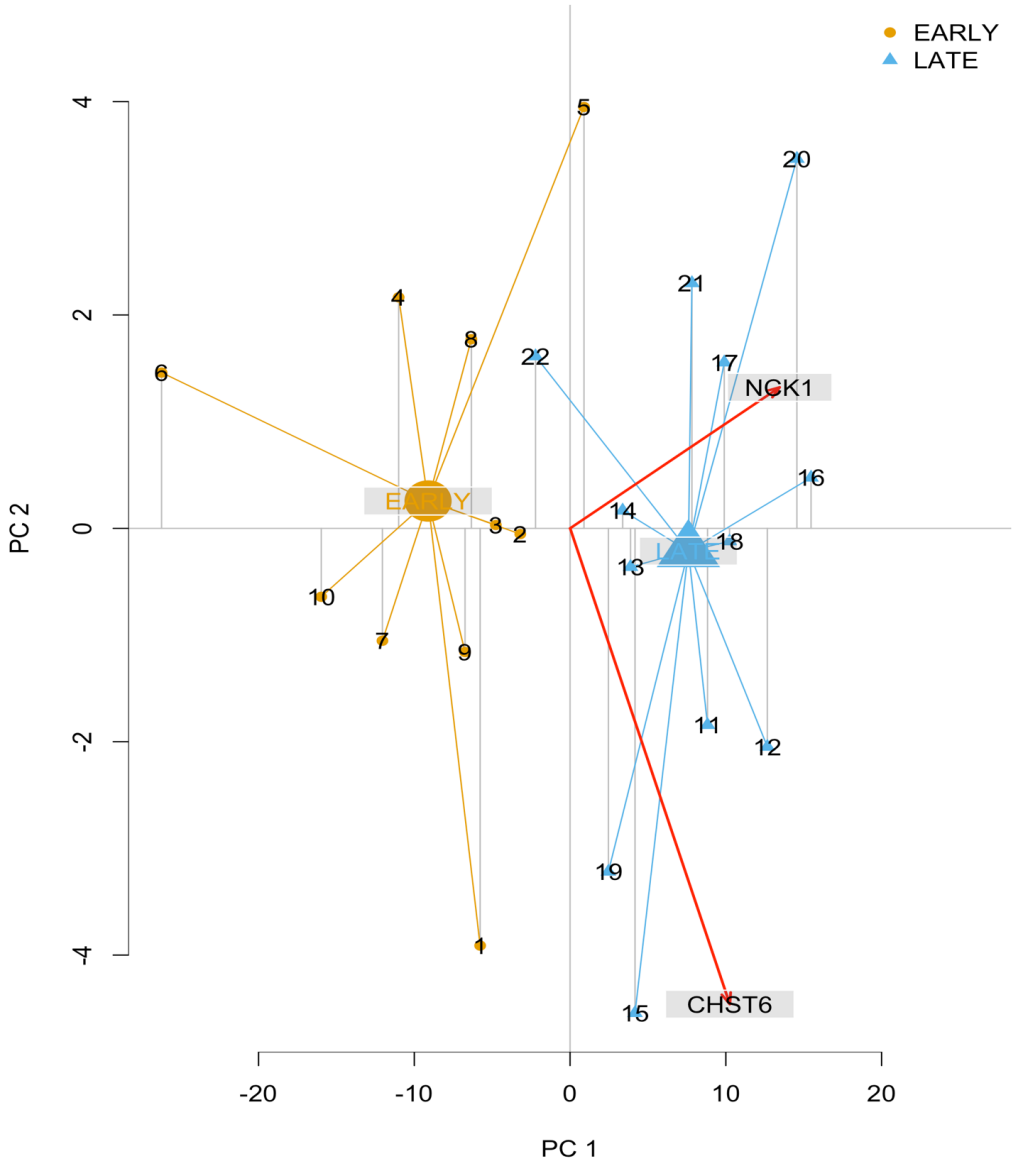


Figure 4-13 - Observation score plot of PCA of quantile normalised SF proteins from comparison of hip early OA and late OA groups.Plot shows group centroids (large coloured shapes to correspond with groups). Orange arrows show the first two orthogonal variables. Courtesy of Dr. A. Chanalaris.

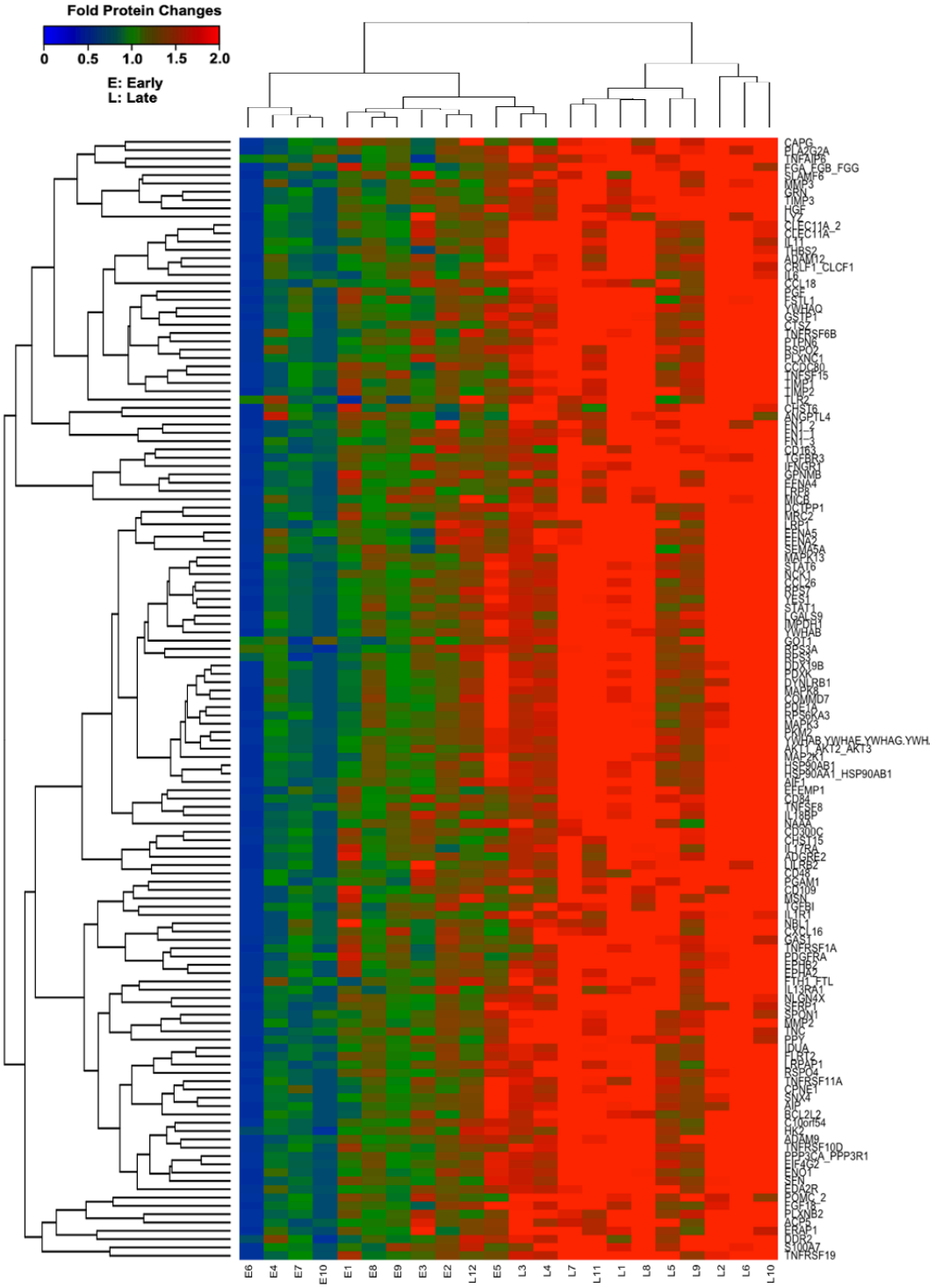


Figure 4-14 - Heatmap and hierarchical clustering of PCA of quantile normalised SF proteins from comparison of hip early OA and late OA groups.Heatmap and hierarchical clustering based on first 5 principal components showing samples cluster into two main groups with the right group composed predominantly of late group samples Courtesy of Dr. A. Chanalaris.

4.7 Network Analysis

Network analysis of the proteins identified by the LASSO analysis, performed using the STRING v11 database [249], demonstrates the protein – protein interactions evidenced in the literature (Figure 4-15). Support and graphical representations provided by Dr. A. Chanalaris. In the case of the hip markers discussed above, the network appears to be centred on the TSG-6 / TIMP-1 axis similar to the knee study. There are numerous metalloproteinase interactions that implicate these markers in the pathogenesis of OA via cartilage degradation and complex extracellular matrix signaling pathways. A further protein-protein interaction network reconstruction was performed, based only on experimental data and with high significance cut off, using PCA markers producing a network of 786 proteins with 212 connections. The expected number of connections with 786 random proteins is 83 => p <10e-6, therefore indicating this network of proteins may be biologically related.

The network formed by the LASSO markers alone (Figure 4-15) shows a matrix degradation pattern centred on TIMP-1 and MMP-2, 9 & 13 all of which are established enzymes that are active in dysregulation of ECM turnover in OA [349–351]. The STRING interactome (Figure 4-16) based on the PCA markers also supports this pathway of activity as well as clusters of mitogen activated protein kinases (MAPK) and ephrins. The MAPKs have been shown to regulate proinflammatory cytokines and MMPs in OA [119] and ephrins have been shown to play a role in bone remodelling in disease [352].

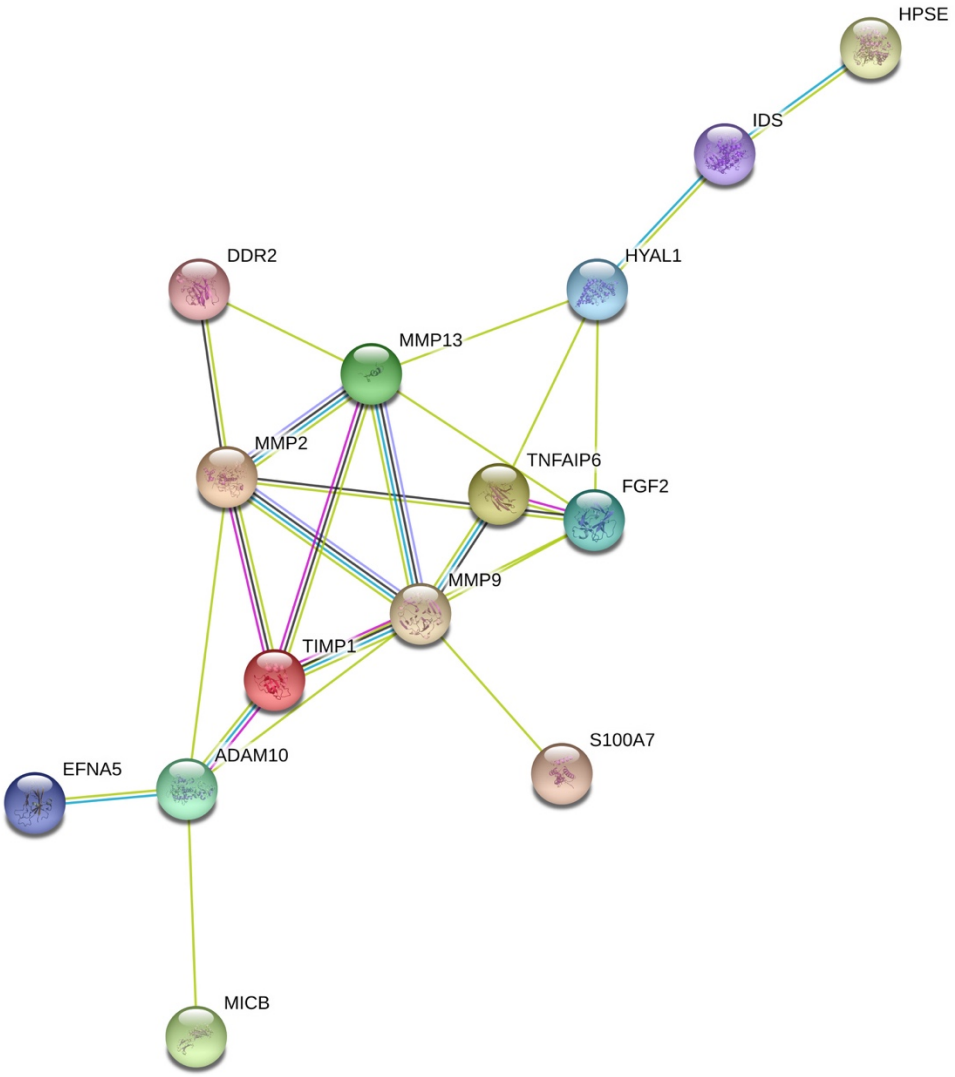


Figure 4-15 – Focused network map of protein-protein interactions based solely on hip LASSO markers.Coloured nodes indicate query proteins and proteins from the first shell of interactions. Connection lines indicate number of different sources supporting the interaction.

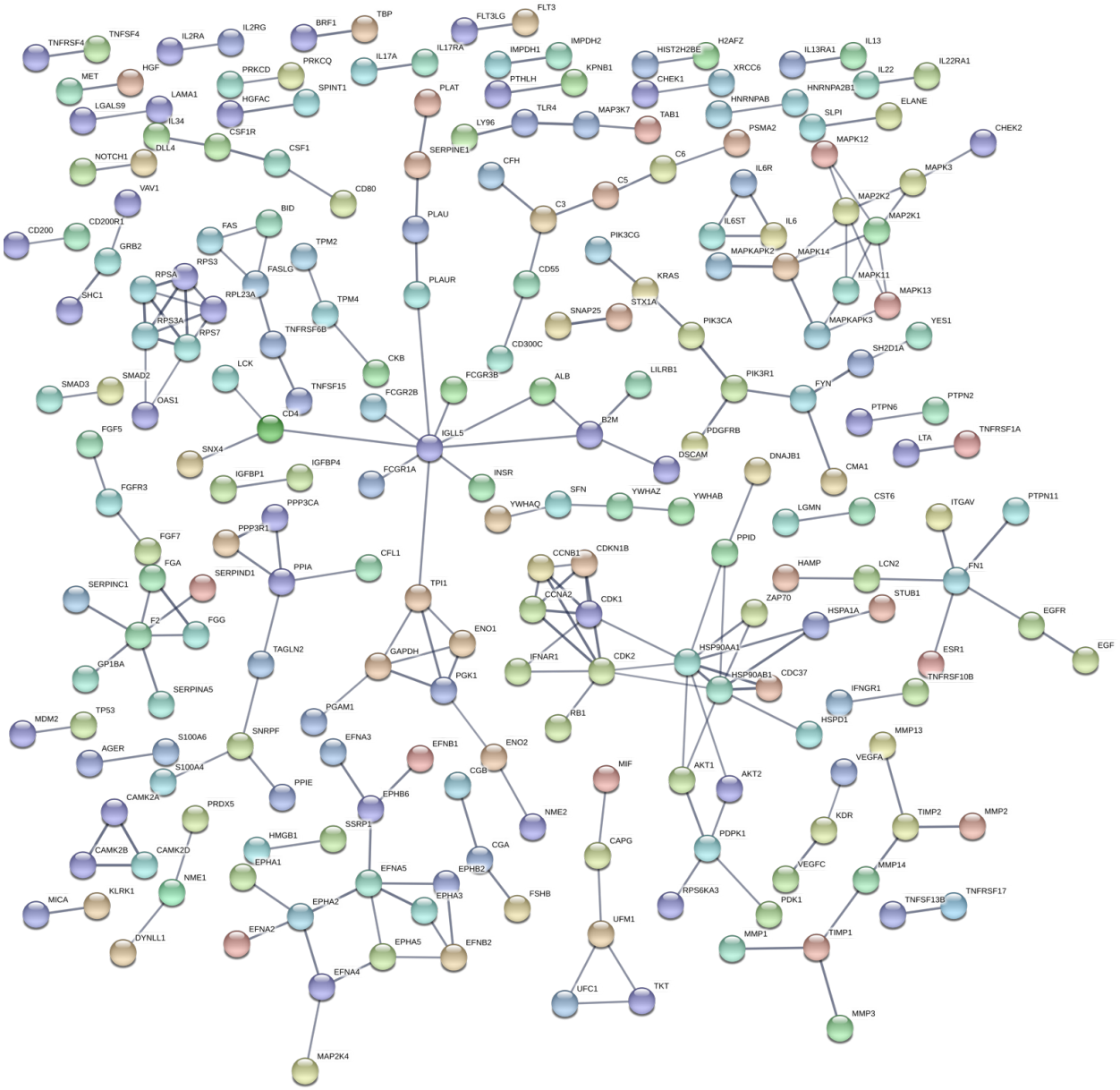


Figure 4-16 – Interactome network of protein - protein interactions.Based on the PCA top loading markers from the first 5 principal components of PCA of quantile normalised SF proteins from comparison of knee injury and late OA groups using the whole SOMAscan panel of 1310 proteins.Coloured nodes indicate query proteins and proteins from the first shell of interactions. Thickness of connection lines indicate the strength of the interaction. Courtesy of Dr. A. Chanalaris.

4.8 Discussion

This study aimed to identify biomarker fingerprints of different clinical and structural disease states of hip OA groups using multivariate analyses of SOMAscan data. Before this could begin the collection of the early OA hip cohort required a bespoke acquisition method due to the dramatically lower SF volume seen in patients with FAI. The optimisation of this method, allied to the relative rarity of these patients when compared to the knee, and the low sample volume illustrates the novelty of this cohort. During the SF validation work it was noted that the dynamic range of the hip SF appeared to be higher than that seen in the knee. This was observed across the whole SOMAscan and it is suspected that this related to the different systems of collection and processing. All hip samples were centrifuged and snap frozen immediately after collection by the same researcher and subsequent HAse processing performed as a single batch with the same reagents whilst the knee samples were HAse treated serially. We hypothesise that there may be a processing difference but are unable to attribute this. However, this should not cause alarm since no between group analyses were performed to compare the hip groups with knee groups, only within joint comparisons were performed. Descriptive graphical plots of individual marker quantification are shown as combined plots for both joints for interest but did not affect multivariate models.

The results in this chapter have suggested that the early and late hip OA groups appear to be biologically distinct groups with differing underlying biology. Similar

to the knee study there are a greater number of markers describing the late hip OA group and all LASSO markers identified in this study were elevated in the late hip OA group. It is interesting that the early and late OA comparison produced superior separation in the hip, as compared with the knee. It is possible that this relates to the batch differences seen which, it is postulated, may be attributable to possible handling differences. It is further worthy of note that there is no comparable hip group to the knee injury cohort. Nevertheless the suggestion, in this work, of a difference in the pathways enriched between hip and knee may support the clinical suggestion that the early hip group is not directly comparable to the knee injury cohort.

The different supervised and unsupervised analyses have highlighted common candidate markers of interest in the description of these clinical groups and their biology must be considered.

The biological relevance of TSG-6 to OA pathogenesis is discussed above in paragraph 3.6.1. In summary it is stimulated by pro-inflammatory cytokines, binds hyaluronan and has been associated with OA progression in both murine and human *in vivo* models.

Discoidin domain-containing receptor 2 (DDR-2) is a cell surface tyrosine kinase receptor that is activated by type II collagen and is involved in the control of cellular responses to the ECM by regulation of multiple MMPs in chondrocytes and synovial fibroblasts [345,353,354]. Up regulation and activation of DDR-2 has

been shown to induce expression of MMP-13 by chondrocytes and this has been specifically associated with type II collagen degradation in murine models of OA [355,356]. Furthermore, in a murine destabilisation of medial meniscus (DMM) model using DDR-2 knockout mice there was significantly delayed progression of degeneration in the DDR-2 deficient mice compared to controls [357]. DDR-2 has also been shown to influence osteoblast differentiation [358] via phosphorylation of Runx2 and to interact with type I collagen in a diabetic OA model promoting crosslinking via induction of lysyl oxidase in osteoblasts [359].

Iduronate 2 sulfatase (IDS) is a lysosomal enzyme required for the degradation of GAGs including, most notably, heparan sulfate and dermatan sulfate [311]. Heparan sulfate expression and function is important for the healthy function of joints playing an important role in the regulation of cartilage homeostasis by regulation of TIMP-3 and its affinity for ADAMTS-5 [360]. Heparan sulfate biosynthesis has been shown to be markedly disrupted in OA cartilage when compared to normal cartilage [291]. This mechanism may be implicated in the progression of OA via the receptor for advanced glycation end products (RAGE) as a result heparan sulfate facilitation of RAGE binding the high mobility group box 1 (HMGB-1) [237].

S100 calcium binding protein A7, also termed psoriasin, is an alarmin and member of the S100 family of proteins shown to have a role in the pathogenesis of psoriasis [361]. The S100 family comprises 25 calcium binding proteins involved in a number of post secretory extra cellular functions with S100A7 itself having been shown to

play a role in wound healing and immune modulation [330]. S100A7 acts via RAGE and its expression is elevated in inflammatory conditions [362]. It has also been shown to enhance the activity of MMP-2 and MMP-9 via the S100A7-RAGE signal transduction pathway in osteosarcoma [363]. Another closely related member of the S100 family, S100A8/A9 or calprotectin, has previously been shown to play a role in the initiation of cartilage degradation in OA by upregulation of MMPs and aggrecanases in a murine model of early OA [364]. In a human study S100A8, S100A9 and S100A12 were all shown to be upregulated in RA as compared with a minor inflammatory arthropathy cohort [365]. Of note, in a recent case-control study calprotectin has been shown to be significantly higher in knee OA patients than in healthy controls but at its highest levels in the serum of patients with the earliest stages of clinical and radiographic OA [366].

The Ephrins are a large family of plasma membrane bound proteins. They are signaling ligands for tyrosine kinase receptors influencing cell adhesion [318] with Ephrin A5 mediating cell signaling via ADAM-10 cleavage [367].

MHC class I polypeptide-related sequence B (MIC-B) is part of an immunoglobulin superfamily that are recognised by T-cell receptors [368]. It is cleaved by ADAM-10 and via this pathway has been mechanistically implicated in numerous malignancies [369].

NCK1 is a tyrosine kinase adaptor protein, mediating cortactin driven actin remodelling in exosome production; involved with glucose tolerance and insulin signaling, bone mass and cellular remodelling [370].

CHST6 is a sugar sulfotransferase important in the production of the proteoglycan keratan sulfate (KS) [371].

This study represents one of the first studies to examine hip SF from any non-end stage cohort. Three studies were identified in the literature that describe SF analyses in hip OA cohorts with potential similarity to this study. However, two of these studies described the use of saline lavage as the method of SF acquisition [372,373] while presenting no data to show how they handled the resulting dilutions or the effects on the interpretation of results. The third study describes SF analysis of a clinically defined hip OA cohort with inclusion criteria including K&L grade of 1 or greater although they did not recruit any patients to their hip cohort with K&L grade below 3 [374]. Accordingly, this study may be the first to examine the proteomic analysis of undiluted SF from a unique human cohort of phenotypically defined patients established as an early OA cohort. There are reports of examination of SF harvested from patients at surgery for THR, peri-prosthetic joint infection or revision arthroplasty for other causes than infection and in these cases the only broad spectrum analysis used was mass spectrometry [375–378].

The findings of this study suggest some novel proteins that may help deliver insights into the underlying biological processes in hip OA including S100A7 an alarmin with an established role in response to wounding and immune modulation as well as IDS which is known to be required for degradation of heparan sulfate, a key extracellular matrix component. These links are supported by network analysis showing enriched proinflammatory and injury response to pathways as well as literature indicating potential inflammatory OA subtypes [379–381]. This study also suggests that differences appear to exist between the early and late disease states in hip OA. The different proteomic profiles seen further suggest that biological differences exist between hip and knee OA with a greater proportion of proinflammatory cytokines seen in the hip as well as fewer coagulation and complement cascade proteins. However, there remain common thematic processes of matrix degradation pathways that are supported in the literature as playing roles in the progression of degeneration.

This study, similar to chapter 3 discussed above, is limited by a low sample number. However, it should be noted that this limitation results from the extremely high cost of the SOMAscan assay and its a-priori unproven utility for SF analysis. Furthermore, in the case of the hip, the inherent challenges of sample collection make large SF cohorts impossible to collect meaning this study has contributed both a vital pilot study and a novel biomarker discovery project in the early hip OA field. There is, again, inherent group age difference with the early

OA hip group being significantly younger than the late OA hip group which, as discussed in chapter 3, is unavoidable due to the nature of disease timing.

The work presented in this chapter has defined a novel SF cohort of early hip OA and suggested that biological differences appear to exist between the early and late hip OA groups. The network analysis has shown that the markers identified are acting in known pathways in concert with matrix degradation mediators and signal pathways that are activated in the pathogenesis of hip OA from early to late disease. This work contributes to the biological insights into the pathogenesis of early hip OA and potentially suggests viable targets for future study in larger cohorts using multiplex analysis techniques. This utility may also be ideally suited to the investigation of longitudinal cohorts where so-called "rapid progressor" patients can be monitored serially to identify the active pathways that may identify the key drivers of disease and how such markers modulate over the course of its progression.

Chapter 5 Variations in Treatment Pathways to Access Care

5.1 Introduction

OA is one of the most common diseases worldwide. It is a chronic, progressive, disabling disease with many factors influencing its development and progression. In 2013, OA alone accounted for 12.8 million years lived with disease [382].

Joint replacement surgery is an extremely effective treatment for OA [383]. In 2019 95,677 primary THR operations and 103,617 primary TKR operations were performed in England, Wales and Northern Ireland [384]. This number continues to grow with demand rising due to an ageing, and increasingly obese, population. It has been projected that the number of THR and TKR procedures performed annually could rise to beyond 96,000 and 119,000 respectively by 2035 [385]. However, the associated costs of these procedures make this continued growth unfeasible, requiring means of identification of those at risk and implementation of cost-effective preventative measures.

There remains a critical problem that existing diagnostic tools are capable only of identifying established, late-stage disease. If a tool were available to aid earlier diagnosis the issue remains that a clearer understanding of the journeys taken by patients prior to referral for interventions is required to identify where such an instrument would be best placed. Relatively little is still known about how patient experiences and socio-economic status affects access to care and outcome of

treatment. Further understanding the referral patterns and different routes of access to care for OA patients and their wider characteristics can inform, and enhance guidance aimed at personalising treatment, thereby helping patients access the right care at the right time.

5.2 Aims

The aims of this study were:

1. To identify patients that undergo THR and examine their prior treatment pathway, relating to the order and timings of receiving NICE recommended non-surgical treatments i.e. analgesia (including paracetamol, NSAIDs, partial opiates and full opiate medications), physiotherapy and intermediate surgical treatments (intra-articular injection and arthroscopy).
2. To identify whether there is any variation in the pathway taken, according to patient factors, including age, sex, social deprivation and BMI.

5.3 Proposed pathways

To answer the study aims patients were considered as having been treated along one of three "Treatment Pathways" in the community by means of the following routes:

1. **Analgesic pathway:** Patients that received prescriptions of paracetamol, topical and parenteral NSAIDs, or opiates. What were the most common combinations of analgesic prescriptions from paracetamol (acetaminophen) to morphine via partial opiates (tramadol, codeine) and NSAID (topical and parenteral)? The date from first prescription to date of surgery was used to examine pathway duration.

2. **Physiotherapy pathway:** Patients that received referrals for physiotherapy consultations. The duration from first referral to surgery date was used to examine pathway duration.

3. **Intermediate surgery:** Patients that underwent intermediate surgical interventions including joint injections and arthroscopy. The duration from first referral to surgery date was used to examine pathway duration.

Pathways were constructed based on the NICE guidance document clinical guideline for care and management of OA (CG 177) [10] and those treatments which are coded in CPRD. It is accepted that a holistic approach to community care may involve a variety of alternative procedures including orthotics, acupuncture, and herbal treatments however these were not factors investigated in this study.

5.4 Materials and methods

5.4.1 Data source

This descriptive population based study summarised the variations in routes of access to orthopaedic care for patients who initially present to their GP with joint pain and who ultimately undergo arthroplasty. To achieve this the study used electronic health records to identify a population comprising all patients coded as having received a joint replacement in the Clinical Practice Research Datalink (CPRD) between 1995 and 2016. CPRD is a research database that covers primary health care records from over 11 million patients across 674 GP surgeries that represent approximately 7% of all patients in the UK population. It is considered to be a representative sample of the UK population and has been validated for epidemiological research in many disciplines including musculoskeletal research [386–388]. Prior to beginning this work a request for access to the CPRD database was made. The study protocol was written with the stated study aims, rationale and analysis plan which was submitted to the independent scientific advisory committee (ISAC). The application was approved and registered (ref: protocol number 17_127R).

5.4.2 Case identification

To identify cases, a data matrix was compiled to identify all variables and exposures for investigation. The index exposure for patient identification was a record of primary total hip or knee replacement. To identify this, a data dictionary

of code lists was compiled using primary care Read codes and the OPCS Classification of Interventions and Procedures version 4 (OPCS-4) and ICD-10 coding systems. These codes are specific to procedures or events recorded in the primary care record or the hospital episode statistics (HES) record. Although they are specific, they do not take into account side of surgery. In the case of subsequent exposures, including physiotherapy referrals and interval surgical procedures, such as arthroscopy or intra-articular steroid injections, they also do not take the site into account. Accordingly, it is necessary to identify the relevant codes for anatomical site and side to then combine the codes for specific exposures with location or site codes. For medication prescriptions, physiotherapy, and intermediate surgical procedures, there is also the further need to identify when multiple events have occurred. The data dictionary was therefore compiled collaboratively with the data administrator such that the data extract contained a complete and comprehensive list of clinical codes that identified all events of interest in a manner that could be analysed to achieve the study aims. A sample OPCS-4 code list for THR identification is shown in appendix D.1.

5.4.3 Analyses

The overall duration of a patient's pathway from first presentation with hip or knee joint pain to date of arthroplasty surgery was calculated. The treatments that patients received during this time, and the total number and relative proportions of these events were used to identify which treatment pathway each patient belonged to. Using descriptive statistics, the average number and length of the

treatments was identified and the proportions, and numbers, of patients following each pathway (i.e. analgesic, physiotherapy, intermediate surgery) was explored to determine whether timing from presentation to arthroplasty differed according to the relevant pathway. Characteristics are reported as frequencies, in percentage terms, for categorical data, means with standard deviations (SD) for normal continuous data and medians with inter quartile range (IQR) for non-normal continuous data. All analyses were performed with STATA MP version 14.2 (Stata Corp., College Station, TX).

5.5 Results

5.5.1 Demographics

In total, a cohort of 72,339 patients were identified from the CPRD database as having undergone THR between 1995 and 2016, of which 44,333 were female and 28,006 were male. Overall mean age at time of THR in the cohort was 69 years (SD 11). When stratified by gender the mean age at THR was 70 years (SD 11) for females and 67 years (SD 11) for males.

5.5.2 Analgesic Pathway

Data were recorded for paracetamol, NSAID, cyclo-oxygenase (COX) inhibitors, partial opiate and full opiate medications. All combinations of medications available from those coded within the data were recorded separately and the commonest five pathways identified (Figure 5-1) showing significant variation in

the proportion of patients following each pathway with increasing use of full opioid medications in recent years. A total of 68,068 (94.1%) patients were prescribed at least one of the classes of analgesic medication consisting of 42,083 females with mean age 70 years (SD 11) and 25,985 males with mean age 67 (SD 11). Over the time period investigated the rates of prescription of different classes of analgesics have changed. Whilst the rates of prescriptions for simple analgesics, including paracetamol, NSAIDs and partial opioid analgesics have remained constant; between 1995 and 2016 the proportion of patients that received a prescription for a full opioid medication more than doubled from 18% in 1995 to 44% in 2016 (Figure 5-2). When considering the number of patients taking each individual analgesic pathway however, there appears to be a marked change over the period of the study. In 1995 47% of patients were treated with a mixture of paracetamol, NSAID, and partial opiate medications, making this the most common combination of treatments and pathway taken to surgery. The next most common of the analgesic pathways was the combination of paracetamol, NSAID, partial opiate and full opiate medications, with another 16% taking this pathway and receiving a full opiate in addition. However, by 2015 this had shifted dramatically to 27% and 30% on those pathways respectively (Figure 5-1).

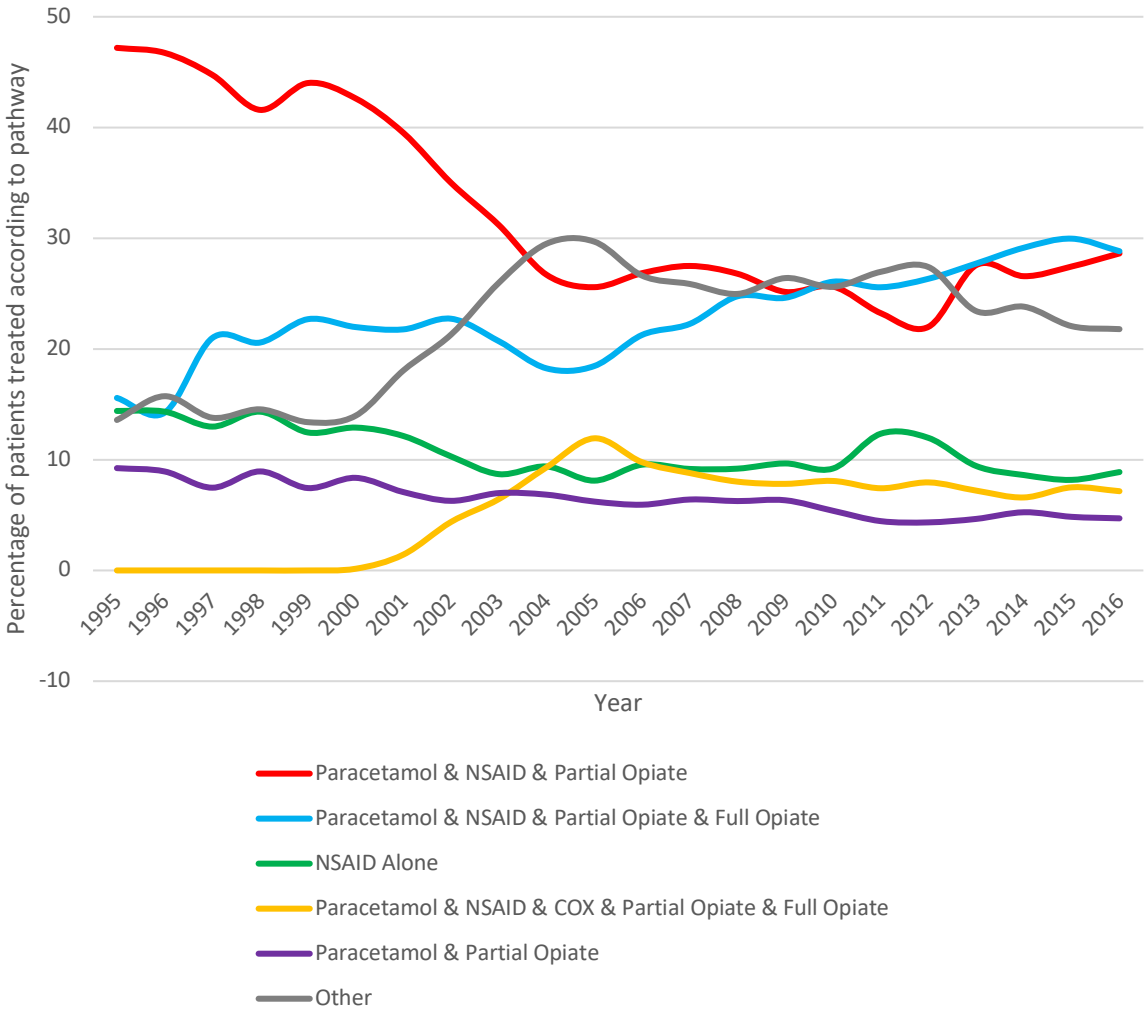


Figure 5-1 - Variation in analgesic pathway over time showing increase in numbers of patients on pathways that include full opiate analgesia.

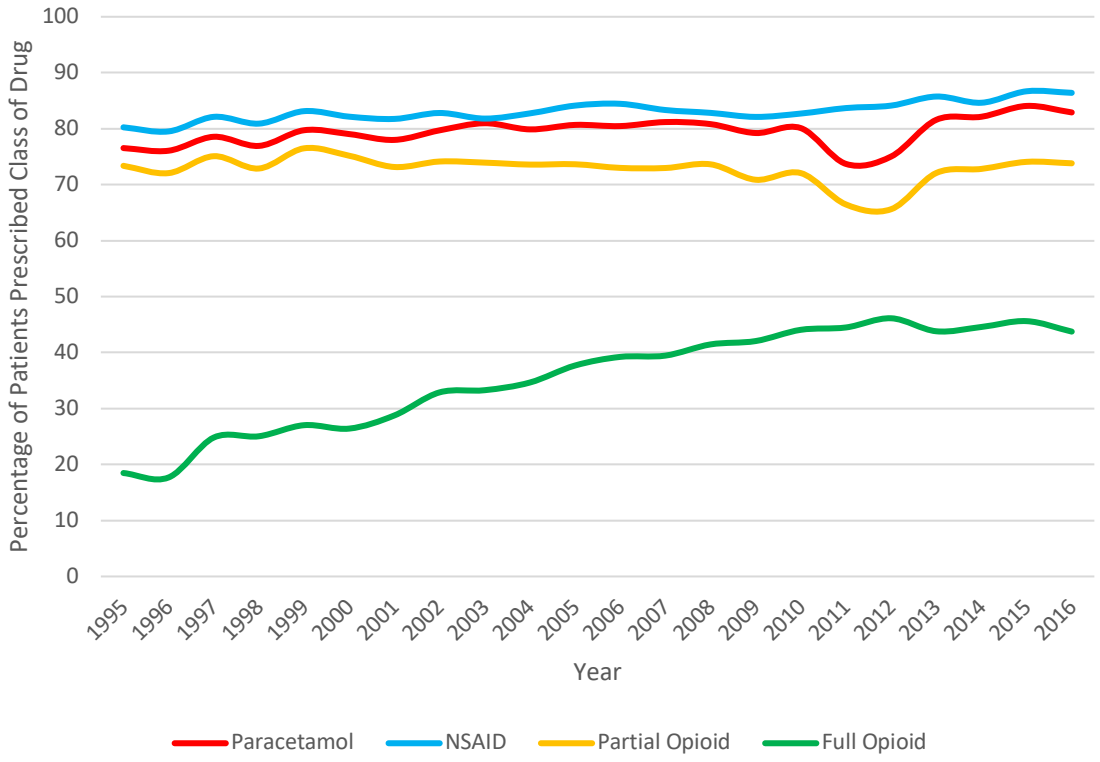


Figure 5-2 - Change in rates of prescription by class of medication over time

The interval between last prescription of medications and surgery was most commonly less than 1 month with 36% of patients receiving a prescription for paracetamol and 35% receiving a prescription for a full opioid medication within the month prior to surgery. Figure 5-3 shows the interval between patients' last prescription for paracetamol, NSAIDs, partial opioid and full opioids in the year prior to THR. When considering the analgesic pathways the cumulative interval from first prescription of a medication included in that pathway to surgery was markedly prolonged for those pathways that included full opiate medication, as compared to the pathways consisting of less potent analgesic medications as shown in Figure 5-4 and Table 5-1.

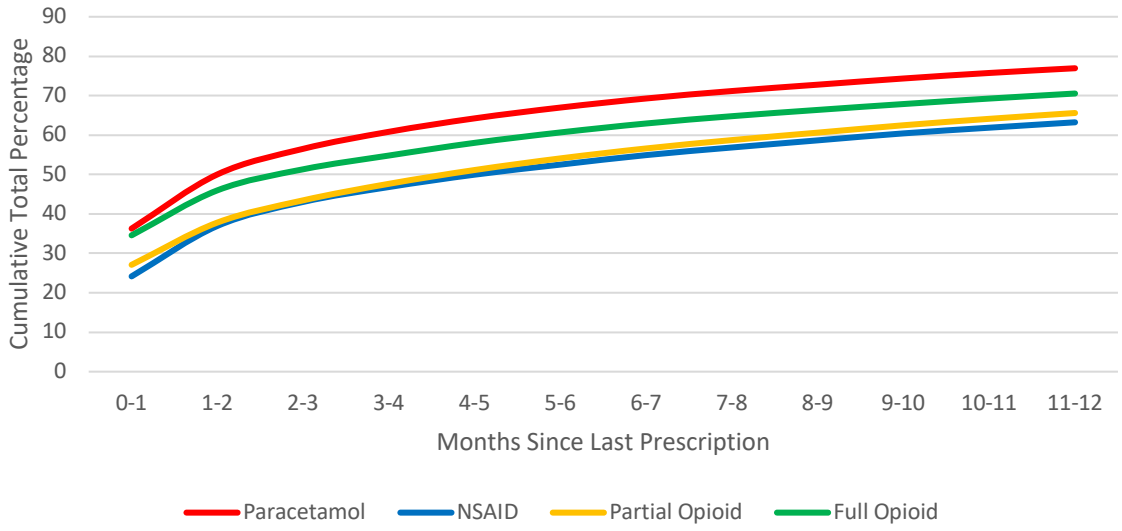


Figure 5-3 - Cumulative interval (months) between last analgesic prescription and THR

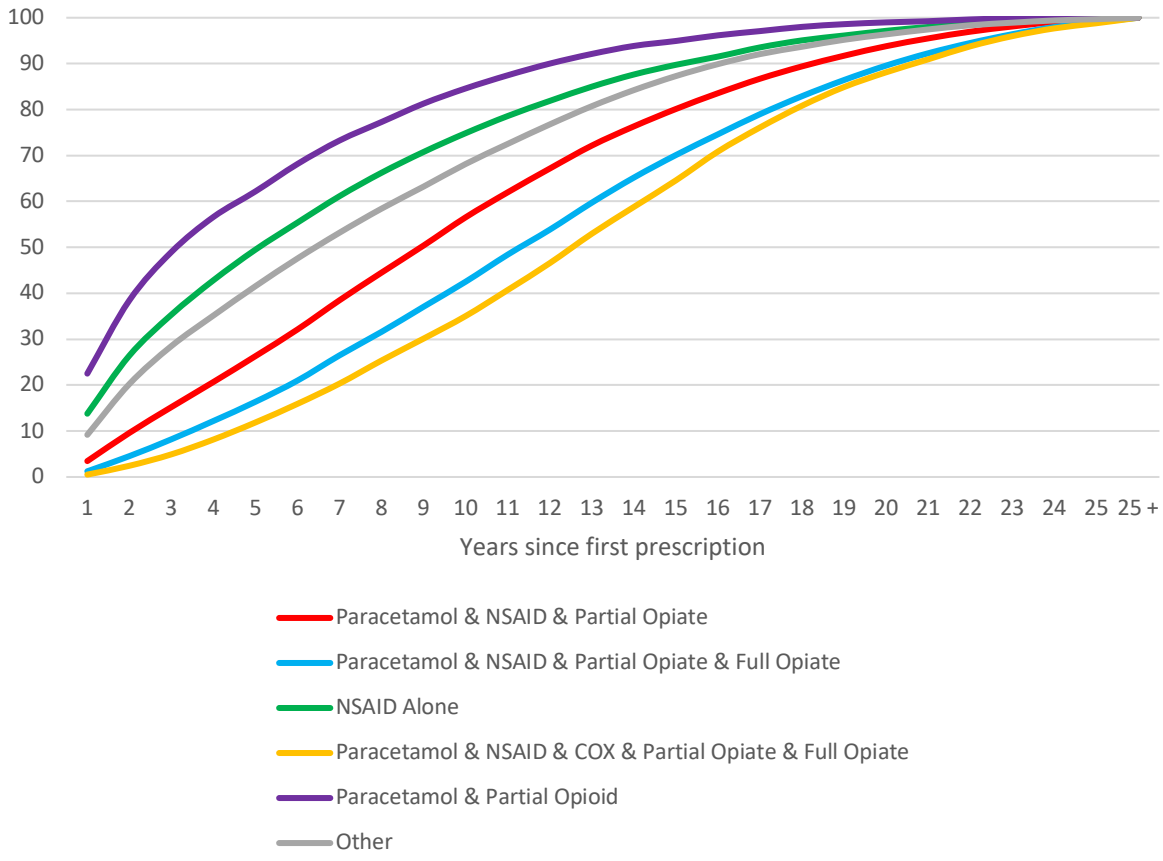


Figure 5-4 - Cumulative interval (years) from first analgesic prescription to THR by pathway

Table 5-1 – Interval (years) between first prescription and eventual surgery by pathway (Mean and SD)

Analgesic Pathway	Mean	SD
Paracetamol & NSAID & Partial Opiate	9.01	5.91
Paracetamol & NSAID & Partial Opiate & Full Opiate	11.37	6.12
NSAID Alone	4.20	5.06
Paracetamol & NSAID & COX & Partial Opiate & Full Opiate	12.38	5.92
Paracetamol & Partial Opioid	3.55	4.43
Other	7.04	5.83

5.5.3 Physiotherapy Pathway

Data were recorded on referral to physiotherapy for 23,028 (32%) patients consisting of 14,984 (65%) females with mean age 69 years (SD 11) and 8,044 (35%) males with mean age 67 years (SD 11) who subsequently underwent THR. Over the duration of the study rates of referral to physiotherapy have risen significantly with fewer than 10% of patients referred in 1995. This proportion climbed to close to 50% over the subsequent 20 years (Figure 5-5).

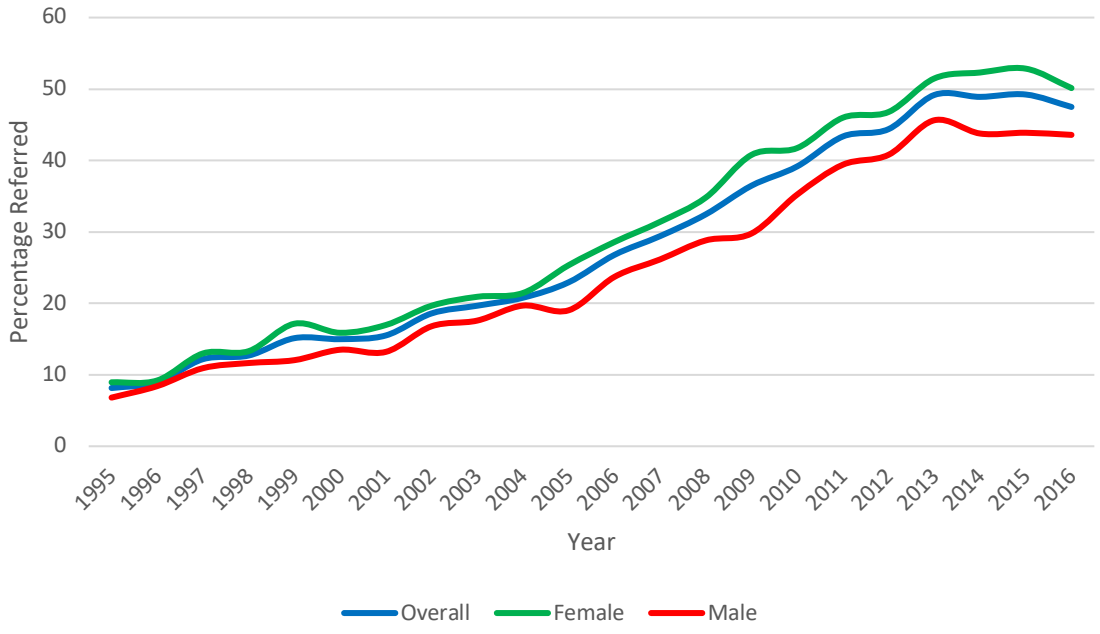


Figure 5-5 - Trend in rate of referral for physiotherapy prior to THR over time

The median number of referrals to physiotherapy was 2 (range 1 – 143, IQR 1 – 3) and median duration of physiotherapy pathway was 3 years to surgery (range 27 days – 24 years, IQR 1 – 6 years) with extended time to surgery shown in Figure 5-6.

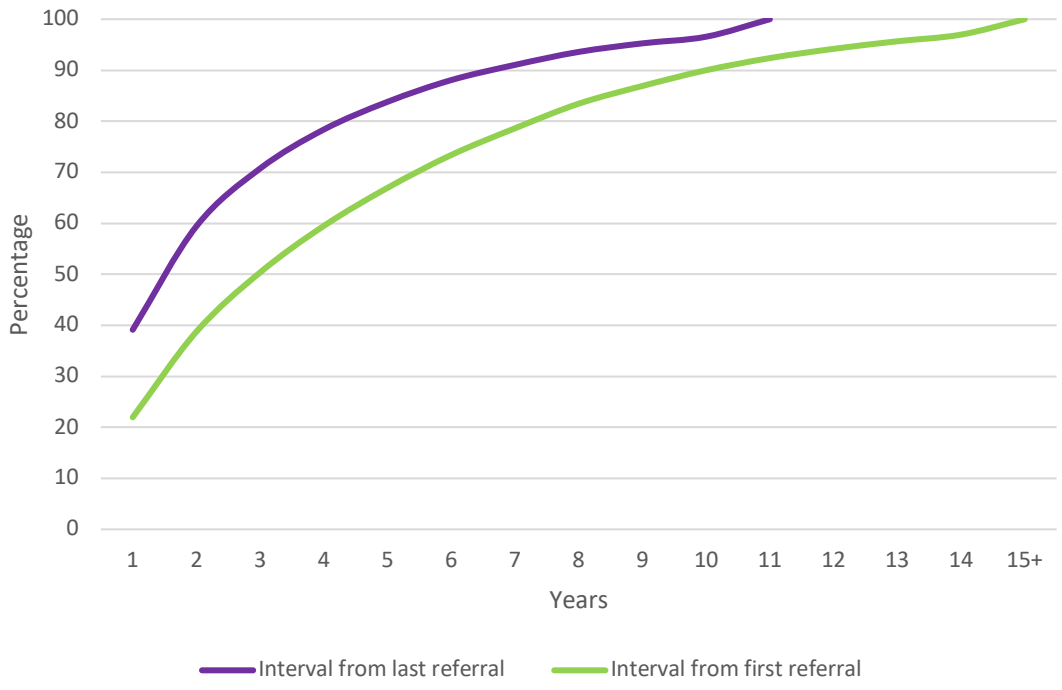


Figure 5-6 – Cumulative interval between first and last referral to physiotherapy and THR

5.5.4 Intermediate Surgery Pathway

Only 78 patients that underwent THR were referred for hip arthroscopy during the time period under investigation in this study making it impossible to draw any conclusions from them as a specific pathway. Of the patients that underwent THR, 11,324 (16%) received intra-articular steroid injections at some stage prior to surgery. These consisted of 7,321 females with mean age 71 years (SD 10) and 4,003 males with mean age 69 years (SD 10). The proportion of patients receiving intra-

articular steroid injections prior to THR increased from 8% in 1995 to 18% in 2016 (Figure 5-7).

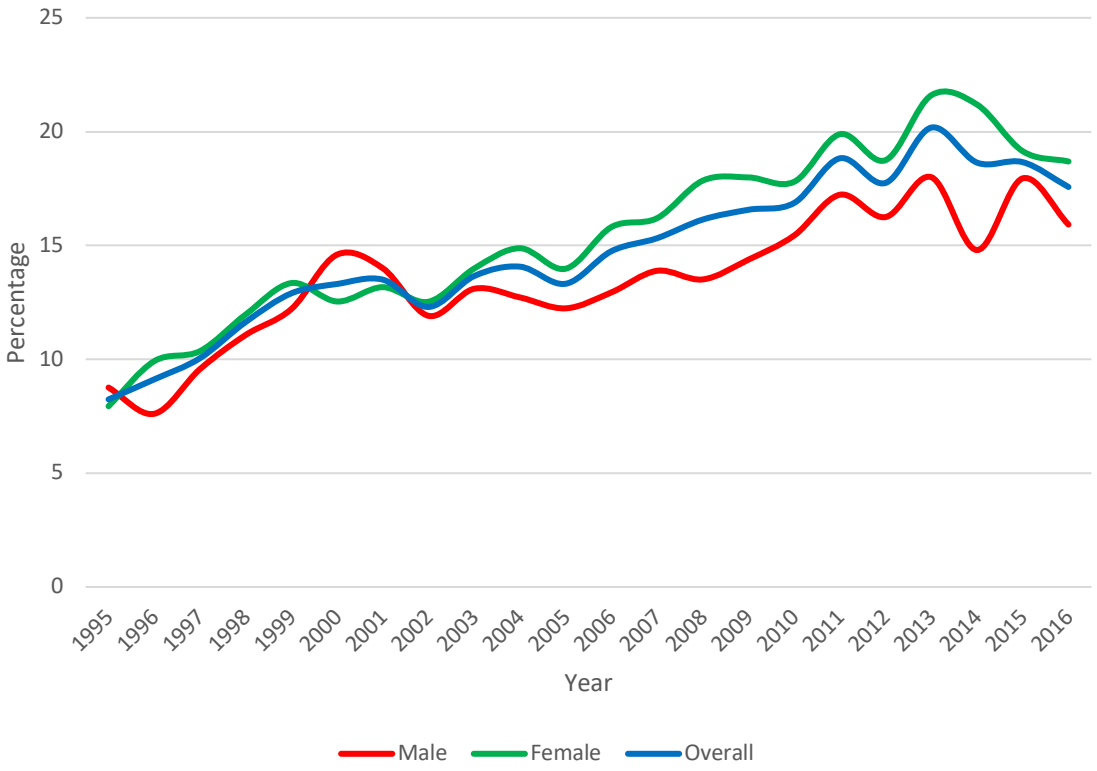


Figure 5-7 - Trend in referral for steroid injection over time

The median number of injections was 1 (range 1 – 52, IQR 1 – 3) and median duration of intermediate surgery pathway was 5 years to surgery (range 1 day – 29 years, IQR 2 – 10 years) with extended time to surgery shown in Figure 5-8.

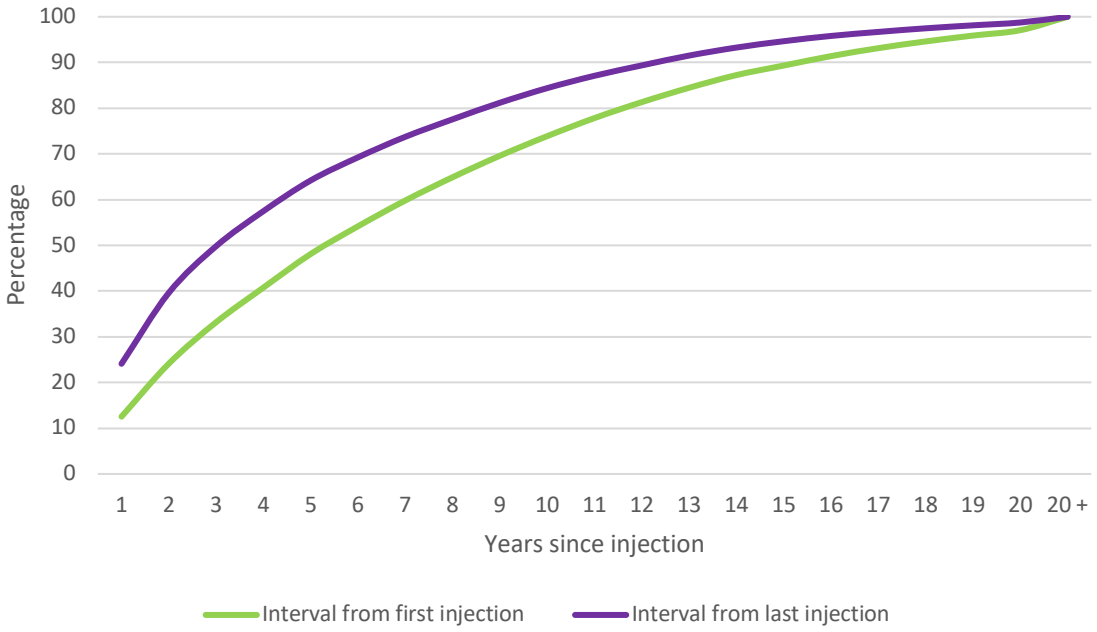


Figure 5-8 - Interval between first and last intra-articular steroid injection and THR

5.5.5 Inequalities in access to care

Over the period of the study the rates of prescription of full opioid analgesia rose. The sharpest rise was seen in those with higher BMI and of most deprived status. BMI was stratified in accordance with the WHO groups and index of multiple deprivation (IMD) combined into 5 groups where group 1 represents the least deprived and group 5 represents the most deprived. Rates of prescription for full opioid analgesia increased with increasing BMI and deprivation (Figure 5-9 & Figure 5-10). Individuals with BMI over 35 and the most deprived were most likely to receive a prescription for full opioid analgesia. Figure 5-11 highlights this discrepancy between prescription rates for the most and least deprived patients in this cohort and the appearances suggest the trend has widened over time.

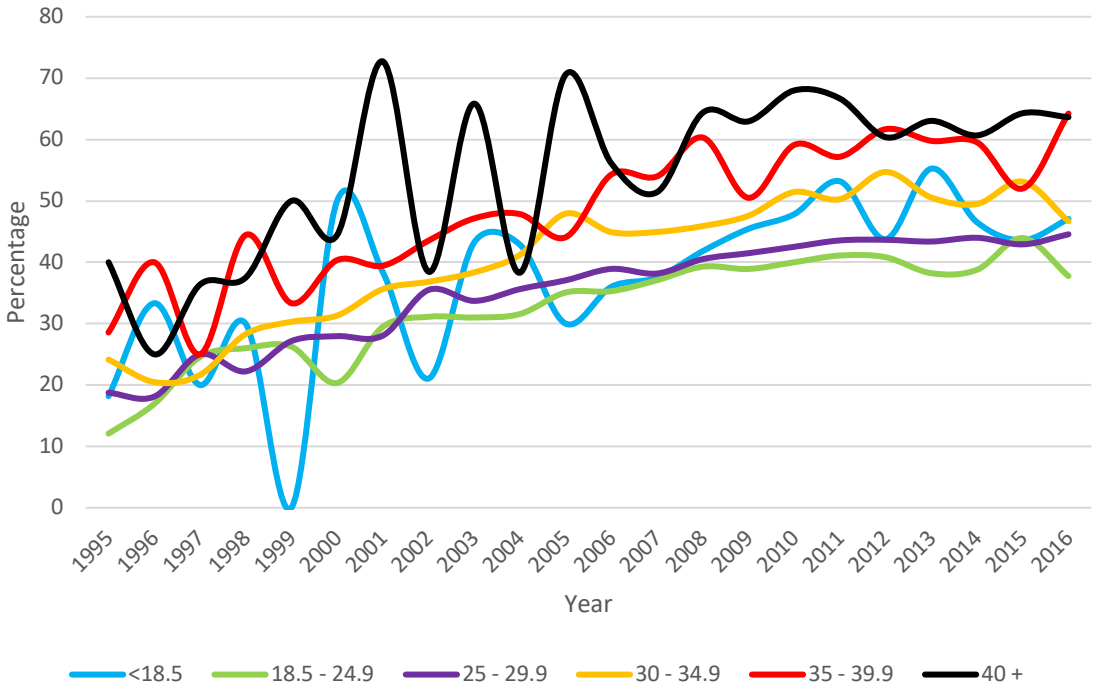


Figure 5-9 - Trend in rates of full opioid prescription over time by WHO BMI group

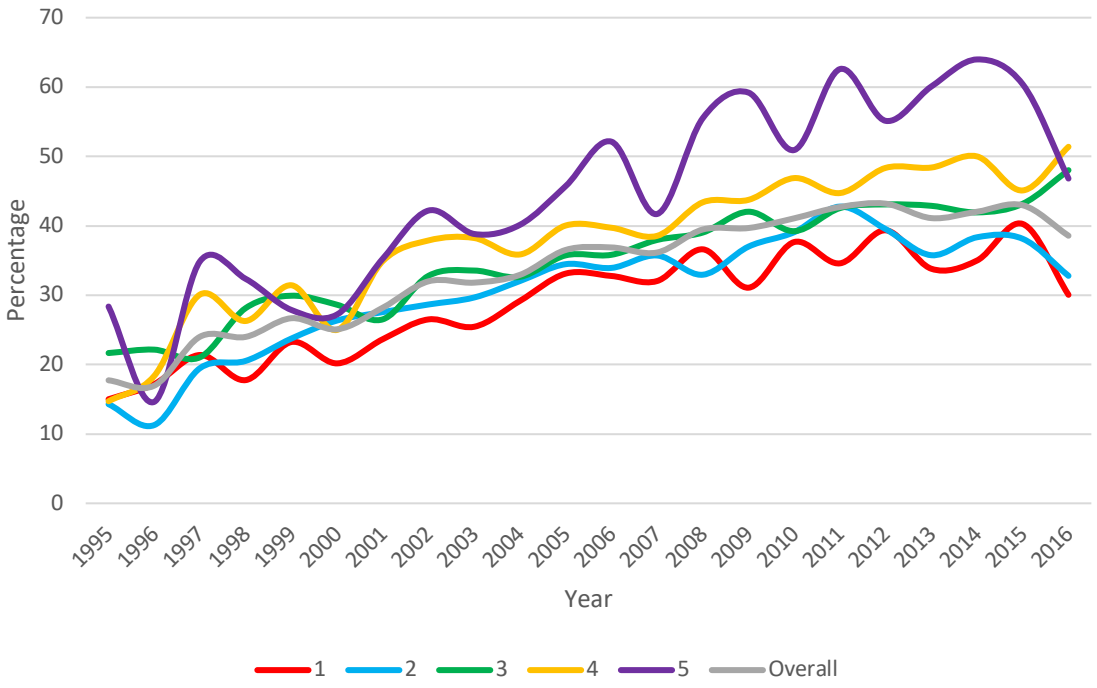


Figure 5-10 - Trend in rates of full opioid prescription over time by IMD class

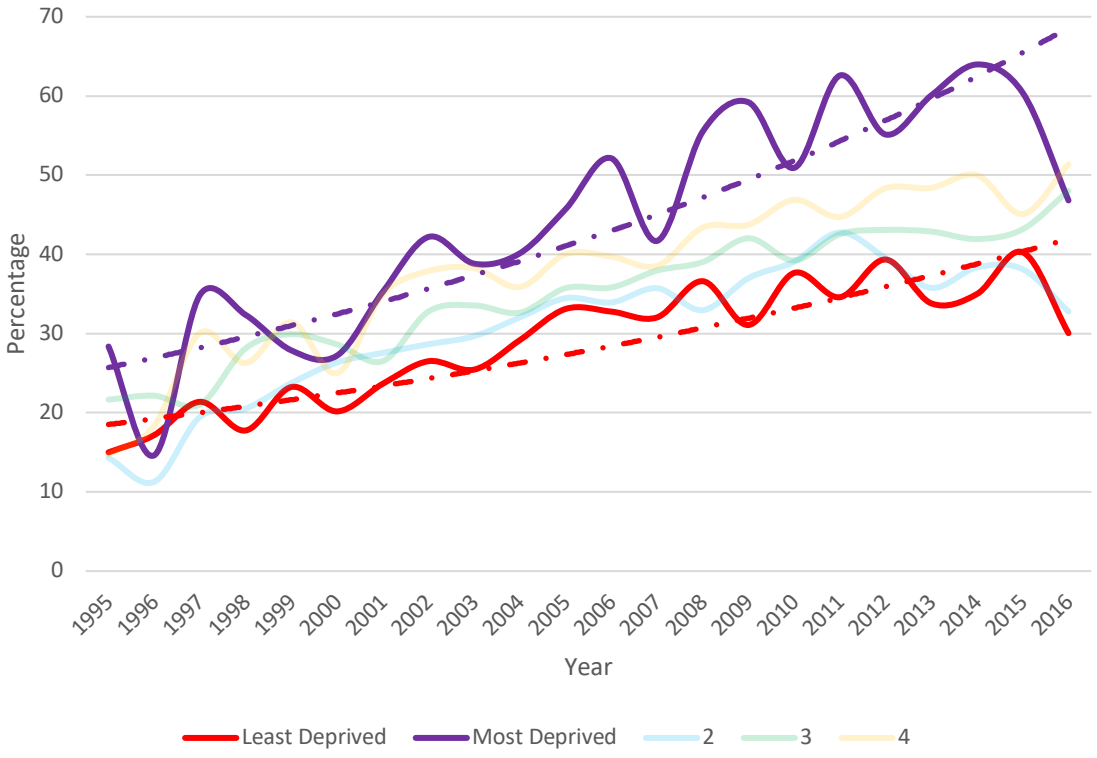


Figure 5-11 - Trend in rates of full opioid prescription highlighting the most and least deprived patient groups over time by IMD class

The mean age of patients undergoing intra-articular steroid injections rose over the duration of the study from 68 years (SD 10) in 1995 to 71 years (SD 10) in 2016 whilst the mean age of patients referred for physiotherapy remained stable from 68 (SD 13) years in 1995 to 68 (SD 11) in 2016. When age was stratified into groups, over time there emerges a pattern showing lower rates of intra-articular steroid injection in the younger age groups, in particular those under 50 years of age (Figure 5-12).

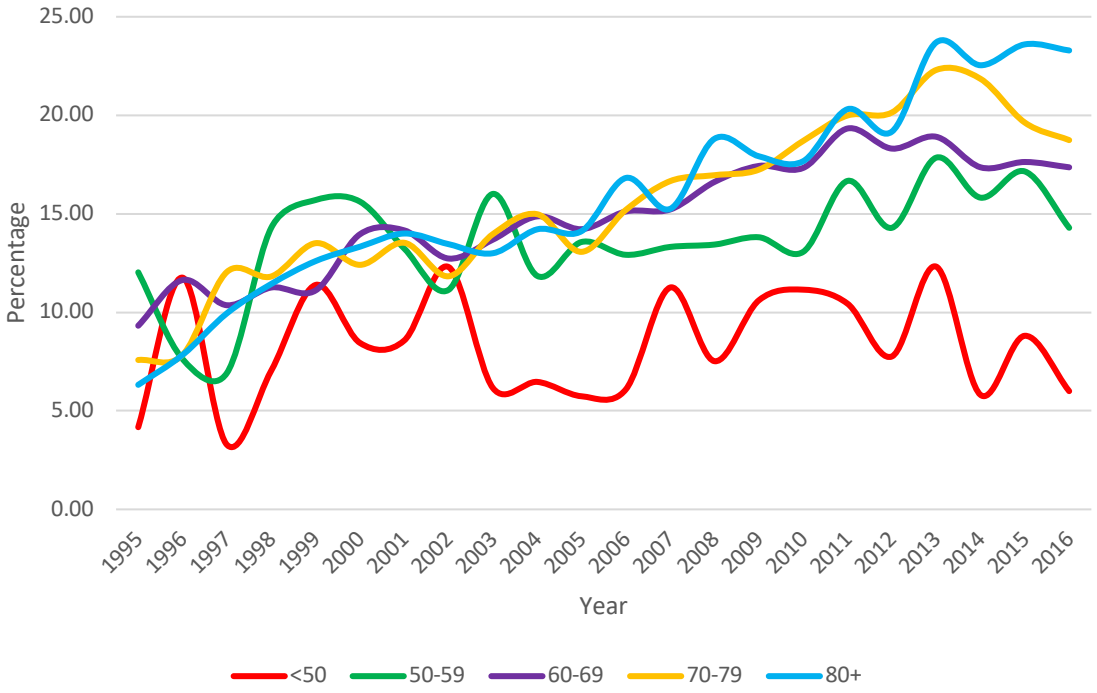


Figure 5-12 - Trend in stratified age of patients referred for intra-articular steroid injection over time

The mean BMI of patients referred to physiotherapy rose during the period of the study from 26 (SD 4) in 1995 to 28 (SD 5) in 2016 whilst the mean BMI of patients undergoing intra-articular steroid injections remained stable at 28 between 1995 (SD 4) and 2016 (SD 5). Figure 5-13 shows the change in BMI between 1995 and 2016 overall and individually for the physio and intermediate surgery (injection) pathways.

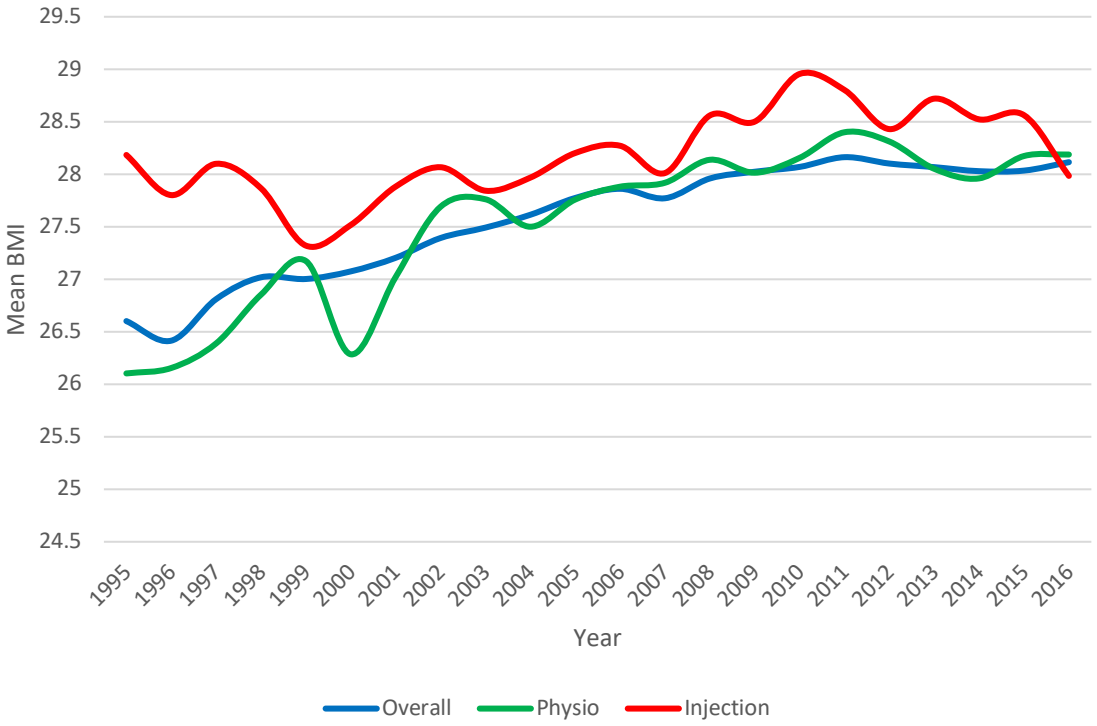


Figure 5-13 - Trend in BMI of patients undergoing THR over time

The rates of referral for physiotherapy (Figure 5-14 & Figure 5-15) and intra-articular steroid injection (Figure 5-16 & Figure 5-17) did not vary by BMI group or IMD group over the period of the study with annualised rates reflecting the overall referral rates. Greater variation was seen across years in the classes at the extremes of BMI, where fewer patient numbers were seen. However, the overall trend was seen to rise in line with the overall trend.

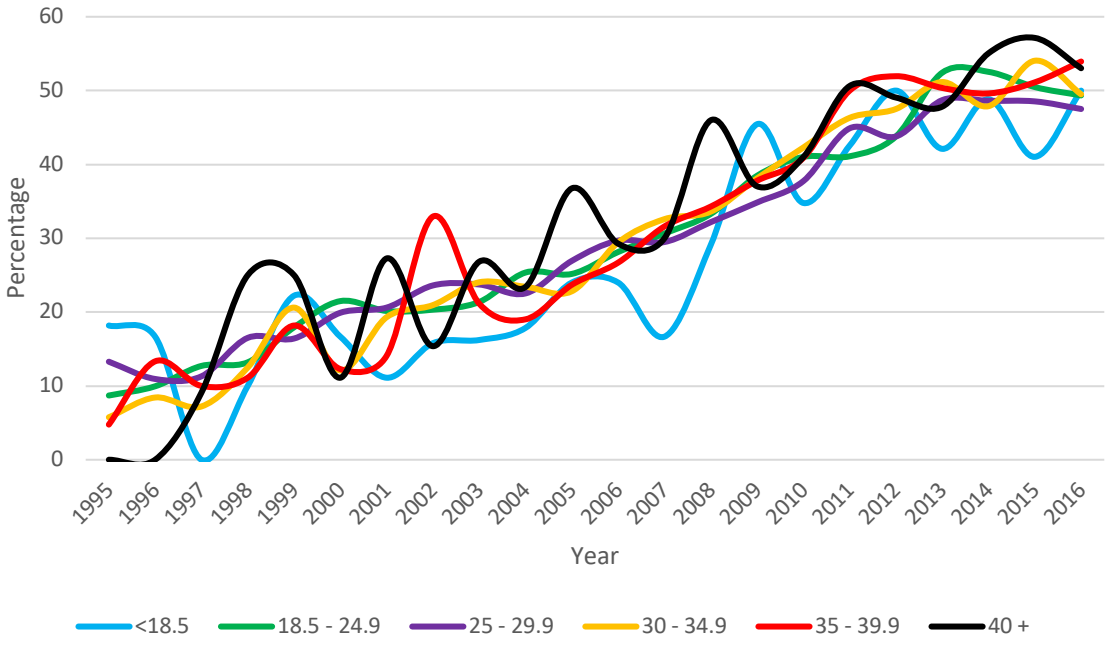


Figure 5-14 - Percentage of patients referred for physiotherapy by WHO BMI groups

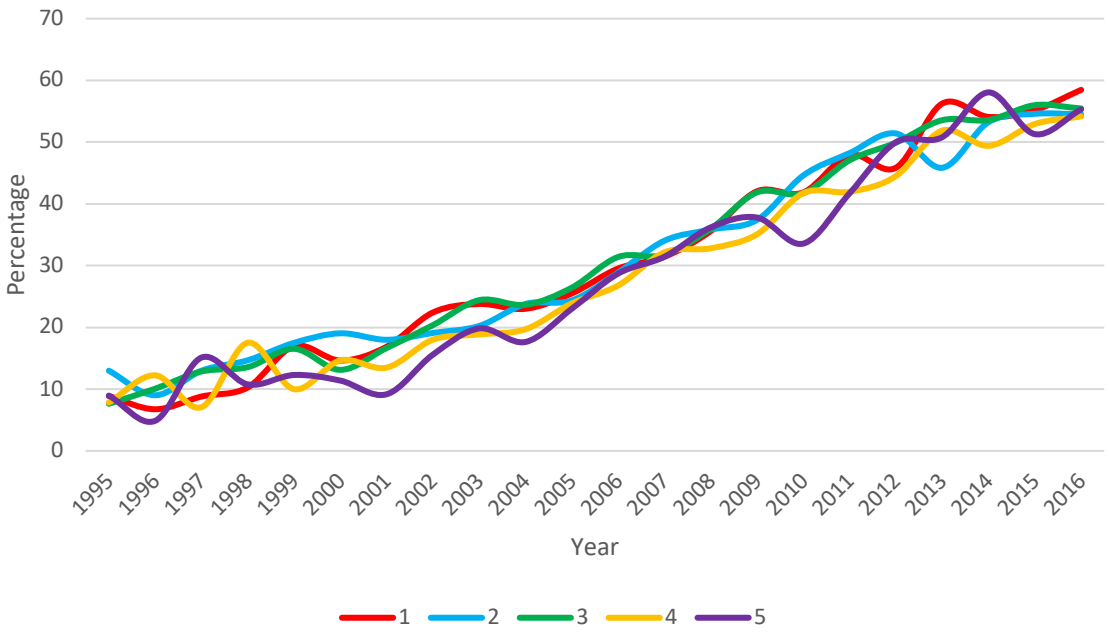


Figure 5-15 - Percentage of patients referred for physiotherapy by IMD class

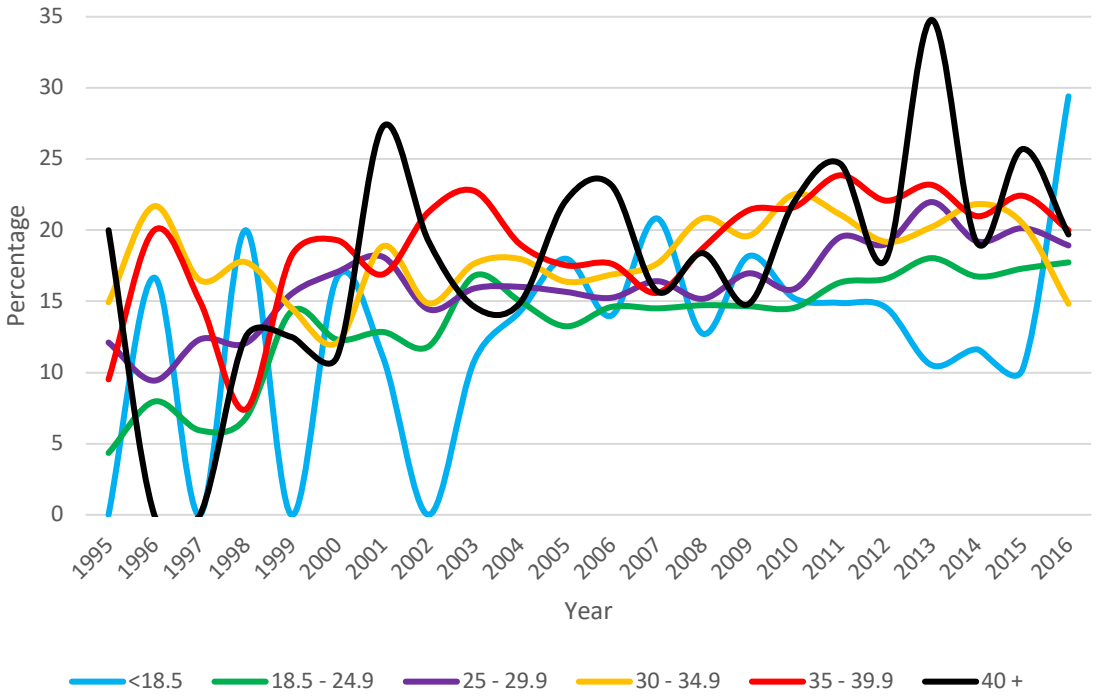


Figure 5-16 - Percentage of patients referred for steroid injection by WHO BMI group

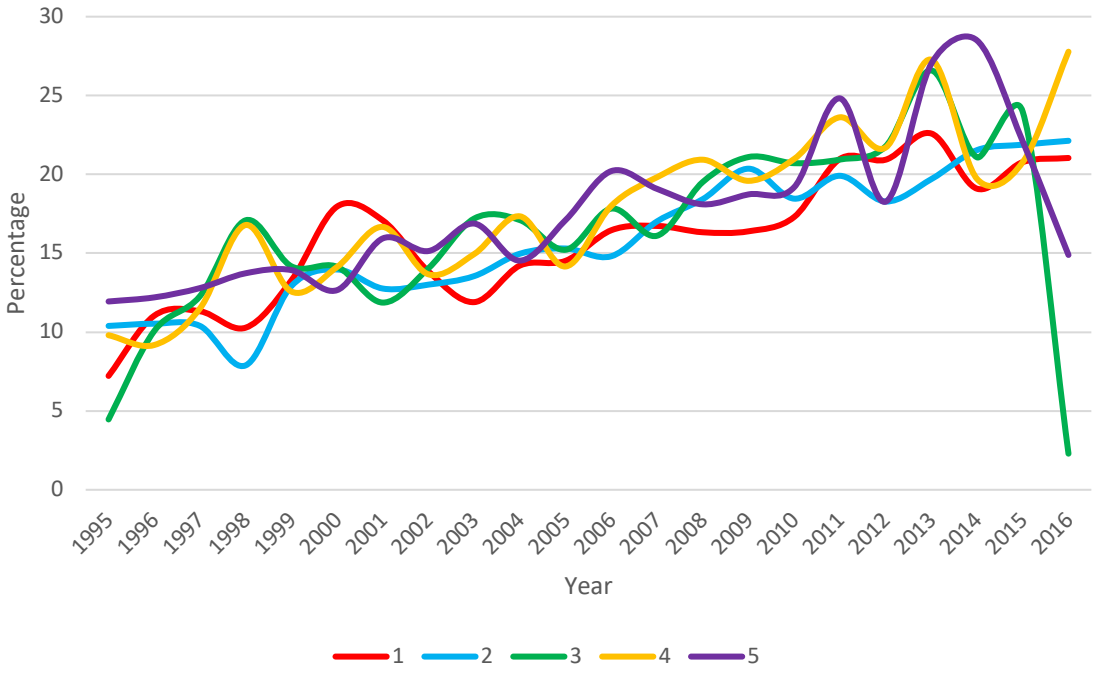


Figure 5-17 - Percentage of patients referred for physiotherapy by IMD class

5.6 Discussion

The work in this chapter aimed to identify the pathways taken by patients on their journey from presentation in primary care with the earliest features of OA until their eventual arthroplasty, and to identify reasons for variations in access to that definitive care. Analgesics, physiotherapy, and intermediate surgical treatment pathways were investigated and variations in the pathway taken, according to patient factors including age, sex, social deprivation and BMI were considered.

The analyses and results presented in this chapter demonstrate that there has been significant change over the study period in the volume of patients receiving joint injections, pain medications, and physiotherapy referrals for joint pain as a result of OA that goes on to be treated with arthroplasty. This is likely to be due to a wide variety of factors but, notably, the marked increase in the use of full opioid medication demonstrated a shift in the primary treatment of OA pain. This study not only demonstrated an overall increase in the use of full opioids but also that this is most common after they have already been prescribed paracetamol and NSAIDs, suggesting an escalation of analgesic prescription prior to receiving surgery. Furthermore, these data suggest that patients who go on to receive prescriptions for full opioid analgesics have a longer interval from first prescription to surgery indicating a delay to surgery. These are concerning findings since full opioid analgesia has been shown to be ineffective in OA. The high use of opioids has been identified in a Swedish population based cohort study showing that one out of every four OA patients receives a prescription for opioids

over a 12 month period [389] whilst an epidemiological study in the United States of America demonstrated that between 2007 & 2014 13.4% of patients with hip OA received a prescription for opioid medication [390]. More recently a randomised trial investigating the efficacy of opioid vs. non-opioid medications for the treatment of hip or knee OA demonstrated no improvement in pain related function over a 12 month study period [391]. Despite the growing literature demonstrating limited effectiveness in alleviating OA pain, added to high rates of dependency and morbidity, many commonly used leading clinical guidelines continue to recommend the use of opioid analgesics [392].

The results of this study appear to show that patients with higher BMI are more likely to be prescribed opioid analgesia and, as a consequence, are more likely to experience a delay to surgery. This appears to demonstrate inequality in access to care in this data over patients in lower BMI groups. This is a concern because although higher BMI has been associated with persistent knee and hip pain [393] and is an established risk factor for the development and progression of OA [394] it has also been shown that patients with BMI greater than 30 require arthroplasty at a younger age than those in lower BMI groups [395]. One possible interpretation of these results is that in addition to the link shown in the evidence between BMI and joint pain in OA, there may also be a dose response in the interaction between BMI and OA pain meaning that increasing BMI may require greater analgesic use. An alternative interpretation, however, may be that patients with higher BMI are more likely to be discouraged from surgery or be systematically denied surgery

due to service rationing using BMI cutoffs. These methods have been shown to be ineffective in isolation [396,397] and, when implemented, mean patients with higher BMI experience longer periods of analgesic use prior to surgery with these patients seeing an escalation in pain medication to opioids. Over the duration of this study national guidance on the care and management of OA evolved. A notable change came between the 2008 (CG59) [398] and 2014 (CG177) [10] editions of the NICE guidance at which time guidance was amended to remove recommendation for glucosamine supplements, and hyaluronan intra-articular injections, due to a lack of evidence supporting their efficacy. However, over time there has been very little change to the guidance on pharmacological management, in particular the analgesic options suggested. A holistic approach to care, beginning with a wider assessment of the patients' health, and progressing of management from lifestyle modification through medication and adjuncts prior to referral to surgery has remained the core principle of care.

The results of this study also appear to suggest that patients from lower socio-economic groups are more likely to be prescribed full opioid analgesia. This, allied to the findings that patients receiving full opioids are more likely to experience a longer interval before surgery, suggests that there may be socioeconomic inequalities in access to orthopaedic care. This is supported by literature evidence that deprived areas have higher rates of OA [399]. There is a greater need for surgery in these groups, with this need magnified by patient level factors including obesity [400]. This may also be related to psychological and social factors which

are not simply secondary aspects of persistent pain, rather they interact with the pain due to impacts on relationships, work and wider health [401].

Over time there has been a progression in the number of patients receiving intra-articular steroid injections in the hip and being referred to physiotherapy prior to THR, which is in line with NICE guidance [10] and represents a response to the evidence base. However, this is tempered by the observation that the rise in injections is not seen in the younger age group, in whom the rates of injection have remained largely stable over the study period, suggesting further inequality with respect to injections, since no age effect was seen in referral to physiotherapy or analgesic prescriptions. Alternatively, it remains possible that intra-articular injections are less effective in patients in the younger age group, however this was beyond the scope of this study. It is also worth noting that the overall proportion of patients receiving both physiotherapy and steroid injections remains relatively low overall with only approximately half of patients being referred for physiotherapy by the end of the study period.

The data shows that only 78 patients in this cohort underwent hip arthroscopy prior to THR. This represents an interesting observation due to the contrasting literature surrounding knee arthroscopy [402]. It should be noted that hip arthroscopy is a significantly less common procedure than knee arthroscopy due to its technical challenges and the fact that it is not commissioned in all regions. The relative rarity of the procedure, particularly prior to 2010, may account for this. Earlier work in our group has shown that between 2002 & 2013 only 11,329

hip arthroscopies took place in England over the whole study, with half of those procedures after 2010 [403].

The primary limitation of this study is that where medication prescriptions, and physiotherapy referrals, were made and recorded the compliance with these measures cannot be ascertained from the CPRD data. It must also be observed that whether the medication prescriptions were completed, or whether patients attended physiotherapy, the subsequent compliance of patients with the regimen could not be confirmed. However, CPRD has been extensively used for research of this nature and it has been shown to have highly complete prescription data which very closely represents the rates of prescriptions actually provided to patients [404]. Furthermore, when considering that patients have attended their GP seeking interventions to treat symptoms that ultimately resulted in arthroplasty surgery it follows that they would comply with less significant treatments offered beforehand.

In summary, the work presented in this chapter suggests that there are significant variations in routes to access orthopaedic care for patients with OA. In particular there appears to be a concerning rise in the use of opioid analgesia in OA which may be having the most pronounced effect on those patients with greatest need. There appears to be wide variation in the nature, frequency and duration of the treatments offered, as well as inequalities based upon patient level factors including age, BMI and socioeconomic status.

Chapter 6 Thesis Discussion

The aim of this thesis was to examine the biological signature of different disease states of OA in synovial fluid (SF) from the hip and knee of patients with different clinical states of hip OA. This was performed using a novel, broad spectrum, proteomic platform called the SOMAscan to identify the SF signatures of the different disease states.

The research contained in this thesis was motivated by the need for measurable markers of OA that have the capacity not only to detect OA, but also to aid in the discrimination between different disease states. This gap in the current understanding means that disease modifying OA medications and treatments cannot be developed without tools that can be used to monitor disease progression and serve as outcomes for future clinical studies.

The complexity of OA, and its multi-factorial aetiology, highlights that it has become clear that the traditional concept of a biomarker is not a deliverable goal. A single marker that can be used to diagnose the disease with response to any subsequent treatments, then titrated against the serial measurement of the same marker to gauge efficacy, cannot be expected in OA.

In order to explore the role of biomarkers in detecting OA and discriminating between the different states of disease, it is important to identify how any such future output of these experiments could ever be introduced into clinical practice.

A complete analysis of the nature, and the duration, of the patient journey, from onset of symptoms through to eventual treatment, will also be required to provide the necessary context upon which the utilisation of any such future tool can be directed. Without this context we cannot identify how, when, or where novel tools and treatments must be targeted. This required an epidemiological assessment of the current patient journey from presentation to primary care through to eventual orthopaedic care, in order to identify the barriers that exist in identifying patients at early disease states.

Brief summary discussions of the studies of the chapters of this thesis, including their main findings and implications, are presented below.

6.1 Chapter 2 – Validation of SF for SOMAscan

Whilst this chapter is not strictly an experimental chapter it contains a critically important and necessary step in the genesis of this thesis. SF has not previously been used on the SOMAscan due to its highly viscous nature. Accordingly, before completing the analysis of the SOMAscan data, it was important to validate that the SF matrix was compatible. To this end all SF samples were treated with HAse prior to analysis. Subsequently six markers were chosen to assay, in house, using ELISA and MSD based upon findings from previous experiments completed within our group (Jayadev et al. 2020 [147]). Assay data was compared against the SOMAscan data to determine if significant differences existed between the performance of these two assay methods and the SOMAscan data output. These

included ADAMTS-4 (ELISA) and TIMP-1, MCP-1, IP-10, TGF-β1, IL-6 (MSD). The assays of five of these markers performed well returning good data, however the TGF-β1 assay failed to provide any usable information. The assays of TIMP-1, MCP-1, IP-10 and IL-6 showed good correlation with the SOMAscan whilst the ADAMTS-4 did not, although there was a trend towards a correlation in the hip group samples. There were appearances of different dynamic range between the hip and knee group samples. Although they were treated according to the same protocol there appears to be a batch difference between the two cohorts. However, this finding does not affect the analysis since no comparisons between joints were undertaken. This work has, for the first time, directly validated the SOMAscan platform for the analysis of HAse digested SF against traditional laboratory bench assays. To examine SF to its maximum potential and gain the greater insight that the broad proteomic platform of SOMAscan can deliver, it was critical to validate the platform with this fluid and its particular matrix properties.

6.2 Chapter 3 – Knee SF SOMAscan Analysis

Advances in broad scope proteomics platforms, such as the SOMAscan, have permitted the analysis of large numbers of proteins using only small amounts of sample material. The work in this chapter aimed to identify SF biomarker fingerprints of different clinical and structural disease states of knee OA groups using multivariate analyses of SOMAscan with particular focus on the early OA group. Accordingly, multivariate techniques were required that were developed

for use in situations where there are numerous "input" variables that may describe changes in the "output" variables, in this case markers, that may contribute to class separation.

The groups in this study are necessarily small and it is accepted that this is the primary limitation of this work. However, this limitation is a direct consequence of the extremely high cost of this platform running at over £1000 per sample. The strength in the penalistic nature of the LASSO analysis supporting the PLS-DA and PCA methods can give reassurance, meaning the view of the message should not be diminished. This chapter has harnessed the opportunity to accelerate biomarker discovery and use the targets identified to perform protein-protein interaction network analysis, rather than striving for a single protein. The data suggest knee injury and late OA groups appeared to be distinct groups. There exist robust cases in the literature to support the potential role of the markers identified in the associated disease processes. It was surprising to see the lack of separation between early OA and injury groups. This is most likely to be due to the small group sizes under investigation which were necessary for this exploratory study and differences may be detectable with large group sizes. Future follow up studies will clarify whether the early group constitute a chronic injury pattern with some repair potential or if the injury group are, in fact, already committed to an end stage pathway. Notwithstanding this, the identification of markers that identify the early OA group as distinct from the end stage group will generate a means of selection of candidates for clinical trials of early interventions.

Recent work on knee joint distraction in patients with radiographic knee OA has shown longitudinal change in SF concentrations of activin A, TGF-β1, FGF-2 and MCP-1 were associated with PROM score and for all except activin A, the magnitude of change correlated positively with the magnitude of PROM score improvement [405]. SF analysis of a knee injury cohort with samples taken within 8 weeks after injury has also shown elevated IL-6, MCP-1, MMP-3, TIMP-1, activin A and TSG-6 with IL-6, TIMP-1 and TSG-6 associated with poorer PROM scores [261]. The marker biology identified supports the findings in this chapter and this evidence suggests that knee injury, and OA, are bioactive disease states which are modifiable, and demonstrate short term changes in measurable biology over time. These approaches demonstrate that with bigger studies on larger cohorts and using multiplexed approaches there are potential therapeutic options that may be viable treatments prior to arthroplasty if the correct patients who might benefit can be identified. Randomised trials with biological end points may be possible with such data.

The findings in this chapter show that multiplexed assay techniques with multivariate statistical analyses can discriminate between different disease states of OA in the knee. Of interest they have suggested there may be biological similarity between clinically distinct injury, and early OA, knee cohorts and that there appear to be discernible differences between injury and late OA cohorts. This work further supports the value of large-scale proteomic approaches and specifically demonstrate the utility of the SOMAscan for anlaysis of SF.

6.3 Chapter 4 – Hip SF SOMAscan Analysis

This chapter aimed to identify SF biomarker fingerprints of different clinical and structural disease states of hip OA groups using multivariate analyses of SOMAscan. However, before the analysis was possible it was necessary to devise an entirely new method to acquire the SF samples from the early hip OA group. The hip joint does not give up its fluid with ease and hence a cohort of human early hip OA SF has not previously been collected. However, this challenge was met, and a bespoke system was optimised to identify SF under direct vision at hip arthroscopy and achieve SF collection simultaneously prior to surgery. The low volumes yielded benefitted from the wide spectrum proteomic analysis of the SOMAscan.

The findings of this study suggest that biological differences appear to exist between early and late hip OA as well as suggesting hip OA may be biologically different to knee OA. In both the knee and hip studies, the majority of markers described the late group. The most descriptive markers in the knee were from complement, and coagulation, cascade pathways, vesicle pathways, and response to wounding pathways. Aggrecan, which is implicated in ECM organisation, and released in degradation, was seen to be elevated in SF from injury group patients in the knee study. However, it was a weak loader that was not selected by LASSO analysis and was not seen in the hip study. The findings of the hip study suggested some novel proteins that may help deliver insights into the underlying biological processes in hip OA including S100A7 which has an established role in response

to wounding and immune modulation as well as IDS which is known to be required for degradation of heparan sulfate, a key ECM component. Network analysis further showed enriched proinflammatory, cartilage degradation and ECM signaling pathways. The proteomic profiles show differences between hip and knee OA in these cohorts. A greater proportion of proinflammatory cytokines were seen in the hip whilst there were fewer coagulation and complement cascade proteins when compared to the knee.

Again, due to the extremely high cost of the analysis platform, this study was limited by the low sample number. Whilst this is accepted it is again important to consider that the strongly penalistic nature of the analysis mitigates this risk since the markers identified by the PLS-DA regression models are supported through the agreement with the LASSO analysis. It should not be forgotten that SF is the best source of joint disease information due to its proximity to the bioactivity, and can be considered analogous to CSF, which also bathes the diseased tissue in diseases such as Alzheimer's, where biomarker discovery projects using the SOMAscan have been successfully undertaken with similarly small group sizes [406] as compared with studies using serum samples requiring groups that are an order of magnitude larger [407].

This study has, for the first time, described a human early OA hip SF cohort using the SOMAscan and further provided a method of collection of undiluted SF. This, allied to the proof of utility of the SOMAscan for SF analysis, can be used to explore the underlying biopathogenesis using multivariate analysis techniques. It has

delivered an insight into the pathogenesis, from early to late stage hip OA, and has suggested potentially viable targets for future study in larger, longitudinal cohorts using multiplex analysis techniques. In contrast to the knee, this study has identified pathways enriched for cartilage degradation and ECM signaling as well as suggesting this may be mediated by MAPK and ephrin activity.

6.4 Chapter 5 – Variations in Treatment Pathways

This epidemiological study aimed to examine the variations and inequalities in treatment pathways, from presentation with joint pain to eventual joint replacement and classify the route taken in terms of NICE recommended non-surgical treatments: analgesia, physiotherapy, and intermediate surgical treatments. This was conducted using CPRD data covering over 20 years of UK primary care episodes. The findings demonstrated that there has been a significant increase in the number of patient interactions over the study period together with a marked increase in the use of full opioid medication whilst the use of "simple analgesics" including paracetamol and NSAIDs remained stable, suggesting an evolution over time in the management of pain resulting from OA towards far higher rates of full opioid prescriptions, potentially due to ongoing increases in the number of patients requiring THR. In any event this trend is in contrast to a growing literature showing no value in the use of such medications in OA. The work also showed variations in steroid injections by age and higher rates of opioid prescriptions in patients of low socioeconomic status and with higher BMI. The

work presented in this chapter suggests that there are variations in access to orthopaedic care in OA. In particular the findings appear to suggest a concerning rise in the use of opioid analgesia in OA which may be adversely affecting those patients with greatest need most.

In order to influence the future economic, and healthcare, burden of OA and the eventual need for arthroplasty a cohesive means of managing patients will be required. This will begin in the community with standardised pathways to accessing care. This will permit research studies to access patients that can be clinically stratified into defined groups. Those patients identified at an early disease state will be the target cohorts for exploratory biomarker studies. Broad spectrum, multiplexed approaches such as the SOMAscan will be the ideal tools to gain the greatest biological insights. These cohorts may then be monitored, longitudinally, after their biological appearances have been defined, to identify the features that predict progression, and subsequently, those that predict rapid progression.

6.5 Clinical Relevance

This thesis brings together a group of studies that investigate a gap in the current understanding of how we, as a scientific and medical community, diagnose, monitor, and ultimately understand OA. The results of this work have suggested that there may be biological differences between the underlying processes seen to be activated in OA of the hip and the knee. It has, for the first time, been able to

investigate an early hip OA cohort using SF which required an entirely new sample collection method. This work has validated the SOMAscan, a new platform in the analysis of SF, and harnessed it to conduct analyses that has suggested early and late hip OA may be biologically distinct disease entities and that the SF phenotype appears to be different to that seen in the knee. It has shown that there remain significant challenges translating that information from within the joint into tools to diagnose OA, monitor disease progression, and assess response to intervention.

The studies within this thesis have identified a large number of transmembrane proteins that appear to be differentially expressed in the various disease states of both hip and knee OA. This leads to the consideration of the emerging area of exosomal study. Exosomes are membrane-bound extracellular lipid bilayer vesicles that are produced and secreted by most cells in both healthy and diseased states. Their roles are varied and have been shown to be important in intracellular communication and protein homeostasis [408,409]. They have been described in most body fluids including SF [410] and over 300 exosomal proteins were identified in the earlier 1129 protein SOMAscan [177]. Whilst the work herein did not investigate whether the proteins identified may be exosomal, this represents an emerging area of investigation in OA due to the possible utility of exosomal targets for mechanistic investigation and therapeutics, such as has been used in other fields, including cancer and cardiovascular disease [411,412].

Proteins of the complement and coagulation cascades were seen in both hip and knee cohorts; in keeping with an established literature showing complement

dysregulation plays a key role in synovial joint OA [137]. In the knee the complement and coagulation proteins were the strongest pathway contributing significantly to the late OA samples including C3, which has recently been shown to correlate with clinical symptoms of knee OA in SF but not serum [413]. Interestingly, in the hip, there were fewer complement and coagulation proteins found, with those markers tending to load towards the early group in the analysis. Conversely there was a greater number of cytokines and chemokines seen in the hip group, which may support a different biological profile of hip OA, as compared with the knee injury cohort, since the early hip OA cohort were collected at FAI surgery whilst the knee injury cohort were sampled after specific ACL/meniscal injury. This is supported by the literature showing that innate and adaptive immune pathways are implicated in the response to injury in the pathogenesis of OA. Hence understanding these processes will be central to the identification of novel therapeutic targets [256].

In summary, these findings demonstrate that SF analysis remains the most informative medium for the biological investigation of the global joint changes seen in OA. The validation of newer multiplex platforms for bioanalysis has proven these systems to be able to handle SF samples, generate meaningful insight, and suggest novel targets for investigation. Finally, this thesis has shown that there are numerous and varied different routes taken by patients to access orthopaedic care, and that the durations of the pathways taken, as well as the treatments offered, vary significantly. In the study in this thesis, many more patients received

analgesia prior to arthroplasty than received alternative interventions including physiotherapy, intra-articular steroid injections and arthroscopy. These variations suggest there may be barriers to accessing appropriate patients sufficiently early in the pathogenesis of OA to be suitable for potential future studies of DMOADs.

Chapter 7 Future Directions

The work in this thesis was motivated by an appreciation of the lack of clinically useful biomarkers in OA. Now, at its conclusion, reflecting on this journey, it is even more apparent that the development of any such tool will require the clinical and scientific community to work together, rather than as individuals in adversarial competition, to deliver meaningful insights that will enlighten the search.

Laboratory experiments and clinical trials are needed to unite the mechanistic underpinnings of OA to inform its treatment. These studies cannot begin until patients can be identified and stratified clearly. This issue presents the fundamental, seemingly circular, challenge in OA. We must identify patients earlier and more clearly, so that we may develop a tool for diagnosis, investigation and monitoring.

It is hoped that this work will contribute to the field such that future researchers may be able to maximise the value of the samples they are gifted by their study participants, and to achieve the best value for the money generously invested in their work by the charitable funders, such as those who have generously supported this work.

Throughout this work it has become clear that a critically important next step will need to come from longitudinal studies with large patient groups investigating the

SF from well characterised, and clinically phenotyped, cohorts of patients to perform multivariate analyses of data from proteomic platforms such as the SOMAscan. Such a study, with the huge associated costs, could not be conceived without the evidence to clearly demonstrate the utility of the system to reliably generate meaningful data, and be cost effective, which this work has provided.

One area that the entire joint disease field has hitherto struggled to develop is the acquisition of a cohort of normal SF. Whilst SF provides the best assessment of the joint in its disease state, the biomarker discovery process continues to suffer from the inability to define the appearances of normal SF, and therefore must infer that the clinically proven disease accounts for the appearances seen. However, a cohort of normal SF from healthy controls may never be an achievable goal, since the natural volume contained within a healthy joint is only that which is required to lubricate the joint surfaces.

Notwithstanding the above, it should be kept in mind throughout all clinical research, who this work is truly for. We investigate OA with the patient at heart. The fervent desire is to deliver better outcomes for the patient community we serve, and it will only be with their continued help and support that work such as that contained within this thesis may continue. It should be borne in mind that the patients presenting complaint is not "I have K&L grade 2" or "I have a cam lesion", it is "I have hip pain". This pain may prevent them playing sport or, simply, sleeping at night, but we must remember that the aim shall always be to deliver the best care possible at the right time. For this reason, whilst the end goal for the

researcher may be DMOADs, no such treatments whether medical, surgical or otherwise will be possible without a proven means of accurate diagnosis, stratification and progression monitoring with which to identify candidates and test treatments.

A future, large cohort, longitudinal study can now be undertaken to investigate the SF appearances of OA using the SOMAscan. Large patient numbers will allow greater resolution to identify group differences between different disease states and to correlate changes in biomarker concentrations over time with clinical and radiological parameters to infer mechanistic changes that drive clinical phenotypic changes over time in the OA population. Without the proof that the SOMAscan can examine SF and differentiate between groups such a project would present to great a financial gamble to justify the investment.

Bibliography

1. Johnson VL, Giuffre BM, Hunter DJ. Osteoarthritis: what does imaging tell us about its etiology? *Semin Musculoskelet Radiol* 2012;**16**:410–8.

2. Burleigh A, Chanalaris A, Gardiner MD *et al.* Joint immobilization prevents murine osteoarthritis and reveals the highly mechanosensitive nature of protease expression in vivo. *Arthritis Rheum* 2012;**64**:2278–88.

3. Martel-Pelletier J, Barr AJ, Cicuttini FM *et al.* Osteoarthritis. *Nat Rev Dis Prim* 2016;**2**:16072.

4. Martel-Pelletier J, Boileau C, Pelletier JP *et al.* Cartilage in normal and osteoarthritis conditions. *Best Pract Res Clin Rheumatol* 2008;**22**:351–84.

5. March LM, Bachmeier CJ. Economics of osteoarthritis: a global perspective. *Baillieres Clin Rheumatol* 1997;**11**:817–34.

6. James SL, Abate D, Abate KH *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018;**392**:1789–858.

7. Turkiewicz A, Petersson IF, Björk J *et al.* Current and future impact of osteoarthritis on health care: A population-based study with projections to year 2032. *Osteoarthr Cartil* 2014;**22**:1826–32.

8. Prieto-Alhambra D, Judge A, Javaid MK *et al.* Incidence and risk factors for clinically diagnosed knee, hip and hand osteoarthritis: Influences of age, gender and osteoarthritis affecting other joints. *Ann Rheum Dis* 2014;**73**:1659–64.

9. Cibere J. Do we need radiographs to diagnose osteoarthritis? *Best Pract Res Clin Rheumatol* 2006;**20**:27–38.

10. NICE. Osteoarthritis: care and management. 2014:8, 10.

11. Arthritis Case Definitions.

12. Kuettner K, Goldberg V. *Osteoarthritic Disorders*. Kuettner K, Goldberg V (eds.). Rosemont: American Academy of Orthopaedic Surgeons, 1995.

13. Kraus VB, Blanco FJ, Englund M *et al.* Call for standardized definitions of osteoarthritis and risk stratification for clinical trials and clinical use. *Osteoarthr Cartil* 2015;**23**:1233–41.

14. Altman R, Alarcón G, Appelrouth D *et al.* The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;**34**:505–14.

15. Altman R, Asch E, Bloch D *et al.* Development of criteria for the classification and reporting of osteoarthritis: Classification of osteoarthritis of the knee. *Arthritis Rheum* 1986;**29**:1039–49.

16. Cooper C, Snow S, McAlindon TE *et al.* Risk factors for the incidence and progression of radiographic knee osteoarthritis. *Arthritis Rheum* 2000;**43**:995.

17. Oliveria SA, Felson DT, Reed JI *et al.* Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. *Arthritis Rheum* 1995;**38**:1134–41.

18. Lawrence RC, Felson DT, Helmick CG *et al.* Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. *Arthritis Rheum* 2008;**58**:26–35.

19. Murphy L, Schwartz TA, Helmick CG *et al.* Lifetime risk of symptomatic knee osteoarthritis. *Arthritis Care Res* 2008;**59**:1207–13.

20. Woolf A, Pfleger B. Burden of major musculoskeletal conditions. *Bull World Health Organ* 2003;**81**:646–56.

21. Agricola R, Waarsing JH, Arden NK *et al.* Cam impingement of the hip--a risk factor for hip osteoarthritis. *Nat Rev Rheumatol* 2013;**9**:630–4.

22. Agricola R, Heijboer MP, Roze RH *et al.* Pincer deformity does not lead to osteoarthritis of the hip whereas acetabular dysplasia does: Acetabular coverage and development of osteoarthritis in a nationwide prospective cohort study (CHECK). *Osteoarthr Cartil* 2013;**21**:1514–21.

23. Bartley EJ, Palit S, Staud R. Predictors of Osteoarthritis Pain: the Importance of Resilience. *Curr Rheumatol Rep* 2017;**19**:57.

24. Campbell CM, Buenaver LF, Finan P *et al.* Sleep, Pain Catastrophizing, and Central Sensitization in Knee Osteoarthritis Patients with and Without Insomnia.

Arthritis Care Res 2015;**67**:1387–96.

25. Knight JB, Callahan LF, Luong M-LN *et al.* The association of disability and pain with individual and community socioeconomic status in people with hip osteoarthritis. *Open Rheumatol J* 2011;**5**:51–8.

26. Taruc-Uy RL, Lynch SA. Diagnosis and treatment of osteoarthritis. *Prim Care Clin Off Pract* 2013;**40**:821–36.

27. Brandt KD, Dieppe P, Radin EL. Etiopathogenesis of osteoarthritis. *Rheum Dis Clin North Am* 2008;**34**:531–59.

28. Engel GL. The need for a new medical model: a challenge for biomedicine. *Science* 1977;**196**:129–36.

29. Carr AJ. Beyond disability: Measuring the social and personal consequences of osteoarthritis. *Osteoarthr Cartil* 1999;**7**:230–8.

30. Hawker GA, Stewart L, French MR *et al.* Understanding the pain experience in hip and knee osteoarthritis - an OARSI/OMERACT initiative. *Osteoarthr Cartil* 2008;**16**:415–22.

31. Geenen R, Overman CL, Christensen R *et al.* EULAR recommendations for the health professional's approach to pain management in inflammatory arthritis and osteoarthritis. *Ann Rheum Dis* 2018;**77**:797–807.

32. Aresti N, Kassam J, Nicholas N *et al.* Hip osteoarthritis. *BMJ* 2016;**354**:1–10.

33. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthr Cartil* 2007;**15**:A1–56.

34. Ahlbäck S. Osteoarthrosis of the knee. A radiographic investigation. *Acta Radiol Diagn* 1968;**Suppl 277**:7–62.

35. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957;**16**:494–502.

36. Felson DT, Niu J, Guermazi A *et al.* Defining radiographic incidence and progression of knee osteoarthritis: Suggested modifications of the Kellgren and Lawrence scale. *Ann Rheum Dis* 2011;**70**:1884–6.

37. Gossec L, Jordan JM, Lam MA *et al.* Comparative evaluation of three semi-quantitative radiographic grading techniques for hip osteoarthritis in terms of validity and reproducibility in 1404 radiographs: report of the OARSI-OMERACT Task Force. *Osteoarthr Cartil* 2009;**17**:182–7.

38. Gossec L, Jordan JM, Mazzuca SA *et al.* Comparative evaluation of three semi-quantitative radiographic grading techniques for knee osteoarthritis in terms of validity and reproducibility in 1759 X-rays: report of the OARSI-OMERACT task force. Extended report. *Osteoarthr Cartil* 2008;**16**:742–8.

39. Javaid MK, Lynch JA, Tolstykh I *et al.* Pre-radiographic MRI findings are associated with onset of knee symptoms: the most study. *Osteoarthr Cartil* 2010;**18**:323–8.

40. Guermazi A, Niu J, Hayashi D *et al.* Prevalence of abnormalities in knees detected by MRI in adults without knee osteoarthritis: Population based observational study (Framingham Osteoarthritis Study). *BMJ* 2012;**345**:1–13.

41. Conaghan PG, Hunter DJ, Maillefert JF *et al.* Summary and recommendations of the OARSI FDA osteoarthritis Assessment of Structural Change Working Group. *Osteoarthr Cartil* 2011;**19**:606–10.

42. Lawrence RC, Helmick CG, Arnett FC *et al.* Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis Rheum* 1998;**41**:778–99.

43. Palmer a, Pollard T, Mcnally E *et al.* Delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) predicts the pattern of hip osteoarthritis progression at five-year follow-up. *Arthrosc - J Arthrosc Relat Surg* 2013;**29**:e199.

44. Hesper T, Bulat E, Bixby S *et al.* Both 3-T dGEMRIC and acetabular-femoral T2 difference may detect cartilage damage at the chondrolabral junction. *Clin Orthop Relat Res* 2017;**475**:1058–65.

45. Hirose J, Nishioka H, Nakamura E *et al.* T 1ρ and T 2 mapping of the proximal tibiofibular joint in relation to aging and cartilage degeneration. *Eur J Radiol* 2012;**81**:2776–82.

46. Nishioka H, Hirose J, Nakamura E *et al.* T1ρ and T2 mapping reveal the in vivo extracellular matrix of articular cartilage. *J Magn Reson Imaging* 2012;**35**:147–55.

47. Bittersohl B, Miese FR, Hosalkar HS *et al.* T2* mapping of acetabular and femoral hip joint cartilage at 3 T: A prospective controlled study. *Invest Radiol* 2012;**47**:392–7.

48. Friedrich KM, Mamisch TC, Plank C *et al.* Diffusion-weighted imaging for the follow-up of patients after matrix-associated autologous chondrocyte transplantation. *Eur J Radiol* 2010;**73**:622–8.

49. Fernquest S, Palmer A, Gammer B *et al.* Compositional MRI of the hip: reproducibility, effect of joint unloading, and comparison of T2 relaxometry with delayed gadolinium-enhanced magnetic resonance imaging of cartilage. *Cartilage* 2019.

50. Madelin G, Babb JS, Xia D *et al.* Reproducibility and repeatability of quantitative sodium magnetic resonance imaging in vivo in articular cartilage at 3 T and 7 T. *Magn Reson Med* 2012;**68**:841–9.

51. Borthakur A, Shapiro EM, Beers J *et al.* Sensitivity of MRI to proteoglycan depletion in cartilage: Comparison of sodium and proton MRI. *Osteoarthr Cartil* 2000;**8**:288–93.

52. Luyten FP, Denti M, Filardo G *et al.* Definition and classification of early osteoarthritis of the knee. *Knee Surg Sport Traumatol Arthrosc* 2012;**20**:401–6.

53. Ryd L, Brittberg M, Eriksson K *et al.* Pre-osteoarthritis: definition and diagnosis of an elusive clinical entity. *Cartilage* 2015;**6**:156–65.

54. Frobell RB, Lohmander LS, Roos HP. Acute rotational trauma to the knee: Poor agreement between clinical assessment and magnetic resonance imaging findings. *Scand J Med Sci Sport* 2007;**17**:109–14.

55. Wilder F V., Hall BJ, Barrett JP *et al.* History of acute knee injury and osteoarthritis of the knee: A prospective epidemiological assessment. The clearwater osteoarthritis study. *Osteoarthr Cartil* 2002;**10**:611–6.

56. Spindler KP, Wright RW. Anterior cruciate ligament tear. *N Engl J Med* 2008;**359**:2135–42.

57. Brown TD, Johnston RC, Saltzman CL *et al.* Posttraumatic osteoarthritis: A first estimate of incidence, prevalence, and burden of disease. *J Orthop Trauma* 2006;**20**:739–44.

58. Nicholls AS, Kiran A, Pollard TCB *et al.* The association between hip morphology parameters and nineteen-year risk of end-stage osteoarthritis of the hip: A nested case-control study. *Arthritis Rheum* 2011;**63**:3392–400.

59. Palmer AJR, Ayyar Gupta V, Fernquest S *et al.* Arthroscopic hip surgery compared with physiotherapy and activity modification for the treatment of symptomatic femoroacetabular impingement: Multicentre randomised controlled trial. *BMJ* 2019;**364**.

60. Thomas GER, Palmer AJR, Batra RN *et al.* Subclinical deformities of the hip are significant predictors of radiographic osteoarthritis and joint replacement in

women. A 20 year longitudinal cohort study. *Osteoarthr Cartil* 2014.

61. Clohisy JC, Dobson MA, Robison JF *et al.* Radiographic structural abnormalities associated with premature, natural hip-joint failure. *J Bone Jt Surg - Ser A* 2011;**93**:3–9.

62. Hawker GA, Croxford R, Bierman AS *et al.* All-cause mortality and serious cardiovascular events in people with hip and knee osteoarthritis: A population based cohort study. *PLoS One* 2014;**9**:1–12.

63. National Joint Registry. 13th Annual Report. 2016.

64. Judge A, Welton NJ, Sandhu J *et al.* Equity in access to total joint replacement of the hip and knee in England: Cross sectional study. *BMJ* 2010;**341**:1–10.

65. Sierra RJ, Trousdale RT. Labral reconstruction using the ligamentum teres capitis: report of a new technique. *Clin Orthop Relat Res* 2009;**467**:753–9.

66. Ferguson SJ, Bryant JT, Ganz R *et al.* The acetabular labrum seal: a poroelastic finite element model. *Clin Biomech* 2000;**15**:463–8.

67. Greaves LL, Gilbart MK, Yung AC *et al.* Effect of acetabular labral tears, repair and resection on hip cartilage strain: A 7T MR study. *J Biomech* 2010;**43**:858–63.

68. McCarthy JC, Noble PC, Schuck MR *et al.* The Otto E. Aufranc Award: The role of labral lesions to development of early degenerative hip disease. *Clin Orthop Relat Res* 2001:25–37.

69. Nade S, Newbold PJ. Factors determining the level and changes in intra-articular pressure in the knee joint of the dog. *J Physiol* 1983;**338**:21–36.

70. Narvani a a, Tsiridis E, Kendall S *et al.* A preliminary report on prevalence of acetabular labrum tears in sports patients with groin pain. *Knee Surg Sports Traumatol Arthrosc* 2003;**11**:403–8.

71. Philippon MJ, Briggs KK, Fagrelius T *et al.* Labral refixation: current techniques and indications. *HSS J* 2012;**8**:240–4.

72. Link TM, Schwaiger BJ, Zhang AL. Regional articular cartilage abnormalities of the hip. *Am J Roentgenol* 2015;**205**:502–12.

73. Yacovelli S, Parvizi J. Return to sports after joint preservation hip surgery. *Orthop Clin North Am* 2020;**51**:427–39.

74. Saberi Hosnijeh F, Kavousi M, Boer CG *et al.* Development of a prediction model for future risk of radiographic hip osteoarthritis. *Osteoarthr Cartil* 2018;**26**:540–6.

75. Lohmander LS, Englund PM, Dahl LL *et al.* The long-term consequence of anterior cruciate ligament and meniscus injuries: Osteoarthritis. *Am J Sports Med* 2007;**35**:1756–69.

76. Chu CR, Williams AA, Coyle CH *et al.* Early diagnosis to enable early treatment of pre-osteoarthritis. *Arthritis Res Ther* 2012;**14**:212.

77. Warren W ed. Gray's Anatomy 20th Edition CC BY-NC-SA 3.0. Lea & Febiger

Philadelphia and New York., 1918.

78. Mathiessen A, Conaghan PG. Synovitis in osteoarthritis: Current understanding with therapeutic implications. *Arthritis Res Ther* 2017;**19**:1–9.

79. Smith MD, Barg E, Weedon H *et al.* Microarchitecture and protective mechanisms in synovial tissue from clinically and arthroscopically normal knee joints. *Ann Rheum Dis* 2003;**62**:303–7.

80. D. Smith M. The Normal Synovium. *Open Rheumatol J* 2011;**5**:100–6.

81. Filer A, Ward LSC, Kemble S *et al.* Identification of a transitional fibroblast function in very early rheumatoid arthritis. *Ann Rheum Dis* 2017;**76**:2105–12.

82. Robinson WH, Lepus CM, Wang Q *et al.* Low-grade inflammation as a key mediator of the pathogenesis of osteoarthritis. *Nat Rev Rheumatol* 2016;**12**:580–92.

83. Croft AP, Campos J, Jansen K *et al.* Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature* 2019;**570**:246–51.

84. Cai S, Ming B, Ye C *et al.* Similar transition processes in synovial fibroblasts from rheumatoid arthritis and osteoarthritis: A single-cell study. *J Immunol Res* 2019;**2019**:4080735.

85. Fam H, Bryant J, Kontopoulou M. Rheological properties of synovial fluids. *Biorheology* 2007;**44**:59–74.

86. Martin-Alarcon L, Schmidt TA. Rheological effects of macromolecular

interactions in synovial fluid. *Biorheology* 2016;**53**:49–67.

87. Lee JH, Jung JH, Kim J *et al.* Proteomic analysis of human synovial fluid reveals potential diagnostic biomarkers for ankylosing spondylitis. *Clin Proteomics* 2020;**17**:1–11.

88. Athanasiou KA, Darling EM, DuRaine GD *et al. Articular Cartilage*. 1st ed. CRC Press LLC, 2013.

89. Eggli PS, Herrmann W, Hunziker EB *et al.* Matrix compartments in the growth plate of the proximal tibia of rats. *Anat Rec* 1985;**211**:246–57.

90. Sophia Fox AJ, Bedi A, Rodeo SA. The basic science of articular cartilage: Structure, composition, and function. *Sports Health* 2009;**1**:461–8.

91. Knudson CB, Knudson W. Cartilage proteoglycans. *Semin Cell Dev Biol* 2001;**12**:69–78.

92. Mow VC, Guo XE. Mechano-electrochemical properties of articular cartilage: Their inhomogeneities and anisotropies. *Annu Rev Biomed Eng* 2002;**4**:175–209.

93. Eyre DR. The collagens of articular cartilage. *Semin Arthritis Rheum* 1991;**21**:2–11.

94. Luo Y, Sinkeviciute D, He Y *et al.* The minor collagens in articular cartilage. *Protein Cell* 2017;**8**:560–72.

95. Eyre D. Collagen of articular cartilage. - PubMed - NCBI. *Arthritis Res Ther*

2002;**4**:30–5.

96. Eyre DR. Collagens and cartilage matrix homeostasis. *Clin Orthop Relat Res* 2004:118–22.

97. Rice SJ, Beier F, Young DA *et al.* Interplay between genetics and epigenetics in osteoarthritis. *Nat Rev Rheumatol* 2020;**16**:268–81.

98. Outerbridge RE. The etiology of chondromalacia patellae. *J Bone Joint Surg Br* 1961;**43-B**:752–7.

99. Buckwalter JA, Mankin HJ, Grodzinsky AJ. Articular cartilage and osteoarthritis. *Instr Course Lect* 2005;**54**:465–80.

100. Goldring MB, Marcu KB. Cartilage homeostasis in health and rheumatic diseases. *Arthritis Res Ther* 2009;**11**.

101. Loeser RF, Goldring SR, Scanzello CR *et al.* Osteoarthritis: A disease of the joint as an organ. *Arthritis Rheum* 2012;**64**:1697–707.

102. Fridman R. Preface — Matrix metalloproteinases. *Biochim Biophys Acta - Mol Cell Res* 2017;**1864**:1925–6.

103. Nagase H, Woessner JF. Matrix metalloproteinases. *J Biol Chem* 1999;**274**:21491–4.

104. Song RH, Tortorella MD, Malfait AM *et al.* Aggrecan degradation in human articular cartilage explants is mediated by both ADAMTS-4 and ADAMTS-5.

Arthritis Rheum 2007;**56**:575–85.

105. Kashiwagi M, Tortorella M, Nagase H *et al.* TIMP-3 is a potent inhibitor of aggrecanase 1 (ADAM-TS4) and aggrecanase 2 (ADAM-TS5). *J Biol Chem* 2001;**276**:12501–4.

106. Tortorella MD, Liu RQ, Burn T *et al.* Characterization of human aggrecanase 2 (ADAM-TS5): Substrate specificity studies and comparison with aggrecanase 1 (ADAM-TS4). *Matrix Biol* 2002;**21**:499–511.

107. Yang CY, Chanalaris A, Troeberg L. ADAMTS and ADAM metalloproteinases in osteoarthritis - looking beyond the "usual suspects." *Osteoarthr Cartil* 2016,25:1000-1009.

108. Takaishi H, Kimura T, Dalal S *et al.* Joint diseases and matrix metalloproteinases: A Role for MMP-13. *Curr Pharm Biotechnol* 2008.

109. Burrage PS, Mix KS, Brinckerhoff CE. Matrix metalloproteinases: Role in arthritis. *Front Biosci* 2006.

110. Lin E a, Liu C-J. The role of ADAMTSs in arthritis. *Protein Cell* 2010;**1**:33–47.

111. Yamamoto K, Murphy G, Troeberg L. Extracellular regulation of metalloproteinases. *Matrix Biol* 2015;**44–46**:255–63.

112. Khokha R, Murthy A, Weiss A. Metalloproteinases and their natural inhibitors in inflammation and immunity. *Nat Rev Immunol* 2013;**13**:649–65.

113. Arpino V, Brock M, Gill SE. The role of TIMPs in regulation of extracellular matrix proteolysis. *Matrix Biol* 2015;**44–46**:247–54.

114. Glyn-Jones S, Palmer a JR, Agricola R *et al.* Osteoarthritis. *Lancet* 2015;**6736**:1–12.

115. Lefebvre V, Peeters-Joris C, Vaes G. Modulation by interleukin 1 and tumor necrosis factor α of production of collagenase, tissue inhibitor of metalloproteinases and collagen types in differentiated and dedifferentiated articular chondrocytes. *BBA - Mol Cell Res* 1990;**1052**:366–78.

116. Kaneko S, Satoh T, Chiba J *et al.* Interleukin–6 and interleukin–8 levels in serum and synovial fluid of patients with osteoarthritis. *Cytokines Cell Mol Ther* 2000;**6**:71–9.

117. Alaaeddine N, Di Battista JA, Pelletier JP *et al.* Differential effects of IL-8, LIF (pro-inflammatory) and IL-11 (anti-inflammatory) on TNF-α-induced PGE2 release and on signalling pathways in human OA synovial fibroblasts. *Cytokine* 1999;**11**:1020–30.

118. Scott JL, Gabrielides C, Davidson RK *et al.* Superoxide dismutase downregulation in osteoarthritis progression and end-stage disease. *Ann Rheum Dis* 2010;**69**:1502–10.

119. Chow YY, Chin KY. The Role of Inflammation in the pathogenesis of osteoarthritis. *Mediators Inflamm* 2020;**2020**.

120. Bacon K. Chemokine / Chemokine Receptor Nomenclature * International Union of Immunological Societies / World Health Organization. *J Leukoc Biol* 2001;**70**:465–6.

121. Raghu H, Lepus CM, Wang Q *et al.* CCL2/CCR2, but not CCL5/CCR5, mediates monocyte recruitment, inflammation and cartilage destruction in osteoarthritis. *Ann Rheum Dis* 2017;**76**:914–22.

122. Borzí RM, Mazzetti I, Cattini L *et al.* Human chondrocytes express functional chemokine receptors and release matrix-degrading enzymes in response to C-X-C and C-C chemokines. *Arthritis Rheum* 2000;**43**:1734–41.

123. Uguccioni M, D'Apuzzo M, Loetscher M *et al.* Actions of the chemotactic cytokines MCP-1, MCP-2, MCP-3, RANTES, MIP-1α and MIP-1β on human monocytes. *Eur J Immunol* 1995;**25**:64–8.

124. Manferdini C, Paolella F, Gabusi E *et al.* Osteoarthritic milieu affects adipose-derived mesenchymal stromal cells. *J Orthop Res* 2020;**38**:336–47.

125. Zhou X, Cao H, Yuan Y *et al.* Biochemical signals mediate the crosstalk between cartilage and bone in osteoarthritis. *Biomed Res Int* 2020;**2020**:1–8.

126. Dieppe PA, Lohmander LS. Pathogenesis and management of pain in osteoarthritis. *Lancet* 2005;**365**:965–73.

127. Correa D, Somoza RA, Lin P *et al.* Sequential exposure to fibroblast growth factors (FGF) 2, 9 and 18 enhances hMSC chondrogenic differentiation. *Osteoarthr*

Cartil 2015;**23**:443–53.

128. Hellingman CA, Koevoet W, Kops N *et al.* Fibroblast growth factor receptors in in vitro and in vivo chondrogenesis: Relating tissue engineering using adult mesenchymal stem cells to embryonic development. *Tissue Eng - Part A* 2010;**16**:545–56.

129. Convill JG, Tawy GF, Freemont AJ *et al.* Clinically relevant molecular biomarkers for use in human knee osteoarthritis: A systematic review. *Cartilage* 2020.

130. Boffa A, Merli G, Andriolo L *et al.* Synovial fluid biomarkers in knee osteoarthritis: A systematic review and quantitative evaluation using BIPEDs criteria. *Cartilage* 2020.

131. Sacitharan PK. Ageing and Osteoarthritis. In: Harris JR, Korolchuk VI (eds.). *Biochemistry and Cell Biology of Ageing: Part II Clinical Science*. Singapore: Springer Singapore, 2019, 123–59.

132. Watt FE. Osteoarthritis biomarkers: year in review. *Osteoarthr Cartil* 2018;**26**:312–8.

133. Roberts S, Evans H, Wright K *et al.* ADAMTS-4 activity in synovial fluid as a biomarker of inflammation and effusion. *Osteoarthr Cartil* 2015;**23**:1622–6.

134. Ruiz-Romero C, Blanco FJ. Proteomics role in the search for improved diagnosis, prognosis and treatment of osteoarthritis. *Osteoarthr Cartil* 2010;**18**:500–

9.

135. Balakrishnan L, Nirujogi RS, Ahmad S *et al.* Proteomic analysis of human osteoarthritis synovial fluid. *Clin Proteomics* 2014;**11**:6.

136. Yamagiwa H, Sarkar G, Charlesworth MC *et al.* Two-dimensional gel electrophoresis of synovial fluid: method for detecting candidate protein markers for osteoarthritis. *J Orthop Sci* 2003;**8**:482–90.

137. Wang Q, Rozelle AL, Lepus CM *et al.* Identification of a central role for complement in osteoarthritis. *Nat Med* 2011;**17**:1674–9.

138. Kamphorst JJ, van der Heijden R, DeGroot J *et al.* Profiling of endogenous peptides in human synovial fluid by NanoLC−MS: Method validation and peptide identification. *J Proteome Res* 2007;**6**:4388–96.

139. Ritter SY, Subbaiah R, Bebek G *et al.* Proteomic analysis of synovial fluid from the osteoarthritic knee: Comparison with transcriptome analyses of joint tissues. *Arthritis Rheum* 2013;**65**:981–92.

140. Fernández-Puente P, Calamia V, González-Rodríguez L *et al.* Multiplexed mass spectrometry monitoring of biomarker candidates for osteoarthritis. *J Proteomics* 2017;**152**:216–25.

141. Atkinson A.J. J, Colburn WA, DeGruttola VG *et al.* Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;**69**:89–95.

142. Lin CH, Chiu SI, Chen TF *et al.* Classifications of neurodegenerative disorders using a multiplex blood biomarkers-based machine learning model. *Int J Mol Sci* 2020;**21**:1–15.

143. Zizzari IG, Napoletano C, Di Filippo A *et al.* Exploratory pilot study of circulating biomarkers in metastatic renal cell carcinoma. *Cancers (Basel)* 2020;**12**:1–15.

144. Ng E, Le AK, Nguyen MH *et al.* Early multiplexed detection of cirrhosis by giant magnetoresistive biosensors with protein biomarkers. *ACS sensors* 2020.

145. Mayeux R. Biomarkers: potential uses and limitations. *NeuroRx* 2004;**1**:182–8.

146. Bauer DC, Hunter DJ, Abramson SB *et al.* Classification of osteoarthritis biomarkers: a proposed approach. *Osteoarthritis Cartilage* 2006;**14**:723–7.

147. Jayadev C, Hulley P, Swales C *et al.* Synovial fluid fingerprinting in end-stage knee osteoarthritis a novel biomarker concept. *Bone Joint Res* 2020;**9**:623–32.

148. Roughley PJ, Mort JS. The role of aggrecan in normal and osteoarthritic cartilage. *J Exp Orthop* 2014;**1**:8.

149. Meulenbelt I, Kloppenburg M, Kroon HM *et al.* Urinary CTX-II levels are associated with radiographic subtypes of osteoarthritis in hip, knee, hand, and facet joints in subject with familial osteoarthritis at multiple sites: the GARP study. *Ann Rheum Dis* 2006;**65**:360–5.

150. Attur M, Krasnokutsky-Samuels S, Samuels J *et al.* Prognostic biomarkers in

osteoarthritis. *Curr Opin Rheumatol* 2013;**25**:136–44.

151. van Spil WE, DeGroot J, Lems WF *et al.* Serum and urinary biochemical markers for knee and hip-osteoarthritis: A systematic review applying the consensus BIPED criteria. *Osteoarthr Cartil* 2010;**18**:605–12.

152. Stiell IG, Greenberg GH, McKnight RD *et al.* A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Ann Emerg Med* 1992;**21**:384–90.

153. Anderson KM, Odell PM, Wilson PWF *et al.* Cardiovascular disease risk profiles. *Am Heart J* 1991;**121**:293–8.

154. Anderson KM, Wilson PWF, Odell PM *et al.* An updated coronary risk profile. A statement for health professionals. *Circulation* 1991;**83**:356–62.

155. Collins GS, Reitsma JB, Altman DG *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015;**162**:55.

156. Kiddle SJ, Steves CJ, Mehta M *et al.* Plasma protein biomarkers of Alzheimer's disease endophenotypes in asymptomatic older twins: early cognitive decline and regional brain volumes. *Transl Psychiatry* 2015;**5**:e584.

157. Belongie KJ, Ferrannini E, Johnson K *et al.* Identification of novel biomarkers to monitor β-cell function and enable early detection of type 2 diabetes risk. *PLoS One* 2017;**12**:1–18.

158. Ganz P, Heidecker B, Hveem K *et al.* Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. *JAMA* 2016;**315**:2532–41.

159. Parolo S, Marchetti L, Lauria M *et al.* Combined use of protein biomarkers and network analysis unveils deregulated regulatory circuits in Duchenne muscular dystrophy. *PLoS One* 2018;**13**:e0194225.

160. SomaLogic. Technical White Paper SOMAscan ™ Proteomic Assay. 2013:1–24.

161. Chirinos JA, Zhao L, Jia Y *et al.* Reduced apolipoprotein M and adverse outcomes across the spectrum of human heart failure. *Circulation* 2020:1463–76.

162. Mcardle A, Qasim Butt A, Szentpetery A *et al.* Developing clinically relevant biomarkers in inflammatory arthritis: A multiplatform approach for serum candidate protein discovery. *Proteomics - Clin Appl* 2016;**10**:691–8.

163. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 1990;**249**:505–10.

164. Ellington a D, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. *Nature* 1990;**346**:818–22.

165. Brody E, Gold L, Mehan M *et al.* Life's simple measures: Unlocking the proteome. *J Mol Biol* 2012;**422**:595–606.

166. Eaton BE. The joys of in vitro selection: Chemically dressing oligonucleotides

to satiate protein targets. *Curr Opin Chem Biol* 1997;**1**:10–6.

167. Gold L, Ayers D, Bertino J *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* 2010;**5**.

168. Rohloff JC, Gelinas AD, Jarvis TC *et al.* Nucleic acid ligands with protein-like side chains: Modified aptamers and their use as diagnostic and therapeutic agents. *Mol Ther - Nucleic Acids* 2014;**3**:e201.

169. Keeney TR, Bock C, Gold L *et al.* Automation of the SomaLogic proteomics assay: A platform for biomarker discovery. *J Lab Autom* 2009;**14**:360–6.

170. Ostroff RM, Bigbee WL, Franklin W *et al.* Unlocking biomarker discovery: Large scale application of aptamer proteomic technology for early detection of lung cancer. *PLoS One* 2010;**5**:1–10.

171. Billing AM, Ben Hamidane H, Bhagwat AM *et al.* Complementarity of SOMAscan to LC-MS/MS and RNA-seq for quantitative profiling of human embryonic and mesenchymal stem cells. *J Proteomics* 2017;**150**:86–97.

172. Sierra-Sánchez Á, Garrido-Martín D, Lourido L *et al.* Screening and validation of novel biomarkers in osteoarticular pathologies by comprehensive combination of protein array technologies. *J Proteome Res* 2017;**16**:1890–9.

173. Kim CH, Tworoger SS, Stampfer MJ *et al.* Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. *Sci Rep* 2018;**8**:1–10.

174. Williams SA, Kivimaki M, Langenberg C *et al.* Plasma protein patterns as

comprehensive indicators of health. *Nat Med* 2019;**25**:1851–7.

175. Malekzadeh A, Leurs C, van Wieringen W *et al.* Plasma proteome in multiple sclerosis disease progression. *Ann Clin Transl Neurol* 2019;**6**:1582–94.

176. King JD, Rowland G, Villasante Tezanos AG *et al.* Joint fluid proteome after Anterior cruciate ligament rupture reflects an acute posttraumatic inflammatory and chondrodegenerative state. *Cartilage* 2018.

177. Webber J, Stone TC, Katilius E *et al.* Proteomics analysis of cancer exosomes using a novel modified aptamer-based array (somascantm) platform. *Mol Cell Proteomics* 2014;**13**:1050–64.

178. Mueller SK, Nocera AL, Dillon ST *et al.* Noninvasive exosomal proteomic biosignatures, including cystatin SN, peroxiredoxin-5, and glycoprotein VI, accurately predict chronic rhinosinusitis with nasal polyps. *Int Forum Allergy Rhinol* 2019;**9**:177–86.

179. Nocairi H, Qannari EM, Vigneau E *et al.* Discrimination on latent components with respect to patterns. Application to multicollinear data. *Comput Stat Data Anal* 2005;**48**:139–47.

180. Indahl UG, Martens H, Næs T. From dummy regression to prior probabilities in PLS-DA. *J Chemom* 2007;**21**:529–36.

181. Biancolillo A, Marini F. Chemometric methods for spectroscopy-based pharmaceutical analysis. *Front Chem* 2018;**6**:1–14.

182. Jolliffe IT, Cadima J, Cadima J. Principal component analysis : a review and recent developments Subject Areas : Author for correspondence : 2016.

183. Ringnér M. What is principal component analysis? *Nat Biotechnol* 2008;**26**:303–4.

184. Giuliani A. The application of principal component analysis to drug discovery and biomedical data. *Drug Discov Today* 2017;**22**:1069–76.

185. David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol Biol* 2014;**1084**:193–226.

186. Spiegelberg J, Rusz J. Can we use PCA to detect small signals in noisy data? *Ultramicroscopy* 2016;**172**:40–6.

187. Wold S, Sjöström M, Eriksson L. PLS-regression: A basic tool of chemometrics. *Chemom Intell Lab Syst* 2001;**58**:109–30.

188. Kuligowski J, Pérez-Guaita D, Escobar J *et al.* Evaluation of the effect of chance correlations on variable selection using Partial Least Squares-Discriminant Analysis. *Talanta* 2013;**116**:835–40.

189. Brereton RG, Lloyd GR. Partial least squares discriminant analysis: Taking the magic away. *J Chemom* 2014;**28**:213–25.

190. Gromski PS, Muhamadali H, Ellis DI *et al.* A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding. *Anal Chim Acta* 2015;**879**:10–23.

191. Rodríguez-Pérez R, Fernández L, Marco S. Overoptimism in cross-validation when using partial least squares-discriminant analysis for omics data: a systematic study. *Anal Bioanal Chem* 2018;**410**:5981–92.

192. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J R Stat Soc Ser B* 1996;**58**:267–88.

193. Pavlou M, Ambler G, Seaman SR *et al.* How to develop a more accurate risk prediction model when there are few events. *BMJ* 2015;**351**:h3868.

194. Bujak R, Daghir-Wojtkowiak E, Kaliszan R *et al.* PLS-based and regularization-based methods for the selection of relevant variables in non-targeted metabolomics data. *Front Mol Biosci* 2016;**3**:1–10.

195. Steyerberg EW, Eijkemans MJC, Habbema JDF. Application of shrinkage techniques in logistic regression analysis: A case study. *Stat Neerl* 2001;**55**:76–88.

196. Kloppenburg M, Berenbaum F. Osteoarthritis year in review 2019: epidemiology and therapy. *Osteoarthr Cartil* 2020;**28**:242–8.

197. Menetski JP, Hoffmann SC, Cush SS *et al.* The foundation for the national institutes of health biomarkers consortium: Past accomplishments and new strategic direction. *Clin Pharmacol Ther* 2019;**105**:829–43.

198. Pereira D, Peleteiro B, Araújo J *et al.* The effect of osteoarthritis definition on prevalence and incidence estimates: A systematic review. *Osteoarthr Cartil* 2011;**19**:1270–85.

199. Leal J, Murphy J, Garriga C *et al.* Costs of joint replacement in osteoarthritis: a study using the National Joint Registry and Clinical Practice Research Datalink datasets. *Arthritis Care Res (Hoboken)* 2020.

200. Zamborsky R, Danisovic L. Surgical techniques for knee cartilage repair: An updated large-scale systematic review and network meta-analysis of randomized controlled trials. *Arthrosc - J Arthrosc Relat Surg* 2020;**36**:845–58.

201. van Spil WE, Szilagyi IA. Osteoarthritis year in review 2019: biomarkers (biochemical markers). *Osteoarthr Cartil* 2020;**28**:296–315.

202. Ebell MH. Osteoarthritis: rapid evidence review. *Am Fam Physician* 2018;**97**:523–6.

203. Ackerman IN, Kemp JL, Crossley KM *et al.* Hip and knee osteoarthritis affects younger people, too. *J Orthop Sports Phys Ther* 2017;**47**:67–79.

204. Litwic A, Edwards MH, Dennison EM *et al.* Epidemiology and burden of osteoarthritis. *Br Med Bull* 2013;**105**:185–99.

205. Birn J, Pruente R, Avram R *et al.* Sonographic evaluation of hip joint effusion in osteoarthritis with correlation to radiographic findings. *J Clin Ultrasound* 2014;**42**:205–11.

206. Jayadev C, Rout R, Price A *et al.* Hyaluronidase treatment of synovial fluid to improve assay precision for biomarker research using multiplex immunoassay platforms. *J Immunol Methods* 2012;**386**:22–30.

207. McCarty WJ, Luan A, Siddiqui M *et al.* Biomechanical properties of mixtures of blood and synovial fluid. *J Orthop Res* 2011;**29**:240–6.

208. Candia J, Cheung F, Kotliarov Y *et al.* Assessment of variability in the SOMAscan Assay. *Sci Rep* 2017;**7**:1–13.

209. van den Berg RA, Hoefsloot HCJ, Westerhuis JA *et al.* Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics* 2006;**7**:1–15.

210. Esbensen KH, Geladi P. Principles of proper validation: Use and abuse of re-sampling for validation. *J Chemom* 2010;**24**:168–87.

211. Kim KA, Zhao J, Andarmani S *et al.* R-spondin proteins: A novel link to β-catenin activation. *Cell Cycle* 2006;**5**:23–6.

212. Baker AH, Edwards DR, Murphy G. Metalloproteinase inhibitors: Biological actions and therapeutic opportunities. *J Cell Sci* 2002;**115**:3719–27.

213. Chou CH, Attarian DE, Wisniewski HG *et al.* TSG-6 – a double-edged sword for osteoarthritis (OA). *Osteoarthr Cartil* 2018;**26**:245–54.

214. Robertson DM, Burger HG, Fuller PJ. Inhibin/activin and ovarian cancer. *Endocr Relat Cancer* 2004;**11**:35–49.

215. Xue X, Wu J, Ricklin D *et al.* Regulator-dependent mechanisms of C3b processing by factor i allow differentiation of immune responses. *Nat Struct Mol Biol* 2017;**24**:643–51.

216. Mosnier LO, Griffin JH. Protein C anticoagulant activity in relation to anti-inflammatory and anti-apoptotic activities. *Front Biosci* 2006;**11**:2381–99.

217. Mihara E, Hirai H, Yamamoto H *et al.* Active and water-soluble form of lipidated wnt protein is maintained by a serum glycoprotein afamin/α-albumin. *Elife* 2016;**5**:1–19.

218. Tan K, Lawler J. The structure of the Ca2+-binding, glycosylated F-spondin domain of F-spondin - A C2-domain variant in an extracellular matrix protein. *BMC Struct Biol* 2011;**11**.

219. Autieri M V. IL-19 and other IL-20 family member cytokines in vascular inflammatory diseases. *Front Immunol* 2018;**9**:1–7.

220. Kazanietz MG, Durando M, Cooke M. CXCL13 and its receptor CXCR5 in cancer: Inflammation, immune response, and beyond. *Front Endocrinol (Lausanne)* 2019;**10**:1–15.

221. Rawal N, Pangburn MK. Structure/function of C5 convertases of complement. *Int Immunopharmacol* 2001;**1**:415–22.

222. Blaum BS, Hannan JP, Herbert AP *et al.* Structural basis for sialic acid-mediated self-recognition by complement factor H. *Nat Chem Biol* 2015;**11**:77–82.

223. Schmidt J, Bosserhoff A-K. Processing of MIA protein during melanoma cell migration. *Int J cancer* 2009;**125**:1587–94.

224. Martina E, Degen M, Ruegg C *et al.* Tenascin-W is a specific marker of glioma-

associated blood vessels and stimulates angiogenesis in vitro . *FASEB J* 2010;**24**:778–87.

225. Zhou Y, Do DC, Ishmael FT *et al.* Mannose receptor modulates macrophage polarization and allergic inflammation through miR-511-3p. *J Allergy Clin Immunol* 2018;**141**:350-364.e8.

226. Ling TY, Chen CL, Huang YH *et al.* Identification and characterization of the acidic pH binding sites for growth regulatory ligands of low density lipoprotein receptor-related protein-1. *J Biol Chem* 2004;**279**:38736–48.

227. Lee GH, Fujita M, Takaoka K *et al.* A GPI processing phospholipase A2, PGAP6, modulates Nodal signaling in embryos by shedding CRIPTO. *J Cell Biol* 2016;**215**:705–18.

228. Ryan E, Shen D, Wang X. Structural studies reveal an important role for the pleiotrophin C-terminus in mediating interactions with chondroitin sulfate. *FEBS J* 2016;**283**:1488–503.

229. Venkatraman Girija U, Furze CM, Gingras AR *et al.* Molecular basis of sugar recognition by collectin-K1 and the effects of mutations associated with 3MC syndrome. *BMC Biol* 2015;**13**.

230. Kläning E, Christensen B, Bajic G *et al.* Multiple low-affinity interactions support binding of human osteopontin to integrin αxβ2. *Biochim Biophys Acta - Proteins Proteomics* 2015;**1854**:930–8.

231. Van 't Veer C, Kalafatis M, Bertina RM *et al.* Increased tissue factor-initiated prothrombin activation as a result of the Arg506 → Gln mutation in factor V(LEIDEN). *J Biol Chem* 1997;**272**:20721–9.

232. Osorio-Conles Ó, Guitart M, Moreno-Navarrete JM *et al.* Adipose tissue and serum CCDC80 in obesity and its association with related metabolic disease. *Mol Med* 2017;**23**:225–34.

233. Saunders JT, Schwarzbauer JE. Fibronectin matrix as a scaffold for procollagen proteinase binding and collagen processing. *Mol Biol Cell* 2019;**30**:2218–26.

234. Letronne F, Laumet G, Ayral AM *et al.* ADAM30 Downregulates APP-Linked Defects Through Cathepsin D Activation in Alzheimer's Disease. *EBioMedicine* 2016;**9**:278–92.

235. Mehta-D'souza P, Klopocki AG, Oganesyan V *et al.* Glycan bound to the selectin low affinity state engages Glu-88 to stabilize the high affinity state under force. *J Biol Chem* 2017;**292**:2510–8.

236. Wu M, Chen G, Li YP. TGF-β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res* 2016;**4**.

237. Rosenberg JH, Rai V, Dilisio MF *et al.* Osteoarthritis : Potentially novel therapeutic targets. 2018;**434**:171–9.

238. Alto LT, Terman JR. Semaphorins and their signaling mechanisms. *Methods Mol Biol* 2017;**1493**:1–25.

239. Wang Y, Nakayama M, Pitulescu ME *et al.* Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* 2010;**465**:483–6.

240. Towne JE, Garka KE, Renshaw BR *et al.* Interleukin (IL)-1F6, IL-1F8, and IL-1F9 signal through IL-1Rrp2 and IL-1RAcP to activate the pathway leading to NF-κB and MAPKs. *J Biol Chem* 2004;**279**:13677–88.

241. Sparrow E, Bodman-Smith MD. Granulysin: The attractive side of a natural born killer. *Immunol Lett* 2020;**217**:126–32.

242. Fu P, Yang Z, Bach LA. Prohibitin-2 binding modulates insulin-like growth factor-binding protein-6 (IGFBP-6)-induced rhabdomyosarcoma cell migration. *J Biol Chem* 2013;**288**:29890–900.

243. Thysen S, Cailotto F, Lories R. Osteogenesis induced by frizzled-related protein (FRZB) is linked to the netrin-like domain. *Lab Investig* 2016;**96**:570–80.

244. Banerji S, Ni J, Wang SX *et al.* LYVE-1, a new homologue of the CD44 glycoprotein, is a lymph-specific receptor for hyaluronan. *J Cell Biol* 1999;**144**:789–801.

245. Li K, Zhang Y, Zhang Y *et al.* Tyrosine kinase Fyn promotes osteoarthritis by activating the β-catenin pathway. *Ann Rheum Dis* 2018;**77**:935–43.

246. McNally RS, Davis BK, Clements CM *et al.* DJ-1 enhances cell survival through the binding of Cezanne, a negative regulator of NF-κB. *J Biol Chem* 2011;**286**:4098–106.

247. Dondelinger F, Mukherjee S. The joint lasso: high-dimensional regression for group structured data. *Biostatistics* 2020;**21**:219–35.

248. Lê S, Josse J, Husson F. FactoMineR : An R Package for Multivariate Analysis. *J Stat Softw* 2008;**25**:253–8.

249. Szklarczyk D, Gable AL, Lyon D *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;**47**:D607–13.

250. Jeyaraman M, Muthu S, Gulati A *et al.* Mesenchymal stem cell–derived exosomes: A potential therapeutic avenue in knee osteoarthritis. *Cartilage* 2020.

251. Domenis R, Zanutel R, Caponnetto F *et al.* Characterization of the proinflammatory profile of synovial fluid-derived exosomes of patients with osteoarthritis. *Mediators Inflamm* 2017;**2017**.

252. Kolhe R, Hunter M, Liu S *et al.* Gender-specific differential expression of exosomal miRNA in synovial fluid of patients with osteoarthritis. *Sci Rep* 2017;**7**:2029.

253. Lepetsos P, Papavassiliou AG. ROS/oxidative stress signaling in osteoarthritis. *Biochim Biophys Acta - Mol Basis Dis* 2016;**1862**:576–91.

254. Arnold N, Rehm M, Büchele G *et al.* Growth differentiation factor-15 as a potent predictor of long-term mortality among subjects with osteoarthritis. *J Clin Med* 2020;**9**:3107.

255. Goutas A, Papathanasiou I, Mourmoura E *et al.* Oxidative stress response is mediated by overexpression and spatiotemporal regulation of caveolin-1. *Antioxidants* 2020;**9**:1–23.

256. Woodell-May JE, Sommerfeld SD. Role of inflammation and the immune system in the progression of osteoarthritis. *J Orthop Res* 2020;**38**:253–7.

257. Lien-Iversen T, Morgan DB, Jensen C *et al.* Does surgery reduce knee osteoarthritis, meniscal injury and subsequent complications compared with non-surgery after ACL rupture with at least 10 years follow-up? A systematic review and meta-analysis. *Br J Sports Med* 2019:592–8.

258. Englund M, Lohmander LS. Risk factors for symptomatic knee osteoarthritis fifteen to twenty-two years after meniscectomy. *Arthritis Rheum* 2004;**50**:2811–9.

259. Lee TH, Klampfer L, Shows TB *et al.* Transcriptional regulation of TSG6, a tumor necrosis factor- and interleukin-1-inducible primary response gene coding for a secreted hyaluronan-binding protein. *J Biol Chem* 1993;**268**:6154–60.

260. Chou CH, Lee MTM, Song IW *et al.* Insights into osteoarthritis progression revealed by analyses of both knee tibiofemoral compartments. *Osteoarthr Cartil* 2015;**23**:571–80.

261. Watt FE, Paterson E, Freidin A *et al.* Acute molecular changes in synovial fluid following human knee injury: Association with early clinical outcomes. *Arthritis Rheumatol* 2016;**68**:2129–40.

262. Wisniewski HG, Colón E, Liublinska V *et al.* TSG-6 activity as a novel biomarker of progression in knee osteoarthritis. *Osteoarthr Cartil* 2014;**22**:235–41.

263. Wisniewski HG, Maier R, Lotz M *et al.* TSG-6: a TNF-, IL-1-, and LPS-inducible secreted glycoprotein associated with arthritis. *J Immunol* 1993;**151**:6593–601.

264. Lee TH, Wisniewski HG, Vilcek J. A novel secretory tumor necrosis factor-inducible protein (TSG-6) is a member of the family of hyaluronate binding proteins, closely related to the adhesion receptor CD44. *J Cell Biol* 1992;**116**:545–57.

265. Mahoney DJ, Swales C, Athanasou NA *et al.* TSG-6 inhibits osteoclast activity via an autocrine mechanism and is functionally synergistic with osteoprotegerin. *Arthritis Rheum* 2011;**63**:1034–43.

266. Milner CM, Higman VA, Day AJ. TSG-6: A pluripotent inflammatory mediator? *Biochem Soc Trans* 2006;**34**:446–50.

267. Park Y, Jowitt TA, Day AJ *et al.* Nuclear magnetic resonance insight into the multiple glycosaminoglycan binding modes of the link module from human TSG-6. *Biochemistry* 2016;**55**:262–76.

268. Daghestani HN, Kraus VB. Inflammatory biomarkers in osteoarthritis. *Osteoarthr Cartil* 2015;**23**:1890–6.

269. Atukorala I, Kwoh CK, Guermazi A *et al.* Synovitis in knee osteoarthritis: A precursor of disease? *Ann Rheum Dis* 2016;**75**:390–5.

270. Brew K, Dinakarpandian D, Nagase H. Tissue inhibitors of

metalloproteinases: Evolution, structure and function. *Biochim Biophys Acta - Protein Struct Mol Enzymol* 2000;**1477**:267–83.

271. Naito K, Takahashi M, Kushida K *et al.* Measurement of matrix metalloproteinases (MMPs) and tissue inhibitor of metalloproteinases-1 (TIMP-1) in patients with knee osteoarthritis: Comparison with generalized osteoarthritis. *Rheumatology* 1999;**38**:510–5.

272. Cattano NM, Driban JB, Balasubramanian E *et al.* Biochemical comparison of osteoarthritic knees with and without effusion. *BMC Musculoskelet Disord* 2011;**12**.

273. Baenziger NL, Brodie GN, Majerus PW. A thrombin-sensitive protein of human platelet membranes. *Proc Natl Acad Sci U S A* 1971;**68**:240–3.

274. Lebensohn AM, Rohatgi R. R-spondins can potentiate WNT signaling without LGRs. *Elife* 2018;**7**:1–18.

275. Krönke G, Uderhardt S, Kim KA *et al.* R-spondin 1 protects against inflammatory bone damage during murine arthritis by modulating the Wnt pathway. *Arthritis Rheum* 2010;**62**:2303–12.

276. Choi BY, Chang SH, Cho HJ *et al.* The association of radiographic progression with serum R-spondin 1 (RSPO1) levels or Dickkopf-1 (DKK1)/RSPO1 ratios in rheumatoid arthritis patients: Clinical evidence for reciprocal inhibition between DKK1 and RSPO1. *Scand J Rheumatol* 2014;**43**:453–61.

277. Abed É, Chan TF, Delalandre A *et al.* R-spondins are newly recognized players

in osteoarthritis that regulate Wnt signaling in osteoblasts. *Arthritis Rheum* 2011;**63**:3865–75.

278. Ohkawara B, Glinka A, Niehrs C. Rspo3 binds syndecan 4 and induces Wnt/PCP signaling via clathrin-mediated endocytosis to promote morphogenesis. *Dev Cell* 2011;**20**:303–14.

279. Park S, Cui J, Yu W *et al.* Differential activities and mechanisms of the four r-spondins in potentiating wnt/-catenin signaling. *J Biol Chem* 2018;**293**:9759–69.

280. Ge JJ, Lu WK, Zhu N *et al.* Effect of LGR4 on synovial cells and inflammatory factors in rats with traumatic osteoarthritis. *Eur Rev Med Pharmacol Sci* 2019;**23**:6783–90.

281. Surendran S, Kim SH, Jee BK *et al.* Anti-apoptotic Bcl-2 gene transfection of human articular chondrocytes protects against nitric oxide-induced apoptosis. *J Bone Jt Surg - Ser B* 2006;**88**:1660–5.

282. Zhu M, Chen M, Zuscik M *et al.* Inhibition of β-catenin signaling in articular chondrocytes results in articular cartilage destruction. *Arthritis Rheum* 2008;**58**:2053–64.

283. Iannone F, De Bari C, Scioscia C *et al.* Increased Bcl-2/p53 ratio in human osteoarthritic cartilage: A possible role in regulation of chondrocyte metabolism. *Ann Rheum Dis* 2005;**64**:217–21.

284. Tomasello F, Messina A, Lartigue L *et al.* Outer membrane VDAC1 controls

permeability transition of the inner mitochondrial membrane in cellulo during stress-induced apoptosis. *Cell Res* 2009;**19**:1363–76.

285. Wojtuszkiewicz A, Schuurhuis GJ, Kessler FL *et al.* Exosomes secreted by apoptosis-resistant Acute Myeloid Leukemia (AML) blasts harbor regulatory network proteins potentially involved in antagonism of apoptosis. *Mol Cell Proteomics* 2016;**15**:1281–98.

286. Nagase Y, Iwasawa M, Akiyama T *et al.* Anti-apoptotic Molecule Bcl-2 Regulates the Differentiation , Activation , and Survival of Both Osteoblasts and Osteoclasts *. 2009;**284**:36659–69.

287. Moriishi T, Fukuyama R, Miyazaki T *et al.* Overexpression of BCLXL in osteoblasts inhibits osteoblast apoptosis and increases bone volume and strength. *J Bone Miner Res* 2016;**31**:1366–80.

288. Willimott S, Merriam T, Wagner SD. Apoptosis induces Bcl-XS and cleaved Bcl-XL in chronic lymphocytic leukaemia. *Biochem Biophys Res Commun* 2011;**405**:480–5.

289. Cole CL, Rushton G, Jayson GC *et al.* Ovarian cancer cell heparan sulfate 6- O -sulfotransferases regulate an angiogenic program induced by heparin-binding epidermal growth factor (EGF) -like growth factor / EGF receptor signaling. 2014;**289**:10488–501.

290. Gu KN, Bang SY, Lee HS *et al.* Deletion at 2q14.3 is associated with worse

response to TNF-α blockers in patients with rheumatoid arthritis. *Arthritis Res Ther* 2019;**21**:3–9.

291. Chanalaris A, Clarke H, Guimond SE *et al.* Heparan sulfate proteoglycan synthesis is dysregulated in human osteoarthritic cartilage. *Am J Pathol* 2019;**189**:632–47.

292. Wang Z, Lin L, Zhang JS *et al.* Editorial: The fibroblast growth factor signaling pathway in metabolic regulation, development, disease, and repair after injury. *Front Pharmacol* 2020;**11**:10–1.

293. Nakayama F, Yasuda T, Umeda S *et al.* Fibroblast growth factor-12 (FGF12) translocation into intestinal epithelial cells is dependent on a novel cell-penetrating peptide domain: Involvement of internalization in the in vivo role of exogenous FGF12. *J Biol Chem* 2011;**286**:25823–34.

294. Minina E, Schneider S, Rosowski M *et al.* Expression of Fgf and Tgf b signaling related genes during embryonic endochondral ossification. 2005;**6**:102–9.

295. Huang QL, Zhou FJ, Wu C Bin *et al.* Circulating biomarkers for predicting infliximab response in rheumatoid arthritis: A systematic bioinformatics analysis. *Med Sci Monit* 2017;**23**:1849–55.

296. Williams SE, Brown TI, Roghanian A *et al.* SLPI and elafin: One glove, many fingers. *Clin Sci* 2006;**110**:21–35.

297. Fransès RE, McWilliams DF, Mapp PI *et al.* Osteochondral angiogenesis and

increased protease inhibitor expression in OA. *Osteoarthr Cartil* 2010;**18**:563–71.

298. Majchrzak-Gorecka M, Majewski P, Grygier B *et al.* Secretory leukocyte protease inhibitor (SLPI), a multifunctional protein in the host defense response. *Cytokine Growth Factor Rev* 2016;**28**:79–93.

299. Cassuto J, Folestad A, Göthlin J *et al.* Concerted actions by MMPs, ADAMTS and serine proteases during remodeling of the cartilage callus into bone during osseointegration of hip implants. *Bone Reports* 2020;**13**:100715.

300. Ichijo H, Rönnstrand L, Miyagawa K *et al.* Purification of transforming growth factor-beta 1 binding proteins from porcine uterus membranes. *J Biol Chem* 1991;**266**:22459–64.

301. Kilpatrick DC, Chalmers JD. Human L-Ficolin (Ficolin-2) and Its Clinical Significance. 2012;**2012**.

302. Garlatti V, Belloy N, Martin L *et al.* Structural insights into the innate immune recognition specificities of L- and H-ficolins. *EMBO J* 2007;**26**:623–33.

303. Ambrosio AR, Bavia L, Borges BS *et al.* Novel findings on the role of ficolins and colectins in the innate response against Leishmania braziliensis. *Acta Trop* 2020;**212**:105673.

304. Angeletti A, Cantarelli C, Petrosyan A *et al.* Loss of decay-accelerating factor triggers podocyte injury and glomerulosclerosis. *J Exp Med* 2020;**217**.

305. Karpus ON, Kiener HP, Niederreiter B *et al.* CD55 deposited on synovial

collagen fibers protects from immune complex-mediated arthritis. *Arthritis Res Ther* 2015;**17**:1–12.

306. Hoek RM, De Launay D, Kop EN *et al.* Deletion of either CD55 or CD97 ameliorates arthritis in mouse models. *Arthritis Rheum* 2010;**62**:1036–42.

307. Jaeger PA, Lucin KM, Britschgi M *et al.* Network-driven plasma proteomics expose molecular changes in the Alzheimer's brain. *Mol Neurodegener* 2016;**11**:31.

308. Galasko D. Expanding the repertoire of biomarkers for Alzheimer's disease: Targeted and non-targeted approaches. *Front Neurol* 2015;**6**:1–13.

309. Capurro M, Martin T, Shi W *et al.* Glypican-3 binds to Frizzled and plays a direct role in the stimulation of canonical Wnt signaling. *J Cell Sci* 2014;**127**:1565–75.

310. Freeman ME, Kanyicska B, Lerant A *et al.* Prolactin: structure, function, and regulation of secretion. *Physiol Rev* 2000;**80**:1523–631.

311. Wilson PJ, Meaney CA, Hopwood JJ *et al.* Sequence of the human iduronate 2-sulfatase (IDS) gene. *Genomics* 1993;**17**:773–5.

312. Brömme D, Li Z, Barnes M *et al.* Human cathepsin V functional expression, tissue distribution, electrostatic surface potential, enzymatic characterization, and chromosomal localization. *Biochemistry* 1999;**38**:2377–85.

313. Lopez S, Stuhl L, Fichelson S *et al.* NACA is a positive regulator of human erythroid-cell differentiation. *J Cell Sci* 2005;**118**:1595–605.

314. Brown GD. Dectin-1: a signalling non-TLR pattern-recognition receptor. *Nat Rev Immunol* 2006;**6**:33–43.

315. Sepasi Tehrani H, Moosavi-Movahedi AA. Catalase and its mysteries. *Prog Biophys Mol Biol* 2018;**140**:5–12.

316. Hezam K, Jiang J, Sun F *et al.* Artemin promotes oncogenicity, metastasis and drug resistance in cancer cells. *Rev Neurosci* 2017;**29**:93–8.

317. Kozhemyakina E, Lassar AB, Zelzer E. A pathway to bone: Signaling molecules and transcription factors involved in chondrocyte development and maturation. *Dev* 2015;**142**:817–31.

318. Mancia F, Shapiro L. ADAM and Eph: How ephrin-signaling Cclls become detached. *Cell* 2005;**123**:185–7.

319. Shao Q, Yang T, Huang H *et al.* Uncoupling of UNC5C with polymerized TUBB3 in microtubules mediates netrin-1 repulsion. *J Neurosci* 2017;**37**:5620–33.

320. Briganti F, Mangani S, Scozzafava A *et al.* Carbonic anhydrase catalyzes cyanamide hydration to urea: Is it mimicking the physiological reaction? *J Biol Inorg Chem* 1999;**4**:528–36.

321. Murín R, Schaer A, Kowtharapu BS *et al.* Expression of 3-hydroxyisobutyrate dehydrogenase in cultured neural cells. *J Neurochem* 2008;**105**:1176–86.

322. Dulberger CL, McMurtrey CP, Hölzemer A *et al.* Human leukocyte antigen F presents peptides and regulates immunity through interactions with NK cell

receptors. *Immunity* 2017;**46**:1018-1029.e7.

323. López-Arbesu R, Ballina-García FJ, Alperi-López M *et al.* MHC class I chain-related gene B (MICB) is associated with rheumatoid arthritis susceptibility. *Rheumatology* 2007;**46**:426–30.

324. Hall ET, Cleverdon ER, Ogden SK. Dispatching sonic hedgehog: molecular mechanisms controlling deployment. *Trends Cell Biol* 2019;**29**:385–95.

325. Mesters JR, Barinka C, Li W *et al.* Structure of glutamate carboxypeptidase II, a drug target in neuronal damage and prostate cancer. *EMBO J* 2006;**25**:1375–84.

326. Fujita M, Zhu K, Fujita CK *et al.* Proinflammatory secreted phospholipase A2 type IIA (sPLA-IIA) induces integrin activation through direct binding to a newly identified binding site (site 2) in integrins αvβ3, α4β1, and α5β1. *J Biol Chem* 2015;**290**:259–71.

327. Goepfert A, Lehmann S, Blank J *et al.* Structural analysis reveals that the cytokine IL-17F forms a homodimeric complex with receptor IL-17RC to drive IL-17RA-independent signaling. *Immunity* 2020;**52**:499-512.e5.

328. Oberlies J, Watzl C, Giese T *et al.* Regulation of NK cell function by human granulocyte arginase. *J Immunol* 2009;**182**:5259–67.

329. Surmann-Schmitt C, Widmann N, Dietz U *et al.* Wif-1 is expressed at cartilage-mesenchyme interfaces and impedes Wnt3a-mediated inhibition of chondrogenesis. *J Cell Sci* 2009;**122**:3627–37.

330. D'Amico F, Skarmoutsou E, Granata M *et al.* S100A7: A rAMPing up AMP molecule in psoriasis. *Cytokine Growth Factor Rev* 2016;**32**:97–104.

331. Lei X, Shi F, Basu D *et al.* Proteolytic processing of angiopoietin-like protein 4 by proprotein convertases modulates its inhibitory effects on lipoprotein lipase activity. *J Biol Chem* 2011;**286**:15747–56.

332. Kovács KB, Pataki I, Bárdos H *et al.* Molecular characterization of p.Asp77Gly and the novel p.Ala163Val and p.Ala163Glu mutations causing protein C deficiency. *Thromb Res* 2015;**135**:718–26.

333. Kanarek N, Keys HR, Cantor JR *et al.* Histidine catabolism is a major determinant of methotrexate sensitivity. *Nature* 2018;**559**:632–6.

334. Nolan K, Kattamuri C, Luedeke DM *et al.* Structure of neuroblastoma suppressor of tumorigenicity 1 (NBL1) insights for the functional variability across bone morphogenetic protein (BMP) antagonists. *J Biol Chem* 2015;**290**:4759–71.

335. Wu Y, Bressette D, Carrell JA *et al.* Tumor necrosis factor (TNF) receptor superfamily member TACI is a high affinity receptor for TNF family members APRIL and BLyS. *J Biol Chem* 2000;**275**:35478–85.

336. Chen JW, Dodia C, Feinstein SI *et al.* 1-Cys peroxiredoxin, a bifunctional enzyme with glutathione peroxidase and phospholipase A2 activities. *J Biol Chem* 2000;**275**:28421–7.

337. Kelly JD, Haldeman BA, Grant FJ *et al.* Platelet-derived growth factor (PDGF)

stimulates PDGF receptor subunit dimerization and intersubunit trans-phosphorylation. *J Biol Chem* 1991;**266**:8987–92.

338. Scafuri B, Varriale A, Facchiano A *et al.* Binding of mycotoxins to proteins involved in neuronal plasticity: A combined in silico/wet investigation. *Sci Rep* 2017;**7**:1–11.

339. Zlotnik A, Tsesis S, Gruenbaum BF *et al.* Relationship between glutamate, GOT and GPT levels in maternal and fetal blood: A potential mechanism for fetal neuroprotection. *Early Hum Dev* 2012;**88**:773–8.

340. Logan T, Bendor J, Toupin C *et al.* α-Synuclein promotes dilation of the exocytotic fusion pore. *Nat Neurosci* 2017;**20**:681–9.

341. Woods CG, Stricker S, Seemann P *et al.* Mutations in WNT7A cause a range of limb malformations, including fuhrmann syndrome and Al-Awadi/Raas-Rothschild/Schinzel phocomelia syndrome. *Am J Hum Genet* 2006;**79**:402–8.

342. Krijgsveld J, Zaat SAJ, Meeldijk J *et al.* Thrombocidins, microbicidal proteins from human blood platelets, are C- terminal deletion products of CXC chemokines. *J Biol Chem* 2000;**275**:20374–81.

343. Pilatova K, Greplova K, Demlova R *et al.* Role of platelet chemokines, PF-4 and CTAP-III, in cancer biology. *J Hematol Oncol* 2013;**6**:2–9.

344. Loughney K, Hill TR, Florio VA *et al.* Isolation and characterization of cDNAs encoding PDE5A, a human cGMP-binding, cGMP-specific 3',5'-cyclic nucleotide

phosphodiesterase. *Gene* 1998;**216**:139–47.

345. Shrivastava A, Radziejewski C, Campbell E *et al.* An orphan receptor tyrosine kinase family whose members serve as nonintegrin collagen receptors. *Mol Cell* 1997;**1**:25–34.

346. Benedict CA, Ware CF. Trail: Not just for tumors anymore? *J Exp Med* 2012;**209**:1903–6.

347. Fosang AJ, Rogerson FM, East CJ *et al.* ADAMTS-5: The story so far. *Eur Cells Mater* 2008;**15**:11–26.

348. Martel-Pelletier J, Tardif G, Rousseau Trépanier J *et al.* The ratio adipsin/MCP-1 is strongly associated with structural changes and CRP/MCP-1 with symptoms in obese knee osteoarthritis subjects: data from the Osteoarthritis Initiative. *Osteoarthr Cartil* 2019;**27**:1163–73.

349. Wang M, Sampson ER, Jin H *et al.* MMP13 is a critical target gene during the progression of osteoarthritis. *Arthritis Res Ther* 2013;**15**:1–11.

350. Zeng GQ, Chen AB, Li W *et al.* High MMP-1, MMP-2, and MMP-9 protein levels in osteoarthritis. *Genet Mol Res* 2015;**14**:14811–22.

351. Chevalier X, Conrozier T, Gehrmann M *et al.* Tissue inhibitor of metalloprotease-1 (TIMP-1) serum level may predict progression of hip osteoarthritis. *Osteoarthr Cartil* 2001;**9**:300–7.

352. Matsuo K, Otaki N. Bone cell interactions through Eph/ephrin: Bone

modeling, remodeling and associated diseases. *Cell Adhes Migr* 2012;**6**:148–56.

353. Vogel W, Gish GD, Alves F *et al.* The discoidin domain receptor tyrosine kinases are activated by collagen. *Mol Cell* 1997;**1**:13–23.

354. Wang J, Lü H, Liu X *et al.* Functional analysis of discoidin domain receptor 2 in synovial fibroblasts in rheumatoid arthritis. *J Autoimmun* 2002;**19**:161–8.

355. Xu L, Peng H, Wu D *et al.* Activation of the discoidin domain receptor 2 induces expression of matrix metalloproteinase 13 associated with osteoarthritis in mice. *J Biol Chem* 2005;**280**:548–55.

356. Xu L, Polur I, Servais JM *et al.* Intact pericellular matrix of articular cartilage is required for unactivated discoidin domain receptor 2 in the mouse model. *Am J Pathol* 2011;**179**:1338–46.

357. Manning LB, Li Y, Chickmagalur NS *et al.* Discoidin domain receptor 2 as a potential therapeutic target for development of disease-modifying osteoarthritis drugs. *Am J Pathol* 2016;**186**:3000–10.

358. Zhang Y, Su J, Yu J *et al.* An essential role of discoidin domain receptor 2 (DDR2) in osteoblast differentiation and chondrocyte maturation via modulation of Runx2 activation. *J Bone Miner Res* 2011;**26**:604–17.

359. Khosravi R, Sodek KL, Faibish M *et al.* Collagen advanced glycation inhibits its Discoidin Domain Receptor 2 (DDR2)-mediated induction of lysyl oxidase in osteoblasts. *Bone* 2014;**58**:33–41.

360. Troeberg L, Lazenbatt C, Anower-E-Khuda MF *et al.* Sulfated glycosaminoglycans control the extracellular trafficking and the activity of the metalloprotease inhibitor timp-3. *Chem Biol* 2014;**21**:1300–9.

361. Watson PH, Leygue ER, Murphy LC. Psoriasin (S100A7). *Int J Biochem Cell Biol* 1998;**30**:567–71.

362. Wolf R, Howard OMZ, Dong H *et al.* Chemotactic activity of S100A7 (psoriasin) is mediated by RAGE and potentiates inflammation with highly homologous but functionally distinct S100A15. *J Immunol* 2008;**181**:1499–506.

363. Kataoka K, Ono T, Murata H *et al.* S100A7 promotes the migration and invasion of osteosarcoma cells via the receptor for advanced glycation end products. *Oncol Lett* 2012;**3**:1149–53.

364. Zreiqat H, Belluoccio D, Smith MM *et al.* S100A8 and S100A9 in experimental osteoarthritis. *Arthritis Res Ther* 2010;**12**:1–13.

365. Baillet A, Trocmé C, Berthier S *et al.* Synovial fluid proteomic fingerprint: S100A8, S100A9 and S100A12 proteins discriminate rheumatoid arthritis from other inflammatory joint diseases. *Rheumatology (Oxford)* 2010;**49**:671–82.

366. Safa A, Bagherifard A, Hadi Al-Baseesee H *et al.* Serum calprotectin as a blood-based biomarker for monitoring knee osteoarthritis at early but not late stages. *Cartilage* 2020:1947603520961161.

367. Saftig P, Lichtenthaler SF. The alpha secretase ADAM10: A metalloprotease

with multiple functions in the brain. *Prog Neurobiol* 2015;**135**:1–20.

368. Pellet P, Vaneensberghe C, Debré P *et al.* MIC genes in non-human primates. *Eur J Immunogenet* 1999;**26**:239–41.

369. Smith TM, Tharakan A, Martin RK. Targeting ADAM10 in cancer and autoimmunity. *Front Immunol* 2020;**11**:1–12.

370. Hu H, Li H, Feng X. Downregulation of lncRNA NCK1-AS1 inhibits cancer cell migration and invasion in nasopharyngeal carcinoma by upregulating mir-135a. *Cancer Manag Res* 2019;**11**:10531–7.

371. Groux-Degroote S, Krzewinski-Recchi MA, Cazet A *et al.* IL-6 and IL-8 increase the expression of glycosyltransferases and sulfotransferases involved in the biosynthesis of sialylated and/or sulfated LewisX epitopes in the human bronchial mucosa. *Biochem J* 2008;**410**:213–23.

372. Xie J, Naito M, Maeyama A. Intracapsular pressure and interleukin-1β cytokine in hips with acetabular dysplasia. *Acta Orthop* 2010;**81**:189–92.

373. Abrams GD, Safran MR, Shapiro LM *et al.* Fibronectin-aggrecan complex as a marker for cartilage degradation in non-arthritic hips. *Knee Surgery, Sport Traumatol Arthrosc* 2014;**22**:768–73.

374. Ren G, Lutz I, Railton P *et al.* Serum and synovial fluid cytokine profiling in hip osteoarthritis: Distinct from knee osteoarthritis and correlated with pain. *BMC Musculoskelet Disord* 2018;**19**:1–11.

375. Higuera CA, Zmistowski B, Malcom T *et al.* Synovial fluid cell count for diagnosis of chronic periprosthetic hip infection. *J Bone Jt Surg - Am Vol* 2017;**99**:753–9.

376. Snelling SJB, Bas S, Puskas GJ *et al.* Presence of IL-17 in synovial fluid identifies a potential inflammatory osteoarthritic phenotype. *PLoS One* 2017;**12**:1–13.

377. Catelas I, Lehoux EA, Ning Z *et al.* Differential proteomic analysis of synovial fluid from hip arthroplasty patients with a pseudotumor vs. periprosthetic osteolysis. *J Orthop Res* 2018;**36**:1849–59.

378. Abe H, Sakai T, Ando W *et al.* Synovial joint fluid cytokine levels in hip disease. *Rheumatol (United Kingdom)* 2014;**53**:165–72.

379. Scanzello CR. Role of low-grade inflammation in osteoarthritis. *Curr Opin Rheumatol* 2017;**29**:79–85.

380. Greene MA, Loeser RF. Aging-related inflammation in osteoarthritis. *Osteoarthr Cartil* 2015;**23**:1966–71.

381. van den Bosch MHJ. Inflammation in osteoarthritis: is it time to dampen the alarm(in) in this debilitating disease? *Clin Exp Immunol* 2019;**195**:153–66.

382. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet (London, England)* 2015;**386**:743–

800.

383. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. *Lancet* 2019;**393**:1745–59.

384. *National Joint Registry. 17th Annual Report.*, 2020.

385. Culliford D, Maskell J, Judge A *et al.* Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink. *Osteoarthr Cartil* 2015;**23**:594–600.

386. Conrad N, Judge A, Tran J *et al.* Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *Lancet* 2018;**391**:572–80.

387. Herrett E, Gallagher AM, Bhaskaran K *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;**44**:827–36.

388. Ferguson RJ, Prieto-Alhambra D, Walker C *et al.* Validation of hip osteoarthritis diagnosis recording in the UK Clinical Practice Research Datalink. *Pharmacoepidemiol Drug Saf* 2019;**28**:187–93.

389. Thorlund JB, Turkiewicz A, Prieto-Alhambra D *et al.* Opioid use in knee or hip osteoarthritis: a region-wide population-based cohort study. *Osteoarthr Cartil* 2019;**27**:871–7.

390. DeMik DE, Bedard NA, Dowdle SB *et al.* Are we still prescribing opioids for osteoarthritis? *J Arthroplasty* 2017;**32**:3578-3582.e1.

391. Krebs EE, Gravely A, Nugent S *et al.* Effect of opioid vs nonopioid medications on pain-related function in patients with chronic back pain or hip or knee osteoarthritis pain the SPACE randomized clinical trial. *JAMA - J Am Med Assoc* 2018;**319**:872–82.

392. Yip K, Oettinger J. Why are we still using opioids for osteoarthritis? *Int J Clin Pract* 2020;**74**:1–4.

393. Van Tunen JAC, Peat G, Bricca A *et al.* Association of osteoarthritis risk factors with knee and hip pain in a population-based sample of 29-59 year olds in Denmark: A cross-sectional analysis. *BMC Musculoskelet Disord* 2018;**19**:1–11.

394. Wills AK, Black S, Cooper R *et al.* Life course body mass index and risk of knee osteoarthritis at the age of 53 years: Evidence from the 1946 British birth cohort study. *Ann Rheum Dis* 2012;**71**:655–60.

395. Martinez-Cano JP, Zamudio-Castilla L, Chica J *et al.* Body mass index and knee arthroplasty. *J Clin Orthop Trauma* 2020;**11**:S711–6.

396. Anis HK, Emara AK, Klika AK *et al.* The potential effects of imposing a body mass index threshold on patient-reported outcomes after total knee arthroplasty. *J Arthroplasty* 2020.

397. Lash H, Hooper G, Hooper N *et al.* Should a patients BMI status be used to restrict access to total hip and knee arthroplasty? Functional outcomes of arthroplasty relative to BMI - single centre retrospective review. *Open Orthop J*

2013;**7**:594–9.

398. Conaghan PG, Dickson J, Grant RL. Guidelines: Care and management of osteoarthritis in adults: Summary of NICE guidance. *Bmj* 2008;**336**:502–3.

399. Reyes C, Garcia-Gil M, Elorza JM *et al.* Socio-economic status and the risk of developing hand, hip or knee osteoarthritis: A region-wide ecological study. *Osteoarthr Cartil* 2015;**23**:1323–9.

400. Judge A, Welton NJ, Sandhu J *et al.* Modeling the need for hip and knee replacement surgery. Part 1. A two-stage cross-cohort approach. *Arthritis Care Res* 2009;**61**:1657–66.

401. Meints SM, Edwards RR. Evaluating psychosocial contributions to chronic pain outcomes. *Prog Neuro-Psychopharmacology Biol Psychiatry* 2018;**87**:168–82.

402. Abram SGF, Judge A, Beard DJ *et al.* Rates of knee arthroplasty within one-year of undergoing arthroscopic partial meniscectomy in England: temporal trends, regional and age-group variation in conversion rates. *Osteoarthr Cartil* 2019;**27**:1420–9.

403. Palmer AJR, Malak TT, Broomfield J *et al.* Past and projected temporal trends in arthroscopic hip surgery in England between 2002 and 2013. *BMJ open Sport Exerc Med* 2016;**2**:e000082.

404. Prieto-Alhambra D, Javaid MK, Judge A *et al.* Association between bisphosphonate use and implant survival after primary total arthroplasty of the

knee or hip: population based retrospective cohort study. *BMJ* 2011;**343**:d7222.

405. Watt FE, Hamid B, Garriga C *et al.* The molecular profile of synovial fluid changes upon joint distraction and is associated with clinical response in knee osteoarthritis. *Osteoarthr Cartil* 2020;**28**:324–33.

406. Westwood S, Liu B, Baird AL *et al.* The influence of insulin resistance on cerebrospinal fluid and plasma biomarkers of Alzheimer's pathology. *Alzheimer's Res Ther* 2017;**9**:1–11.

407. Sattlecker M, Kiddle SJ, Newhouse S *et al.* Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. *Alzheimer's Dement* 2014;**10**:724–34.

408. He C, Zheng S, Luo Y *et al.* Exosome theranostics: Biology and translational medicine. *Theranostics* 2018;**8**:237–55.

409. Kaushik S, Cuervo AM. Proteostasis and aging. *Nat Med* 2015;**21**:1406–15.

410. Li Z, Wang Y, Xiao K *et al.* Emerging role of exosomes in the joint diseases. *Cell Physiol Biochem* 2018;**47**:2008–17.

411. Gilligan KE, Dwyer RM. Engineering exosomes for cancer therapy. *Int J Mol Sci* 2017;**18**.

412. Luo Q, Guo D, Liu G *et al.* Exosomes from MiR-126-overexpressing adscs are therapeutic in relieving acute myocardial ischaemic injury. *Cell Physiol Biochem* 2018;**44**:2105–16.

413. Cheng TH, Yoon SH, Lee P *et al.* Knee synovial fluid complement C3-β chain levels correlate with clinical symptoms of knee osteoarthritis. *Int J Rheum Dis* 2020;**23**:569–75.

414. SOMAlogic. SOMAmer Reagents Generated to Human Proteins. 2015:1–20.

Appendix A

A.1 shows an example of the Oxford Musculoskeletal Biobank (OMB) patient consent form, completed with the patient prior to harvesting of synovial fluid samples. Approved by research ethics committee and compliant with Human Tissue Association (HTA) standards. A copy was retained by the patient, a copy was held in the biobank and a copy was placed in the patient's medical health record and cross-checked at the point of sampling.

A.2 shows an example of the femoroacetabular impingement trial (FAIT) study consent form which is completed with the patient prior inclusion in the trial. SF was harvested during hip arthroscopy in those participants randomised to the operative arm of the trial. Approved by research ethics committee and compliant with Human Tissue Association (HTA) standards. A copy was retained by the patient, a copy was held in the trial master file and a copy was placed in the patient's medical health record and cross-checked at the point of sampling.

A.1 – Sample OMB Patient Consent Form

OMB-F 002 v7.0 Dated 07/03/2014 Approved 00/00/0000

Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences.
Nuffield Orthopaedic Centre, Headington
Oxford, OX3 7HE United Kingdom
Telephone: 01865 737418 Fax: 01865 737640
E-mail: omb-manager@ndorms.ox.ac.uk

Title of Project: **Oxford Musculoskeletal Biobank**
Oxford REC C Reference: **09/H0606/11**

[Patient sticky here]

OXFORD MUSCULOSKELETAL BIOBANK (OMB) CONSENT FORM

	If you would like to take part in this research, please initial the boxes below and sign the form		
1	I confirm that I have read, understood and have had time to consider the information sheet Version 5.0 and have been given a copy to keep. I have had the opportunity to ask questions about this project.		Initial
2	I understand that any samples of blood, urine and/or tissue taken for research are gifted to the OMB as detailed in the patient information sheet.		Initial
3	I understand that any prosthetic implant removed during surgery may be gifted to the OMB as detailed in the patient information sheet. OMB may destroy gifted implants without further permission being sought from me when no longer needed for research.		Initial
4	I understand that OMB staff can collect and store information, including x-rays and scans, from my health care records for research. I understand the OMB will keep my information confidential.		Initial
5	I understand that any samples/implants I have gifted to the OMB, and the information gathered about me, can be stored by the OMB for possible use in future projects. I understand that some of these projects may be carried out by researchers other than the University of Oxford, including researchers working for commercial companies. This may include research undertaken in the European Union and the rest of the world. Information will only be passed to researchers in an anonymous way that protects my identity.		Initial
6	I understand that any samples/implants I have gifted to the OMB, and the information gathered about me, is for the purpose of research. If a commercial product were developed as a result of this research, I will not profit financially from such a product.		Initial
7	I understand that results from research tests on samples might be medically important to me. I understand that my hospital consultant and GP will be informed, and that relevant experimental findings can be discussed with me.		Initial
8	I understand how the samples/implants and information gathered about me will be taken, that my participation is voluntary, and that I am free at any time to withdraw my permission for the storage and distribution of the samples/implants and information, providing that they have not already been used in research.		Initial
	Genetic research – please initial the box below		
9	I understand that samples may be used in genetic research aimed at understanding the genetic influences on diseases, and that the results of these investigations are unlikely to have any implications for me personally.		Initial
	Options – please make your wishes clear and initial the boxes below		
10	I agree that the OMB may contact me in the future in order to follow my progress during/after treatment.	YES / NO	Initial
11	I agree that the OMB may contact me in the future to inform me of other research projects that I may be interested in. The OMB will not pass on my contact details without my explicit permission.	YES / NO	Initial

Name of patient	Signature	Date
I have discussed the study with this patient who has agreed to give informed consent.		
Name of OMB-registered consenter	Signature	Date

Consent Form: White copy (original) to Medical Notes, pink copy to OMB Consent Files, blue copy to patient.

SAMPLE

A.2 – Sample FAIT Consent Form

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences
University of Oxford
Nuffield Orthopaedic Centre
Oxford
OX3 7LD
Telephone: 01865 223488 Email: fait@ndorms.ox.ac.uk

Patient Consent Form

REC Reference: 13/SC/0154

A Randomised Controlled Trial of Surgical versus Non-Surgical Treatment of Femoroacetabular Impingement

	Please Initial
I confirm that I have read and understand the information sheet for the above study version ____ date _____. I have had the opportunity to consider the information, ask questions, and have had these answered satisfactorily.	☐
I understand that my participation is voluntary and that I am free to withdraw at any time, without giving reason, and without my medical or legal rights being affected.	☐
I understand that personal data collected during the study will be held securely in password-protected computers or lockable cabinets, and used solely for the purpose of contacting me about this or future studies. I give permission for these individuals to see my personal data.	☐
I understand that any blood or urine collected may be retained for analysis as part of this or future research. Should I undergo surgery while participating in this study, tissue that is removed during the routine treatment of my condition may also be retained. I consider these samples a gift to the University of Oxford and I understand I will not gain any direct personal benefit.	☐
I agree to having x-rays of my hip joints during this study. I understand that in doing so I will be exposed to a small dose of ionising radiation equivalent to the increased exposure of moving from Oxfordshire to Cornwall for 1 year.	☐
I am not pregnant to the best of my knowledge.	☐
I agree for data and samples collected in this study to be given to researchers, including those working outside of the EU, for assistance with analysis and data interpretation. I understand that any of my data that leaves our centre will be fully anonymised so that I cannot be identified.	☐
I am not I understand that sections of my medical notes will be used to extract the data for the research; I give permission to the study team to do so.	☐
I understand that sections of my medical notes may be accessed by responsible members of the University of Oxford and the NHS Trust for the purpose of monitoring or audit. I give permission for these individuals to access my records.	☐
I agree to take part in the above study.	☐
I agree to my GP being informed of my participation in the study.	☐
I would be happy to be contacted again for further osteoarthritis studies. *(circle one)* Yes / No	

Name of Participant:	Signature:	Date:
Name of Researcher:	Signature:	Date:

Appendix B

B.1 – SOMAscan Assay 1310 Content

The summary document in this section shows the proteins contained in the 1310 panel SOMAscan, used for the experiments in this thesis, including the full analyte name and UNIPROT ID. Reproduced from SOMAscan menu document "SOMAscan assay 1.3k content" (SSM-045) [414].

The exact content of any pre-specified menu offered by SomaLogic may be altered on an ongoing basis, including the addition of SOMAmer® reagents as they are created, and the removal of others if deemed necessary, as we continue to improve the performance of the SOMAscan assay. However, the client will know the exact content at the time of study contracting. SomaLogic reserves the right to alter the menu at any time in its sole discretion.

Number	SOMAmer SeqID	Analyte Name	UniProt ID
1	5227-60_3	[Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 1, mitochondrial	Q15118
2	14156-33_3	14-3-3 protein beta/alpha	P31946
3	14157-21_3	14-3-3 protein epsilon	P62258
4	4179-57_3	14-3-3 protein family	P31946, P62258, P61981, Q04917, P27348, P63104, P31947
5	4829-43_6	14-3-3 protein sigma	P31947
6	7625-27_3	14-3-3 protein theta	P27348
7	5858-6_5	14-3-3 protein zeta/delta	P63104
8	4995-16_1	15-hydroxyprostaglandin dehydrogenase [NAD(+)]	P15428
9	4563-61_2	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1	P19174
10	10361-25_3	2'-5'-oligoadenylate synthase 1	P00973
11	3898-5_2	26S proteasome non-ATPase regulatory subunit 7	P51665
12	5230-99_3	3-hydroxy-3-methylglutaryl-coenzyme A reductase	P04035
13	4217-49_3	3-hydroxyacyl-CoA dehydrogenase type-2	Q99714
14	5861-78_3	3-hydroxyanthranilate 3,4-dioxygenase	P46952
15	4693-72_2	3-hydroxyisobutyrate dehydrogenase, mitochondrial	P31937
16	4460-8_2	3-phosphoinositide-dependent protein kinase 1	O15530
17	5026-66_1	40S ribosomal protein S3	P23396
18	5484-63_3	40S ribosomal protein S3a	P61247
19	3864-5_2	40S ribosomal protein S7	P62081
20	4957-1_2	40S ribosomal protein SA	P08865
21	2682-68_1	60 kDa heat shock protein, mitochondrial	P10809
22	4187-49_2	6-phosphogluconate dehydrogenase, decarboxylating	P52209
23	4160-49_1	72 kDa type IV collagenase	P08253
24	3174-2_1	A disintegrin and metalloproteinase with thrombospondin motifs 1	Q9UHI8
25	3175-51_5	A disintegrin and metalloproteinase with thrombospondin motifs 13	Q76LX8
26	4533-76_2	A disintegrin and metalloproteinase with thrombospondin motifs 15	Q8TE58
27	2809-25_2	A disintegrin and metalloproteinase with thrombospondin motifs 4	O75173
28	3168-8_2	A disintegrin and metalloproteinase with thrombospondin motifs 5	Q9UNA0
29	5261-13_3	Abelson tyrosine-protein kinase 2	P42684
30	4771-10_3	Acid sphingomyelinase-like phosphodiesterase 3a	Q92484
31	4194-26_3	Acidic leucine-rich nuclear phosphoprotein 32 family member B	Q92688
32	3758-68_3	Activated Protein C	P04070
33	2806-49_2	Activin receptor type-1B	P36896
34	4976-57_1	Adapter molecule crk	P46108
35	5012-67_1	Adenylate kinase isoenzyme 1	P00568
36	5023-23_1	Adenylosuccinate lyase	P30566
37	4546-27_3	Adhesion G protein-coupled receptor E2	Q9UHX3
38	4551-72_3	Adhesion G-protein coupled receptor G5	Q8IZF4
39	3554-24_1	Adiponectin	Q15848
40	11513-92_3	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1	P28907
41	4535-50_2	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 2	Q10588
42	14115-34_3	Adrenomedullin	P35318
43	4125-52_2	Advanced glycosylation end product-specific receptor, soluble	Q15109
44	4763-31_3	Afamin	P43652
45	4188-1_2	Aflatoxin B1 aldehyde reductase member 2	O43488
46	3280-49_2	Aggrecan core protein	P16112
47	2813-11_2	Agouti-related protein	O00253
48	3839-60_2	AH receptor-interacting protein	O00170
49	3709-4_2	Alanine aminotransferase 1	P24298
50	4192-10_2	Alcohol dehydrogenase [NADP(+)]	P14550
51	2795-23_3	Alkaline phosphatase, tissue-nonspecific isozyme	P05186
52	2849-49_1	Allograft inflammatory factor 1	P55008
53	4549-78_2	Alpha-(1,3)-fucosyltransferase 5	Q11128
54	2879-9_2	Alpha-1-antichymotrypsin	P01011
55	4153-11_2	Alpha-1-antichymotrypsin complex	P01011
56	3580-25_8	Alpha-1-antitrypsin	P01009
57	3024-18_2	Alpha-2-antiplasmin	P08697
58	3581-53_3	Alpha-2-HS-glycoprotein	P02765
59	3708-62_1	Alpha-2-macroglobulin	P01023
60	3640-14_3	alpha-2-macroglobulin receptor-associated protein	P30533
61	11105-171_3	Alpha-enolase	P06733
62	5792-8_2	alpha-Fetoprotein	P02771
63	3169-70_2	Alpha-L-iduronidase	P35475

Number	SOMAmer SeqID	Analyte Name	UniProt ID
64	4292-5_3	Alpha-soluble NSF attachment protein	P54920
65	8458-111_3	Alpha-synuclein	P37840
66	3343-1_4	Aminoacylase-1	Q03154
67	5183-53_3	AMP Kinase (alpha1beta1gamma1)	Q13131 Q9Y478 P54619
68	5245-40_5	AMP Kinase (alpha2beta2gamma1)	P54646 O43741 P54619
69	2970-60_2	Amphiregulin	P15514
70	14134-49_3	Amphoterin-induced protein 2	Q86SJ2
71	3171-57_2	Amyloid beta A4 protein	P05067
72	4874-3_1	Angiogenin	P03950
73	2811-27_1	Angiopoietin-1	Q15389
74	3773-15_4	Angiopoietin-1 receptor, soluble	Q02763
75	2602-2_2	Angiopoietin-2	O15123
76	2500-2_3	Angiopoietin-4	Q9Y264
77	3281-19_1	Angiopoietin-related protein 3	Q9Y5C1
78	3796-79_2	Angiopoietin-related protein 4	Q9BY76
79	3710-49_2	Angiostatin	P00747
80	2805-6_2	Angiotensin-converting enzyme 2	Q9BYF1
81	3484-60_2	Angiotensinogen	P01019
82	7624-19_3	Ankyrin-2	Q01484
83	4960-72_1	Annexin A1	P04083
84	4961-17_1	Annexin A2	P07355
85	14158-17_3	Annexin A5	P08758
86	5335-73_3	Annexin A6	P08133
87	4959-2_1	Anterior gradient protein 2 homolog	O95994
88	4413-3_2	Antileukoproteinase	P03973
89	3344-60_4	Antithrombin-III	P01008
90	2750-3_2	Apolipoprotein A-I	P02647
91	2797-56_2	Apolipoprotein B	P04114
92	4712-28_2	Apolipoprotein D	P05090
93	2418-55_9	Apolipoprotein E	P02649
94	5312-49_3	Apolipoprotein E (isoform E2)	P02649
95	2937-10_2	Apolipoprotein E (isoform E3)	P02649
96	2938-55_2	Apolipoprotein E (isoform E4)	P02649
97	11510-31_3	Apolipoprotein L1	O14791
98	14125-5_3	Apolipoprotein M	O95445
99	3412-7_1	Apoptosis regulator Bcl-2	P10415
100	8447-11_3	Appetite-regulating hormone	Q9UBU3
101	5867-60_3	Arginase-1	P05089
102	3538-26_1	Aromatic-L-amino-acid decarboxylase	P20711
103	2939-10_2	Artemin	Q5T4W7
104	3583-54_4	Arylsulfatase A	P15289
105	3172-28_2	Arylsulfatase B	P15848
106	5452-71_3	Asialoglycoprotein receptor 1	P07306
107	4912-17_1	Aspartate aminotransferase, cytoplasmic	P17174
108	4965-27_1	ATP synthase subunit beta, mitochondrial	P06576
109	9202-309_3	ATP synthase subunit O, mitochondrial	P48047
110	5460-60_3	ATP-dependent RNA helicase DDX19B	Q9UMR2
111	5443-62_2	Atrial natriuretic factor	P01160
112	3875-62_1	AT-rich interactive domain-containing protein 3A	Q99856
113	3091-70_2	Aurora kinase A	O14965
114	3346-72_2	Aurora kinase B	Q96GD4
115	2751-16_2	Azurocidin	P20160
116	4126-22_1	Bactericidal permeability-increasing protein	P17213
117	4973-18_1	Baculoviral IAP repeat-containing protein 3	Q13489
118	3472-40_2	Baculoviral IAP repeat-containing protein 5	O15392
119	4561-65_2	Baculoviral IAP repeat-containing protein 7 Isoform beta	Q96CA5
120	2816-50_2	Basal Cell Adhesion Molecule	P50895
121	3585-54_4	Basigin	P35613
122	13111-79_3	B-cell lymphoma 6 protein	P41182
123	2891-1_3	B-cell receptor CD22	P20273
124	5870-23_2	Bcl2-associated agonist of cell death	Q92934
125	4423-77_2	Bcl-2-like protein 1	Q07817
126	13097-11_3	Bcl-2-like protein 2	Q92843
127	3413-50_2	Bcl-2-related protein A1	Q16548
128	4866-59_2	BDNF/NT-3 growth factors receptor	Q16620
129	3485-28_2	Beta-2-microglobulin	P61769
130	3347-9_2	beta-adrenergic receptor kinase 1	P25098
131	5456-59_2	Beta-Ala-His dipeptidase	Q96KN2

Number	SOMAmer SeqID	Analyte Name	UniProt ID
132	13088-397_3	Betacellulin	P35070
133	2558-51_3	Beta-endorphin	P01189
134	5801-72_3	beta-nerve growth factor	P01138
135	5798-3_3	BH3-interacting domain death agonist	P55957
136	3284-75_1	Biglycan	P21810
137	3654-27_4	BMP-binding endothelial regulator protein	Q8N8U9
138	3348-49_2	Bone morphogenetic protein 1	P13497
139	3587-53_2	Bone morphogenetic protein 10	O95393
140	8459-10_3	Bone morphogenetic protein 6	P22004
141	2972-57_2	Bone morphogenetic protein 7	P18075
142	4859-6_1	Bone morphogenetic protein receptor type-1A	P36894
143	4862-63_1	Bone morphogenetic protein receptor type-2	Q13873
144	3415-61_2	Bone sialoprotein 2	P21815
145	3723-1_2	Brain natriuretic peptide 32	P16860
146	2421-7_3	Brain-derived neurotrophic factor	P23560
147	4534-10_2	Brain-specific serine protease 4	Q9GZN4
148	5262-57_3	Breast cancer anti-estrogen resistance protein 3	O75815
149	3461-58_1	Brevican core protein	Q96GW7
150	4328-2_2	Brother of CDO	Q9BWV1
151	4900-8_1	C3a anaphylatoxin	P01024
152	2755-8_2	C3a anaphylatoxin des Arginine	P01024
153	2851-63_3	C5a anaphylatoxin	P01031
154	2501-51_3	Cadherin-1	P12830
155	3656-9_1	Cadherin-12	P55289
156	5410-53_3	Cadherin-15	P55291
157	3797-1_1	Cadherin-2	P19022
158	2643-57_2	Cadherin-3	P22223
159	2819-23_2	Cadherin-5	P33151
160	3591-51_3	Cadherin-6	P55285
161	4903-72_1	Calcineurin	Q08209 P63098
162	3657-74_3	Calcineurin subunit B type 1	P63098
163	5253-1_1	Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1A	P54750
164	3877-67_2	Calcium/calmodulin-dependent protein kinase kinase 1	Q8N5S9
165	3592-4_3	Calcium/calmodulin-dependent protein kinase type 1	Q14012
166	3418-12_2	Calcium/calmodulin-dependent protein kinase type 1D	Q8IU85
167	3350-53_2	Calcium/calmodulin-dependent protein kinase type II subunit alpha	Q9UQM7
168	3351-1_1	Calcium/calmodulin-dependent protein kinase type II subunit beta	Q13554
169	3419-49_2	Calcium/calmodulin-dependent protein kinase type II subunit delta	Q13557
170	2449-1_4	Calcium-dependent phospholipase A2	P39877
171	2668-70_2	Calpain I	P07384 P04632
172	3026-5_2	Calpastatin	P20810
173	5264-65_3	Calreticulin	P27797
174	3466-8_2	cAMP-dependent protein kinase catalytic subunit alpha	P17612
175	4963-19_1	cAMP-regulated phosphoprotein 19	P56211
176	5255-22_3	cAMP-specific 3',5'-cyclic phosphodiesterase 4D	Q08499
177	4469-78_2	Carbohydrate sulfotransferase 15	Q7LFX5
178	4428-1_2	Carbohydrate sulfotransferase 2	Q9Y4C5
179	4429-51_2	Carbohydrate sulfotransferase 6	Q9GZX3
180	4969-2_1	Carbonic anhydrase 1	P00915
181	3420-21_2	Carbonic anhydrase 13	Q8N1Q1
182	4970-55_1	Carbonic anhydrase 2	P00918
183	3799-11_2	Carbonic anhydrase 3	P07451
184	3177-49_2	Carbonic anhydrase 4	P22748
185	3352-80_3	Carbonic anhydrase 6	P23280
186	3356-50_1	Carbonic anhydrase 7	P43166
187	3798-71_1	Carbonic anhydrase 9	Q16790
188	3289-19_2	Carbonic anhydrase-related protein 10	Q9NS85
189	3518-54_2	Carboxypeptidase B2	Q96IY4
190	5343-74_3	Carboxypeptidase E	P16870
191	2889-37_2	Cardiotrophin-1	Q16619
192	5225-50_3	Casein kinase II 2-alpha:2-beta heterotetramer	P68400 P67870
193	5226-36_3	Casein kinase II 2-alpha':2-beta heterotetramer	P19784 P67870
194	3427-63_2	Casein kinase II subunit alpha	P68400
195	5340-24_3	Caspase-10	Q92851
196	4904-7_1	Caspase-2	P42575
197	3593-72_3	Caspase-3	P42574
198	3488-64_2	Catalase	P04040
199	3061-61_2	Cathepsin B	P07858

Number	SOMAmer SeqID	Analyte Name	UniProt ID
200	5508-62_3	Cathepsin D	P07339
201	3594-6_5	Cathepsin E	P14091
202	9212-22_3	Cathepsin F	Q9UBX1
203	2431-17_3	Cathepsin G	P08311
204	8465-52_3	Cathepsin H	P09668
205	3364-76_2	Cathepsin L2	O60911
206	3181-50_2	Cathepsin S	P25774
207	4971-1_1	Cathepsin Z	Q9UBR2
208	3676-15_3	Cation-independent mannose-6-phosphate receptor	P11717
209	2770-51_2	C-C motif chemokine 1	P22362
210	4144-13_1	C-C motif chemokine 13	Q99616
211	2900-53_3	C-C motif chemokine 14	Q16627
212	3509-1_1	C-C motif chemokine 15	Q16663
213	4913-78_1	C-C motif chemokine 16	O15467
214	3519-3_2	C-C motif chemokine 17	Q92583
215	3044-3_2	C-C motif chemokine 18	P55774
216	4922-13_1	C-C motif chemokine 19	Q99731
217	2578-67_2	C-C motif chemokine 2	P13500
218	2468-62_3	C-C motif chemokine 20	P78556
219	2516-57_3	C-C motif chemokine 21	O00585
220	3508-78_3	C-C motif chemokine 22	O00626
221	2913-1_2	C-C motif chemokine 23	P55773
222	4128-27_2	C-C motif chemokine 24	O00175
223	2705-5_2	C-C motif chemokine 25	O15444
224	9168-31_3	C-C motif chemokine 26	Q9Y258
225	2192-63_10	C-C motif chemokine 27	Q9Y4X3
226	2890-59_2	C-C motif chemokine 28	Q9NRJ3
227	3040-59_1	C-C motif chemokine 3	P10147
228	2783-18_2	C-C motif chemokine 3-like 1	P16619
229	2781-63_2	C-C motif chemokine 4-like	Q8NHW4
230	5480-49_3	C-C motif chemokine 5	P13501
231	4886-3_1	C-C motif chemokine 7	P80098
232	2785-15_2	C-C motif chemokine 8	P80075
233	10337-83_3	CCAAT/enhancer-binding protein beta	P17676
234	3290-50_2	CD109 antigen	Q6YHK3
235	5451-1_3	CD166 antigen	Q13740
236	13116-25_3	CD177 antigen	Q8N6Q3
237	3029-52_2	CD209 antigen	Q9NNX6
238	5062-60_3	CD226 antigen	Q15762
239	5412-53_3	CD27 antigen	P26842
240	3534-14_2	CD40 ligand	P29965
241	3292-75_1	CD48 antigen	P09326
242	3293-2_4	CD5 antigen-like	O43866
243	11514-196_3	CD59 glycoprotein	P13987
244	9190-7_3	CD63 antigen	P08962
245	5807-77_3	CD70 antigen	P32970
246	5065-8_2	CD83 antigen	Q01151
247	2822-56_2	CD97 antigen	P48960
248	3326-58_2	Cell adhesion molecule 1	Q5BY67
249	3630-27_4	Cell adhesion molecule 3	Q8N126
250	4541-49_2	Cell adhesion molecule-related/down-regulated by oncogenes	Q4KMG0
251	5103-30_3	Cell surface glycoprotein CD200 receptor 1	Q8TD46
252	6152-111_3	Cellular tumor antigen p53	P04637
253	4962-52_1	Cerebral dopamine neurotrophic factor	Q49AH0
254	5246-64_3	cGMP-dependent 3',5'-cyclic phosphodiesterase	O00408
255	5254-69_3	cGMP-inhibited 3',5'-cyclic phosphodiesterase A	Q14432
256	5256-86_3	cGMP-specific 3',5'-cyclic phosphodiesterase	O76074
257	11104-13_3	Chitinase-3-like protein 1	P36222
258	3600-2_3	Chitotriosidase-1	Q13231
259	5013-2_1	Chloride intracellular channel protein 1	O00299
260	13117-232_3	Choline/ethanolamine kinase	Q9Y259
261	3362-61_2	Chordin-like protein 1	Q9BU40
262	13103-125_3	Chorionic somatomammotropin hormone	P0DML2 P0DML3
263	4540-11_2	Chromobox protein homolog 5	P45973
264	8476-11_3	Chromogranin-A	P10645
265	3423-59_2	Chymase	P23946
266	3489-9_2	Ciliary neurotrophic factor	P26441
267	2711-6_2	Ciliary neurotrophic factor receptor subunit alpha	P26992

Number	SOMAmer SeqID	Analyte Name	UniProt ID
268	3028-36_2	Ck-beta-8-1	P55773
269	4542-24_2	Clusterin	P10909
270	5066-134_3	CMRF35-like molecule 6	Q08708
271	4905-63_1	Coactosin-like protein	Q14019
272	4876-32_1	Coagulation factor IX	P00740
273	5307-12_3	Coagulation factor IXab	P00740
274	4906-35_1	Coagulation Factor V	P12259
275	3184-25_2	Coagulation factor VII	P08709
276	4878-3_1	Coagulation Factor X	P00742
277	3077-66_2	Coagulation factor Xa	P00742
278	2190-55_1	Coagulation Factor XI	P03951
279	4203-50_2	Cofilin-1	P23528
280	3234-23_2	Coiled-coil domain-containing protein 80	Q76M96
281	4807-13_3	Collagen alpha-1(VIII) chain	P27658
282	4543-65_2	Collagen alpha-1(XXIII) chain	Q86Y22
283	4925-54_2	Collagenase 3	P45452
284	4430-44_3	Collectin-11	Q9BWP8
285	5457-5_2	Collectin-12	Q5KU26
286	2823-7_1	COMM domain-containing protein 7	Q86VX2
287	2753-2_2	Complement C1q subcomponent	P02745 P02746 P02747
288	3285-23_2	Complement C1r subcomponent	P00736
289	3590-8_3	Complement C1s subcomponent	P09871
290	3186-2_2	Complement C2	P06681
291	2754-50_2	Complement C3	P01024
292	4480-59_2	Complement C3b	P01024
293	2683-1_2	Complement C3b, inactivated	P01024
294	5803-24_3	Complement C3d fragment	P01024
295	4481-34_2	Complement C4	P0C0L4, P0C0L5
296	2182-54_1	Complement C4b	P0C0L4 P0C0L5
297	2381-52_4	Complement C5	P01031
298	4482-66_2	Complement C5b-C6 complex	P01031,P13671
299	4967-1_1	Complement component 1 Q subcomponent-binding protein, mitochondrial	Q07021
300	14136-234_3	Complement component C1q receptor	Q9NPY3
301	4127-75_1	Complement component C6	P13671
302	2888-49_2	Complement component C7	P10643
303	2429-27_4	Complement component C8	P07357,P07358,P07360
304	3060-43_2	Complement component C9	P02748
305	5069-9_3	Complement decay-accelerating factor	P08174
306	4129-72_1	Complement factor B	P00751
307	2946-52_2	Complement factor D	P00746
308	4159-130_1	Complement factor H	P08603
309	3666-17_4	Complement factor H-related protein 5	Q9BXR6
310	2567-5_6	Complement factor I	P05156
311	2975-19_2	Connective tissue growth factor	P29279
312	4544-4_3	Connective tissue-activating peptide III	P02775
313	2974-61_2	Contactin-1	Q12860
314	3296-92_2	Contactin-2	Q02246
315	3298-52_2	Contactin-4	Q8IWV2
316	3299-29_2	Contactin-5	O94779
317	5346-24_3	Copine-1	Q99829
318	4785-30_3	Corticosteroid-binding globulin	P08185
319	4890-10_1	Corticotropin	P01189
320	4337-49_2	C-reactive protein	P02741
321	3800-71_2	Creatine kinase B-type	P12277
322	2670-67_4	Creatine kinase M-type	P06732
323	3714-49_2	Creatine kinase M-type:Creatine kinase B-type heterodimer	P12277 P06732
324	3294-55_2	Cryptic protein	P0CG37
325	4332-6_2	C-type lectin domain family 1 member B	Q9P126
326	3361-26_2	C-type lectin domain family 4 member K	Q9UJ71
327	3030-3_2	C-type lectin domain family 4 member M	Q9H2X3
328	3603-60_5	C-type lectin domain family 7 member A	Q9BXN2
329	3041-55_2	C-type mannose receptor 2	Q9UBG0
330	4141-79_1	C-X-C motif chemokine 10	P02778
331	3038-9_2	C-X-C motif chemokine 11	O14625
332	3487-32_2	C-X-C motif chemokine 13	O43927
333	2436-49_4	C-X-C motif chemokine 16	Q9H2A7
334	2979-8_2	C-X-C motif chemokine 5	P42830
335	3495-15_2	C-X-C motif chemokine 6	P80162

Number	SOMAmer SeqID	Analyte Name	UniProt ID
336	9188-119_3	C-X-C motif chemokine 9	Q07325
337	3422-4_2	Cyclin-dependent kinase 1:G2/mitotic-specific cyclin-B1 complex	P06493 P14635
338	3357-67_2	Cyclin-dependent kinase 2:Cyclin-A2 complex	P24941 P20248
339	3358-51_2	Cyclin-dependent kinase 5:Cyclin-dependent kinase 5 activator 1 complex	Q00535 Q15078
340	3359-11_2	Cyclin-dependent kinase 8:Cyclin-C complex	P49336 P24863
341	3719-2_2	Cyclin-dependent kinase inhibitor 1B	P46527
342	2609-59_2	Cystatin-C	P01034
343	3803-10_2	Cystatin-D	P28325
344	3302-58_1	Cystatin-F	O76096
345	3303-23_2	Cystatin-M	Q15828
346	3802-50_1	Cystatin-S	P01036
347	4324-33_2	Cystatin-SA	P09228
348	5459-33_3	Cystatin-SN	P01037
349	9171-11_3	Cysteine and glycine-rich protein 3	P50461
350	3187-52_2	Cysteine-rich secretory protein 3	P54108
351	7628-40_3	Cysteine-rich with EGF-like domain protein 1	Q96HD1
352	2942-50_2	Cytochrome c	P99999
353	2943-5_2	Cytochrome P450 3A4	P08684
354	2634-2_2	Cytokine receptor common subunit gamma	P31785
355	2607-54_2	Cytokine receptor-like factor 1:Cardiotrophin-like cytokine factor 1 Complex	O75462 Q9UBD9
356	2746-56_1	Cytokine receptor-like factor 2	Q9HC73
357	5014-49_1	Cytoplasmic protein NCK1	P16333
358	3414-40_2	Cytoplasmic tyrosine-protein kinase BMX	P51813
359	5345-51_3	Cytoskeleton-associated protein 2	Q8WWK9
360	3192-3_2	Cytosolic non-specific dipeptidase	Q96KP4
361	5068-54_3	Cytotoxic and regulatory T-cell molecule	O95727
362	3022-4_2	Cytotoxic T-lymphocyte protein 4	P16410
363	4314-12_2	dCTP pyrophosphatase 1	Q9H773
364	4907-56_1	D-dimer	P02671 P02675 P02679
365	4355-13_1	Death-associated protein kinase 2	Q9UIK4
366	2666-53_2	Decorin	P07585
367	5349-69_3	Delta-like protein 1	O00548
368	3305-6_1	Delta-like protein 4	Q9NR61
369	4979-34_2	Dermatopontin	Q07507
370	4389-2_1	Desert hedgehog protein N-product	O43323
371	13126-52_3	Desmocollin-2	Q02487
372	4981-6_1	Desmocollin-3	Q14574
373	2976-58_2	Desmoglein-1	Q02413
374	5071-3_3	Desmoglein-2	Q14126
375	3122-6_2	Diablo homolog, mitochondrial	Q9NR28
376	3644-5_4	Dickkopf-like protein 1	Q9UK85
377	3535-84_1	Dickkopf-related protein 1	O94907
378	3607-71_1	Dickkopf-related protein 3	Q9UBP4
379	3365-7_2	Dickkopf-related protein 4	Q9UBT3
380	3178-5_2	Dipeptidyl peptidase 1	P53634
381	3608-12_5	Dipeptidyl peptidase 2	Q9UHL4
382	3606-2_5	Discoidin domain-containing receptor 2	Q16832
383	4420-7_2	Disintegrin and metalloproteinase domain-containing protein 12	O43184
384	3795-6_2	Disintegrin and metalloproteinase domain-containing protein 9	Q13443
385	2871-73_2	DNA repair protein RAD51 homolog 1	Q06609
386	2876-74_2	DNA topoisomerase 1	P11387
387	3852-19_2	DnaJ homolog subfamily B member 1	P25685
388	9175-48_3	Down syndrome cell adhesion molecule	O60469
389	4978-54_2	Drebrin-like protein	Q9UJU6
390	5252-33_3	Dual 3',5'-cyclic-AMP and -GMP phosphodiesterase 11A	Q9HCR9
391	2864-2_3	Dual specificity mitogen-activated protein kinase kinase 1	Q02750
392	3628-3_4	Dual specificity mitogen-activated protein kinase kinase 2	P36507
393	6151-18_3	Dual specificity mitogen-activated protein kinase kinase 3	P46734
394	5242-37_3	Dual specificity mitogen-activated protein kinase kinase 4	P45985
395	3480-7_1	Dual specificity protein phosphatase 3	P51452
396	4359-87_1	Dual specificity tyrosine-phosphorylation-regulated kinase 3	O43781
397	5879-51_3	Dynactin subunit 2	Q13561
398	3881-49_2	Dynein light chain 1, cytoplasmic	P63167
399	3845-51_2	Dynein light chain roadblock-type 1	Q9NP97
400	10342-55_3	E3 SUMO-protein ligase PIAS4	Q8N2W9
401	10336-3_3	E3 ubiquitin-protein ligase CHIP	Q9UNE7
402	4245-80_2	E3 ubiquitin-protein ligase Mdm2	Q00987
403	14120-2_3	E3 ubiquitin-protein ligase RNF43	Q68DV7

Number	SOMAmer SeqID	Analyte Name	UniProt ID
404	14122-132_3	E3 ubiquitin-protein ligase ZNRF3	Q9ULT6
405	2826-53_2	Ectodysplasin-A, secreted form	Q92838
406	3182-38_2	Ectonucleoside triphosphate diphosphohydrolase 1	P49961
407	4436-1_2	Ectonucleoside triphosphate diphosphohydrolase 3	O75355
408	4437-56_3	Ectonucleoside triphosphate diphosphohydrolase 5	O75356
409	4435-66_2	Ectonucleotide pyrophosphatase/phosphodiesterase family member 7	Q6UWV6
410	8480-29_3	EGF-containing fibulin-like extracellular matrix protein 1	Q12805
411	4982-54_1	Elafin	P19957
412	5882-34_2	Elongation factor 1-beta	P24534
413	4908-6_1	Endoglin	P17813
414	4964-67_1	Endoplasmic reticulum aminopeptidase 1	Q9NZ08
415	4983-6_1	Endoplasmic reticulum resident protein 29	P30040
416	2201-17_6	Endostatin	P39060
417	2981-9_3	Endothelial cell-selective adhesion molecule	Q96AP7
418	3805-16_2	Endothelial cell-specific molecule 1	Q9NQ30
419	2714-78_2	Endothelial monocyte-activating polypeptide 2	Q12904
420	3611-70_4	Endothelin-converting enzyme 1	P42892
421	3189-61_2	Enteropeptidase	P98073
422	5301-7_3	Eotaxin	P51671
423	3431-54_2	Ephrin type-A receptor 1	P21709
424	5076-53_2	Ephrin type-A receptor 10	Q5JZY3
425	4834-61_2	Ephrin type-A receptor 2	P29317
426	3432-21_2	Ephrin type-A receptor 3	P29320
427	3806-55_1	Ephrin type-A receptor 5	P54756
428	5077-28_3	Ephrin type-B receptor 2	P29323
429	3612-6_4	Ephrin type-B receptor 4	P54760
430	5078-82_3	Ephrin type-B receptor 6	O15197
431	14124-6_3	Ephrin-A2	O43921
432	14153-8_3	Ephrin-A3	P52797
433	2614-28_2	Ephrin-A4	P52798
434	2615-60_2	Ephrin-A5	P52803
435	13104-32_3	Ephrin-B1	P98172
436	14131-37_3	Ephrin-B2	P52799
437	2514-65_3	Ephrin-B3	Q15768
438	5509-7_3	Epidermal growth factor	P01133
439	2677-1_1	Epidermal growth factor receptor	P00533
440	4212-5_3	Epidermal growth factor receptor substrate 15-like 1	Q9UBC2
441	5328-33_37	Epidermal growth factor receptor variant III	P00533
442	4956-2_1	Epiregulin	O14944
443	4122-12_2	Epithelial discoidin domain-containing receptor 1	Q08345
444	5813-58_3	Erythropoietin	P01588
445	2715-25_2	Erythropoietin receptor	P19235
446	3470-1_2	E-selectin	P16581
447	4708-3_2	Estradiol 17-beta-dehydrogenase 1	P14061
448	2945-25_1	Estrogen receptor	P03372
449	4997-19_1	Eukaryotic initiation factor 4A-III	P38919
450	4230-1_2	Eukaryotic translation initiation factor 4 gamma 2	P78344
451	4184-43_3	Eukaryotic translation initiation factor 4E-binding protein 2	Q13542
452	5885-55_3	Eukaryotic translation initiation factor 4H	Q15056
453	2612-5_2	Eukaryotic translation initiation factor 5	P55010
454	5888-29_3	Eukaryotic translation initiation factor 5A-1	P63241
455	3366-51_2	Extracellular matrix protein 1	Q16610
456	8463-2_3	Extracellular superoxide dismutase [Cu-Zn]	P08294
457	5032-64_1	FACT complex subunit SSRP1	Q08945
458	4985-11_1	Fatty acid-binding protein, epidermal	Q01469
459	5437-63_3	Fatty acid-binding protein, heart	P05413
460	11516-7_3	Fatty acid-binding protein, liver	P07148
461	4440-15_2	Fc receptor-like protein 3	Q96P31
462	5934-1_3	Ferritin	P02794 P02792
463	3367-8_3	Fetuin-B	Q9UGM5
464	2796-62_2	Fibrinogen	P02671 P02675 P02679
465	4989-7_1	Fibrinogen gamma chain	P02679
466	3486-58_2	Fibroblast growth factor 1	P05230
467	2441-2_4	Fibroblast growth factor 10	O15520
468	4392-54_2	Fibroblast growth factor 12	P61328
469	4393-3_1	Fibroblast growth factor 16	O43320
470	3494-71_3	Fibroblast growth factor 17	O60258
471	2761-49_2	Fibroblast growth factor 18	O76093

Number	SOMAmer SeqID	Analyte Name	UniProt ID
472	2762-30_2	Fibroblast growth factor 19	O95750
473	3025-50_1	Fibroblast growth factor 2	P09038
474	2763-66_2	Fibroblast growth factor 20	Q9NP95
475	3807-1_2	Fibroblast growth factor 23	Q9GZV9
476	4123-60_2	Fibroblast growth factor 4	P08620
477	3065-65_1	Fibroblast growth factor 5	P12034
478	4130-71_1	Fibroblast growth factor 6	P10767
479	4487-1_1	Fibroblast growth factor 7	P21781
480	4394-71_2	Fibroblast growth factor 8 isoform A	P55075
481	2443-10_4	Fibroblast growth factor 8 isoform B	P55075
482	2764-20_2	Fibroblast growth factor 9	P31371
483	5532-53_3	Fibroblast growth factor receptor 1	P11362
484	3808-76_2	Fibroblast growth factor receptor 2	P21802
485	3809-1_2	Fibroblast growth factor receptor 3	P22607
486	4988-49_2	Fibroblast growth factor receptor 4	P22455
487	4131-72_2	Fibronectin	P02751
488	3434-34_1	Fibronectin Fragment 3	P02751
489	3435-53_2	Fibronectin Fragment 4	P02751
490	3613-62_5	Ficolin-1	O00602
491	3313-21_2	Ficolin-2	Q15485
492	5462-62_3	Ficolin-3	O75636
493	3053-49_2	Fms-related tyrosine kinase 3 ligand	P49771
494	4986-59_1	Focal adhesion kinase 1	Q05397
495	3032-11_2	Follicle stimulating hormone	P01215, P01225
496	4132-27_2	Follistatin	P19883
497	13112-179_3	Follistatin-related protein 1	Q12841
498	3438-10_2	Follistatin-related protein 3	O95633
499	9213-24_3	Formimidoyltransferase-cyclodeaminase	O95954
500	2827-23_2	Fractalkine	P78423
501	5864-10_3	Fructose-bisphosphate aldolase A	P04075
502	5347-59_3	G2/mitotic-specific cyclin-B1	P14635
503	4548-4_2	Galactoside 3(4)-L-fucosyltransferase	P21217
504	11094-104_3	Galectin-10	Q05315
505	3033-57_1	Galectin-2	P05162
506	3066-12_1	Galectin-3	P17931
507	5000-52_1	Galectin-3-binding protein	Q08380
508	2982-82_2	Galectin-4	P56470
509	9196-8_3	Galectin-7	P47929
510	4909-68_1	Galectin-8	O00214
511	9197-4_3	Galectin-9	O00182
512	10339-48_3	Gamma-enolase	P09104
513	5897-58_3	Gastrin-releasing peptide	P07492
514	3314-74_2	GDNF family receptor alpha-1	P56159
515	2515-14_3	GDNF family receptor alpha-2	O00451
516	2505-49_3	GDNF family receptor alpha-3	O60609
517	4775-34_3	Gelsolin	P06396
518	3217-74_2	Glia-derived nexin	P07093
519	5822-22_3	Glial cell line-derived neurotrophic factor	P39905
520	3034-1_2	Glial fibrillary acidic protein	P14136
521	4891-50_1	Glucagon	P01275
522	2857-70_2	Glucocorticoid receptor	P04150
523	5223-59_3	Glucokinase regulatory protein	Q14397
524	4272-46_2	Glucose-6-phosphate isomerase	P06744
525	5478-50_2	Glutamate carboxypeptidase 2	Q04609
526	4993-16_1	Glutathione S-transferase A3	Q16772
527	4911-49_2	Glutathione S-transferase P	P09211
528	3848-14_2	Glyceraldehyde-3-phosphate dehydrogenase	P04406
529	11081-1_3	Glycerol-3-phosphate dehydrogenase [NAD(+)], cytoplasmic	P21695
530	3441-64_2	Glycogen synthase kinase-3 alpha/beta	P49840 P49841
531	5196-7_3	Glycylpeptide N-tetradecanoyltransferase 1	P30419
532	3315-15_1	Glypican-2	Q8N158
533	4842-62_2	Glypican-3	P51654
534	4991-12_1	Glypican-5	P78333
535	5350-14_2	Glypican-6	Q9Y625
536	4792-51_2	gp41 C34 peptide, HIV	Q70626
537	4992-49_1	Granulins	P28799
538	4840-73_1	Granulocyte colony-stimulating factor	P09919
539	2719-3_4	Granulocyte colony-stimulating factor receptor	Q99062

Number	SOMAmer SeqID	Analyte Name	UniProt ID
540	4697-59_2	Granulocyte-macrophage colony-stimulating factor	P04141
541	3195-50_2	Granulysin	P22749
542	3440-7_2	Granzyme A	P12544
543	4133-54_2	Granzyme B	P10144
544	3373-5_2	Granzyme H	P20718
545	5265-12_3	GRB2-related adapter protein 2	O75791
546	4449-67_3	Gremlin-1	O60565
547	2986-49_2	Gro-beta/gamma	P19876 P19875
548	2949-6_2	Group 10 secretory phospholipase A2	O15496
549	2447-7_4	Group IIE secretory phospholipase A2	Q9NZK7
550	5463-22_3	Growth arrest-specific protein 1	P54826
551	5464-52_3	Growth factor receptor-bound protein 2	P62993
552	2948-58_2	Growth hormone receptor	P10912
553	12060-28_3	Growth/differentiation factor 11	O95390
554	2765-4_3	Growth/differentiation factor 11/8	O95390 O14793
555	4374-45_2	Growth/differentiation factor 15	Q99988
556	4880-21_1	Growth/differentiation factor 2	Q9UK05
557	2752-62_2	Growth/differentiation factor 5	P43026
558	14583-49_3	Growth/differentiation factor 8	O14793
559	3067-67_1	Growth/differentiation factor 9	O60383
560	2985-35_1	Growth-regulated alpha protein	P09341
561	5193-51_3	GTPase KRas	P01116
562	4282-3_2	GTP-binding nuclear protein Ran	P62826
563	3054-3_2	Haptoglobin	P00738
564	4124-24_2	Heat shock 70 kDa protein 1A	P0DMV8
565	5903-91_2	Heat shock cognate 71 kDa protein	P11142
566	11103-24_3	Heat shock protein beta-1	P04792
567	2625-53_4	Heat shock protein HSP 90-alpha/beta	P07900 P08238
568	5467-15_3	Heat shock protein HSP 90-beta	P08238
569	2622-18_1	Heme oxygenase 2	P30519
570	11096-57_3	HemK methyltransferase family member 2	Q9Y5N5
571	4915-64_2	Hemoglobin	P69905, P68871
572	3332-57_1	Hemojuvelin	Q6ZVN8
573	2768-56_2	Hemopexin	P02790
574	5465-32_3	Heparan-sulfate 6-O-sulfotransferase 1	O60243
575	3316-58_1	Heparin cofactor 2	P05546
576	4134-4_2	Heparin-binding EGF-like growth factor	Q99075
577	5134-52_2	Hepatitis A virus cellular receptor 2	Q8TDQ0
578	2681-23_2	Hepatocyte growth factor	P14210
579	3617-80_4	Hepatocyte growth factor activator	Q04756
580	2837-3_2	Hepatocyte growth factor receptor	P08581
581	4407-10_1	Hepatocyte growth factor-like protein	P26927
582	4553-65_3	Hepatoma-derived growth factor-related protein 2	Q7Z4V5
583	3504-58_2	Hepcidin	P81172
584	14132-21_3	HERV-H LTR-associating protein 2	Q9UM44
585	4450-26_2	Heterogeneous nuclear ribonucleoprotein A/B	Q99729
586	4994-178_1	Heterogeneous nuclear ribonucleoprotein K	P61978
587	4224-7_2	Heterogeneous nuclear ribonucleoprotein Q	O60506
588	5351-52_3	Heterogeneous nuclear ribonucleoproteins A2/B1	P22626
589	13131-5_3	Hexokinase-1	P19367
590	13130-150_3	Hexokinase-2	P52789
591	5178-5_3	High affinity cAMP-specific 3',5'-cyclic phosphodiesterase 7A	Q13946
592	5201-50_4	High affinity cGMP-specific 3',5'-cyclic phosphodiesterase 9A	O76083
593	3312-64_1	High affinity immunoglobulin gamma Fc receptor I	P12314
594	3477-63_2	High affinity nerve growth factor receptor	P04629
595	2524-56_3	High mobility group protein B1	P09429
596	5900-11_2	Histidine triad nucleotide-binding protein 1	P49773
597	4996-66_1	Histidine-rich glycoprotein	P04196
598	2865-77_2	Histone acetyltransferase KAT6A	Q92794
599	2858-29_2	Histone acetyltransferase type B catalytic subunit	O14929
600	2859-69_2	Histone deacetylase 8	Q9BY41
601	2987-37_3	Histone H1.2	P16403
602	14144-3_3	Histone H2A type 3	Q7L7L0
603	4163-5_2	Histone H2A.z	P0C0S5
604	14143-8_3	Histone H2B type 2-E	Q16778
605	14146-92_3	Histone H3.1	P68431
606	5843-60_3	Histone-lysine N-methyltransferase EHMT2	Q96KQ7
607	2732-58_3	Homeobox protein NANOG	Q9H9S0

Number	SOMAmer SeqID	Analyte Name	UniProt ID
608	3443-61_2	Homeodomain-interacting protein kinase 3	Q9H422
609	3879-50_1	Hsp90 co-chaperone Cdc37	Q16543
610	4914-10_1	Human Chorionic Gonadotropin	P01215,P01233
611	3196-6_2	Hyaluronan and proteoglycan link protein 1	P10915
612	13089-6_3	Hypoxia-inducible factor 1-alpha	Q16665
613	5061-27_3	ICOS ligand	O75144
614	3198-4_1	Iduronate 2-sulfatase	P22304
615	11089-7_3	Immunoglobulin A	P01876 P01877
616	4987-17_1	Immunoglobulin alpha Fc receptor	P24071
617	4916-2_1	Immunoglobulin D	P01880
618	4135-84_2	Immunoglobulin E	P01854
619	2744-57_2	Immunoglobulin G	P01857
620	3069-52_3	Immunoglobulin M	P01871
621	13124-20_3	Immunoglobulin superfamily containing leucine-rich repeat protein 2	Q6UXK2
622	2860-19_2	Importin subunit alpha-1	P52292
623	3887-90_2	Importin subunit beta-1	Q14974
624	10358-33_3	Induced myeloid leukemia cell differentiation protein Mcl-1	Q07820
625	2988-57_2	Inducible T-cell costimulator	Q9Y6W8
626	2748-3_2	Inhibin beta A chain	P08476
627	8467-9_3	Inhibin beta A chain:Inhibin beta B chain heterodimer	P08476 P09529
628	3888-8_1	Inhibitor of growth protein 1	Q9UK53
629	5021-13_1	Inorganic pyrophosphatase	Q15181
630	5229-90_3	Inosine-5'-monophosphate dehydrogenase 1	P20839
631	5250-53_3	Inosine-5'-monophosphate dehydrogenase 2	P12268
632	4883-56_2	Insulin	P01308
633	3448-13_2	Insulin receptor	P06213
634	3197-70_2	Insulin-degrading enzyme	P14735
635	4232-19_2	Insulin-like growth factor 1 receptor	P08069
636	2952-75_2	Insulin-like growth factor I	P05019
637	2771-35_2	Insulin-like growth factor-binding protein 1	P08833
638	8469-41_3	Insulin-like growth factor-binding protein 2	P18065
639	2571-12_3	Insulin-like growth factor-binding protein 3	P17936
640	2950-57_2	Insulin-like growth factor-binding protein 4	P22692
641	2685-21_2	Insulin-like growth factor-binding protein 5	P24593
642	2686-67_2	Insulin-like growth factor-binding protein 6	P24592
643	3320-49_2	Insulin-like growth factor-binding protein 7	Q16270
644	3503-4_2	Integrin alpha-I: beta-1 complex	P56199, P05556
645	8488-33_3	Integrin alpha-IIb: beta-3 complex	P08514 P05106
646	4917-62_1	Integrin alpha-V: beta-5 complex	P06756, P18084
647	4811-33_3	Inter-alpha-trypsin inhibitor heavy chain H4	Q14624
648	4342-10_3	Intercellular adhesion molecule 1	P05362
649	5486-73_3	Intercellular adhesion molecule 2	P13598
650	2649-77_2	Intercellular adhesion molecule 3	P32942
651	5124-69_3	Intercellular adhesion molecule 5	Q9UMF0
652	9183-7_3	Interferon alpha/beta receptor 1	P17181
653	14128-121_3	Interferon alpha-10	P01566
654	3497-13_2	Interferon alpha-2	P01563
655	14129-1_3	Interferon alpha-7	P01567
656	14127-240_3	Interferon beta	P01574
657	14147-50_3	Interferon gamma	P01579
658	5825-49_3	Interferon gamma receptor 1	P15260
659	9180-6_3	Interferon gamma receptor 2	P38484
660	4396-54_1	Interferon lambda-1	Q8IU54
661	4397-26_2	Interferon lambda-2	Q8IZJ0
662	10351-51_3	Interferon regulatory factor 1	P10914
663	4851-25_1	Interleukin-1 alpha	P01583
664	3037-62_1	Interleukin-1 beta	P01584
665	2630-12_2	Interleukin-1 Receptor accessory protein	Q9NPH3
666	5353-89_2	Interleukin-1 receptor antagonist protein	P18510
667	2991-9_2	Interleukin-1 receptor type 1	P14778
668	14133-93_3	Interleukin-1 receptor type 2	P27930
669	4234-8_2	Interleukin-1 receptor-like 1	Q01638
670	2994-71_2	Interleukin-1 receptor-like 2	Q9HB29
671	2773-50_2	Interleukin-10	P22301
672	10344-334_3	Interleukin-10 receptor subunit alpha	Q13651
673	2631-50_2	Interleukin-10 receptor subunit beta	Q08334
674	4493-92_1	Interleukin-11	P20809
675	10367-62_3	Interleukin-12	P29459, P29460

Number	SOMAmer SeqID	Analyte Name	UniProt ID
676	2632-5_2	Interleukin-12 receptor subunit beta-1	P42701
677	3815-14_1	Interleukin-12 receptor subunit beta-2	Q99665
678	3072-4_2	Interleukin-13	P35225
679	2633-52_2	Interleukin-13 receptor subunit alpha-1	P78552
680	3445-53_2	Interleukin-15 receptor subunit alpha	Q13261
681	2774-10_3	Interleukin-16	Q14005
682	2992-59_2	Interleukin-17 receptor A	Q96F46
683	5084-154_3	Interleukin-17 receptor B	Q9NRM6
684	5468-67_3	Interleukin-17 receptor C	Q8NAC3
685	3376-49_2	Interleukin-17 receptor D	Q8NFM7
686	9170-24_3	Interleukin-17A	Q16552
687	3499-77_2	Interleukin-17B	Q9UHF5
688	4136-40_2	Interleukin-17D	Q8TAD2
689	2775-54_2	Interleukin-17F	Q96PD4
690	3446-7_2	Interleukin-18 receptor 1	Q13478
691	2993-1_2	Interleukin-18 receptor accessory protein	O95256
692	3073-51_2	Interleukin-18-binding protein	O95998
693	3035-80_2	Interleukin-19	Q9UHD0
694	3070-1_2	Interleukin-2	P60568
695	3151-6_1	Interleukin-2 receptor subunit alpha	P01589
696	4138-25_2	Interleukin-20	Q9NYY1
697	5085-18_3	Interleukin-20 receptor subunit alpha	Q9UHF4
698	2778-10_2	Interleukin-22	Q9GZX6
699	3620-67_2	Interleukin-22 receptor subunit alpha-1	Q8N6P7
700	5087-5_3	Interleukin-22 receptor subunit alpha-2	Q969I5
701	10365-132_3	Interleukin-23	P29460, Q9NPF7
702	5088-175_3	Interleukin-23 receptor	Q5VWK5
703	3321-2_2	Interleukin-24	Q13007
704	4137-57_2	Interleukin-25	Q9H293
705	2829-19_2	Interleukin-27	Q8NEV9 Q14213
706	5132-71_3	Interleukin-27 receptor subunit alpha	Q6UWB1
707	4717-55_2	Interleukin-3	P08700
708	4490-65_2	Interleukin-3 receptor subunit alpha	P26951
709	4556-10_2	Interleukin-34	Q6ZMJ4
710	14150-7_3	Interleukin-36 alpha	Q9UHA7
711	14149-9_3	Interleukin-36 beta	Q9NZH7
712	2723-9_2	Interleukin-37	Q9NZH6
713	2906-55_3	Interleukin-4	P05112
714	3055-54_2	Interleukin-4 receptor subunit alpha	P24394
715	11071-1_3	Interleukin-5	P05113
716	4491-4_2	Interleukin-5 receptor subunit alpha	Q01344
717	4673-13_2	Interleukin-6	P05231
718	4139-71_2	Interleukin-6 receptor subunit alpha	P08887
719	2620-4_2	Interleukin-6 receptor subunit beta	P40189
720	4140-3_2	Interleukin-7	P13232
721	5089-11_3	Interleukin-7 receptor subunit alpha	P16871
722	3447-64_2	Interleukin-8	P10145
723	5834-18_3	Interleukin-9	P15248
724	4924-32_1	Interstitial collagenase	P03956
725	2997-8_1	Junctional adhesion molecule B	P57087
726	2998-53_2	Junctional adhesion molecule C	Q9BX67
727	5094-62_3	Junctional adhesion molecule-like	Q86YT9
728	2831-29_1	Kallikrein-11	Q9UBX7
729	3199-54_2	Kallikrein-12	Q9UKR0
730	3200-49_2	Kallikrein-13	Q9UKR3
731	3681-87_3	Kallikrein-14	Q9P0G3
732	2833-20_1	Kallikrein-4	Q9Y5K2
733	3201-49_2	Kallikrein-5	Q9Y337
734	3450-4_2	Kallikrein-6	Q92876
735	3378-49_2	Kallikrein-7	P49862
736	2834-54_1	Kallikrein-8	O60259
737	3449-58_2	Kallistatin	P29622
738	8485-7_3	Kelch-like ECH-associated protein 1	Q14145
739	5354-11_3	Keratin, type I cytoskeletal 18	P05783
740	5095-21_3	Killer cell immunoglobulin-like receptor 2DL4	Q99706
741	5096-51_3	Killer cell immunoglobulin-like receptor 3DL2	P43630
742	5097-14_3	Killer cell immunoglobulin-like receptor 3DS1	Q14943
743	5098-79_3	Killer cell lectin-like receptor subfamily F member 1	Q9NZS2

Number	SOMAmer SeqID	Analyte Name	UniProt ID
744	4557-61_2	Kin of IRRE-like protein 3	Q8IZU9
745	5228-25_2	Kinesin-like protein KIF23	Q02241
746	4495-33_2	Kininogen-1	P01042
747	3202-28_2	Kremen protein 2	Q8NCW0
748	2828-82_2	Kunitz-type protease inhibitor 1	O43278
749	2843-13_2	Kunitz-type protease inhibitor 2	O43291
750	4559-64_2	Kynureninase	Q16719
751	4455-89_2	Lactadherin	Q08431
752	4801-13_3	Lactoperoxidase	P22079
753	2780-35_2	Lactotransferrin	P02788
754	3889-64_2	Lamin-B1	P20700
755	2728-62_2	Laminin	P25391, P07942, P11047
756	13133-73_3	Latent-transforming growth factor beta-binding protein 4	Q8N2S1
757	2635-61_2	Layilin	Q6UX15
758	3622-33_4	Legumain	Q99538
759	8484-24_3	Leptin	P41159
760	5400-52_3	Leptin receptor	P48357
761	4237-70_3	Leucine carboxyl methyltransferase 1	Q9UIC8
762	10990-21_3	Leucine-rich repeat serine/threonine-protein kinase 2	Q5S007
763	4452-9_2	Leucine-rich repeat transmembrane neuronal protein 1	Q86UE6
764	4453-83_3	Leucine-rich repeat transmembrane neuronal protein 3	Q86VH5
765	4547-59_2	Leucine-rich repeat transmembrane protein FLRT1	Q9NZU1
766	13122-19_3	Leucine-rich repeat transmembrane protein FLRT2	O43155
767	13123-3_3	Leucine-rich repeat transmembrane protein FLRT3	Q9NZU0
768	3322-52_2	Leucine-rich repeats and immunoglobulin-like domains protein 3	Q6UXM1
769	5837-49_3	Leukemia inhibitory factor receptor	P42702
770	5090-49_2	Leukocyte immunoglobulin-like receptor subfamily B member 1	Q8NHL6
771	5091-28_3	Leukocyte immunoglobulin-like receptor subfamily B member 2	Q8N423
772	6653-58_3	Leukocyte surface antigen CD47	Q08722
773	3204-2_2	Leukotriene A-4 hydrolase	P09960
774	4304-18_2	Ligand-dependent nuclear receptor corepressor-like protein	Q8N3X6
775	2999-6_2	Limbic system-associated membrane protein	Q13449
776	3074-6_2	Lipopolysaccharide-binding protein	P18428
777	13095-51_3	Lithostathine-1-alpha	P05451
778	3890-8_2	L-lactate dehydrogenase B chain	P07195
779	3291-30_2	Low affinity immunoglobulin epsilon Fc receptor	P06734
780	3309-2_2	Low affinity immunoglobulin gamma Fc region receptor II-a	P12318
781	3310-62_1	Low affinity immunoglobulin gamma Fc region receptor II-b	P31994
782	3311-27_1	Low affinity immunoglobulin gamma Fc region receptor III-B	O75015
783	3858-5_1	Low molecular weight phosphotyrosine protein phosphatase	P24666
784	13129-40_3	Low-density lipoprotein receptor	P01130
785	9182-3_3	Low-density lipoprotein receptor-related protein 1, soluble	Q07954
786	7640-29_3	Low-density lipoprotein receptor-related protein 1B	Q9NZR2
787	3323-37_1	Low-density lipoprotein receptor-related protein 8	Q14114
788	4831-4_2	L-Selectin	P14151
789	13114-50_3	Lumican	P51884
790	2953-31_2	Luteinizing hormone	P01215, P01229
791	13107-9_3	Ly6/PLAUR domain-containing protein 3	O95274
792	3206-4_2	Lymphatic vessel endothelial hyaluronic acid receptor 1	Q9Y5Y7
793	5099-14_3	Lymphocyte activation gene 3 protein	P18627
794	3623-84_4	Lymphocyte antigen 86	O95711
795	4143-74_2	Lymphotactin	P47992
796	3505-6_2	Lymphotoxin alpha1:beta2	P01374, Q06643
797	3506-49_1	Lymphotoxin alpha2:beta1	P01374, Q06643
798	4703-87_2	Lymphotoxin-alpha	P01374
799	3179-51_2	Lysosomal protective protein	P10619
800	5100-53_3	Lysosome membrane protein 2	Q14108
801	4920-10_1	Lysozyme C	P61626
802	3738-54_4	Macrophage colony-stimulating factor 1	P09603
803	2638-12_2	Macrophage colony-stimulating factor 1 receptor	P07333
804	2637-77_2	Macrophage mannose receptor 1	P22897
805	4496-60_2	Macrophage metalloelastase	P39900
806	5356-2_3	Macrophage migration inhibitory factor	P14174
807	3684-78_3	Macrophage scavenger receptor types I and II	P21757
808	4968-50_1	Macrophage-capping protein	P40121
809	2640-3_2	Macrophage-stimulating protein receptor	Q04912
810	3853-56_1	Malate dehydrogenase, cytoplasmic	P40925
811	5001-6_2	Mammaglobin-B	O75556

Number	SOMAmer SeqID	Analyte Name	UniProt ID
812	3605-77_4	Mannan-binding lectin serine protease 1	P48740
813	3000-66_1	Mannose-binding protein C	P11226
814	3820-68_2	MAP kinase-activated protein kinase 2	P49137
815	3822-54_2	MAP kinase-activated protein kinase 3	Q16644
816	3821-28_1	MAP kinase-activated protein kinase 5	Q8IW41
817	2475-1_3	Mast/stem cell growth factor receptor Kit	P10721
818	3325-2_2	Matrilin-2	O00339
819	3208-2_4	Matrilin-3	O15232
820	2789-26_2	Matrilysin	P09237
821	3209-69_1	Matrix extracellular phosphoglycoprotein	Q9NQ76
822	5002-76_1	Matrix metalloproteinase-14	P50281
823	5268-49_3	Matrix metalloproteinase-16	P51512
824	2838-53_1	Matrix metalloproteinase-17	Q9ULZ9
825	2579-17_5	Matrix metalloproteinase-9	P14780
826	5470-69_2	Mediator of RNA polymerase II transcription subunit 1	Q15648
827	3823-9_1	Megakaryocyte-associated tyrosine-protein kinase	P42679
828	2687-2_1	Melanoma-derived growth regulatory protein	Q16674
829	3685-53_1	Membrane frizzled-related protein	Q9BY79
830	3627-71_4	Membrane metallo-endopeptidase-like 1	Q495T6
831	3893-64_1	Mesothelin	Q13421
832	2211-9_6	Metalloproteinase inhibitor 1	P01033
833	2278-61_4	Metalloproteinase inhibitor 2	P16035
834	2480-58_3	Metalloproteinase inhibitor 3	P35625
835	3210-1_2	Methionine aminopeptidase 1	P53582
836	3170-6_1	Methionine aminopeptidase 2	P50579
837	3891-56_1	Methyl-CpG-binding domain protein 4	O95243
838	2730-58_2	MHC class I polypeptide-related sequence A	Q29983
839	5102-55_3	MHC class I polypeptide-related sequence B	Q29980
840	5854-60_3	Microtubule-associated protein tau	P10636
841	2911-27_2	Midkine	P21741
842	5280-68_5	Mitochondrial glutamate carrier 2	Q9H1K4
843	4545-53_3	Mitochondrial import inner membrane translocase subunit TIM14	Q96DA6
844	3115-64_2	Mitogen-activated protein kinase 1	P28482
845	5004-69_1	Mitogen-activated protein kinase 11	Q15759
846	5005-4_1	Mitogen-activated protein kinase 12	P53778
847	5006-71_1	Mitogen-activated protein kinase 13	O15264
848	5007-1_1	Mitogen-activated protein kinase 14	Q16539
849	2855-49_2	Mitogen-activated protein kinase 3	P27361
850	3825-18_2	Mitogen-activated protein kinase 8	P45983
851	5244-12_3	Mitogen-activated protein kinase 9	P45984
852	5259-2_3	Mitogen-activated protein kinase kinase kinase 7:TGF-beta-activated kinase 1 and MAP3K7-binding protein 1 fusion	O43318 Q15750
853	5009-11_1	Moesin	P26038
854	10364-6_3	Mothers against decapentaplegic homolog 2	Q15796
855	10363-13_3	Mothers against decapentaplegic homolog 3	P84022
856	9176-3_3	Mucin-1	P15941
857	4923-79_1	Muellerian-inhibiting factor	P03971
858	10362-35_3	Myc proto-oncogene protein	P01106
859	3514-49_2	Myeloblastin	P24158
860	3166-92_1	Myeloid cell surface antigen CD33	P20138
861	2580-83_2	Myeloperoxidase	P05164
862	3042-7_2	Myoglobin	P02144
863	7648-9_3	Myosin-binding protein C, slow-type	Q00872
864	3894-15_2	N-acetyl-D-glucosamine kinase	Q9UJ70
865	3616-3_5	N-acetylglucosamine-6-sulfatase	P15586
866	3173-49_2	N-acylethanolamine-hydrolyzing acid amidase	Q02083
867	5030-52_1	NAD-dependent protein deacetylase sirtuin-2	Q8IXJ6
868	2731-29_2	NADPH--cytochrome P450 reductase	P16435
869	3854-24_1	Nascent polypeptide-associated complex subunit alpha	Q13765
870	5104-57_3	Natural cytotoxicity triggering receptor 1	O76036
871	2734-49_4	Natural cytotoxicity triggering receptor 2	O95944
872	3003-29_2	Natural cytotoxicity triggering receptor 3	O14931
873	5063-12_2	Natural killer cell receptor 2B4	Q9BZW8
874	5139-32_3	Netrin receptor UNC5C	O95185
875	5140-56_3	Netrin receptor UNC5D	Q6UXZ4
876	6649-51_3	Netrin-1	O95631
877	3327-27_1	Netrin-4	Q9HB63
878	4498-62_2	Neural cell adhesion molecule 1, 120 kDa isoform	P13591

Number	SOMAmer SeqID	Analyte Name	UniProt ID
879	4246-40_2	Neural cell adhesion molecule L1	P32004
880	3601-54_3	Neural cell adhesion molecule L1-like protein	O00533
881	9178-30_3	Neuregulin-1	Q02297
882	14139-16_3	Neuregulin-4	Q8WWG1
883	5110-84_3	Neurexin-1-beta	P58400
884	5111-15_3	Neurexin-3-beta	Q9HDB5
885	4562-1_2	Neurexophilin-1	P58417
886	2944-66_2	Neuroblastoma suppressor of tumorigenicity 1	P41271
887	5107-7_2	Neurogenic locus notch homolog protein 1	P46531
888	5106-52_3	Neurogenic locus notch homolog protein 2	Q04721
889	5108-72_3	Neurogenic locus notch homolog protein 3	Q9UM47
890	5357-60_3	Neuroligin-4, X-linked	Q8N0W4
891	5109-24_3	Neuronal cell adhesion molecule	Q92823
892	13109-82_3	Neuronal growth regulator 1	Q7Z3B1
893	5542-22_2	Neuropilin-1	O14786
894	4145-58_2	Neurotrophin-3	P20783
895	4146-58_2	Neurotrophin-4	P34130
896	3212-30_3	Neutral ceramidase	Q9NR71
897	9172-69_3	Neutrophil collagenase	P22894
898	5308-89_2	Neutrophil elastase	P08246
899	2836-68_2	Neutrophil gelatinase-associated lipocalin	P80188
900	2790-54_2	Neutrophil-activating peptide 2	P02775
901	5011-11_1	Nicotinamide phosphoribosyltransferase	P43490
902	3213-65_2	Nidogen-1	P14543
903	3633-70_5	Nidogen-2	Q14112
904	3081-70_2	NKG2D ligand 1	Q9BZM6
905	3082-9_2	NKG2D ligand 2	Q9BZM5
906	2747-3_2	NKG2D ligand 3	Q9BZM4
907	3056-11_1	NKG2-D type II integral membrane protein	P26718
908	5846-24_3	Noggin	Q13253
909	9187-2_3	Non-histone chromosomal protein HMG-14	P05114
910	5260-80_3	Non-receptor tyrosine-protein kinase TYK2	P29597
911	4250-23_3	NSFL1 cofactor p47	Q9UNZ2
912	2658-27_1	NT-3 growth factor receptor	Q16288
913	7655-11_3	N-terminal pro-BNP	P16860
914	5236-2_3	Nuclear receptor subfamily 1 group D member 1	P20393
915	5909-51_3	Nucleoside diphosphate kinase A	P15531
916	4249-64_2	Nucleoside diphosphate kinase B	P22392
917	4254-6_2	NudC domain-containing protein 3	Q8IVD9
918	3635-76_4	OCIA domain-containing protein 1	Q9NX40
919	4769-10_3	Olfactomedin-4	Q6UX06
920	2693-20_3	Oncostatin-M	P13725
921	3634-5_4	Opioid-binding protein/cell adhesion molecule	Q14982
922	11067-13_3	Osteocalcin	P02818
923	5358-3_3	Osteomodulin	Q99983
924	13113-7_3	Osteopontin	P10451
925	5112-73_3	OX-2 membrane glycoprotein	P41217
926	3636-37_4	Oxidized low-density lipoprotein receptor 1	P78380
927	4588-1_2	Pancreatic hormone	P01298
928	4148-49_2	Pappalysin-1	Q13219
929	5954-62_3	Parathyroid hormone	P01270
930	2962-50_2	Parathyroid hormone-related protein	P12272
931	3727-35_1	Peptide YY	P10082
932	3329-14_2	Peptidoglycan recognition protein 1	O75594
933	3844-2_1	Peptidyl-prolyl cis-trans isomerase A	P62937
934	4718-5_2	Peptidyl-prolyl cis-trans isomerase B	P23284
935	5202-4_3	Peptidyl-prolyl cis-trans isomerase D	Q08752
936	5238-26_3	Peptidyl-prolyl cis-trans isomerase E	Q9UNP9
937	5248-68_2	Peptidyl-prolyl cis-trans isomerase F, mitochondrial	P30405
938	3457-57_1	Periostin	Q15063
939	3855-56_1	Peroxiredoxin-1	Q06830
940	5017-19_1	Peroxiredoxin-5, mitochondrial	P30044
941	5018-68_1	Peroxiredoxin-6	P30041
942	5915-58_3	Peroxisomal targeting signal 1 receptor	P50542
943	2696-87_2	Persephin	O60542
944	3847-56_2	Persulfide dioxygenase ETHE1, mitochondrial	O95571
945	4267-81_3	Pescadillo homolog	O00541
946	4276-10_2	Phosphatidylethanolamine-binding protein 1	P30086

Number	SOMAmer SeqID	Analyte Name	UniProt ID
947	3831-21_1	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN	P60484
948	3391-10_2	Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform	P48736
949	9173-21_3	Phosphoglucomutase-1	P36871
950	5020-50_1	Phosphoglycerate kinase 1	P00558
951	3896-5_2	Phosphoglycerate mutase 1	P18669
952	4910-21_1	Phospholipase A2	P04054
953	2692-74_2	Phospholipase A2, membrane associated	P14555
954	9211-19_3	Pigment epithelium-derived factor	P36955
955	3390-72_2	PIK3CA/PIK3R1	P42336 P27986
956	14114-18_3	PILR alpha-associated neural protein	Q8IYJ0
957	8446-4_3	Pituitary adenylate cyclase-activating polypeptide 27	P18509
958	8450-36_3	Pituitary adenylate cyclase-activating polypeptide 38	P18509
959	3078-1_2	Placenta growth factor	P49763
960	4152-58_2	Plasma kallikrein	P03952
961	4479-14_2	Plasma protease C1 inhibitor	P05155
962	3389-7_2	Plasma serine protease inhibitor	P05154
963	4150-75_2	Plasmin	P00747
964	4151-6_2	Plasminogen	P00747
965	2925-9_1	Plasminogen activator inhibitor 1	P05121
966	2695-25_1	Platelet endothelial cell adhesion molecule	P16284
967	2697-7_2	Platelet factor 4	P02776
968	2973-15_2	Platelet glycoprotein 4	P16671
969	4990-87_1	Platelet glycoprotein Ib alpha chain	P07359
970	3194-36_2	Platelet glycoprotein VI	Q9HCN6
971	14123-34_3	Platelet receptor Gi24	Q9H7M9
972	5015-15_1	Platelet-activating factor acetylhydrolase	Q13093
973	2642-4_1	Platelet-activating factor acetylhydrolase IB subunit beta	P68402
974	3828-54_2	Platelet-derived growth factor C	Q9NRA1
975	10366-11_3	Platelet-derived growth factor receptor alpha	P16234
976	3459-49_2	Platelet-derived growth factor receptor beta	P09619
977	4499-21_1	Platelet-derived growth factor subunit A	P04085
978	4149-8_2	Platelet-derived growth factor subunit B	P01127
979	3045-72_2	Pleiotrophin	P21246
980	9216-100_3	Plexin-B2	O15031
981	4564-2_2	Plexin-C1	O60486
982	3216-2_2	Polymeric immunoglobulin receptor	P01833
983	6641-60_3	PolyUbiquitin K48-linked	P0CG47
984	6647-55_3	PolyUbiquitin K63-linked	P0CG48
985	4271-75_2	Prefoldin subunit 5	Q99471
986	5060-62_3	Programmed cell death 1 ligand 1	Q9NZQ7
987	3004-67_2	Programmed cell death 1 ligand 2	Q9BQ51
988	2247-20_11	Prokineticin-1	P58294
989	2585-2_5	Prolactin	P01236
990	5114-65_3	Prolactin receptor	P16471
991	2915-6_2	Proliferating cell nuclear antigen	P12004
992	4258-15_2	Proliferation-associated protein 2G4	Q9UQ80
993	5029-3_1	Prolyl endopeptidase FAP	Q12884
994	9204-33_3	Pro-opiomelanocortin	P01189
995	2960-66_2	Properdin	P27918
996	4459-68_2	Proprotein convertase subtilisin/kexin type 7	Q16549
997	5231-79_3	Proprotein convertase subtilisin/kexin type 9	Q8NBP7
998	3761-4_2	Prostaglandin G/H synthase 2	P35354
999	8468-19_3	Prostate-specific antigen	P07288
1000	5918-5_3	Proteasome activator complex subunit 1	Q06323
1001	5204-13_3	Proteasome activator complex subunit 3	P61289
1002	3859-50_2	Proteasome subunit alpha type-1	P25786
1003	4280-47_2	Proteasome subunit alpha type-2	P25787
1004	3860-7_2	Proteasome subunit alpha type-6	P60900
1005	4706-17_2	Protein 4.1	P11171
1006	4322-28_3	Protein amnionless	Q9BXJ7
1007	5016-61_2	Protein deglycase DJ-1	Q99497
1008	4278-14_3	Protein disulfide-isomerase	P07237
1009	4719-58_2	Protein disulfide-isomerase A3	P30101
1010	2623-54_4	Protein E7_HPV16	P03129
1011	2624-31_2	Protein E7_HPV18	P06788
1012	2760-2_2	Protein FAM107A	O95990
1013	4774-62_3	Protein FAM107B	Q9H098

Number	SOMAmer SeqID	Analyte Name	UniProt ID
1014	9177-6_3	Protein FAM3B	P58499
1015	13102-1_3	Protein FAM3D	Q96BQ1
1016	5092-51_3	Protein jagged-1	P78504
1017	5093-47_3	Protein jagged-2	Q9Y219
1018	2644-11_2	Protein kinase C alpha type	P17252
1019	5475-10_3	Protein kinase C beta type (splice variant beta-II)	P05771
1020	2869-68_3	Protein kinase C delta type	Q05655
1021	5476-66_3	Protein kinase C gamma type	P05129
1022	3379-29_1	Protein kinase C iota type	P41743
1023	3817-18_1	Protein kinase C theta type	Q04759
1024	2645-54_1	Protein kinase C zeta type	Q05513
1025	4238-4_2	Protein lin-7 homolog B	Q9HAP6
1026	2737-22_2	Protein NOV homolog	P48745
1027	2769-3_2	Protein Rev_HV2BE	P18093
1028	5852-6_3	Protein S100-A12	P80511
1029	14116-129_3	Protein S100-A4	P26447
1030	13090-17_3	Protein S100-A6	P06703
1031	5921-58_3	Protein S100-A7	P31151
1032	5339-49_3	Protein S100-A9	P06702
1033	5364-7_3	Protein SET	Q01105
1034	4889-82_1	Protein Wnt-7a	O00755
1035	13119-26_3	Protein Z-dependent protease inhibitor	Q9UK55
1036	4471-50_2	Protein-glutamine gamma-glutamyltransferase E	Q08188
1037	3832-51_1	Protein-tyrosine kinase 6	Q13882
1038	5316-54_3	Prothrombin	P00734
1039	3220-40_2	Proto-oncogene tyrosine-protein kinase receptor Ret	P07949
1040	5488-74_3	Proto-oncogene tyrosine-protein kinase Src	P12931
1041	5275-28_2	Proto-oncogene vav	P15498
1042	4154-57_2	P-selectin	P16109
1043	4414-69_2	Pulmonary surfactant-associated protein D	P35247
1044	11098-1_3	Pyridoxal kinase	O00764
1045	3897-61_1	Pyridoxal phosphate phosphatase	Q96GD0
1046	4240-31_2	Pyruvate kinase PKM	P14618
1047	9207-60_3	Quinone oxidoreductase-like protein 1	O95825
1048	2647-66_2	Rab GDP dissociation inhibitor beta	P50395
1049	3392-68_2	RAC-alpha/beta/gamma serine/threonine-protein kinase	P31749 P31751 Q9Y243
1050	5360-9_2	RAC-beta serine/threonine-protein kinase	P31751
1051	5481-16_3	Ras GTPase-activating protein 1	P20936
1052	2870-29_2	Ras-related C3 botulinum toxin substrate 1	P63000
1053	5271-5_1	Ras-related C3 botulinum toxin substrate 3	P60763
1054	2616-23_18	Receptor tyrosine-protein kinase erbB-2	P04626
1055	2617-56_35	Receptor tyrosine-protein kinase erbB-3	P21860
1056	2618-10_2	Receptor tyrosine-protein kinase erbB-4	Q15303
1057	3437-80_3	Receptor-type tyrosine-protein kinase FLT3	P36888
1058	11102-22_3	Regenerating islet-derived protein 4	Q9BYZ8
1059	14135-3_3	Relaxin receptor 1	Q9HBX9
1060	3396-54_2	Renin	P00797
1061	5483-1_3	Repulsive guidance molecule A	Q96B86
1062	3046-31_1	Resistin	Q9HD89
1063	5118-74_2	Reticulon-4	Q9NQC3
1064	5105-2_3	Reticulon-4 receptor	Q9BZR6
1065	5024-67_1	Retinoblastoma-associated protein	P06400
1066	3079-62_2	Retinoic acid receptor responder protein 2	Q99969
1067	5482-61_3	Retinol-binding protein 4	P02753
1068	3331-8_1	RGM domain family member B	Q6NW40
1069	8470-213_3	Ribonuclease H1	O60930
1070	3469-74_1	Ribosomal protein S6 kinase alpha-3	P51812
1071	3865-53_1	Ribosomal protein S6 kinase alpha-5	O75582
1072	3866-7_2	Ribosome maturation protein SBDS	Q9Y3A5
1073	4284-18_3	RNA-binding protein 39	Q14498
1074	5116-62_2	Roundabout homolog 2	Q9HCK4
1075	5117-14_3	Roundabout homolog 3	Q96MS0
1076	4566-24_3	R-spondin-2	Q6UXX9
1077	13094-75_3	R-spondin-3	Q9BXY4
1078	8464-31_3	R-spondin-4	Q2I0M5
1079	5129-12_3	Scavenger receptor class F member 1	Q14162
1080	5130-67_3	Scavenger receptor class F member 2	Q96GP6
1081	5028-59_1	Scavenger receptor cysteine-rich type 1 protein M130	Q86VB7

Number	SOMAmer SeqID	Analyte Name	UniProt ID
1082	13101-60_3	Sclerostin	Q9BQB4
1083	13093-6_3	Secreted and transmembrane protein 1	Q8WVN6
1084	3221-54_1	Secreted frizzled-related protein 1	Q8N474
1085	2841-13_2	Secreted frizzled-related protein 3	Q92765
1086	3728-52_2	Secretin	P09683
1087	3867-49_1	Seizure 6-like protein 2	Q6UXD5
1088	3222-11_2	Semaphorin-3A	Q14563
1089	5363-51_3	Semaphorin-3E	O15041
1090	13132-14_3	Semaphorin-5A	Q13591
1091	5122-92_2	Semaphorin-6A	Q9H2E6
1092	5121-3_2	Semaphorin-6B	Q9H3T3
1093	3624-3_4	Serine protease 27	Q9BQR3
1094	3317-33_1	Serine protease HTRA2, mitochondrial	O43464
1095	3471-49_1	Serine/threonine-protein kinase 16	O75716
1096	5249-31_3	Serine/threonine-protein kinase 17B	O94768
1097	2853-68_2	Serine/threonine-protein kinase Chk1	O14757
1098	3360-50_4	Serine/threonine-protein kinase Chk2	O96017
1099	3629-60_4	Serine/threonine-protein kinase MRCK beta	Q9Y5S2
1100	3387-1_2	Serine/threonine-protein kinase PAK 3	O75914
1101	3827-22_1	Serine/threonine-protein kinase PAK 6	Q9NQU5
1102	3388-58_2	Serine/threonine-protein kinase PAK 7	Q9P286
1103	5359-65_3	Serine/threonine-protein kinase pim-1	P11309
1104	3394-81_2	Serine/threonine-protein kinase PLK1	P53350
1105	2969-11_1	Serine/threonine-protein kinase receptor R3	P37023
1106	3400-49_2	Serine/threonine-protein kinase TBK1	Q9UHD2
1107	5493-17_3	Serine/threonine-protein kinase WNK3	Q9BYP7
1108	4162-54_2	Serotransferrin	P02787
1109	3707-12_2	Serum albumin	P02768
1110	4336-2_1	Serum amyloid A-1 protein	P0DJI8
1111	2474-54_5	Serum amyloid P-component	P02743
1112	4261-55_2	Serum paraoxonase/arylesterase 1	P27169
1113	4929-55_1	Sex hormone-binding globulin	P04278
1114	4984-83_1	S-formylglutathione hydrolase	P10768
1115	4567-82_2	SH2 domain-containing protein 1A	O60880
1116	5272-55_2	SHC-transforming protein 1	P29353
1117	5125-6_3	Sialic acid-binding Ig-like lectin 14	Q08ET2
1118	2741-22_2	Sialic acid-binding Ig-like lectin 6	O43699
1119	2742-68_2	Sialic acid-binding Ig-like lectin 7	Q9Y286
1120	3007-7_2	Sialic acid-binding Ig-like lectin 9	Q9Y336
1121	4464-10_2	Sialoadhesin	Q9BZZ2
1122	10370-21_3	Signal transducer and activator of transcription 1-alpha/beta	P42224
1123	10346-5_3	Signal transducer and activator of transcription 3	P40763
1124	10372-18_3	Signal transducer and activator of transcription 6	P42226
1125	3642-4_1	SLAM family member 5	Q9UIB8
1126	5128-53_3	SLAM family member 6	Q96DU3
1127	5487-7_3	SLAM family member 7	Q9NQ25
1128	3643-90_2	SLIT and NTRK-like protein 1	Q96PX8
1129	4568-17_2	SLIT and NTRK-like protein 5	O94991
1130	3868-8_1	Small glutamine-rich tetratricopeptide repeat-containing protein alpha	O43765
1131	5494-52_3	Small nuclear ribonucleoprotein F	P62306
1132	9169-14_3	Small ubiquitin-related modifier 3	P55854
1133	5957-30_3	Somatostatin-28	P61278
1134	8462-18_3	Somatotropin	P01241
1135	2743-5_2	Sonic hedgehog protein	Q15465
1136	3903-49_2	Sorting nexin-4	O95219
1137	3043-49_2	SPARC	P09486
1138	4467-49_2	SPARC-like protein 1	Q14515
1139	13118-5_3	SPARC-related modular calcium-binding protein 1	Q9H4F8
1140	5031-10_1	Spectrin alpha chain, non-erythrocytic 1	Q13813
1141	3902-21_2	S-phase kinase-associated protein 1	P63208
1142	4294-16_2	Sphingosine kinase 1	Q9NYA1
1143	4468-21_2	Sphingosine kinase 2	Q9NRA0
1144	4297-62_3	Spondin-1	Q9HCB6
1145	3399-31_2	Stabilin-2	Q8WWQ8
1146	4930-21_1	Stanniocalcin-1	P52823
1147	4500-50_2	Stem cell growth factor-alpha	Q9Y240
1148	2966-65_2	Stem Cell Growth Factor-beta	Q9Y240
1149	5489-18_3	Stress-induced-phosphoprotein 1	P31948

Number	SOMAmer SeqID	Analyte Name	UniProt ID
1150	3516-60_2	Stromal cell-derived factor 1	P48061
1151	2788-55_1	Stromelysin-1	P08254
1152	8479-4_3	Stromelysin-2	P09238
1153	2877-3_1	SUMO-conjugating enzyme UBC9	P63279
1154	2794-60_2	Superoxide dismutase [Cu-Zn]	P00441
1155	5008-51_1	Superoxide dismutase [Mn], mitochondrial	P04179
1156	13105-7_3	Synaptosomal-associated protein 25	P60880
1157	3766-51_4	Syntaxin-1A	Q16623
1158	3232-28_2	Tartrate-resistant acid phosphatase type 5	P13686
1159	2875-15_2	TATA-box-binding protein	P20226
1160	3143-3_1	T-cell surface glycoprotein CD4	P01730
1161	4155-3_2	Tenascin	P24821
1162	5810-25_3	Teratocarcinoma-derived growth factor 1	P13385
1163	5490-53_3	Testican-1	Q08629
1164	5491-12_3	Testican-2	Q92563
1165	5133-17_3	TGF-beta receptor type-2	P37173
1166	4815-25_3	Thioredoxin domain-containing protein 12	O95881
1167	4157-2_1	Thrombin	P00734
1168	5947-90_3	Thrombopoietin	P40225
1169	3473-78_2	Thrombopoietin Receptor	P40238
1170	3474-19_2	Thrombospondin-1	P07996
1171	3339-33_1	Thrombospondin-2	P35442
1172	3340-53_1	Thrombospondin-4	P35443
1173	3010-53_2	Thymic stromal lymphopoietin	Q969D9
1174	4301-58_2	Thymidine kinase, cytosolic	P04183
1175	5035-7_1	Thymidylate synthase	P04818
1176	4165-2_1	Thyroglobulin	P01266
1177	3873-51_2	Thyroid peroxidase	P07202
1178	3521-16_2	Thyroid Stimulating Hormone	P01215 P01222
1179	2706-69_2	Thyroxine-binding globulin	P05543
1180	4931-59_1	Tissue Factor	P13726
1181	3336-50_1	Tissue factor pathway inhibitor	P10646
1182	2212-69_1	Tissue-type plasminogen activator	P00750
1183	2603-61_2	T-lymphocyte activation antigen CD80	P33681
1184	5337-64_3	T-lymphocyte activation antigen CD86	P42081
1185	3324-51_1	T-lymphocyte surface antigen Ly-9	Q9HBG7
1186	3835-11_2	Toll-like receptor 2	O60603
1187	11101-18_3	Toll-like receptor 4	O00206
1188	3647-49_4	Toll-like receptor 4:Lymphocyte antigen 96 complex	O00206 Q9Y6Y9
1189	10356-21_3	Transcription factor AP-1	P05412
1190	5384-67_3	Transcription factor IIIB 90 kDa subunit	Q92994
1191	3009-3_2	Transforming growth factor beta receptor type 3	Q03167
1192	2333-72_1	Transforming growth factor beta-1	P01137
1193	4156-74_1	Transforming growth factor beta-2	P61812
1194	3520-58_1	Transforming growth factor beta-3	P10600
1195	3283-21_1	Transforming growth factor-beta-induced protein ig-h3	Q15582
1196	9201-13_3	Transgelin-2	P37802
1197	4306-4_2	Transketolase	P29401
1198	3872-2_1	Translationally-controlled tumor protein	P13693
1199	5080-131_3	Transmembrane glycoprotein NMB	Q14956
1200	9185-15_3	Trefoil factor 1	P04155
1201	9191-8_3	Trefoil factor 2	Q03403
1202	4721-54_2	Trefoil factor 3	Q07654
1203	4309-59_3	Triosephosphate isomerase	P60174
1204	5033-27_1	Tropomyosin alpha-1 chain	P09493
1205	7660-21_3	Tropomyosin alpha-4 chain	P67936
1206	4472-5_2	Tropomyosin beta chain	P07951
1207	5441-67_3	Troponin I, cardiac muscle	P19429
1208	5440-26_3	Troponin I, fast skeletal muscle	P48788
1209	5315-22_3	Troponin T, cardiac muscle	P45379
1210	3049-61_2	Trypsin-1	P07477
1211	5034-79_1	Trypsin-2	P07478
1212	3479-71_2	Trypsin-3	P35030
1213	3403-1_2	Tryptase beta-2	P20231
1214	3404-51_2	Tryptase gamma	Q9NRR2
1215	5936-53_3	Tumor necrosis factor	P01375
1216	2917-3_2	Tumor necrosis factor ligand superfamily member 11	O14788
1217	5939-42_3	Tumor necrosis factor ligand superfamily member 12	O43508

Number	SOMAmer SeqID	Analyte Name	UniProt ID
1218	3059-50_2	Tumor necrosis factor ligand superfamily member 13B	Q9Y275
1219	5355-69_3	Tumor necrosis factor ligand superfamily member 14	O43557
1220	2968-61_1	Tumor necrosis factor ligand superfamily member 15	O95150
1221	2708-54_2	Tumor necrosis factor ligand superfamily member 18	Q9UNG2
1222	2839-2_1	Tumor necrosis factor ligand superfamily member 4	P23510
1223	3052-8_2	Tumor necrosis factor ligand superfamily member 6, soluble form	P48023
1224	3421-54_2	Tumor necrosis factor ligand superfamily member 8	P32971
1225	2599-51_2	Tumor necrosis factor ligand superfamily member 9	P41273
1226	4832-75_2	Tumor necrosis factor receptor superfamily member 10A	O00220
1227	5534-49_2	Tumor necrosis factor receptor superfamily member 10B	O14763
1228	14121-24_3	Tumor necrosis factor receptor superfamily member 10D	Q9UBN6
1229	5424-55_3	Tumor necrosis factor receptor superfamily member 11A	Q9Y6Q6
1230	2526-53_3	Tumor necrosis factor receptor superfamily member 11B	O00300
1231	5138-50_3	Tumor necrosis factor receptor superfamily member 12A	Q9NP84
1232	2704-74_1	Tumor necrosis factor receptor superfamily member 13B	O14836
1233	5383-14_3	Tumor necrosis factor receptor superfamily member 13C	Q96RJ3
1234	5352-11_3	Tumor necrosis factor receptor superfamily member 14	Q92956
1235	2665-26_2	Tumor necrosis factor receptor superfamily member 17	Q02223
1236	5526-53_3	Tumor necrosis factor receptor superfamily member 18	Q9Y5U5
1237	5131-15_3	Tumor necrosis factor receptor superfamily member 19	Q9NS68
1238	5115-31_3	Tumor necrosis factor receptor superfamily member 19L	Q969Z4
1239	2654-19_1	Tumor necrosis factor receptor superfamily member 1A	P19438
1240	3152-57_1	Tumor necrosis factor receptor superfamily member 1B	P20333
1241	5404-53_3	Tumor necrosis factor receptor superfamily member 21	O75509
1242	5073-30_2	Tumor necrosis factor receptor superfamily member 25	Q93038
1243	3083-71_1	Tumor necrosis factor receptor superfamily member 27	Q9HAV5
1244	2636-10_2	Tumor necrosis factor receptor superfamily member 3	P36941
1245	3730-81_2	Tumor necrosis factor receptor superfamily member 4	P43489
1246	5392-73_2	Tumor necrosis factor receptor superfamily member 6	P25445
1247	5070-76_3	Tumor necrosis factor receptor superfamily member 6B	O95407
1248	2605-49_2	Tumor necrosis factor receptor superfamily member 8	P28908
1249	2598-9_3	Tumor necrosis factor receptor superfamily member 9	Q07011
1250	2977-7_2	Tumor necrosis factor receptor superfamily member EDAR	Q9UNE0
1251	5036-50_1	Tumor necrosis factor-inducible gene 6 protein	P98066
1252	2619-72_2	Tumor-associated calcium signal transducer 2	P09758
1253	3341-33_4	Tyrosine-protein kinase ABL1	P00519
1254	3416-2_2	Tyrosine-protein kinase BTK	Q06187
1255	3363-31_4	Tyrosine-protein kinase CSK	P41240
1256	4220-39_2	Tyrosine-protein kinase Fer	P16591
1257	3810-50_2	Tyrosine-protein kinase Fgr	P09769
1258	3813-3_2	Tyrosine-protein kinase Fyn	P06241
1259	3374-49_2	Tyrosine-protein kinase HCK	P08631
1260	4998-50_2	Tyrosine-protein kinase JAK2	O60674
1261	3452-17_2	Tyrosine-protein kinase Lck	P06239
1262	3453-87_2	Tyrosine-protein kinase Lyn	P07948
1263	3381-24_2	Tyrosine-protein kinase Lyn, isoform B	P07948
1264	2844-53_2	Tyrosine-protein kinase receptor Tie-1, soluble	P35590
1265	2611-72_2	Tyrosine-protein kinase receptor TYRO3	Q06418
1266	3646-7_4	Tyrosine-protein kinase Tec	P42680
1267	8474-6_3	Tyrosine-protein kinase transmembrane receptor ROR1	Q01973
1268	2878-66_2	Tyrosine-protein kinase Yes	P07947
1269	4476-22_2	Tyrosine-protein kinase ZAP-70	P43403
1270	3005-5_2	Tyrosine-protein phosphatase non-receptor type 1	P18031
1271	3397-7_4	Tyrosine-protein phosphatase non-receptor type 11	Q06124
1272	3401-8_2	Tyrosine-protein phosphatase non-receptor type 2	P17706
1273	4318-12_3	Tyrosine-protein phosphatase non-receptor type 6	P29350
1274	5430-66_3	Tyrosine-protein phosphatase non-receptor type substrate 1	P78324
1275	4474-19_2	Ubiquitin	P62979
1276	9215-117_3	Ubiquitin carboxyl-terminal hydrolase 25	Q9UHP3
1277	5019-16_2	Ubiquitin carboxyl-terminal hydrolase isozyme L1	P09936
1278	2846-24_2	Ubiquitin+1, truncated mutation for UbB	P62979
1279	9199-6_3	Ubiquitin-conjugating enzyme E2 G2	P60604
1280	3874-8_1	Ubiquitin-conjugating enzyme E2 L3	P68036
1281	3905-62_1	Ubiquitin-conjugating enzyme E2 N	P61088
1282	3836-51_2	Ubiquitin-fold modifier 1	P61960
1283	3405-6_2	Ubiquitin-fold modifier-conjugating enzyme 1	Q9Y3C8
1284	14151-4_3	Ubiquitin-like protein ISG15	P05161
1285	4786-58_3	UMP-CMP kinase	P30085

Number	SOMAmer SeqID	Analyte Name	UniProt ID
1286	2652-15_1	Urokinase plasminogen activator surface receptor	Q03405
1287	4158-54_2	Urokinase-type plasminogen activator	P00749
1288	4209-60_2	Vacuolar protein sorting-associated protein VTA1 homolog	Q9NP79
1289	2967-8_1	Vascular cell adhesion protein 1	P19320
1290	2597-8_3	Vascular endothelial growth factor A	P15692
1291	4867-15_2	Vascular endothelial growth factor A, isoform 121	P15692
1292	3132-1_1	Vascular endothelial growth factor C	P49767
1293	13098-93_3	Vascular endothelial growth factor D	O43915
1294	3651-50_5	Vascular endothelial growth factor receptor 2	P35968
1295	2358-19_2	Vascular endothelial growth factor receptor 3	P35916
1296	3522-57_1	Vasoactive Intestinal Peptide	P01282
1297	7638-30_3	Vesicular integral-membrane protein VIP36	Q12907
1298	2961-1_2	Vitamin K-dependent protein C	P04070
1299	2700-56_2	Vitamin K-dependent protein S	P07225
1300	13125-45_3	Vitronectin	P04004
1301	3050-7_2	von Willebrand factor	P04275
1302	4569-52_2	VPS10 domain-containing receptor SorCS2	Q96PQ0
1303	3191-50_2	WAP, kazal, immunoglobulin, kunitz and NTR domain-containing protein 1	Q96NZ8
1304	3235-50_2	WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 2	Q8TEU8
1305	2848-2_2	Wnt inhibitory factor 1	Q9Y5W5
1306	3057-55_1	WNT1-inducible-signaling pathway protein 1	O95388
1307	5927-4_3	WNT1-inducible-signaling pathway protein 3	O95389
1308	3481-87_1	Xaa-Pro aminopeptidase 1	Q9NQW7
1309	5082-51_3	X-linked interleukin-1 receptor accessory protein-like 2	Q9NP60
1310	2835-1_4	X-ray repair cross-complementing protein 6	P12956

Appendix C

C.1 - Sample R code for LASSO analysis

A sample of the R code demonstrating the method for LASSO analysis. Annotated to show descriptions of the sequence and data treatment and handling steps.

```
# 1 - Load data with first row headings from global environment

Whole.Soma.Raw <- read.csv("~/Documents/Work/Oxford/Biomarkers/R Analyses/Whole Soma Raw.csv")


# 2 - Data Preparation


#Lines below tell R to ignore groups A,B &C

Whole.Soma.Raw <- Whole.Soma.Raw[Whole.Soma.Raw$Group != "Group A",]

Whole.Soma.Raw <- Whole.Soma.Raw[Whole.Soma.Raw$Group != "Group B",]

Whole.Soma.Raw <- Whole.Soma.Raw[Whole.Soma.Raw$Group != "Group C",]


#Lines below remove the space in "Group D"
```

```
Whole.Soma.Raw$Group <- as.character(Whole.Soma.Raw$Group)

Whole.Soma.Raw$Group2 <- Whole.Soma.Raw$Group

Whole.Soma.Raw$Group2[Whole.Soma.Raw$Group2 == 'Group D'] <- 'GroupD'

Whole.Soma.Raw$Group2[Whole.Soma.Raw$Group2 == 'Group E'] <- 'GroupE'



#Next lines create an empty matrix and then a loop in order to allow scale & centre
commands

#to be written once but work for all variables

scaled.markers <- matrix(ncol = 1310, nrow = nrow(Whole.Soma.Raw))

for (i in 1:1310){

 scaled.markers[, i] <- scale(Whole.Soma.Raw[,i+1], center = FALSE, scale =
TRUE)

}

#Lines above have filled the matrix with data and scaled but not centered it

#Line below then log transforms the scaled data

log.markers <- log(scaled.markers)
```

```
#The above matrix will have "X-" column headings. The line below saves the
correct

#column headings to be later used

quant_markers <- names(Whole.Soma.Raw)[2:1311]

#Line below allows output data to be viewed as a data frame rather than a string
of data

log.markers <- data.frame(log.markers)



#Line below adds the newly created column from above "Group2" to the data
frame

DAT <- cbind(Whole.Soma.Raw$Group2, log.markers)

#The line below reapplies the correct heading to each column

names(DAT)[2:1311] <- quant_markers

#Line below renames first column in data frame as "group"

names(DAT)[1] <- "group"

#Line below makes the contents of the column "group" either 0 or 1

DAT$group <- as.numeric(DAT$group) - 1
```

```
# 3 - LASSO Method


library(glmnet)

#Package needs data as a matrix not data frame which is performed below

z <- as.matrix(DAT[2:1311])

#Line below defines the outcome of interest (y) as 0 or 1

y <- as.matrix(DAT[1])

#Line below is the initial phase of analysis

glmmod <- glmnet(z, y, alpha=1, family='binomial')

#Line below produces graph of line above (coefficients vs lambda)

plot(glmmod, xvar='lambda')


#Line below cross-validates by sequentially leaving a variable out and repeating
the model

cv.glmmod <- cv.glmnet(z, y, alpha=1, family='binomial', nfolds=10)

#Line below is plot of above

plot(cv.glmmod)
```

```
#Line below generates the best (minimum) lambda value

best_lambda <- cv.glmmod$lambda.min

#Line below simply tells you the number produced by the line above

best_lambda



#There is a lambda value for each model iteration performed above & we now want to see which

#number model is associated with the lowest lambda

lowest <- data.frame(glmmod$lambda)

#Line below finds the number of the row which corresponds with the best lambda

find <- which(lowest[,1]==cv.glmmod$lambda.min)

#Line below makes R show me which line it is

find

#Line below reports the coefficients of the best model

(coef(glmmod)[,find])[which(coef(glmmod)[,find] != 0)]

#Next line reapplies this model to the original data to predict which group each patient is in
```

```
lasso.pred=predict(glmmod, s=best_lambda, newx=z, type="class")

#Line below makes R show me how each patient classifies based on the model

lasso.pred

#Line below makes R show me which group each is truly in

Y
```

Appendix D

D.1 - Sample OPCS-4 code data dictionary for THR

A sample OPCS4 coding data dictionary demonstrating the codes used to identify the index procedure of total hip arthroplasty in the CPRD cohort studied. Up to 2 OPCS4 codes can be used and combined to identify cases.

OPCS1	OPCS2	Description
W371		Primary total prosthetic replacement of hip joint using cement
W378		Other specified total prosthetic replacement of hip joint using cement
W379		Unspecified total prosthetic replacement of hip joint using cement
W381		Primary total prosthetic replacement of hip joint not using cement
W388		Other specified total prosthetic replacement of hip joint not using cement
W389		Unspecified total prosthetic replacement of hip joint not using cement
W391		Primary total prosthetic replacement of hip joint NEC
W398		Other specified other total prosthetic replacement of hip joint
W399		Unspecified other total prosthetic replacement of hip joint
W431	Z843	Primary total prosthetic replacement of joint using cement NEC - Hip joint
W431	Z761	Primary total prosthetic replacement of joint using cement NEC - Head of femur

W431 Z756 Primary total prosthetic replacement of joint using cement NEC - Acetabulum

W438 Z843 Other specified total prosthetic replacement of joint using cement - Hip joint

W438 Z761 Other specified total prosthetic replacement of joint using cement - Head of femur

W438 Z756 Other specified total prosthetic replacement of joint using cement - Acetabulum

W439 Z843 Unspecified total prosthetic replacement of joint using cement - Hip joint

W439 Z761 Unspecified total prosthetic replacement of joint using cement - Head of femur

W439 Z756 Unspecified total prosthetic replacement of joint using cement - Acetabulum

W441 Z843 Primary total prosthetic replacement of joint not using cement NEC - Hip joint

W441 Z761 Primary total prosthetic replacement of joint not using cement NEC - Head of femur

W441 Z756 Primary total prosthetic replacement of joint not using cement NEC - Acetabulum

W448 Z843 Other specified total prosthetic replacement of joint not using cement – Hip joint

W448 Z761 Other specified total prosthetic replacement of joint not using cement – Head of femur

W448 Z756 Other specified total prosthetic replacement of joint not using cement - Acetabulum

W449 Z843 Unspecified total prosthetic replacement of joint not using cement - Hip joint

W449 Z761 Unspecified total prosthetic replacement of joint not using cement - Head of femur

W449 Z756 Unspecified total prosthetic replacement of joint not using cement - Acetabulum

W451 Z843 Primary total prosthetic replacement of joint NEC - Hip joint

W451 Z761 Primary total prosthetic replacement of joint NEC - Head of femur

W451 Z756 Primary total prosthetic replacement of joint NEC - Acetabulum

W458 Z843 Other specified total prosthetic replacement of joint - Hip joint

W458 Z756 Other specified total prosthetic replacement of joint - Head of femur

W458 Z761 Other specified total prosthetic replacement of joint - Acetabulum

W459 Z843 Unspecified total prosthetic replacement of joint - Hip joint

W459 Z756 Unspecified total prosthetic replacement of joint - Head of femur

W459 Z761 Unspecified total prosthetic replacement of joint - Acetabulum

W521 Z843 Primary prosthetic replacement of articulation of bone using cement NEC - Hip joint

W528 Z843 Other specified prosthetic replacement of articulation of bone using cement NEC - Hip joint

W529 Z843 Unspecified prosthetic replacement of articulation of bone using cement NEC - Hip joint

W531 Z843 Primary prosthetic replacement of articulation of bone not using cement NEC - Hip joint

W538 Z843 Other specified prosthetic replacement of articulation of bone not using cement NEC - Hip joint

W539 Z843 Unspecified prosthetic replacement of articulation of bone not using cement NEC - Hip joint

W541 Z843 Primary prosthetic replacement of articulation of bone NEC - Hip joint

W548 Z843 Other specified prosthetic replacement of articulation of bone NEC - Hip joint

W549 Z843 Unspecified prosthetic replacement of articulation of bone NEC - Hip joint

W581 Z843 Primary resurfacing arthroplasty of joint - Hip joint

W581 Z761 Primary resurfacing arthroplasty of joint - Head of femur

W581 Z756 Primary resurfacing arthroplasty of joint - Acetabulum

W588 Z843 Other specified resurfacing arthroplasty of joint - Hip joint

W588	Z761	Other specified resurfacing arthroplasty of joint - Head of femur
W588	Z756	Other specified resurfacing arthroplasty of joint - Acetabulum
W589	Z843	Unspecified resurfacing arthroplasty of joint - Hip joint
W589	Z761	Unspecified resurfacing arthroplasty of joint - Head of femur
W589	Z756	Unspecified resurfacing arthroplasty of joint - Acetabulum
W931		Primary hybrid prosthetic replacement of hip joint using cemented acetabular component
W938		Other specified hybrid prosthetic replacement of hip joint using cemented acetabular component
W939		Unspecified hybrid prosthetic replacement of hip joint using cemented acetabular component
W941		Primary hybrid prosthetic replacement of hip joint using cemented femoral component
W948		Other specified hybrid prosthetic replacement of hip joint using cemented femoral component
W949		Unspecified hybrid prosthetic replacement of hip joint using cemented femoral component
W951		Primary hybrid prosthetic replacement of hip joint using cement NEC
W958		Other specified hybrid prosthetic replacement of hip joint using cement
W959		Unspecified hybrid prosthetic replacement of hip joint using cement