Identification of whole-blood gene expression signature in primary Sjögren's syndrome-associated lymphoma

Shereen Jawad Kadhim Al-Ali

Musculoskeletal Research Group,
Institute of Cellular Medicine,
Faculty of Medical Sciences
Newcastle University

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

March 2016

Declaration

I certify that this thesis titled '**Identification of whole-blood gene expression signature in primary Sjögren's syndrome-associated lymphoma'** is my own work and has not previously been submitted for a degree or other qualification in this or any other university. I declare that this thesis represents my own unaided work, carried out by myself, except where it is acknowledged otherwise in the thesis text.

Shereen Jawad Kadhim Al-Ali

March 2016

Acknowledgment

This work cannot be accomplished without the help of others, so I would like to say thanks to them all. Firstly I would like to give a special thanks to my supervisor Professor Wan-Fai Ng for his advice, continuous support, motivation throughout my PhD. His patience and support helped me overcome many crisis situations and finish my thesis. I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own .

I also want to thank to my co-supervisor Professor David Young, my thesis committee Professor John Isaacs and professor John Loughlin for their advice. I would like to give a special thanks for the BSU members especially Mr. Andrew Skelton for his great help in analyzing the microarray data and to Dr. Simon J. Cockell who also helped with analyzing the data. I would like to express my gratitude to Cambridge Genomic Services' staff how helped with performing the microarray. Also I am so thankful for Dr. Dennis Lendrem for helping me with building the prediction models. I extend my appreciation to all the UKPSSR team who helped with collecting the samples, my thanks also goes to the primary Sjögren's syndrome patients and healthy controls volunteers who their clinical samples are the base of my project. A special thanks to our collaborators in University of Uppsala, Sweden and Stavanger University, Norway for providing us with some of their samples. Thank you to the MRG members who helped me during my study. A warm thanks to Dr. Jessica Tarn for her help and support since day one, and to Kate Hackett and her wonderful family for their kindness and support, I hope our friendship will last forever even when we live apart. Also I want to thank Dr. Katherine James and Mrs. Janet Herdman and all my group members past and present.

I want to say thanks to a friend who is precious to me Zeenah Atwan, thanks for standing by my side when times get hard and thanks for your encouraging selfies!!. You are my rock, I have been blessed with such a friend. Many thanks to my friend Einas Abood for her support and to Siti Mohammed for her encouragement and to my friends back home Nejwa Mohammed and Athraa abdul Ameer for their support.

My heartfelt thanks and gratitude to my family specially my beloved mother for her care and her continuous prayers for me and to my much-loved father for his endless encouragement (God bless them both). Also many thanks to my beloved sisters Saba and Shaymaa and brothers Saad and Salwan. A special thanks to my uncle Faisal Al-Emarah and his wife Dr. Rabab Hussain for their extraordinary support during my study. I also want to thank my cousin Hiyam Al-Ani for her continues care.

I would like to say thank you for my sponsor the higher Committee for Education Development in Iraq (HCED) for funding my study.

Many thanks

Dedication

"After climbing a great hill, one only finds out that there are many more hills to climb"

Nelson Mandela

This thesis is dedicated to …

The biggest heart I'll ever know… ***My Mother***

The greatest hero I'll ever know ***.. My Father***

My beloved country…….. ***Iraq***

Table of Contents

Table of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Supplementary Figures

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Supplementary tables

Abstract

Primary Sjögren's syndrome (pSS) is an autoimmune disease of the exocrine glands. The syndrome is characterized by many systemic features, including a markedly increased risk of lymphoma development. PSS-associated lymphoma was first reported in 1963, however, the mechanisms and the risk factors of lymphoma development in pSS patients remain incompletely understood. The aim of my project is to identify a whole-blood gene expression signature in pSS-associated lymphoma. To achieve this goal, I first evaluated the effect of the depletion of the abundant globin mRNA in whole-blood samples on microarray analyses of pSS. Then I prepared samples (globin mRNA depleted samples) from a "Discovery cohort" which consisted of five subject groups ("pSS (non-lymphoma)", "pSS-associated lymphoma", "pSS-paraproteinemia", "pSS-other cancers" and "healthy controls") to identify a list of differentially expressed genes (DEGs) between the "pSS (non-lymphoma)" and the "pSS-associated lymphoma" groups. The next step was to confirm the differential expression of these genes using qRT-PCR. This has led to the identification of a potential gene expression signature for pSS-associated lymphoma. To further explore the role of these genes in the pathogenesis of pSS-associated lymphoma, I performed pathway analysis using various algorithms provided by Ingenuity Pathway Analysis (IPA). I also compared the microarray data of different subject groups to investigate whether the potential gene signature was "specific" for pSS-associated lymphoma. I then validated the potential transcriptomic signature "biologically" using an independent cohort (the "Validation cohort") consisting of two subject groups – "pSS (without lymphoma) and "pSS-associated lymphoma". Moreover, the potential biosignature was tested in a group of pSS patients with untreated lymphoma. Prediction modelling was used to identify the important genes within the potential biosignature that best predict the development of pSS-associated lymphoma.

I showed that globin mRNA depletion of whole-blood samples provided potentially more sensitive microarray data compared with paired non–globin RNA depleted samples. From the microarray analysis of the "Discovery cohort", 68 DEGs were identified between the lymphoma and non-lymphoma groups (68-DEGs-Mi). qRT-PCR confirmed the differential expression of 26 genes (26-DEGs-MiPCR). Biological validation with an independent

cohort verified 3 genes (3-gene biosignature), 2 of which were up-regulated (*NUDT14*, *MGST3)* and 1 gene was down-regulated (*BMS1*) in pSS-associated lymphoma. Moreover, 2 genes in addition to *NUDT14* (*DRAP1*, *DYNLL1)* were also differentially expressed in a cohort of pSS patients with untreated lymphoma. Prediction modelling suggested that *NUDT14* was the most important gene in predicting membership in the pSS-associated lymphoma group. Pathway analysis of the differentially expressed genes in pSS-associated lymphoma revealed several canonical pathways such as "Aryl Hydrocarbon Receptor Signalling," "Histamine Degradation," "Unfolded protein response," "Neuregulin Signalling," and "T Cell Receptor Signalling." In addition, the Downstream Effects analysis revealed the biological functions in pSS-associated lymphoma and the Upstream Regulators analysis investigates possible gene regulators. Moreover, comparisons of microarray gene-expression data between other pSS subgroups suggest the DEGs were unique to pSS-associated lymphoma. IPA showed that "Interferon Signalling pathway" was the top canonical pathway in all pSS subgroups. Furthermore, similar patterns were seen in the IPA Downstream Effects analyses for the "pSS (non-lymphoma)", "pSS-paraproteinemia" and "pSS-other cancers" groups, while the "pSS-associated lymphoma" group showed a unique pattern, further indicate that a unique gene expression signature exist in pSS-associated lymphoma.

Abbreviations

26-DEGs-MiPCR	26 differentially expressed genes from microarray and RT-PCR
68-DEGs-Mi	68 differentially expressed genes from microarray and RT-PCR
ACA	Anti-centromere antibodies
ACR	American College of Rheumatology
AECG	American-European Consensus Group
AHR	Aryl hydrocarbon receptor nuclear translocator
AICDA	Activation-induced cytidine deaminase
AID	Activation-induced cytidine deaminase
ALL	Acute lymphoblastic leukemia
AMA	Anti-mitochondrial antibodies
ANA	Antinuclear antibodies
ANOVA	Analysis of variance
Anti-M3R	Anti-muscarinic acetylcholine receptor M3
APRIL	A proliferation-inducing ligand
ARNT	Aryl hydrocarbon receptor nuclear translocator
ASMA	Anti-smooth muscle antibodies
AUC	Area under the curve
BAFF	B-cell activating factor
BLK	B lymphoid kinase
BSU	Bioinformatics support unit
C3	Complement component 3
C4	Complement component 4
CA-I	Carbonic anhydrase I
CA-II	Carbonic anhydrase II
CCP	Cyclic citrullinated peptides
CD	Cluster of differentiation
cDNA	Complementary DNA
CGS	Cambridge Genomic Services
CI	Confidence interval
CLL	Chronic lymphoid leukemia
CMV	Cytomegalovirus

COMPASS	Composite Autonomic Symptom Scale
CRI	Cross-reactive idiotype
cRNA	Complementary RNA
Ct	Threshold cycle
CXCL10	Chemokine (C-X-C Motif) Ligand 10
CXCL13	Chemokine (C-X-C Motif) Ligand 13
CXCL9	Chemokine (C-X-C Motif) Ligand 9
CXCR5	Chemokine (C-X-C motif) receptor 5
DCs	Dendritic cells
DEGs	Differentially expressed genes
DLBCL	Diffused Large B-Cell Lymphoma
DNA	Deoxyribonucleic acid
dNTP	Deoxyribose nucleoside triphosphate
EBV	Epstein–Barr virus
EQ-SD	Euroqol questionnaire
ER	Endoplasmic reticulum
ESR	Erythrocyte sedimentation rate
ESSDAI	EULAR Sjögren's Syndrome Disease Activity Index
ESSPRI	EULAR Sjögren's Syndrome Patient Reported Index
EULAR	European League Against Rheumatism
FAM167	Family with sequence similarity 167 member A
fDC	Follicular dendritic cell
FDR	False discovery rate
FOXP3	Forkhead box P3
GC	Germinal centre
GO TERMS	Gene Ontology terms
GWAS	Genome–wide association study
GWS	Genome-wide significant
H2O	Water
H4	Histamine H_4
H4R	Histamine H_4 receptor
HADs	Hospital Anxiety and Depression Scale

HC	Healthy control
HCC	Hepatocellular carcinoma
HCMV	Human cytomegalovirus
HCV	Hepatitis C virus
HLA-DQ	Human Leukocyte Antigen - DQ subregion
HLA-DR	Human Leukocyte Antigen - antigen D Related
HTLV-1	Human T-cell lymphotropic virus type 1
IFN	Interferon
IFN-β	Interferon- beta
IFN-γ	Interferon gamma
IgA	Immunoglobulin A
IgD	Immunoglobuin D
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IKB	Ingenuity Knowledge Base
IL-	Interleukin
IP-10	Interferon-γ–inducible 10-kd protein
IPA	Ingenuity Pathways Analysis
IPI	International Prognostic Index
IRF-5	Interferon regulatory factor 5
IRs	Incidence rates
KCS	Keratoconjuctivitis sicca
KEGG	Kyoto Encyclopedia of Genes and Genomes
LPL/WM	Lymphoplasmacytic lymphoma/Waldenstrom macroglobulinemia
M3R	Muscarinic 3 receptor
MALT	Mucosa-associated lymphoid tissue
MGB	Minor Groove Binders
MGUS	Monoclonal gammopathy of undetermined significance
MHC	Major histocompatibility complex
MIG	Monokine induced by IFN-γ
miRNA	Micro RNA
MM	Multiple Myeloma

M-MLV RT	Moloney Murine Leukemia Virus Reverse Transcriptase
mRF	Monoclonal rheumatoid factor
mRNA	Messenger RNA
MxA	Myxovirus resistance protein A
MZL	Marginal zone lymphoma.
NF-KB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NFQ	Non-fluorescent quencher
NHL	Non-Hodgkin Lymphoma
NK	natural killer cells
NMZL	Nodal marginal zone lymphoma.
NRGs	Neuregulins
NSAID	Nonsteroidal anti-inflammatory drugs
PAS	Pern-Arnt-Sim family
PBMCs	Peripheral Blood Mononuclear cells
PCA	Principle Component Analysis
pDCs	Plasmacytoid dendritic cells
PNA	Peptide nucleic acids
PROFAD-SSI	Profile of Fatigue and Discomfort--Sicca Symptoms Inventory
PRs	Prevalence rates
pSS	Primary Sjögren’s Syndrome
QC	Quality control
qRT-PCR	Quantitative real time polymerase chain reaction
RA	Rheumatoid Arthritis
RBCs	Red Blood Cells
RCT	Randomised controlled trial
RF	Rheumatoid factor
RIN	RNA integrity number
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
RSN	Robust spline normalization
RTX	Rituximab
SAS JMP	Statistical Analysis System

SCAI	Sjögren's Systemic Clinical Activity Index
SDI	SLE Collaborating Clinics Damage Index
SDS	Data Assist™ Software
SF-36	Short Form 36
SGECs	Salivary gland epithelial cells
SICCA	Sjögren's International Collaborative Clinical Alliance
SIR	Standardized incidence ratio
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
SOCSI	Suppressor gene of cytokine signalling 1
SS	Sjögren's syndrome
SSA	Anti-Sjögren's-syndrome-related antigen A
SSB	Sjögren's syndrome type B antigen
SSc	Systemic sclerosis
SSDDI	Sjögren's Syndrome Disease Activity Index
sSS	Secondary Sjögren's syndrome
STAT4	Signal Transducer and Activator of Transcription protein
TCDD	2,3,7,8-Tetrachlorodibenzo-*p*-dioxin
Tfh	Follicular T helper cells
TGF	Transforming growth factor
TGF-β	Transforming growth factor β
Th	T helper
TLR	Toll-Like receptor
TNF	Tumor necrosis factor
TNFSF13B	Tumour necrosis factor ligand superfamily member 13 B
TNF-α	Transforming growth factor alpha
TNIP1	*TNFAIP3* interactive protein 1
Treg	Regulatory T-cell
UKPSSR	UK primary Sjögren's syndrome Registry
UPR	Unfolded protein response
VAS	Visual analogue scales
VST	Variance stabilizing transformation

WBCs	White Blood cells
WCC	White Cell Count
WHO	World Health Organization

Chapter 1
Introduction and literature review

1.1 Overview of primary Sjögren's Syndrome (pSS):

Sjögren's Syndrome (SS) is an autoimmune disease of the exocrine glands, particularly the salivary and lacrimal glands. The hallmark features of SS are a dry mouth and dry eyes. In addition, a variety of systemic manifestations can be observed. Occurring alone without any other associated autoimmune conditions, it is referred to as primary Sjögren's Syndrome (pSS); when Sjögren's syndrome occurs with other autoimmune disease, it is referred to as secondary Sjögren's Syndrome (sSS) (Kassan and Moutsopoulos, 2004).

Historically, the first case of a dry mouth and dry eyes (sicca syndrome) was reported by W.B. Hadden and J. W. Hutchinson in 1871. Afterward the term 'Mikulicz disease', which describes the association of parotid, submandibular and lacrimal glands enlargement with sicca syndrome, was introduced by Johann von Mikulicz-Radecki in 1888. In 1925, Gougerot's syndrome was described; this syndrome had the three main symptoms of Sjögren's syndrome - a dry mouth, dry eyes and polyarthritis (Ghafoor, 2012). In 1933, the Swedish ophthalmologist Henrik Sjögren demonstrated the clinical symptoms of a dry mouth and dry eyes in association with rheumatoid arthritis in 13 of 19 women, and coined the term 'keratoconjunctivitis sicca' that discriminates the syndrome from xerophthalmia that results from vitamin A deficiency. Since then, the term 'Sjögren's Syndrome' has been widely accepted (Bloch et al., 1992). Henrik Sjögren's work was published in English in 1943 and given recognition for his contributions to the field of medicine (Ghafoor, 2012).

Primary Sjögren's syndrome (pSS) has a population prevalence of about 0.5% with a 9:1 female to male ratio (Bowman et al., 2004). Recently, a meta-analysis study including 21 epidemiological studies of pSS which reported data on incidence rates (IRs), prevalence rates (PRs) and the female to male ratio was conducted by Qin and colleagues (Qin et al., 2015). The pooled IRs of pSS was 6.92 (95% CI: 4.98 to 8.86)/100,000 person-years and the overall PRs was 60.82 (95% CI: 43.69 to 77.94) cases/100,000 inhabitants. Furthermore, within the incidence data the female to male ratio was 9.15 (95% CI: 3.35 to 13.18), while the female to male ratio in the prevalence data was 10.72 (95% CI: 7.35 to 15.62) (Qin et al., 2015).

pSS is characterized by lymphocytic infiltration in the affected glands. T-cells particularly the $CD4^+$ subsets represent the majority of the infiltrating cells and seem to be the dominant cells in mild lesions, while the proportion of B-cells increased in the more severely affected glands (Mavragani and Moutsopoulos, 2010). The production of autoantibodies derived from autoreactive B-cells in both the affected glands and peripheral blood is another characteristic feature of pSS. Indeed, autoantibodies targeting the ribonucleoproteins Ro (SSA) and La (SSB) are included in the classification criteria of pSS by the American-European Consensus Group (Vitali et al., 2002). In addition to the glandular manifestations, extraglandular manifestations also occur including fatigue, musculoskeletal involvement, dermatological involvement, pulmonary involvement, gastroenterological involvement, renal involvement, neurological involvement, Raynaud's phenomenon, liver involvement and lymphoproliferative disease particularly non-Hodgkin lymphoma (NHL) (Kassan and Moutsopoulos, 2004).

Secondary Sjögren's Syndrome (sSS) is defined as the presence of SS with other autoimmune diseases such rheumatoid arthritis (RA), systemic lupus erythromatosus (SLE) and others. The prevalence of sSS in RA has been estimated to be as high as 62% (Coll et al., 1987). Another study reported that 28% of a group of 307 RA patients has at least one positive feature of sicca complex but only 3.6% had sSS (Haga et al., 2012). The prevalence of sSS with SLE varies from 6.5–19% (Pan et al., 2008, Patel and Shahane, 2014). SS may also occur with systemic sclerosis (SSc) and with other systemic and organ-specific autoimmune diseases (Ramos-Casals et al., 2012). The occurrence of SS with other autoimmune disorders increases the number of symptoms and also complicates diagnosis, depending on the type of concurrent autoimmune disease. For instance, sSS patients with rheumatoid arthritis (RA) often differ clinically, pathologically, serologically and genetically from pSS, whereas patients with sSS in association with systemic lupus erythromatosus (SLE) are more similar to pSS patients (Peters and Isenberg, 2012).

1.2 Characteristic features of primary Sjögren's Syndrome:

1.2.1 Histological features of primary Sjögren's Syndrome:

The characteristic histopathological feature of pSS is lymphocytic infiltration in the affected organs (salivary and lacrimal glands). At the initial stages of the disease the

infiltrates consist of $CD4^+$ T-cells, B-cells, macrophages and dendritic cells (DCs), with $CD4^+$ T-cells being the dominant cell type (Mavragani and Moutsopoulos, 2010). B-cell infiltrates in the glands contribute to the secretion of autoantibodies such as anti-Ro/SSA and anti-La/SSB, accompanied by the formation of germinal centre-like structure (GC-like structure (Routsias and Tzioufas, 2010). The germinal centre-like structure can be defined as the aggregation of immune cells (mostly T- and B-cells) to form a microenvironment or structure resembling germinal centres within the affected organs. These GC-like structures were identified in 17% of pSS patients and contributed to the production of autoantibodies, as patients with GC-like structure had increased levels of autoantibodies (Salomonsson et al., 2003). The proportion of different cell types within the cellular infiltrates changes with the degree of lesion severity. Although $CD4^+$ T-cells are the dominant cells in the lymphocytic infiltrates in pSS salivary glands, the proportion of $CD4^+$ T-cells proportion declines with increased lesion severity. In contrast, the proportion of $CD20^+$ B-cells increases with increased severity of the lesion. Furthermore, the proportion of $FOXP3^+$ regulatory T-cells (Treg) increases in lesions with intermediate severity, while the percentages of $CD8^+$ T-cells, follicular dendritic cell (fDC) and natural killer (NK) cells in the inflammatory infiltrate had no correlation with the degree of lesion severity (Christodoulou et al., 2010).

1.2.2 Immunological features of primary Sjögren's syndrome:

Immune cells in pSS

Various immune cells are involved in pSS, although much of the research has focused on T- and B-cells. Many researchers have studied $CD4^+$ T-cells, which consist of T helper cells (Th) such as Th1, Th2, Th17, as well as other subtypes such as regulatory cells (Treg). It has been reported that Th1/Th2 imbalance is associated with pSS severity locally (salivary glands) and systematically (peripheral blood). However, pSS patients cannot be distinguished from those with non-pSS sicca syndrome according to this feature (van Woerkom et al., 2005, Alunno et al., 2013). Th17 cells rely on TGF-β, IL-6 and IL-1β for differentiation. Th17 cells and their products (IL-17, IL-6, IL-23 and IL-12) have also been implicated in pSS, as expansion of Th17 cells in pSS patients and animal models of SS as well as their role in the immunopathology in pSS has been documented (Katsifis et al., 2009, Lin et al., 2014). Moreover, it has been suggested that Th17 cells play a critical role

in the pathology of pSS in animal models and might be a possible target for pSS treatment (Lin et al., 2015). With regard to regulatory T cells, it has been shown that $FOXP3^+$ Tregs were increased in severe lesions in the minor salivary gland of pSS patients. Furthermore, a reduced number of $FOXP3^+$ Treg cells correlated with adverse predictors of lymphoma development such as low C4 levels and salivary gland enlargement (Christodoulou et al., 2008). Another study recently described a subset of $CD4^+$ T-cells, called the follicular T helper cells (Tfh), in the labial salivary gland in pSS patients. Tfh contributes to the progression of pSS, as these cells play an important role in the development of B-cells, and they are associated with the lymphocytic infiltration in pSS and the formation of GC-like structure (Maehara et al., 2012, Gong et al., 2014).

B-cells are among the infiltrating cells in the salivary gland and are also the source of autoantibody production. The main types of infiltrative B-cells are type II transitional B-cells (IgD^{high}, $CD38^{low}$) and the marginal zone-like B-cells (Youinou et al., 2010). Other B-cell types that are found in pSS salivary glands include memory B-cells, naïve B-cells, plasmablast and long-lived plasma cells. The frequency of memory B-cells was less than other types of B-cells while the long-lived plasma cells seemed to have the highest frequency in the salivary gland (Aqrawi et al., 2012). In pSS patients, $CD27^+$ memory B-cells were found to be accumulating in the parotid glands, which might explain the finding of the reduction of these cells in peripheral blood (Hansen et al., 2002).

The interaction between T-cells, B-cells and chemokines is known but not fully understood. Recently, Jin and colleagues have reported that B-cell maturation is promoted by $CD4^+CXCR5^+$Tfh cells, which are $CD4^+$ T-cells that highly express the chemokine receptor CXCR5 which control Tfh migration to the salivary gland. In their study, the number of $CD4^+CXCR5^+$Tfh cells was increased in both the salivary glands and peripheral blood of pSS patients and was associated with increased number of abnormal $CD19^+CD27^+$ memory B cells and $CD19^+CD27^{high}$ plasma cells in the salivary glands, suggesting that these $CD4^+CXCR5^+$Tfh cells might contribute to pSS pathogenesis by promoting B-cells maturation (Jin et al., 2014a). The newly described regulatory B-cells (Bregs), which include $CD19^+CD24^{hi}CD38^{hi}$ IL-10 producing cells, has been implicated in pSS (Furuzawa-Carballeda et al., 2013), but their role in pSS is still not clear.

Other immune cells that may be associated with the pathology of pSS include dendritic cells, monocytes and macrophages. It is likely that all these cells work together with various autoantibodies, cytokines and chemokines leading to the development of this complex inflammatory disease.

Autoantibodies in pSS

The presence of autoantibodies has been well described in autoimmune diseases. In pSS, the most commonly associated autoantibodies are antibodies against Ro/SSA and La/SSA autoantigens. Ro/SSA has two non-homologous proteins, Ro52/TRIM21 and Ro60/TROVE2. The tripartite motif Ro52/TRIM21 participates in many functions such as acting as an intracellular Fc-receptor or E3-ubiquitin ligase. When Ro52/TRIM21 serves as an E3-ubiquitin ligase, it regulates cell proliferation, as well as the activation and induction of cell death. Additionally, it regulates TLR-signaling leading to the production of interferon (IFN) through polyubiquitin-mediated degradation of interferon regulatory factors (IRFs); (reviewed (Oke and Wahren-Herlenius, 2012, Yang et al., 2009, Kyriakidis et al., 2014). The ring-shaped RNA binding protein (Ro60/TROVE2) is involved in RNA degradation, promoting cell survival after UV irradiation (Bollain-y-Goytia et al., 2000).

Researchers have also showed that other type of autoantibodies are linked to pSS. For example, antinuclear antibodies (ANA) are found in 83% of pSS patients (Nardi et al., 2006). Rheumatoid factor (RF) and cryoglobulins are also associated with pSS, present in 36–74% and 9–15% of pSS sera respectively (Fauchais et al., 2010, Martel et al., 2011). The presence of cryoglobulins predicts a poor disease outcome and an increased risk of lymphoma development. Anti-centromere antibodies (ACA) are found in 3.7–27.4 % of pSS patients (Caramaschi et al., 1997, Hsu et al., 2006). Another autoantibody, anti-mitochondrial antibody (AMA), has also been implicated in pSS (Zurgil et al., 1992). Antibodies against muscarinic receptor (anti-M3R), smooth muscle (anti-SMA)), carbonic anhydrase II (anti-CA-II) and cyclic citrullinated peptides (anti-CCP) have also been reported in pSS (Manthorpe et al., 1979, Bacman et al., 1996, Takemoto et al., 2005, Barcelos et al., 2009).

Cytokines and chemokines in pSS

Cytokines are produced by a variety of cells and their role is to regulate both innate and adaptive immunity. pSS is a chronic inflammatory disease and the chronic inflammatory

responses indicate cytokine imbalance in the affected organs and peripheral blood. The predominance of Th2 cytokines occurs in the initial stages of pSS, with disease progression, Th1 cytokines become more prevalent (Mitsias et al., 2002). Cytokines can act as pro-inflammatory or anti-inflammatory agents. Since the association of IFN in pSS has been widely investigated. Type I IFN includes IFN-α and IFN-β, and its upregulation in the salivary gland and peripheral blood of pSS patients has been well documented. It is thought that plasmacytoid dendritic cells (pDCs) found in pSS salivary glands are responsible for the production of IFN-α (Emamian et al., 2009, Oxholm et al., 1992, Gottenberg et al., 2006, Szodoray et al., 2005). In addition to type I IFN, high levels of type II IFN (represented by IFN-γ) have also been documented in both the sera and salivary glands of pSS patients and may play a pathological role in pSS (Willeke et al., 2009).

Another cytokine named Tumour Necrosis Factor alpha (TNF-α), which is produced by epithelial cells and $CD4^+$ T-cells, correlates with the degree of systemic features of pSS (Garcic-Carrasco et al., 2001). Interestingly, both TNF-α and interleukin (IL)-1β are found in high levels in pSS patients and may contribute to the chronic inflammatory features of pSS within the affected glands (Ek et al., 2006). IL-12 and IL-18, both produced by monocytes and macrophages, are elevated in pSS and promote the secretion of IFN-γ (Dinarello, 2007, Manoussakis et al., 2007). IL-6 contributes in many functions, it plays a role in the differentiation and growth of B-cells and is present at high levels in the sera of pSS patients. IL-6 also stimulates T-cells and their transition to cytotoxic T-cells. Finally, IL-6 works with TNF-α in contributing to the inflammatory features in the pSS salivary gland (Ishihara and Hirano, 2002, Vucicevic Boras et al., 2006). The levels of B-cell activating factor (BAFF), known also as the tumour necrosis factor ligand superfamily member 13 B (TNFSF13B), are elevated in the serum and salivary glands of pSS patients (Szodoray et al., 2004). Moreover, it has also been suggested that A proliferation-inducing ligand (APRIL), which is another ligand of the BAFF receptor, may be important in pSS pathogenesis (Vosters et al., 2012). Other pro-inflammatory cytokines such as IL-17, IL-21 and IL-22 and anti-inflammatory cytokines such as IL-4, IL-10 and TGF-β, may also be involved in pSS pathogenesis. More recently, over-expression of IL-22 receptor in pSS and pSS-associated lymphoma and the regulation of IL-22 by IL-18 have been reported (Ciccia et al., 2015).

Chemokines are important in the activation and chemotaxis of leucocytes and maintaining the Th1/Th2 balance in diseases and conditions such as autoimmune diabetes, atopic keratoconjunctivitis and cutaneous lupus erythematosus (Kim et al., 2002, Yamagami et al., 2005, Wenzel et al., 2005). In pSS, they are involved in the process of recruiting inflammatory cells (T-cells) into the salivary gland. For instance, interferon-γ–inducible 10-kd protein (IP-10) also known as (CXCL10) and monokine induced by IFN-γ (MIG) also known as (CXCL9) have been implicated in pSS pathogenesis (Ogawa et al., 2002). Another chemokine reported to be associated with pSS is chemokine (C-X-C Motif) Ligand 13 (CXCL13), which regulates B-cell chemotaxis and serves as a biomarker for pSS (Kramer et al., 2013). More recently, it has been reported that high levels of serum CXCL13 was associated with lymphoma development and disease activity in pSS patients (Nocturne et al., 2015a)

1.3 The pathogenesis of primary Sjögren's Syndrome:

The exact aetiology of pSS is unknown. It is believed that one of the potential disease triggers is viral infection. Various viruses have been implicated, including Epstein–Barr virus (EBV), Human T-Lymphotropic virus 1 (HTLV-1) and retroviruses (Ramos-Casals et al., 2002, James et al., 2001, Terada et al., 1994, Lee et al., 2012). Another aetiological factor is genetic predisposition, in particularly the major histocompatibility complex (MHC) class II molecules (Nakken et al., 2001). There is strong evidence for a relationship between the development of pSS and polymorphisms of MHC class II molecules. In European pSS patients, the production of anti-SSB/La autoantibodies correlates with the presence of human leukocyte antigen (HLA)-DQ heterodimer and is highly associated with HLA-DQB1*02 and HLA-DQA1*0501 allele (Tzioufas et al., 2002). HLA-DR2, DR3, and DQ8 are correlated with T-cell and B-cell responses to the human Ro60 molecules in mice (Paisansinsup et al., 2002). Meta-analyses of STAT4 rs7574856 single nucleotide polymorphism (SNP) showed a significant effect of the T allele on pSS development, implicating the gene in pSS pathogenesis (Palomino-Morales et al., 2010). Furthermore, it has been suggested that IRF-5, in particular the IRF5 rs2004640 T allele, may represent a genetic susceptibility allele but this finding requires confirmation using a bigger population (Miceli-Richard et al., 2007). Lessard and colleagues (2013) conducted a genome–wide association study (GWAS) to identify new risk loci that are important in the development

of pSS. In this study, the association of HLA region at 6p21 was revealed. In addition, new association were established as all of the associations identified surpassed the genome-wide significant (GWS) threshold ($p= 5 \times 10^{-8}$) including the interferon regulatory factor 5 (IRF5), Signal transducer and activator of transcription 4 (STAT4), IL-12A, Family with sequence similarity 167, member A- B lymphoid kinase (*FAM167A-BLK*), chemokine (C-X-C motif) receptor 5 (DDX6-CXCR5) and *TNFAIP3* interactive protein 1 (TNIP1). Moreover, 8 associations were identified (statistically significant in the meta-analysis) but they did not exceed the GWS threshold, these associations were *TNFAIP3, PTTG1, PRDM1, DGKQ, FCGR2A, IRAK1BP1, ITSN2, and PHIP* (Lessard et al., 2013). In another study, GWAS experiment was performed in three stages (discovery and 2 replication stages), using samples from pSS patients and controls in Han Chineses. The previously identified associations of STAT4, *TNFAIP3* and the MHC regions with pSS in Europeans were also confirmed in the Han Chinese samples. In addition, a new susceptibility locus represented by *GTF2I* at 7q11.23 was identified. Moreover, the rs117026326 in *GTF2I* showed the most significant association with pSS (Li et al., 2013). Furthermore, in another study, 2 SNPs in TNIP1 (NF-κB repressor) were identified to be associated with seropositive pSS patients. The 2 identified SNPs are rs3792783 and rs7708392 (Nordmark et al., 2013).

The first stage of pSS pathology is lymphocytic infiltration into the salivary and lacrimal glands. The cellular infiltrate is composed of T cells ($CD4^+$) which form about 50–70% of the total population, other immune cells may also be involved and their proportion vary according to severity of the disease (Christodoulou et al., 2010). However, some patients suffer from glandular dysfunction without having severe inflammatory infiltration. In such cases, glandular dysfunction might be the result of an imbalance between pro-inflammatory (IFNs, IL-12, IL-18, TNF-α, IL-1β, IL-6, BAFF, IL-17 and IL-23) and anti-inflammatory cytokines (Transforming growth factor β (TGFβ), IL-4 and IL-10). The overall inflammatory milieu may decrease fluid secretion as well as cause systemic manifestations and lymphomagenesis (Roescher et al., 2009).

Interestingly, vitamin D, which is known for its immune-modulatory function, may also be important in pSS. In this regard, the association between decreased levels of vitamin D and the presence of neuropathy or lymphoma has been described in pSS as well as in other conditions (Agmon-Levin et al., 2012).

Overall, exposure to the possible etiological factors leads to an imbalance of the immunological and autoimmune responses. Furthermore, glandular epithelial cells activation leads to the production of exosomes and apoptotic blebs that contain intracellular antigens. These self-antigens are then captured by antigen-presenting cells, which in turn may further promote the autoimmune responses. These biological processes eventually lead to glandular inflammation (Tzioufas et al., 2012). Figure 1.1 illustrates a possible model of pSS pathogenesis.

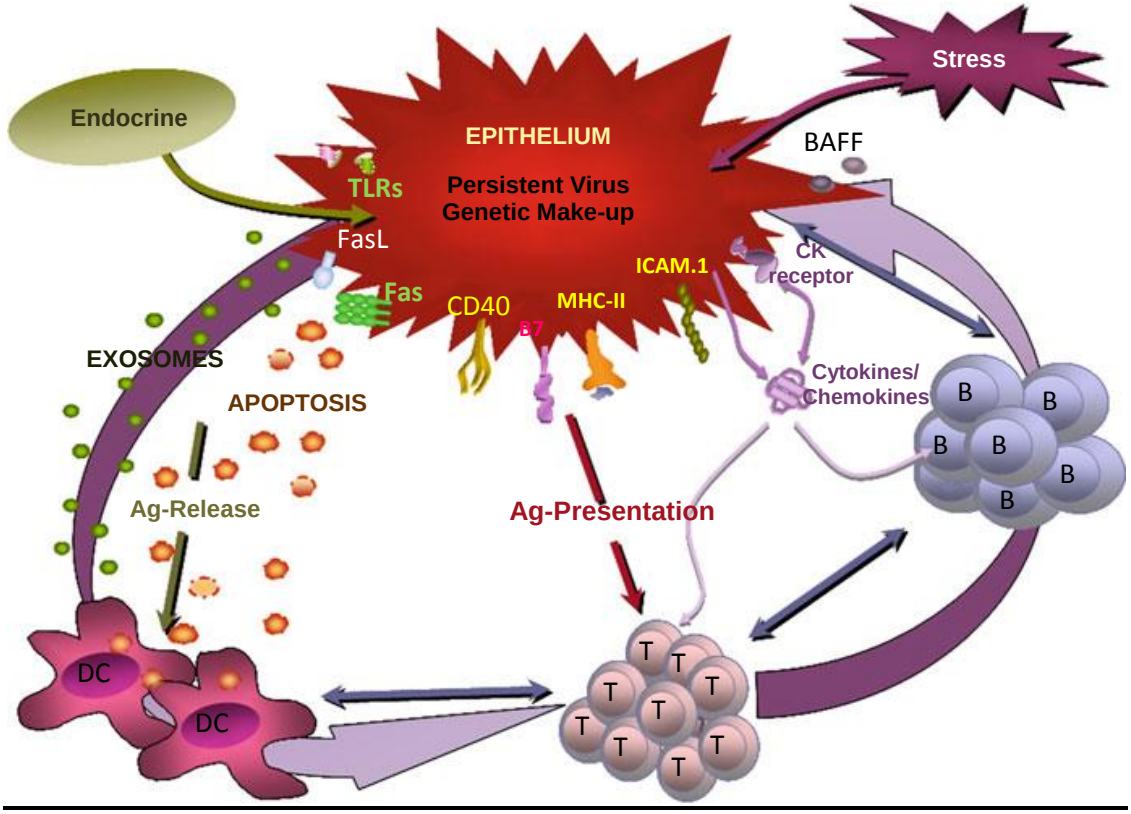


Figure 1.1 A hypothetical model for pSS pathogenesis (Ramos-Casals et al., 2012)

1.4 Biomarkers in pSS:

A biomarker can be defined as a substance or a chemical that indicates a certain biological state. Biomarkers are useful for the assessment of normal vs. pathological states. Definitions of biomarkers in the literature share the same concepts. In 1998, the National Institute of Health Biomarkers Definition Working Group, defined a biomarker as 'a characteristic that can be objectively measured and evaluated as an indicator of normal

biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention'. The World Health Organization (WHO), the United Nations and the International Labour Organization define biomarkers as 'any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease' (reviewed in (Strimbu and Tavel, 2010).

The detection of biomarkers in pSS is useful for several reasons. First, the diagnosis of pSS involves an invasive procedure (biopsy) that may lead to diagnosis delay. Since pSS affects the salivary glands, identification of pSS biomarkers in the saliva is a promising approach. Accordingly, Hu and colleagues (Hu et al., 2007) identified protein and genomic biomarkers in the saliva of pSS patients. In their study, they have discovered 16 protein biomarkers in the whole saliva of pSS patients: 10 were upregulated and 6 were downregulated in pSS in comparison to healthy controls. In addition, they investigated the presence of genomic biomarkers in saliva. They reported the detection of the expression of 27 genes including *GIP2*, which is an IFN-α inducible gene, could differentiate pSS cases and controls.

Recently, a study reported profilins and carbonic anhydrase I (CA-I) as biomarkers in oral fluids in pSS, after removing high-abundance proteins in the samples (Deutsch et al., 2015, Krief et al., 2011). Profilins, which are actin–binding proteins, play a role in organizing microfilaments and cell motility, and regulate the dynamics of the actin polymerization as well as the development of embryos. Profilins contribute to many biological processes such as membrane trafficking and nuclear activity, as well as tumour formation (Witke, 2004, Rawe et al., 2006). Additionally, profilins have also been reported as a biomarker for oral cancer (Hu et al., 2008). CA-I is a metalloenzyme that is involved in many biological processes; its main role is to catalyse the hydration of carbon dioxide (Supuran, 2008). In addition, CA-I was also reported to predict the presence of oral squamous cell carcinoma (Liu et al., 2012). Another recent proteomic study revealed that salivary IL-4, IL-5 and clusterin were biomarkers of pSS (Delaleu et al., 2015).

Along with the salivary glands, pSS also affects the lacrimal glands. Therefore, it has been proposed that tear fluid may be another source for pSS biomarkers. Tear cathepsin S was recently found to be increased significantly in pSS patients. Although cathepsin S activity was elevated in tear samples, it did not correlate with anti-Ro/SSA or anti-La/SSB levels in

pSS patients (Hamm-Alvarez et al., 2014). Furthermore, PAX6, a corneal lineage commitment regulator, is down-regulated in pSS which is associated with ocular surface damage in pSS patients depending on the levels of inflammation. Thus, PAX6 has been suggested to be a biomarker for ocular damage in pSS (McNamara et al., 2014). Several other biomarkers have also been reported in pSS. Many studies have shown that IFN type I activation is a common feature in pSS (Bave et al., 2005, Brkic et al., 2013, Brkic and Versnel, 2014). Moreover, myxovirus resistance protein A (MxA), which is a marker of type I IFN activity, correlates with disease activity and is reduced by IFN inhibitors such as hydroxychloroquine (Maria et al., 2014).

Regarding genomic biomarkers, the contribution of epigenetics was investigated in pSS including DNA methylation, histone modification and microRNA (miRNA) (Konsta et al., 2014). Several studies have been performed in order to identify miRNA biomarkers in pSS. For instance, miRNA-768-3p and miRNA-574 were found to be associated with salivary gland dysfunction (Alevizos et al., 2011). In addition, miRNA-146a was significantly elevated in pSS patients peripheral blood mononuclear cells (PBMCs) as well as in PBMCs, the salivary and lacrimal glands of SS-prone mice and WT C57BL/6J mice (Pauley et al., 2011). Another miRNA, miR-5100, was reported to be a biomarker for salivary gland function in pSS (Tandon et al., 2012).

1.5 The diagnosis of primary Sjögren's Syndrome:

In 1993, the preliminary criteria for the classification of Sjögren's Syndrome were proposed (Vitali et al., 1993). Later on in 2002, the American-European Consensus Group (AECG) revised these to produce a more reliable set of criteria that enable a more accurate diagnosis for the syndrome. The AECG classification criteria include 6 criteria—ocular symptoms, oral symptoms, ocular signs, histopathology of minor salivary gland biopsies, salivary gland involvement and autoantibodies (Table 1.1) (Vitali et al., 2002). To fulfil the AECG criteria, a subject has to fulfil a minimum of 4 out of 6 criteria, of which must include either positive histopathology or autoantibodies or both. A subject fulfilling 3 out of the 4 objective criteria can also be classified as having pSS.

A new set of classification criteria were proposed recently by the American College of Rheumatology (ACR), led by investigators of the Sjögren's International Collaborative

Clinical Alliance (SICCA) (Shiboski et al., 2012). The classification criteria proposed by SICCA are illustrated in Table 1.2. The ACR classification requires the fulfilment of at least 2 out of 3 criteria.

The ACR classification criteria include the addition of IgG4-related disease as an exclusion criterion, as IgG4-related disease mimics pSS in some features (Masaki et al., 2009). The exclusion criteria of the ACR classification criteria include: History of head and neck radiation treatment, Hepatitis C infection, AIDS, Sarcoidosis, Amyloidosis, Graft versus host disease and IgG4-related disease (reviewed in (Fazaa et al., 2014). An ACR-EULAR working group is currently developing a revised set of ACR-EULAR criteria, with the aim of unifying the AECG and ACR criteria.

Table 1.1 AECG classification criteria for primary Sjögren's Syndrome(Vitali et al., 2002)

AECG classification criteria for Sjögren's syndrome
I. Ocular symptoms: a positive response to at least one of the following questions: 1. Have you had daily, persistent, troublesome dry eyes for more than 3 months? 2. Do you have a recurrent sensation of sand or gravel in the eyes? 3. Do you use tear substitutes more than 3 times a day?
II. Oral symptoms: a positive response to at least one of the following questions: 1. Have you had a daily feeling of dry mouth for more than 3 months? 2. Have you had recurrently or persistently swollen salivary glands as an adult? 3. Do you frequently drink liquids to aid in swallowing dry food?
III. Ocular signs—that is, objective evidence of ocular involvement defined as a positive result for at least one of the following two tests: 1. Schirmer's I test, performed without anaesthesia (<5 mm in 5 minutes) 2. Rose bengal score or other ocular dye score (>4 according to van Bijsterveld's scoring system)
IV. Histopathology: In minor salivary glands (obtained through normal-appearing mucosa) focal lymphocytic sialoadenitis, evaluated by an expert histopathologist, with a focus score >1, defined as a number of lymphocytic foci (which are adjacent to normal-appearing mucous acini and contain more than 50 lymphocytes) per 4 mm^2 of glandular tissue
V. Salivary gland involvement: objective evidence of salivary gland involvement defined by a positive result for at least one of the following diagnostic tests: 1. Unstimulated whole salivary flow (<1.5 ml in 15 minutes) 2. Parotid sialography showing the presence of diffuse sialectasias (punctate, cavitary or destructive pattern), without evidence of obstruction in the major ducts 3. Salivary scintigraphy showing delayed uptake, reduced concentration and/or delayed excretion of tracer
VI. Autoantibodies: presence in the serum of the following autoantibodies: 1. Antibodies to Ro(SSA) or La(SSB) antigens, or both

Table 1.2 Classification criteria proposed by the Sjögren International Collaborative Clinical Alliance Group (SICCA). *A patient can be classified as having pSS when there are at least 2 out of the 3 of the following criteria (Shiboski et al., 2012)*

Classification criteria
• Positive serum anti-SSA/Ro and/or anti-SSB/La or (positive rheumatoid factor and antinuclear antibodies with titre ≥1:320) • Labial salivary gland biopsy exhibiting focal lymphocytic sialadenitis with a focus score ≥1 focus/4 mm^2 • Keratoconjunctivitis sicca with ocular staining score ≥ 3 (assuming that individual is not currently using daily eye drops for glaucoma and has not had corneal surgery or cosmetic eyelid surgery in the last 5 years)

1.5.1 The UK primary Sjögren's Syndrome Registry (UKPSSR):

The UK primary Sjögren's Syndrome Registry (UKPSSR) was established in 2009. To date, it consists of a cohort of over 800 pSS patients recruited from 35 centres from the UK. All patients fulfil the AECG classification criteria and the clinical data that have been collected in the registry are provided in Table 1.3. All information is stored in a secured database in an anonymised manner. The samples that have been collected include peripheral blood mononuclear cells, whole blood RNA, DNA and serum samples. In addition, age- and sex-matched healthy controls are also being recruited, with over 350 recruited to date (Ng et al., 2011).

Table 1.3 The clinical and outcome measure data collected for the UKPSSR (Ng et al., 2011)

Clinician's assessment	**Patient-reported outcome**
AECG classification criteria	Symptom assessment
Demographics	PROFAD-SSI
Treatment (pharmacological and non-pharmacological)	ESSPRI
	Epworth Sleepiness Scale
Comorbidity	Orthostatic symptoms scale
Disease activity	Quality of life
ESSDAI	EQ-5D
SCAI	SF-36
SSDAI	Anxiety and depressive symptoms
Disease damage	HADS
SDI	Optional
SSDDI	Autonomic symptoms (COMPASS)
Optional	Cardiovascular risk
Cardiovascular risk assessment	Lifestyle (smoking, physical activity)

1.6 The treatment of primary Sjögren's Syndrome:

Currently there is no effective treatment for the syndrome and that in part is because the exact aetiology that triggers the disease is unknown. It is important for patients to be aware of the symptoms and the consequences so that general measures can be taken by the patients, such as strict maintenance of dental hygiene, avoidance of smoking and dry environments and the application of eye ointments and petroleum jelly on the lips (Venables, 2004). In addition, fluid replacement is often used for alleviating the sicca symptoms (Talal, 1991). Oral pilocarpine may be helpful in improving symptoms but not in objective measures such as Schirmer's test (Tsifetaki et al., 2003). Pilocarpine as well as cevimeline are being used for xerostomia symptoms in pSS. Both medications stimulate the M1 and M3 receptors in the salivary glands and therefore enhance the secretary function (Fox et al., 2001, Fife et al., 2002). Other medications may be used for both glandular and extra-glandular manifestations. These medications include corticosteroids and other

immunomodulatory drugs such as: hydroxychloroquine, which may be effective in reducing salivary gland inflammation (Tishler et al., 1999); methotrexate (Winzer and Aringer, 2010); and cyclophosphamide (Mavragani and Kassan, 2012). Other immunomodulatory drugs such as leflunomide, interferon-α, mizoribine, mycophenolic acid, rebamipide, cladribine and fingolimod have also been tried in pSS treatment (Carsons, 2012) but none of these medications are commonly used in clinic. Finally, the use of biological therapies in pSS is gaining interest among clinicians, researchers and the pharmaceutical industry. For instance, rituximab (RTX), an antibody against CD20 expressed on most B-cell surface and B-cell lymphoma cells, is being investigated as a treatment for pSS with or without lymphoproliferative disorders (Somer et al., 2003, Quartuccio et al., 2009). Seror and colleagues reported that RTX treatment showed a good efficacy in 4 out of 5 pSS-associated lymphoma patients and 9 out 11 pSS patients with systemic features. Moreover, RTX treatment increased the level of BAFF along with B cell biomarkers level's decline (Seror et al., 2007). The efficacy of RTX treatment in pSS patients was evaluated in a placebo-controlled randomised controlled trial (RCT), which consist of 120 patients with recent disease onset. In addition, all patients had visual analogue scales (VAS) > 50 mm in at least 2 out of 4 of global disease, pain, fatigue and dryness. Primary endpoint was measured at 24 weeks. Although RTX treatment did not show significant improvement in disease activity or symptoms over placebo at 24 week, significant improvement of some symptoms were noted in the RTX group at earlier time points (6 weeks) of the trial especially with the fatigue VAS score which was decreased by 30 mm (Devauchelle-Pensec et al., 2014).

1.7 Gene expression and primary Sjögren's Syndrome:

Gene expression studies have been investigated in pSS using different kinds of samples including: exocrine gland biopsies, PBMCs and whole blood. Each of these studies will be described in more detail below.

1.7.1 Gene expression studies in the salivary glands:

The use of microarray profiling to investigate global gene expression in the minor salivary glands has been documented in several studies, implicating several chemokines, cytokines, the MHC molecules, the Bcl-like gene and the type I interferon (IFN) pathways in pSS. These data suggest that both innate and acquired immune system contribute to the pSS

pathology. In one study, *CXCL13* and CD3D were overexpressed in more than 90% of pSS patients. Furthermore, lymphotoxin β, several MHC genes, cytokines and lymphocyte activation factors were also overexpressed. Type I IFNs, which play an important role in the protection against viral infections, were among the top 200 genes that were overexpressed in the minor salivary glands of pSS patients (Hjelmervik et al., 2005). In another study, 23 genes of the IFN pathway, including Toll like receptor 8 and 9 (TLR8 and TLR9), were found to be differentially expressed in salivary gland epithelial cells of pSS patients. Moreover, stimulation of salivary gland epithelial cells (SGECs) in vitro with IFN induced the expression of these genes (Gottenberg et al., 2006). Conversely, some genes were reported to be down-regulated such as carbonic anhydrase II and bcl-2 genes (Hjelmervik et al., 2005).

1.7.2 Gene expression studies in lacrimal glands:

The lacrimal gland is another organ affected in pSS. The main clinical manifestation is the reduction of tear secretion, which in turn results in chronic irritation of the eyes and keratoconjuctivitis sicca (KCS) (Rose and Mackay, 1998). Studying lacrimal glands faces many challenges. First, it is difficult to obtain a biopsy sample from the glands. Second, the lacrimal gland biopsy itself might result in more complications for the patient. As a result, few studies have investigated gene expression in human lacrimal glands. One study has shown that *Fas*, *FasL* and *BAX* genes, involved in promoting lacrimal glands epithelial cells apoptosis (pro-apoptotic genes), were up-regulated in pSS lacrimal glands; at the same time, the anti-apoptotic gene *bcl-2* were down-regulated in the same patients (Wu et al., 2000). Furthermore, in another study increased expression levels of *IFITM1* and *BAFF* (2.5-fold and 3-fold, respectively) were reported in the conjunctival cells of pSS patients (Gottenberg et al., 2006).

The majority of lacrimal gland studies were carried out on animal models, with easier access to lacrimal gland sampling. Data from these gene expression studies implicated abnormality in inflammatory responses, secretory function of the glands and the gland's structure organization in the disease process. In an animal model using C57BL/6.NOD-*Aec1Aec2* mice, Nguyen and colleagues performed lacrimal glands gene expression profiling in the early phase of the disease. Their experiment collected data at intervals ranging from 4–20 weeks and showed that 552 genes were differentially expressed. These

included *ApoE, Baff, Clu, Ctla4, Fas/Fasl, Irf5, Lyzs, Nfkb, Socs3, Stat4, Tap2, Tgfb1, Tnfa,* and *Vcam1*. Most of the differentially expressed genes were related to apoptosis and fatty acid homeostasis (Nguyen et al., 2009). Additionally, changes in gene expression were observed in the genes encoding the inter-epithelial junction proteins and the focal adhesion maturation; these changes led to an increase of infiltration of leukocytes into the lacrimal glands in the early stages of dacryoadenitis in C57BL/6.NOD-*Aec1Aec2* mice (Peck et al., 2011).

1.7.3 Gene expression studies in peripheral blood:

Peripheral blood consists of various types of cells which might provide a source of information about the ongoing biological events in pSS. Few studies have investigated peripheral blood gene expression profiling in pSS and pSS-associated lymphoma. Similar to the findings from gene expression in the salivary glands, activated IFN pathways were documented in these studies (Wildenberg et al., 2008, Emamian et al., 2009, Ogawa, 2010). Furthermore, a marked overexpression of the IFN-inducible gene, IFN-α–inducible protein 27 (*IFI27*), was reported in one study (Kimoto et al., 2011). In another study, the overexpression of the genes that control IFN-α was demonstrated in the blood samples from pSS patients, in parallel with the overexpression of these genes in the labial salivary glands of these patients (Zheng et al., 2009). A very recent study has shown that miR-155 and the suppressor gene of cytokine signalling 1 (SOCS1) were overexpressed in the PBMCs of pSS patients (Chen et al., 2015a). The published peripheral blood gene expression data is summarised in more details in the introductions of Chapters 3 and 4.

1.8 Primary Sjögren's syndrome -associated lymphoma:

The lymphomas associated with pSS not only develop in the affected organs such as the salivary glands, but might also develop in other extra-nodal sites such as the stomach, lungs and skin. The commonest type of lymphoma was the extranodal marginal-zone B-cell lymphoma of the mucosa-associated lymphoid tissue (MALT) (Voulgarelis et al., 1999). Furthermore, different types of non-Hodgkin's lymphoma have been reported in pSS patients in addition to MALT lymphoma including low-grade marginal-zone lymphoma (MZL), and high-grade B cell-lymphomas such as diffuse large B cell lymphoma (DLBCL) (Royer et al., 1997, Kim et al., 2012). Although the first association of lymphoma with pSS was described in 1963, the increased risk of lymphoma development in pSS was not

reported until 1978. Lymphoma was found in 4.3% of pSS patients in a study carried out in 1999 (Voulgarelis et al, 1999). The relative risk (RR) of lymphoma in pSS, especially non-Hodgkin's lymphoma (NHL) has been investigated in several studies and in different pSS cohorts. The first published study estimated up to a 44 times higher risk in pSS patients (Kassan et al., 1978). In two other studies, the RR was 15–20 while another study reported an odds ratio of 6.1 (95% CI: 1.4 to 27) (Zintzaras et al., 2005, Theander et al., 2006, Smedby et al., 2006). Recently, another study evaluated the risk of lymphoma development in pSS in a Norwegian cohort, reporting a 9-time increase in lymphoma development in pSS patients than the general population (Johnsen et al., 2013). Recently, a meta-analysis of the association of NHL with autoimmune diseases showed a smaller standardized incidence ratio (SIR) of 4.9 in pSS patients (Fallah et al., 2014). The SIR of lymphoma was estimated in different cohorts since 1978 and a summary of these studies is demonstrated in Table 1.4 as reviewed in (Nishishinya et al., 2015).

Table 1.4 Standardized incidence ratio of lymphoma in different pSS cohorts (Nishishinya et al., 2015)

Study	Name of the study cohorts	Number of lymphomas	SIR (95% CI)
(Kassan et al., 1978)	Connecticut cancer register	4	44.40 (16.70–118.40)
(Kauppi et al., 1997)	Finnish cancer registry	11	8.70 (4.30–15.50)
(Valesini et al., 1997)	Local cancer registers	9	33.30 (17.30–64.00)
(Davidson et al., 1999)	Cancer registry statistic	3	14.40 (4.70–44.70)
(Pertovaara et al., 2001)	Finnish cancer registry	3	13.00 (2.70–38.00)
(Theander et al., 2006)	National and local registers	11	15.57 (7.80–27.90)
(Lazarus et al., 2006)	Thames cancer registry	11	37.50 (20.70–67.60)
(Zhang et al., 2010)	Local cancer registers	8	48.10 (20.70–94.80)
(Solans-Laque et al., 2011)	GLOBOCAN database	11	15.60 (8.70–28.20)
(Weng et al., 2012)	Nationwide population cohort	23	7.10 (4.25–10.30)
(Johnsen et al., 2013)	Cancer registry of Norway	7	9.00 (7.10–26.30)
(Fallah et al., 2014)	Nationwide cohort (meta-analysis)	143	4.90 (4.20–5.80)

SIR—standardized incidence ratio, CI—confidence interval

1.8.1 Risk factors in pSS-associated lymphoma

Several studies have examined risk factors for NHL development in pSS. These risk factors include: recurrent or constant swelling of salivary glands, lymphadenopathy, cryoglobulinaemia, splenomegaly, and lymphopenia; other factors include low complement factor C4 and C3, skin vasculitis or palpable purpura, monoclonal components in serum or urine, peripheral neuropathy, glomerulonephritis, elevated β_2-microglobulin, $CD4^+$ T-lymphocytopenia, germinal center-like structures in minor salivary gland biopsies, genetic factors, and down-regulation of A20 (Jonsson et al., 2012, Ioannidis et al., 2002).

Lymphadenopathy, which is an abnormal enlargement of the lymph node, is reported to be associated with pSS (Chen et al., 2013). Cryoglobulins, are serum immunoglobulins that precipitate at low temperatures. Three types of cryoglobulins have been described: type I (simple cryoglobulins) are monoclonal immunoglobulins which are mostly IgM. Type I cryoglobulins often accompany haematological disorders and monoclonal gammopathy. Type II and III are "mixed" cryoglobulins. Type II cryoglobulins consist of monoclonal IgM with rheumatoid factor (RF) activity and polyclonal IgG. Type III cryoglobulins also consist of polyclonal IgG, IgM and/or IgA, with one of them having RF activity. Type II and III cryoglobulins are associated mostly with systematic and infectious diseases. Mixed monoclonal cryoglobulinaemia as well as the monoclonal rheumatoid factor (mRF)–associated cross-reactive idiotype (CRI) have been reported as predictive factors in pSS-associated lymphoma (Tzioufas et al., 1996). Consistently, cryoglobulinaemia is associated with pSS-associated lymphoma (in particularly low-grade lymphoma) (Charitaki et al., 2011, Anand et al., 2015). Recently, Quartuccio and colleagues reported the association of lymphoma with cryoglobulinaemic vasculitis (CV), but not with cutaneous vasculitis or hypergammaglobulinaemic vasculitis (HGV) in pSS patients (Quartuccio et al., 2015). Splenomegaly has been reported as a risk factor for lymphoid malignancies (Hall and Kahl, 2015). The association of splenomegaly, cryoglobulinaemia and lymphopenia as lymphoma predictors in pSS has been noted in (Baimpa et al., 2009). Recently, RF was reported along with disease activity to be predictors of lymphoma development in pSS patients. These data was estimated in 101 pSS-associated lymphoma patients (Nocturne et al., 2015c).

1.8.2 Gene expression, biomarkers and predictors in pSS-associated lymphoma:

Several studies have examined gene expression in pSS-associated lymphoma. The main limitation with these studies is their small sample sizes. For instance, Hu and colleagues, comparing the gene expression in the parotid glands with pSS-associated lymphoma and those without, reported eight candidate genes that discriminated pSS from pSS-associated lymphoma (Hu et al., 2009). These genes, listed in order of the highest level of statistical significant (lowest p-value), include *GRB2*, *ARHGDIB*, *CD40*, *PSMB9*, *ALDOA*, *PRDXS*, *PARC* and *PPIA*. The key limitation of this study was the small sample size, with only 8 pSS and 9 pSS-associated lymphoma samples.

Several clinical and serological parameters have been proposed as predictors for lymphoma development in pSS patients. One of the interesting biomarkers in pSS-associated lymphoma is Fms-like tyrosine kinase 3 ligand (Flt-3L). The serum level of Flt-3L was elevated in treated pSS-associated lymphoma (history of lymphoma) as well as at time of lymphoma diagnosis. Moreover, high levels of Flt-3L were associated with the presence of risk factors of lymphoma including low levels of C4 and IgM, high levels of β_2-microglobulin and pSS disease activity, lymphopenia and the presence of purpura (Seror et al., 2010, Tobon et al., 2013). Another biomarker that has been reported in pSS-associated lymphoma is the B-cell homeostatic chemokine CXCL13. This chemokine was reported to be elevated in both the serum and the salivary gland of pSS patients (Kramer et al., 2013, Chen et al., 2015b). Furthermore, CXCL13 is known for its association with ectopic reactive lymphoid tissue in the salivary gland of MALT lymphoma in pSS (Barone et al., 2008). In addition, the serum level of CXCL13 was reported to be significantly high in pSS-associated lymphoma patients in compare with pSS patients without lymphoma (Nocturne et al., 2015a)

The germinal center (GC)–like structure in the labial salivary glands is described as a potential predictive biomarker of lymphoma in pSS: in one study the GC-like structure occurred in 25% of pSS patients (43 patients out of 175); within this subgroup 6 patients who subsequently developed lymphoma were GC+ at the time of diagnosis. In contrast, among those without GC on biopsy at the time of diagnosis (n = 132), only one pSS patient subsequently developed lymphoma (Theander et al., 2011). Recently, the International Prognostic Index (IPI) score, which is a model to predict the outcome of patients with aggressive NHL based on their clinical characteristics prior to treatment, as well as the

EULAR SS disease activity index (ESSDAI) were identified as predictive indicators of pSS-associated lymphoma development (Papageorgiou et al., 2015b).

More recently, the role of TNF alpha induced protein 3 (*TNFAIP3)* in pSS-associated lymphoma development was reported in a French cohort. Germline and somatic genetic variations of *TNFAIP3* were more common in pSS-associated lymphoma compared with pSS patients without lymphoma (Nocturne et al., 2013). The *TNFAIP3* gene, also known as A20, is found on chromosome 6, and encodes an ubiquitin-editing enzyme that regulates the activation of the nuclear factor κB (NF-κB), tumour necrosis factor receptor 1, CD40, toll-like receptors and IL-1 receptor (reviewed in (Ma and Malynn, 2012)). Furthermore, *TNFAIP3* polymorphism (TT>A dinucleotide) is associated with other autoimmune diseases such as SLE and SSc (Adrianto et al., 2011, Koumakis et al., 2012). Data from animal models indicate that A20 deficiency in mice causes a severe inflammatory reaction leading to death, which was thought to be a consequence of failure in A20-mediated termination of TNF-induced NF-κB responses (Lee et al., 2000). Very recently, the germline polymorphism rs2230926 of *TNFAIP3* was assessed in an UK cohort and a French cohort, which showed that the rs2230926G variant was correlated to pSS-associated lymphoma in the UK cohort. Additional confirmation was also obtained by performing meta-analysis of the data from both cohorts (Nocturne et al., 2015b).

There is data suggesting that BAFF might play a role in the formation of ectopic germinal center leading to lymphomagenesis. Gottenberg and colleagues reported that high levels of BAFF and β2-microglobulin in patients who currently have lymphoma or have a history of lymphoma compared with patients without lymphoma (Gottenberg et al., 2013). Moreover, the serum level of BAFF was associated with high systemic disease activity, lymphoproliferative disorders and B-cell clonal expansion in pSS (Seror et al., 2010, Quartuccio et al., 2013). Another mutation in the BAFF receptor (BAFF-R His159Tyr mutation) was recently linked to pSS-associated lymphoma. In the literature, the correlation of BAFF, lymphoproliferative disorders and autoimmune diseases has been documented. In addition, the BAFF-R His159Tyr mutation is also documented in NHL (Ferrer et al., 2009, Hildebrand et al., 2010). Therefore, the assessment of BAFF-R His159Tyr mutation in pSS-associated lymphoma was compelling. Accordingly, this mutation occurs in 8.6% of pSS-associated lymphoma patients compared to 6.2% in pSS patients without lymphoma, and 1.7% in healthy controls. These differences between groups were statistically

significant. Moreover, the prevalence of this mutation was increased in a group of pSS-associated lymphoma patients with disease onset below the age of 40 years (Papageorgiou et al., 2015a). Recently, polymorphisms of BAFF were reported to play a role in the development of lymphoma in pSS, as distinct BAFF gene haplotypes related to the increased risk and development of lymphoma in pSS (Nezos et al., 2014).

Another mutation—the somatic MYD88 Leu265Pro mutation, recently described in Waldenström's macroglobulinemia (WM)—was tested in the peripheral blood of pSS patients and minor salivary glands biopsies of pSS and pSS-associated lymphoma patients. This mutation activates the IRAK-mediated NF-κB signaling. However, pSS and pSS-associated lymphoma patients did not harbor this mutation, suggesting other mechanisms that might be implicated in pSS-associated lymphoma (Voulgarelis et al., 2014).

1.8.3 Mechanisms of lymphoma development in pSS

The mechanism underlying lymphoma development in pSS is still not fully understood. Nonetheless, studies of lymphomagenesis in pSS have provided several key observations and highlighted areas that need further investigation. To date, it is generally believed that the degree of disease severity along with B-cell activation and inflammation leads to pSS-associated lymphoma development (Theander et al., 2006).

Oncogenes may also play a role in lymphomagenesis in pSS. For instance, anti-p53 antibodies were found in pSS-associated lymphoma patients, raising the possibility that the appearance of anti-p53 antibody may be an indicator of malignant transformation. Interestingly, while total inactivation of p53 leads to the progression of high-grade lymphoma, partial inactivation of p53 leads to low-grade lymphoma (Mariette et al., 1999, Du et al., 1995).

The generation of immunoglobulins (Ig) starts at the early stages of B-cell development. The production and maturation of Ig involve breaking and reconnecting DNA (recombination, somatic mutation and isotype switching). These DNA ‘editing’ processes during Ig production increase the risk of translocation of oncogenes. One such sample is the translocation of the *Bcl-2* and *c-myc* oncogenes into the Ig loci in chromosome 14q32 (Sugai et al., 1994). This translocation event leads to the formation of mutagenic B-cells

with a V(H) mutation, which was found to occur in high frequency in the parotid glands in pSS (Zuckerman et al., 2010).

B-cell activation factor (BAFF), which is thought to play an important role in the pathogenesis of pSS, may play an important role in the development of lymphoma in pSS patients. Thus, it has been reported that lymphoma develops more commonly in pSS patients with high levels of BAFF (Groom et al., 2002). Data suggest that germinal centres in the salivary glands of pSS patients can lead to the expansion of oligoclonal B-cell population, which in turn may lead to the development of MALT lymphoma (Voulgarelis and Moutsopoulos, 2001).

Taken together, the existing body of evidence suggests that the interactions between epithelial cells, T cells and B cells in the salivary glands may provide a platform for lymphoma development. Lymphomagenesis is likely to be a multi-step process which evolves from polyclonal lymphoproliferation to monoclonal lymphoproliferation to MALT lymphoma and eventually to high-grade malignant lymphoma. Lymphoma development in pSS may also involve antigen-driven B-cell activation and oncogenic events such as p53 inactivation and bcl-2 activation (Masaki and Sugai, 2004). Figure 1.2 illustrates a hypothetical model of lymphoma development in pSS.

More recently, a new proposed mechanism for the pathophysiology of pSS-associated lymphoma was published by Nocturne and Mariette (Nocturne and Mariette, 2015). This new mechanism suggests that the chronic stimulation of polyclonal RF^+ B-cells (autoreactive B-cells that express a B-cell antigen receptor (BCR) with CDR3 that is strongly homologous to RF) in the salivary gland by auto-antigens (potentially Ro and La or other auto-antigens) might be essential in lymphomagenesis. RF^+ B-cell survival is promoted by BAFF and other cytokines production during pSS. Furthermore, the involvement of *TNFAIP3* in pSS-associated lymphoma might play a role, as mutation and deletion in this gene, which controls NF-κB activation, support RF^+ B-cell survival. Interestingly, genetic mutations of *TNFAIP3* or other NF-κB–controlling genes might affect these B-cells, resulting in their escape as lymphoma B cells. In addition, the formation of GC-like structures might increase the stimulation of autoreactive B-cells that might have an oncogenic mutation featured by NF-κB signalling dysregulation. The failure of elimination of these auto-reactive B-cells within the GC-like structure may also leads to prolonged

activation-induced cytidine deaminase (AICDA or AID) activity (which has a role in RNA-editing and it is the enzyme that mediates somatic hypermutations of the immunoglobulin variable region heavy chain (IgV_H)), increasing the risk of oncogenic somatic mutations that give these B-cells a feature akin to low-grade B-cell lymphoma.

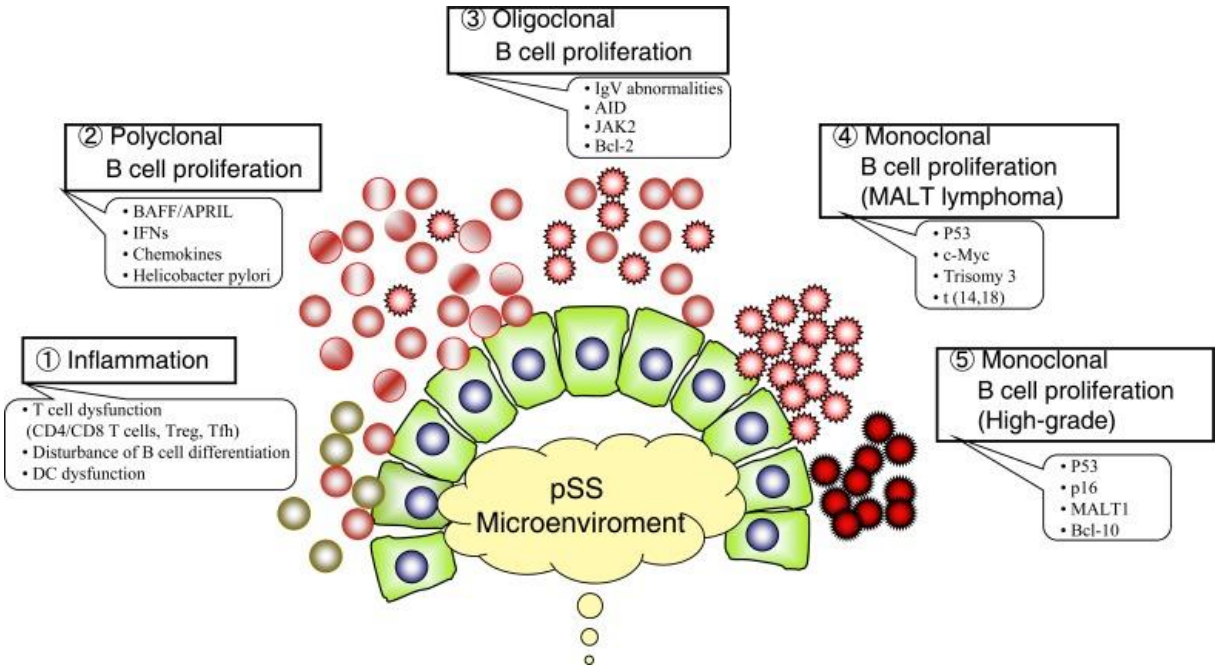


Figure 1.2 Hypothetical model of pSS-associated lymphoma *Chronic stimulation of antigen-specific B cells proliferation by external antigens or autoantigens plays a significant role in the multi-step process of developing lymphoma in pSS. Step 1—Inflammation: $CD4^+$ T cells, memory B cells and dendritic cells infiltration in the minor salivary glands perpetuate chronic inflammation. Step 2—Polyclonal B cell proliferation: Increased production of BAFF and IFNs in pSS patients cause the proliferation of polyclonal B cell and thereby contribute to the characteristic feature of myoepithelial sialadenitis (MESA) or the benign lymphoepithelial lesion. Step 3—Oligoclonal B cell proliferation: BAFF specifically play a role in the regulation and survival of B lymphocyte proliferation, altered the differentiation of B cells. Chronic stimulation by external antigens or autoantigens may drive the proliferation of antigen-specific B cells through the immunoglobulin heavy chain (IgVH-CDR3) restricted usage and increase their transformation frequency. Step 4—Monoclonal B cell proliferation: During the development of B cell, immunoglobulins undergo multiple processes including recombination, somatic mutation and isotype switching. Such events may increase the risk of the translocation of oncogenes (e.g. Bcl-2 and c-Myc) to the immunoglobulin loci (chromosome 14q32). Step 5—Transformation to high grade malignancy: the progression from low-grade MALT lymphoma to high-grade lymphoma may be facilitated by the P53 tumour-suppressor activity defect, amplification of bcl-2 and/or c-Myc, high frequency of t(14,18) translocation and trisomy 3 (Dong et al., 2013).*

1.8.4 Treatment of pSS-associated lymphoma:

There are various therapeutic agents available for the treatment of pSS-associated lymphoma. Ideally, the treatment should target the autoimmunity and the malignancy at the same time. Anti-CD20 monoclonal antibody (e.g., rituximab (RTX)) has been used for the treatment of pSS-associated lymphoma. Rituximab depletes B cells and inhibits B-cell activation but its precise mechanism of action is still unclear (Abdulahad et al., 2012). Another treatment is belimumab, a monoclonal antibody targeting BAFF. Recently, it has been shown that belimumab results in the normalisation of the frequency of peripheral blood B-cells in pSS (Pontarini et al., 2015). Moreover, treatment of pSS with belimumab led to improvement in ESSDAI and ESSPRI scores as well as reduction of B-cell biomarkers after 28 weeks (De Vita et al., 2015). Despite these promising agents, more research is needed to identify the most effective therapeutic strategies for the management of pSS-associated lymphoma.

1.9 Overall design of the project:

Most studies have focused on serum protein biomarkers. However, genetic biomarkers for pSS-associated lymphoma have not been reported. The primary aim of the project is to identify a whole blood gene expression signature in pSS-associated lymphoma that would be useful for diagnosis. In order to accomplish this aim, I have used the following approach.

First, because of the predominance of globin RNA in peripheral blood, I needed to investigate and optimise the effects of globin mRNA depletion on gene expression profiling in pSS. The reason behind performing the optimisation experiment in pSS but not in pSS-associated lymphoma is the availability of gene expression data in pSS in the literature that I can use as a reference. Second, I identified a whole-blood gene expression signature of pSS-associated lymphoma using a whole genome microarray on a set of samples that I refer to as the ‘Discovery cohort.’ The Discovery cohort consisted of five subject groups—four pSS subgroups and a healthy controls group. The four pSS groups were: pSS (without lymphoma), pSS-associated lymphoma, pSS with paraproteinemia and pSS with other cancers. In this experiment, the microarray data were analysed to identify candidate genes in pSS-associated lymphoma by comparing the pSS (without lymphoma) and the pSS-associated lymphoma groups. The differentially expressed genes were then technically

validated using qRT-PCR. A machine learning modelling was performed to predict which genes can serve as candidates to predict lymphoma in pSS. Third, an independent set of samples (the 'Validation cohort') was used to test whether the potential gene expression signature (i.e., the differentially expressed genes that were detected in the microarray and the qRT-PCR data), was also present in the Validation cohort. In this experiment, I also included a set of untreated pSS-associated lymphoma samples to test whether the potential gene expression signature identified in the Discovery cohort was also present in untreated lymphoma samples. Gene-expression levels were measured in the Validation cohort and the untreated pSS-associated lymphoma group using qRT-PCR. In addition, I performed biological pathway analysis in the Discovery cohort to explore the most important biological pathways in pSS-associated lymphoma using the Ingenuity Pathway Analysis (IPA) platform. Finally, I analysed gene expression microarray data and pathway analysis for the other pSS subgroups to ensure that the gene expression signatures and the pathway analysis data were unique to the pSS-associated lymphoma group. Diagram 1.3 illustrates the overall design of the project.

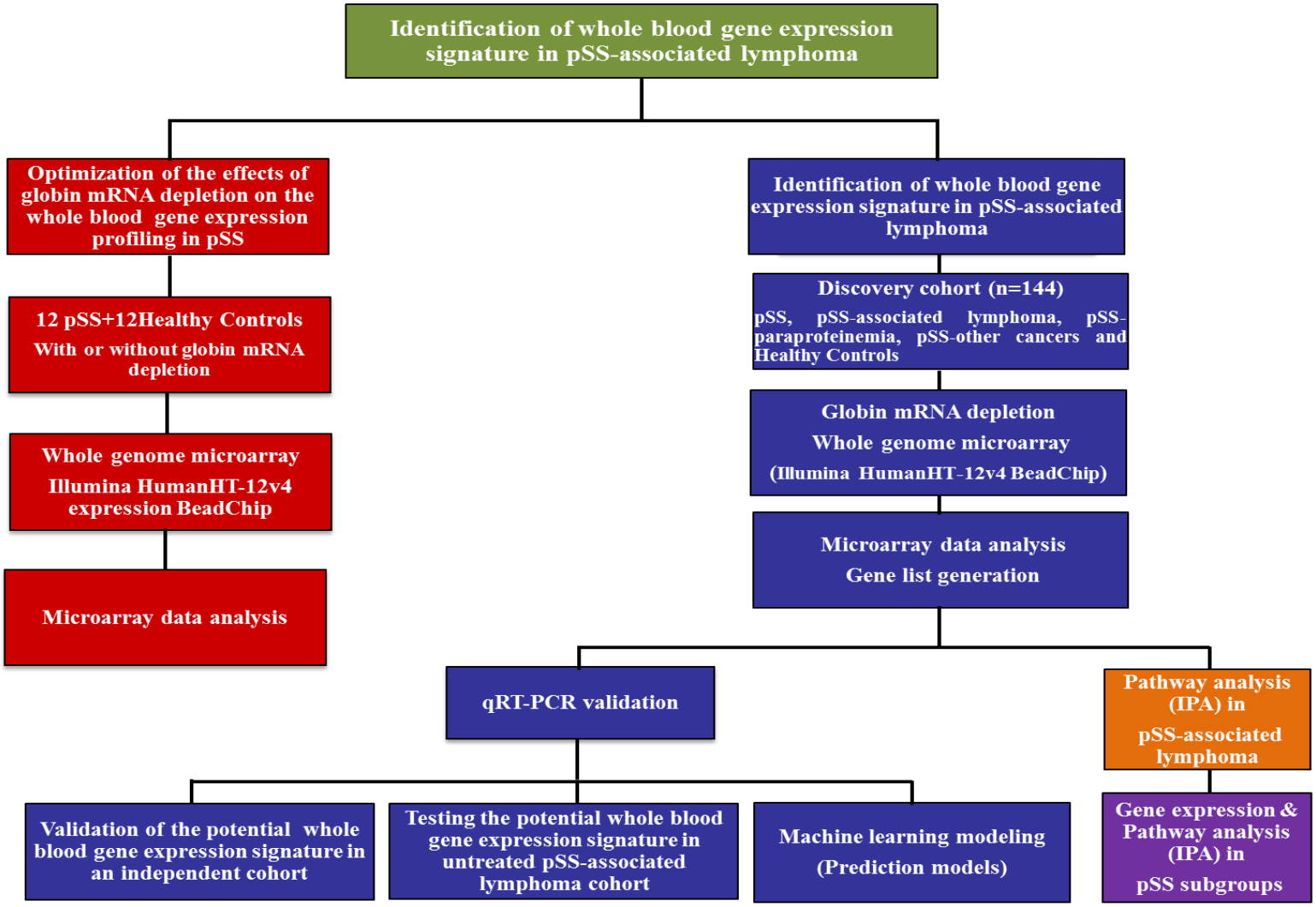


Figure 1.3 Overall design of the project. *Boxes with the same colour are reported in the same chapter*.

Chapter 2
Materials and methods

2.1 UK primary Sjögren's syndrome registry (UKPSSR) and sample collection

The majority of the samples used in this study were taken from the UK Primary Sjögren's Syndrome Registry (UKPSSR) biobank (Ng et al., 2011). All pSS patients fulfil the American European Consensus Group (AECG) classification criteria. Each patient and/or healthy control has various biological samples to be collected including peripheral blood mononuclear cells, serum, RNA and DNA. Throughout my study I have used the RNA samples where the whole blood samples were collected into PAXgene blood RNA tubes (catalogue number 762165, BD, U.S.A.). The tubes contain reagents that stabilise intracellular RNA. The stabilisation of the RNA happens by preventing *in vitro* RNA degradation and gene induction from occurring after blood draw thereby enabling more accurate intracellular RNA analysis of the samples. All samples were kept in the PAXgene blood RNA tubes at –80 ºC until extraction. The following clinical data were extracted from the UKPSSR database for each sample: age, gender, unstimulated oral salivary flow, schirmer's test, presence or absence of anti-Ro/SAA and anti-La/SSB, white cells count, Neutrophils count, lymphocytes count, erythrocyte sedimentation rate (ESR), immunoglobulin G (IgG), complement component 3 (C3), complement component 4 (C4), C-reactive protein (CRP), presence or absence of rheumatoid factor (RF), Sjögren's Syndrome Damage Index (total SSDDI), EULAR Sjögren's Syndrome Patient Reported Index (ESSPRI) and EULAR Sjögren's Syndrome Disease Activity Index score (ESSDAI score). For the samples in the Discovery and the validation cohorts, a Mann-Whitney U test was performed to evaluate any statistical differences on the above clinical data between the subject groups. The clinical data are presented in the relevant result sections of chapter 3 and chapter 4.

2.2 Whole blood RNA extraction

Materials:

1. PAXgene Blood miRNA kit (catalogue number 763134, PreAnalytiX, Switzerland). The kit provides the following materials (some of the buffers were provided as concentrated solutions and they were diluted as instructed in the manufacturer's protocol):
   - PAXgene RNA Spin Columns (red)
   - PAXgene Shredder Spin Columns (lilac)
   - Buffer BM1 (resuspension buffer)
   - Buffer BM2 (binding buffer)
   - Buffer BM3 (washing buffer)
   - Buffer BM4 (washing buffer)
   - Buffer BR5 (elution buffer)
   - Proteinase K
   - RNase-Free DNase set, which includes lyophilized RNase-Free DNase, Buffer RDD and RNase-Free water
   - Secondary Hemogard closures
   - RNase-Free water
   - Microcentrifuge and processing tubes
2. Isopropanol (Catalogue number P/7490/17, Fisher Chemical, UK)

Equipment:

1. Pipettes (Labnet International, U.S.A.)
2. SIGMA 6K15 swing-out rotor centrifuge (Sigma, Germany)
3. Vortex Genie2 (Scientific Industries, U.S.A.)
4. Thermomixer Compact, shaker- incubator (Eppendorf, Germany)
5. Accuspin^{TM} Micro microcentrifuge (Fisher Scientific, Germany)
6. Nano-drop ND-1000 spectrophotometer (Nanodrop Technologies, U.S.A.)

Method:

Whole blood total RNA was extracted using the PAXgene blood miRNA kit according to the manufacturer's protocol. Briefly, frozen samples were thawed at room temperature (15–

25 ºC) for 4 hours to ensure complete lysis of blood cells; the subsequent procedures of RNA extraction were as follows:

1. Samples were pelleted using a swing-out rotor centrifuge for 10 min at 3000–5000 x g, the supernatant was then gently decanted into a container with Virkon solution and the rim of the tube dried with a clean paper towel.
2. The pellets were washed by adding 4 ml RNase-free water, the tubes sealed using a new secondary Hemogard closure, and then the pellets were dissolved by vortex mixing followed by centrifugation for 10 min at 3000–5000 x g, after which the supernatant was discarded by decanting.
3. The pellets were resuspended with buffer BM1 (350 µl) and vortexed until they were visibly dissolved.
4. Samples were transferred into 1.5 ml microcentrifuge tubes and digested by adding buffer BM2 (300 µl) and proteinase K (40 µl) separately, samples then vortexed for 5 s, then an incubation for 10 min at 55 ºC in a shaker-incubator at 1400 rpm.
5. The homogenisation step was then performed by pipetting the samples into PAXgene Shredder Spin Columns (lilac) placed in a 2 ml processing tube, which were spun in a microcentrifuge for 3 min at 13,000 x g. The supernatant of the flow-through was gently transferred into a new microcentrifuge tube to avoid disturbing the pellets.
6. In order to optimise the binding conditions, Isopropanol (700 µl, 100%, purity grade puriss grade (p.a. = 98.5%)) was added and mixed by vortexing.
7. A PAXgene RNA Spin Column (red) placed into a 2 ml processing tube was used at this stage. A portion of the sample (700 µl) was pipetted into the column and centrifuged for 1 min at 13,000 x g; this step was repeated with the remaining sample after replacing the processing tube with a new one. The RNA and miRNA in the sample were bound to the PAXgene silica-membrane in the RNA Spin Column.
8. Buffer BM3 (350 µl) was then added, centrifuged for 15 s at 13,000 x g followed by replacing and discarding the used processing tube.
9. DNase I was prepared by mixing 10 µl of DNase I stock solution (previously prepared by adding 550 µl RNase-free water to 1500 Kunitz units of solid DNase that was provided with the kit), with 70 µl of Buffer RDD for each sample, which was then flicked gently followed by brief centrifugation.

10. Genomic DNA was removed by adding the DNase mixture prepared in step 9 directly onto the PAXgene RNA Spin Column membrane and incubating the tube at room temperature for 15 min.
11. Further washing steps were carried out by adding Buffer BM3 (350 µl) and Buffer BM4 (500 µl), respectively; each buffer addition was followed by centrifugation for 15 s at 13,000 x g. The final washing step was carried out by adding Buffer BM4 (500 µl), followed by centrifugation for 2 min at 13,000 x g. The processing tubes were replaced with a new one after each washing step.
12. To ensure all buffer solution was removed from the RNA Spin Column membrane, the columns were centrifuged for an additional 1 min at 13,000 x g.
13. The RNA was eluted by placing the RNA Spin Columns in a new 1.5 µl microcentrifuge tube and adding 40 µl of Buffer BR5 directly onto the column's membrane. The lid was closed gently and the tube centrifuged for 1 min at 13,000 x g and was repeated using another 40 µl Buffer BR5 in the same microcentrifuge tube, resulting in 80 µl of extracted RNA.
14. The A26/A280 and A260/A230 ratios of the eluent (RNA) were assessed using the Nano-drop ND-1000 spectrophotometer and the RNA samples were then stored at -80 ^{o}C.

2.3 Optimisation of whole blood gene expression signatures in pSS by globin mRNA depletion

2.3.1 RNA clean up and concentration

Materials:

1. RNeasy ® MinElute ® Cleanup Kit (catalogue number 74204, QIAGEN, Netherlands). The kit provides the following materials (some of the buffers were provided as concentrated solutions and they were diluted as instructed in the manufacturer's protocol):
   - RNeasy MinElute Spin Columns
   - Collection tubes with two capacities, 1.5 ml and 2 ml
   - Buffer RLT
   - Buffer RPE

2. RNase-Free water
3. Ethanol 100% (catalogue number E/0650DF/17, Fisher Chemical)

Equipment:

1. Pipettes (Labnet International)
2. Vortex Genie2 (Scientific Industries)
3. Accuspin™ Micro microcentrifuge (Fisher Scientific)
4. Nano-drop ND-1000 spectrophotometer (Nanodrop Technologies)

Method:

The RNeasy ® MinElute ® Cleanup Kit was used in this experiment. Briefly, 1 µg of RNA from each sample was taken and its volume adjusted to 100 µl with RNase-free water. Buffer RLT (350 µl) was added and mixed well, followed by 250 µl of 96–100% ethanol and mixed by pipetting. Each sample was then transferred to an RNeasy MinElute Spin Column that was placed in a 2 ml collection tube and centrifuged for 15 s at 13,000 x g. Each Spin Column was placed in a new collection tube, 500 µl of Buffer RPE added and centrifuged for 15 s at 13,000 x g. This was followed by a washing step with 80% ethanol (500 µl), centrifuged for 2 min at 13,000 x g, and the columns were dried by placing them in new collection tubes and centrifuging them for 5 min at full speed with an opened lid. At this point, the columns were placed in new 1.5 µl microcentrifuge tubes and 14 µl of elution Buffer BR5 was added and centrifuged for 1 min at full speed to elute the concentrated RNA. The purity and concentration of the samples (A260/A280 and A260/A230 ratio) were measured using a Nano-drop ND-1000 spectrophotometer.

2.3.2 Globin mRNA depletion

Materials:

1. Human GLOBINclear kit (catalogue number AM1980, Ambion Inc., U.S.A.) The kit provides the following materials (some of the buffers were provided as concentrated solutions and they were diluted as instructed in the manufacturer's protocol):
   - Globin mRNA depletion reagents, which include Capture Oligo Mix, Streptavidin Magnetic Beads, 2X Hybridization Buffer, Streptavidin Bead buffer and Nuclease-free water.

- RNA purification reagents, which include RNA Binding Beads, RNA wash solution (concentrated), Elution Buffer, RNA Bead Buffer, RNA Binding Buffer (concentrated).

Equipment:

1. Pipettes (Labnet International)
2. Tube Magnetic stand
3. Vortex Genie2 (Scientific Industries)
4. Dry bath system (Star Lab, Taiwan)
5. Accuspin™ Micro microcentrifuge (Fisher Scientific)
6. Nano-drop ND-1000 spectrophotometer (Nanodrop Technologies)

Pre procedure preparations:

1. Streptavidin Magnetic Beads preparation: For each sample, 30 µl of beads were added to a 1.5 ml microcentrifuge tube and washed by subjecting the beads to a magnet to remove the supernatant followed by resuspension with 30 µl of Streptavidin Beads Buffer.
2. Bead Resuspension Mix: For each sample, a total volume of 20 µl was prepared by combining 10 µl of RNA binding beads, 4 µl of RNA Beads Buffer and 6 µl of Isopropanol followed by mixing thoroughly by a vortex mixer.

Method:

The globin RNA was depleted using the Human GLOBINclear kit, which was used according to the manufacturer's protocol, except that a dry bath system instead of an incubator was used to warm up the 2X hybridization buffer and the Streptavidin Bead Buffer as well as other procedures requiring incubations. The kit uses a hybridization capture technology to remove globin mRNA as demonstrated in Figure 2.1. First, the globin mRNA was hybridized with the Globin Capture Oligonucleotides; to each 14 µl of the sample, 1µl of the Capture Oligo mix and 15 µl of pre-warmed 2X hybridization Buffer were added and the mixture allowed to hybridize by incubation for 15 min at 50 ºC. During

this stage the biotinylated oligonucleotides in the Capture Oligo mix bind to the globin mRNA in the sample. Second, the globin mRNA was removed by adding the pre-prepared Streptavidin Magnetic Beads (30 µl). The globin mRNA/biotinylated oligonucleotides were bound to the beads via a 30 min hybridization period at 50 ºC. Third, the bound globin mRNA/biotinylated oligonucleotides were removed by pulling the Streptavidin Magnetic Beads out of suspension using a magnet; the enriched RNA (Globin mRNA–depleted) that was retained in the supernatant was transferred into a new 1.5 ml microcentrifuge tube. Fourth, the RNA samples were washed by adding 100 µl RNA Binding Buffer and 20 µl pre-prepared Bead Resuspension Mix. The enriched RNA bound to the resuspension beads, which were pulled out using a magnet, and the RNA eluted with the Elution Buffer (30 µl) . Finally, the purity and concentration of the RNA was measured using a Nano-drop ND-1000 spectrophotometer and stored at -20 ºC until use.

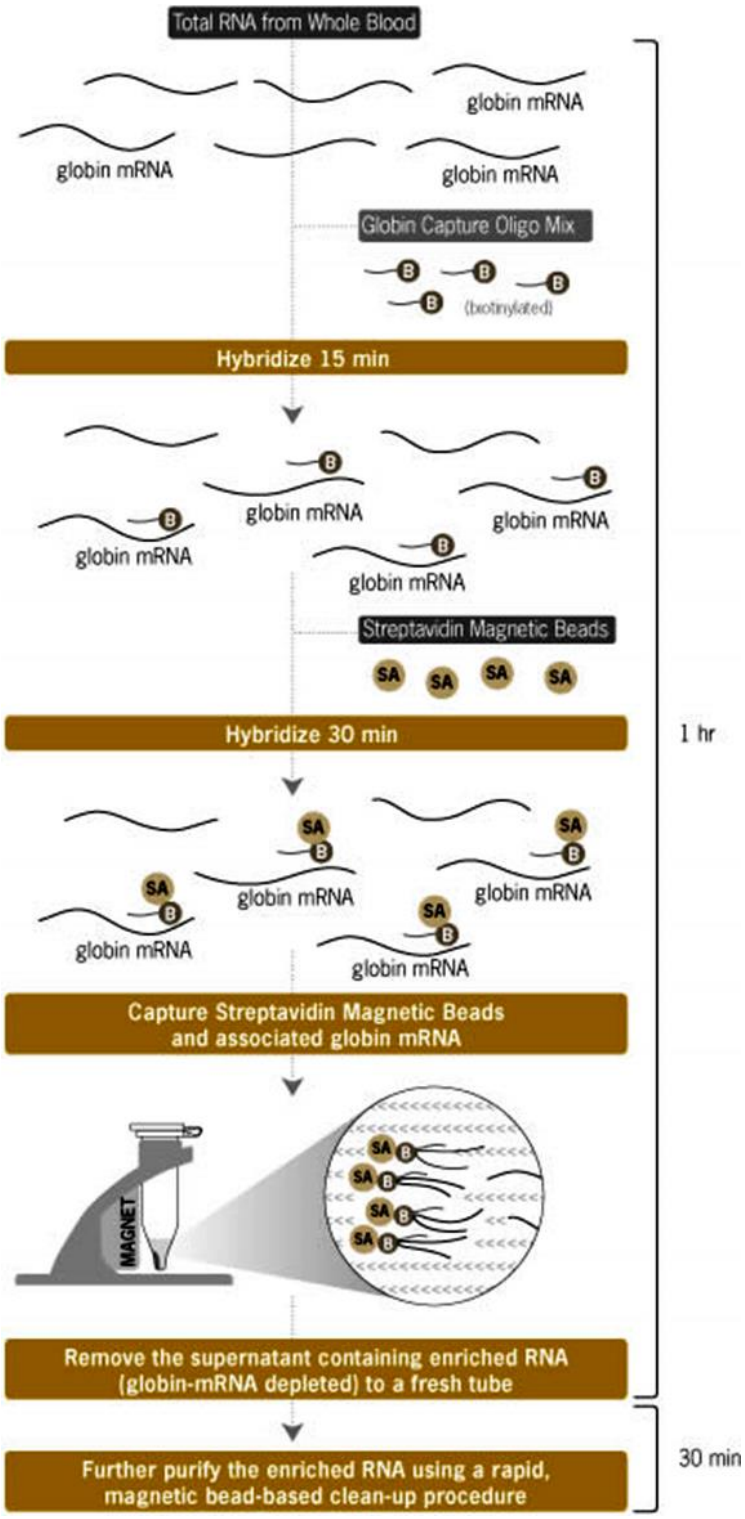


Figure 2.1 Globin mRNA depletion procedure using the Human GLOBINclear kit

2.3.2 Evaluation of β-globin RNA expression levels in total RNA samples with or without Globin mRNA depletion

Materials:

1. M-MLV Reverse Transcriptase kit (Invitrogen, Life Technologies, U.S.A.)
2. DNase-RNase-free water (catalogue number W 4502, SIGMA)
3. TaqMan® Gene Expression Master Mix (catalogue number 4369016, Applied Biosystems, U.S.A.)

Equipment:

1. Pipettes (Labnet International)
2. Vortex Genie2 (Scientific Industries)
3. Accuspin™ Micro microcentrifuge (Fisher Scientific)
4. PTC-200 Peltier Thermal Cycler (MJ Research Inc., U.S.A.)
5. MicroAmp® Fast Optical 96-well (catalogue number 4346906, Applied Biosystems, U.S.A.)
6. MicroAmp® Optical Adhesive Film (catalogue number 4311971, Applied Biosystems, U.S.A.)
7. Heraeus Megafuge 40 (Thermo Scientific, Germany)
8. Applied Biosystems 7900HT Real-Time PCR System, U.S.A.

Method:

The efficacy of globin mRNA depletion was evaluated using TaqMan qRT-PCR. As the amount of total RNA was limited in some samples, we assessed the level of β-globin mRNA in 11 paired samples with or without globin mRNA depletion (2 pSS patients, 9 healthy controls). The forward and reverse primers for the β-globin gene were designed using the Universal probe library assay design and their sequences were 5'-GCACGTGGATCCTGAGAACT-3' and 5'-CACTGGTGGGGTGAATTCTT-3' respectively. The primers were manufactured by SIGMA-Aldrich. The no. 61 Universal probe was used in the RT-PCR reaction. The 18S subunit mRNA sequence was used as a housekeeping gene. The step-by-step protocol is described below:

Step 1: Reverse Transcription PCR. A total of 200 ng of RNA (in 8 µl) from each sample was used to generate cDNA as described in Table 2.1, based on the number of replicates needed for each assay, including an extra amount to allow for pipetting error.

Step 2: A 5 µl aliquot of the resultant cDNA solution from each sample was used for qRT-PCR. The sample's dilutions used were 1:5 for β-globin, and 1:10000 for 18S. To each cDNA sample, 15 µl of RT-PCR master mix were added. The reference volumes of Master mix reagents for a single reaction (20 µl) in the RT-PCR shown in Table 2.2.

Step 3: qRT-PCR data were analysed by normalising the values obtained to the housekeeping gene 18S, and the relative expression level of the β-globin gene in the samples with or without globin mRNA depletion was calculated.

Table 2.1 The M-MLV Reverse Transcriptase PCR reference volumes and their corresponding programs in the thermal cycler

Component	**Volume (µl)/reaction**
Step 1: Master Mix 1	
dNTP	3 µl
Hexamers	1 µl
• **Add 4 µl of Master Mix 1 to each sample** • **Heat to 70 °C for 5 min. using the thermal cycler**	
Step 2: Master Mix 2	
1st strand 5X buffer	4 µl
DTT	2 µl
MMLV (M-MLV RT)	0.5 µl
H_2O	1.5 µl
• **Add 8 µl of Master Mix 2 to each sample (total 20 µl)** • **Heat to 37 °C for 50 min. then 75 °C for 15 min using the thermal cycler**	

Table 2.2 Master mix reagents reference volumes of a single reaction in the RT-PCR .

Component	**Volume (20 µl)/ reaction**
Forward primer	0.4 µl
Reverse primer	0.4 µl
Probe	0.2 µl
TaqMan Gene Expression Master Mix	10 µl
H_2O	4 µl
Total volume	15 µl

2.3.4 Illumina Human HT12-v4 Expression BeadChip (Cambridge Genome Science/University of Cambridge, UK)

A total of 48 samples, 24 globin mRNA-depleted samples (G-depleted group) and 24 non-globin mRNA depleted samples (G non-depleted group) were used in this experiment. Two aliquots from each sample were sent for whole genome microarray analysis at the Cambridge Genomic Services (CGS), University of Cambridge. The first aliquot contained a total of 250 ng of RNA and a second aliquot contained 3 µl from the original sample for quality control (QC). RNA quality was assessed by the Aglient 2100 bioanalyzer using a special Aglient RNA Nano kit. The Bioanalyzer has software (Agilent 2100 Expert Software) for the analysis of the overall integrity of the RNA sample. An RNA Integrity Number (RIN score) was generated for each sample on a scale of 1–10 (1=lowest; 10=highest) as an indication of RNA quality. The 18s/28s ratio and an estimation of RNA concentration were also produced. The Total Prep 96-RNA amplification kit (Ambion, Inc., U.S.A.) was used to amplify the sample by *in vitro* transcription of cDNA generated from the total RNA and the generation of biotin-labelled cRNA. The biotin-labelled cRNA was then used in the direct hybridization assay, which included sequential steps of array hybridization, washing, blocking and streptavidin-Cy3 staining as shown in Figure 2.2. The Illumina iScan was used to detect Cy3 fluorescence emission and the GenomeStudio (version 1.6) was used for the extraction of the raw data.

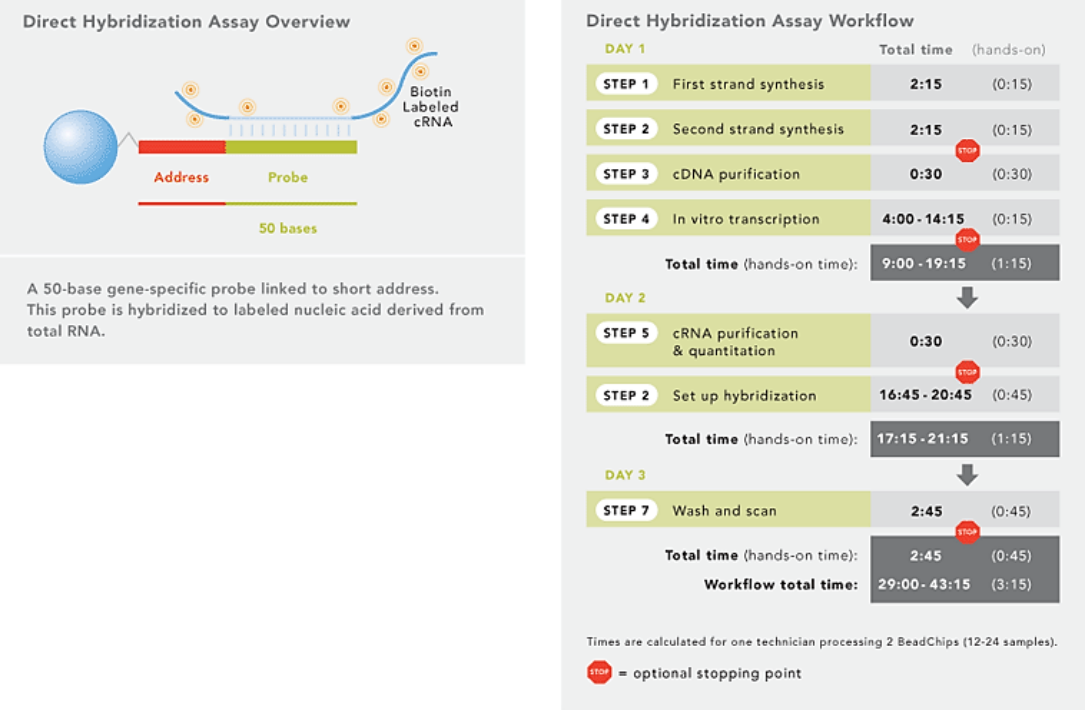


Figure 2.2 Direct Hybridization Assay used in illumina microarray. *The left panel is an overview of the method and on the right panel is an illustration of the workflow* ***(www.illumina.com)***

Microarray data analysis:

Microarray data analysis was carried out with the assistance of the Bioinformatics Support Unit (BSU) at Newcastle University. The raw gene expression data was received from CGS in the form of IDAT files, which are binary data files directly from the microarray scanner. The IDAT files were background-corrected in Illumina's Genome Studio, and exported as a sample probe profile that contained data fields specific to bead-level information such as bead standard error. The subsequent analysis was performed in R, utilising Bioconductor libraries (www.bioconductor.org) (Gentleman et al., 2004). The sample probe profile was read into R and pre-processed using the Lumi package (Du et al., 2008). Pre-processing steps consisted of a dataset transformation using the variance stabilising transformation (VST), and robust spline normalisation (RSN). VST is a log2-like transformation optimised for microarray data, and RSN is a normalisation method to help make samples comparable to one another. After pre-processing, a QC step was carried out to identify problematic or

outlier samples; common metrics were used including: detection p-values, principle component analysis (PCA), and unsupervised hierarchical clustering. If samples were deemed to be outliers or problematic then they were removed from the analysis and the pre-processing stages ran again. Annotation of the array probes was mapped using the lumiHumanAll.db package, specifically nuID, Illumina ID, gene symbol, and description. Differential expression was achieved by fitting linear models using the Limma package (Smyth, 2004). Filters for significance were defined as a greater than absolute log2(1.2)-fold change, and a Benjamini-Hochberg false discovery rate (FDR) corrected p-value of <0.05. The results of this analysis were visualised using volcano plots. Pathway analysis utilised the KEGG REST API as this is still currently maintained, unlike the Bioconductor package. Significant pathways were identified using a hypergeometric test, with an FDR adjusted p-value < 0.05. Additionally, Gene Ontology (GO) terms were identified in the same manner as pathways, for molecular function (MF), cellular components (CC), and biological process (BP). The R scripts for this microarray analysis can be found in supplementary table S1.

2.4 Identification of whole blood gene expression signature of pSS-associated lymphoma

2.4.1 Illumina Human HT12-v4 Expression BeadChip (Cambridge Genome Science/University of Cambridge, UK)

The set of samples used to identify the whole blood gene expression signature of pSS-associated lymphoma is the Discovery cohort, which consisted of 144 globin mRNA–depleted subjects. To reduce the potential influence of gender on the transcriptomic signature of pSS-associated lymphoma, and since pSS predominantly affects females, I chose to use only female subjects. The cohort consists of the following 5 subject groups:

1. 61 pSS patients without history of lymphoma (primary comparator group)
2. 16 pSS-associated lymphoma patients
3. 21 pSS-other cancers patients
4. 23 pSS-paraproteinemia patients
5. 23 healthy controls

All patients in the pSS-associated lymphoma group had a diagnosis of lymphoma in the past and had received treatment for the lymphoma. RNA extraction, depletion of globin mRNA, and a whole genome expression microarray analysis were carried out as described in section 2.3 with some additional steps, which are explained below.

Microarray data analysis:

The analysis of the Discovery cohort microarray data was similar to the protocol used for the Globin study, with some additional steps. An additional QC step was performed to identify outlier samples, using the arrayQualityMetrics package (Kauffmann et al., 2009). Outlier samples were removed from the raw data object and VST, and RSN pre-processing steps as previously described were reapplied. Samples with a RNA integrity number (RIN) of < 7 were removed, except for pSS-associated Lymphoma samples, since the pSS-associated lymphoma group size in this experiment was already relatively small. Following pre-processing, batch effect was identified and this was corrected for using the ComBat function from the SVA package (Johnson et al., 2007). The ComBat function uses an empirical Bayes approach to remove a known batch effect. The resulting matrix from ComBat is used in the linear model to explore differential expression. The primary comparison that I used to identify the whole blood transcriptome of pSS-associated lymphoma was between the pSS and the pSS-associated lymphoma group. In the primary analysis, all samples with RIN < 7 were excluded in the pSS group. On the other hand, as the pSS-associated lymphoma samples were limited, only the technical outliers were excluded as the microarray data of these outliers are markedly different from the remaining samples and therefore might skew the analysis. In order to reduce the risk of overlooking genes that might be important in developing lymphoma in pSS, I performed several sub-analyses in which different criteria were used for inclusion or exclusion of samples within the pSS-associated lymphoma groups based on the quality of the RNA samples and the detection as outliers during PCA analysis. More details about these analyses are described in Chapter 4. The final list of differentially expressed genes in pSS-associated lymphoma was compiled by combining the genes from all sub-analyses and ranking them according to their fold change and p-value.

Finally, additional group comparisons were also performed in order to determine the specificity of the gene expression signature identified, these include:

1. pSS-associated lymphoma vs pSS-other cancers
2. pSS-associated lymphoma vs pSS-paraproteinemia
3. pSS vs pSS-other cancers
4. pSS vs pSS-paraproteinemia
5. pSS-other cancers vs pSS-paraproteinemia
6. pSS vs HC
7. pSS-associated lymphoma vs HC
8. pSS-other cancers vs HC
9. pSS-paraproteinemia vs HC

In addition to identifying a list of differentially expressed genes, I have also conducted pathway analysis using the differentially expressed genes using Ingenuity Pathway Analysis (IPA) for all the comparisons. The methods of these additional analyses are described in the respective result chapters. The R scripts for this microarray analysis can be found in supplementary table S2.

2.4.2 Technical validation of the differentially expressed genes from the microarray data by qRT-PCR

The qRT-PCR was performed using TaqMan® Custom gene expression array plates (Fast plates) from Applied Biosystems (Life Technologies, U.S.A.) for the technical validation of the potential signature identified from the microarray. Because of the limitations on the total amount of RNA available following globin mRNA removal, a total of 61 differentially expressed genes (out of 68) were chosen according to the fold change and p-value (higher fold changes and lower p-values). In addition, the selection of genes in common among the all the analyses (i.e. including the sub-analyses) was considered. Only samples from the pSS without lymphoma and pSS-associated lymphoma groups were used in the technical validation in all the array plates batches, while, samples from all pSS subgroups were used in the first 2 plate batches to evaluate the expression stability of the housekeeping genes.

The TaqMan® Custom gene expression array plates is a technology in which a fast TaqMan gene expression assay (probe and primers sets) is dried-down in a 96-well format plate.

Two formats (32-format and 16-format) of the plates were used. The TaqMan array plate assay involved two steps:

1. Converting the RNA into cDNA by RT-PCR.
2. TaqMan® Custom gene expression array.

1. Converting the total RNA into cDNA by Reverse Transcription PCR

Materials:

1. High-Capacity cDNA Reverse Transcription Kit (catalogue number 4368814, Thermo Fisher Scientific, U.S.A.), which includes 10X RT Buffer, 10X RT Random Primers, 25X dNTP Mix and MultiScribe™ Reverse Transcriptase.
2. DNase-RNase-free water (catalogue number W 4502, SIGMA)

Equipment:

1. Pipettes (Labnet International)
2. Vortex Genie2 (Scientific Industries)
3. Accuspin™ Micro microcentrifuge (Fisher Scientific)
4. Heraeus Megafuge 40 (Thermo Scientific)
5. PTC-200 Peltier Thermal Cycler (MJ Research Inc.)

Method:

1. The kit's components were thawed on ice.
2. As a result of limited RNA entities, both globin mRNA-depleted and non-depleted samples were used to validate the candidate genes. A total of 44 pSS and 15 pSS-associated lymphoma globin mRNA-depleted samples with 250 ng of total RNA were used to validate 49 deferentially expressed genes. A total of 37 pSS and 8 pSS-associated lymphoma non-depleted samples with 200 ng of total RNA were used to validate another 14 deferentially expressed genes. For the protocol, 10 µl of each sample is required. Therefore, for samples with high RNA concentrations, DNase-RNase-free water was added to make up the required volume, while for samples with low RNA concentrations, the RNA samples were concentrated by precipitation to achieve the required volume.

3. The RT master mix was prepared by calculating the component volumes as listed in Table 2.3, taking into account the total number of reactions and margins of error in pipetting.
4. A volume of 10 µl of the 2X RT master mix was added to each 10 µl sample in a PCR micro tube, and mixed well by pipetting. The samples were then centrifuged briefly and placed on ice until RT-PCR was performed.
5. The thermal cycler was programmed according to the conditions listed in Table 2.4.

Table 2.3 The High-Capacity cDNA Reverse Transcription reference volumes of a single reaction

Component	**Volume (µl)/reaction**
10X RT buffer	2
25X dNTP Mix (100mM)	0.8
10X RT Random Primers	2
MultiScribe™ Reverse Transcriptase	1
Nuclease Free water	4.2
Total volume per reaction	10

Table 2.4 The High-Capacity cDNA Reverse Transcription thermal cycler conditions

	Step 1	**Step 2**	**Step 3**	**Step 4**
Temperature	25 ºC	37 ºC	85 ºC	4 ºC
Time	10 min	120 min	5 min	Continually

2. TaqMan® Custom gene expression array

Materials:

1. TaqMan® Gene Expression Master Mix (Applied Biosystems)
2. DNase-RNase-free water (catalogue number W 4502, SIGMA)

Equipment:

1. Pipettes (Labnet International)
2. Vortex Genie2 (Scientific Industries)
3. Accuspin™ Micro microcentrifuge (Fisher Scientific)
4. Heraeus Megafuge 40 (Thermo Scientific)
5. TaqMan® Custom gene expression array plates (Fast plates) (Applied Biosystems, U.S.A.)
6. MicroAmp® Optical Adhesive Film (Applied Biosystems)
7. Applied Biosystems 7900HT Real-Time PCR System

Method:

a. Housekeeping genes selection:

The selection of the housekeeping genes (endogenous controls) for the experiment was performed using NormFinder (http://moma.dk/normfinder-software), an algorithm that ranks a set of candidate normalisation genes according to their expression stability in a given sample set and given experimental design (Andersen et al., 2004). The functionality of NormFinder can be added directly into Microsoft Excel. In my Discovery cohort dataset, 32 housekeeping genes (representing 21 unique genes as each gene has several corresponding probes in the microarray) from the microarray data were selected by NormFinder as candidates. Two genes (*YWHAZ* and *UBC*) were identified as the best candidate genes for normalisation. In addition, in the first and second batches of the qRT-PCR array, I included two other housekeeping genes: the first one was *ACTB*, which had the most stable expression value after the *YWHAZ* and *UBC*. The second one was the 18S subunit, because it is a commonly used housekeeping genes in RT-PCR analysis. After performing the first and the second batches of qRT-PCR, the expression levels of the four housekeeping genes were compared among the five subject groups using the Mann-Whitney U Test. The third and fourth plate batched contain only *YWHAZ* and *ACTB* (the most stable housekeeping genes according to the results of the first two batches).

b. Sample preparations:

The TaqMan® Custom gene expression array plates require a total cDNA amount between 1 and 100 ng. Given the limited amount of RNA samples, I chose to use 5 ng cDNA (in 10 µl) per reaction for the analysis. Equal volumes of cDNA and the TaqMan® Gene Expression Master Mix, as shown in Table 2.5, were mixed together by vortex followed by a brief centrifugation. The total volumes of cDNA and Master Mix needed for the entire experiment was calculated from the total number of TaqMan assays. The preparations were placed on ice until the plates were ready. To each well of the array plate, 20 µl of the mixture was added before loading the plate into a qRT-PCR instrument using the suitable thermal cycling conditions. The genes that I have technically validated for the discovery cohort are listed in Table 2.6.

Table 2.5 TaqMan gene expression Master mix reference volumes of a single reaction

Component	**Volume (20 µl)/ reaction**
5 ng cDNA + DNase-free water	10 µl
2X TaqMan® Gene Expression Master Mix	10 µl
Total volume	20 µl

Table 2.6 TaqMan gene expression assays of the technical validation of the Discovery cohort

Gene symbol	TaqMan assay ID	type of gene	Gene symbol	TaqMan assay ID	Type of gene
18S	Hs99999901_s1	housekeeping gene	*LGALS1*	Hs00355202_m1	DEG
ACTB	Hs99999903_m1	housekeeping gene	*LRFN3*	Hs00225874_m1	DEG
UBC	Hs00824723_m1	housekeeping gene	*LRIG1*	Hs00394267_m1	DEG
YWHAZ	Hs00237047_m1	housekeeping gene	*MAGED1*	Hs00986269_m1	DEG
ALDH9A1	Hs00997881_m1	DEG	*MGST3*	Hs01058946_m1	DEG
ATG12	Hs01047860_g1	DEG	*MYC*	Hs00153408_m1	DEG
ATP1A1	Hs00167556_m1	DEG	*NAT10*	Hs01120371_m1	DEG
BCL11B	Hs01102259_m1	DEG	*NCSTN*	Hs00950933_m1	DEG
BMS1	Hs01036249_m1	DEG	*NUDT14*	Hs00418228_m1	DEG
BTBD11	Hs00537023_m1	DEG	*OAF*	Hs00420156_m1	DEG
C10orf32	Hs00376014_m1	DEG	*PAF1*	Hs00219496_m1	DEG
CBLL1	Hs01128720_m1	DEG	*POM121C*	Hs03406359_mH	DEG
Cd96	Hs00175524_m1	DEG	*PRKCQ*	Hs00989970_m1	DEG
CDR2	Hs00386212_m1	DEG	*PRPF8*	Hs01556852_m1	DEG
CDV3	Hs00250190_m1	DEG	*RAB37*	Hs03988369_g1	DEG
CNPY3	Hs00198139_m1	DEG	*RASGRP1*	Hs00996727_m1	DEG
CYFIP2	Hs00910722_m1	DEG	*RBL2*	Hs00180562_m1	DEG
DDB1	Hs01096550_m1	DEG	*RBP7*	Hs00364812_m1	DEG
DRAP1	Hs01012815_g1	DEG	*RNA28S5*	Hs03654441_s1	DEG
DYNLL1	Hs00853309_g1	DEG	*RPA2*	Hs00358315_m1	DEG
EHBP1L1	Hs00411094_m1	DEG	*RRN3*	Hs01592557_m1	DEG
ESYT1	Hs00248693_m1	DEG	*SEC61G*	Hs00414142_m1	DEG
ETS1	Hs00428293_m1	DEG	*SF3A1*	Hs01066327_m1	DEG
HCFC1R1	Hs01002754_m1	DEG	*SGK223*	Hs00410725_m1	DEG
HLA-DRB1	Hs99999917_m1	DEG	*SLC7A1*	Hs00931450_m1	DEG
HNMT	Hs02759756_s1	DEG	*SMARCA2*	Hs01030846_m1	DEG
HNRNPUL1	Hs00199870_m1	DEG	*SPOCK2*	Hs00360339_m1	DEG
HSP90B1	Hs00427665_g1	DEG	*SRP14*	Hs03055045_g1	DEG
HSPA9	Hs00269818_m1	DEG	*SUN2*	Hs00391446_m1	DEG
ITK	Hs00950637_m1	DEG	*UBXN11*	Hs00377277_m1	DEG
KCTD12	Hs00540818_s1	DEG	*VCP*	Hs00997642_m1	DEG
KHDRBS1	Hs00173141_m1	DEG	*WAC*	Hs00249774_m1	DEG
LEF1	Hs01547250_m1	DEG			

** DEGs= differentially expressed genes*

c. Plate preparations:

The first step of TaqMan® Custom gene expression array plate preparation was to centrifuge the plate shortly, to prevent the loss of the lyophilized primers that are provided within each well when the plate is opened. After removing the plate cover, 20 µl of the cDNA-master mix mixture was dispensed into the appropriate wells, finally the plate was covered with MicroAmp® Optical Adhesive Film and centrifuged shortly.

d. Running the plate in an RT-PCR instrument:

The plate was run using the Applied Biosystems 7900HT Real-Time PCR System. The software that used was the SDS v2.4. The details of the PCR program used for the reaction are shown in Table 2.7.

Table 2.7 TaqMan gene expression experimental parameters. *The reactions were carried out using the Applied Biosystems 7900HT Real-Time PCR System*

Experiment parameters	**Thermal cycling conditions**		
	Stage	**Temp. (ºC)**	**Time (min:sec)**
Reaction volume 20µl	Stage 1	50	2:00
Ramp rate: Standard	Stage 2	95	10:00
	Stage 3	95	0:15
	(40 cycles)	60	1:00

e. RT-PCR data analysis:

The data were analysed using SDS RQ Manager 1.2.1. The analysis consisted of several steps:

i. The AQ file was converted to an RQ file.
ii. In analysing the levels of expression of each gene, the same baseline and threshold were used across all the samples.

iii. The data was normalised using the most stable "housekeeping" gene (*YWHAZ*) in the dataset as determined by the levels of expression across all subject groups. The relative expression level for individual gene to *YWHAZ* was then estimated.

iv. Comparisons of the expression level of each candidate gene in the pSS versus pSS-associated lymphoma groups were made using a non-parametric test (Mann-Whitney U Test) and p-values < 0.05 were considered statistically significant.

2.5 Biological validation of the whole blood gene expression signature in pSS-associated lymphoma

A second set of independent samples (the Validation cohort) was used to provide biological validation of the potential whole blood gene expression signature in pSS-associated lymphoma that were identified in the Discovery cohort. This cohort consisted of 119 pSS and 17 pSS-associated lymphoma samples. Moreover, an additional set of 7 pSS-associated lymphoma samples was included from patients with pSS-associated lymphoma before treatment. Regarding the pSS-associated lymphoma samples, 8 were obtained from collaborators at the University of Uppsala, Sweden (2 pSS-associated lymphoma and 4 untreated pSS-associated lymphoma), and 7 pSS-associated lymphoma samples were obtained from collaborators at Stavanger University, Norway.

2.5.1 RNA clean up and concentration

Materials:

1. glycogen solution (catalogue number G 1767, SIGMA)
2. 3 M sodium acetate (catalogue number S 7899, SIGMA)
3. Ethanol 100% (catalogue number E/0650DF/17, Fisher Chemical)
4. Ethanol 70%
5. Free-DNase-RNase-water (catalogue number W 4502, SIGMA)

Equipment:

1. Accuspin^{TM} Micro microcentrifuge (Fisher Scientific)
2. Nano-drop ND-1000 spectrophotometer (Nanodrop Technologies)

Method:

RNA precipitation was used to concentrate RNA samples. Briefly, 1 µl of glycogen solution was added to each sample. Next, 0.1 volume of 3 M sodium acetate were added followed by 2 volumes of cold Ethanol 100%. After an overnight incubation at -80 ºC the samples were centrifuged (13,000 x g at 4 ºC for 20 min), the supernatant was carefully removed and the pellets were washed with 250 µl of 70% ethanol, followed by another centrifugation (13,000 x g at 4 ºC for 5 min). The supernatant was removed without disturbing the pellets, then the pellets were air-dried for 30 min. The RNA pellets were resuspended with 12 µl of free-DNase-RNase-water. The purity and concentration of the samples (A260/A280 and A260/A230 ratio) were measured by the Nano-drop ND-1000 spectrophotometer.

2.5.2 Biological validation of the microarray data by qRT-PCR

The TaqMan® custom gene expression plates were used as described previously in section 2.4.2. Samples from all the groups including 119 from the pSS and 17 from the pSS-associated lymphoma groups (treated lymphoma) and 7 from the untreated pSS-associated lymphoma (untreated lymphoma) were used to validate 24 genes out of the 26 genes that were differentially expressed in both microarray and qRT-PCR in the Discovery cohort. The genes and their assay IDs are listed in Table 2.8.

The qRT-PCR data were normalised to the most stable housekeeping gene (*YWHAZ*), and the relative expression levels were calculated. A non-parametric test (Mann-Whitney U test) was used to compared the expression levels of the genes tested between the lymphoma groups and the non-lymphoma group.

Table 2.8 TaqMan gene expression assays of the biological validation

Gene symbol	TaqMan assay ID	Type of gene
ACTB	Hs99999903_m1	Housekeeping gene
YWHAZ	Hs00237047_m1	Housekeeping gene
BMS1	Hs01036249_m1	DEG
C10orf32	Hs00376014_m1	DEG
CBLL1	Hs01128720_m1	DEG
CNPY3	Hs00198139_m1	DEG
CYFIP2	Hs00910722_m1	DEG
DRAP1	Hs01012815_g1	DEG
DYNLL1	Hs00853309_g1	DEG
ESYT1	Hs00248693_m1	DEG
HNRNPUL1	Hs00199870_m1	DEG
LEF1	Hs01547250_m1	DEG
LGALS1	Hs00355202_m1	DEG
MAGED1	Hs00986269_m1	DEG
MGST3	Hs01058946_m1	DEG
NUDT14	Hs00418228_m1	DEG
OAF	Hs00420156_m1	DEG
POM121C	Hs03406359_mH	DEG
PRPF8	Hs01556852_m1	DEG
RBP7	Hs00364812_m1	DEG
SEC61G	Hs00414142_m1	DEG
SF3A1	Hs01066327_m1	DEG
SGK223	Hs00410725_m1	DEG
SRP14	Hs03055045_g1	DEG
UBXN11	Hs00377277_m1	DEG
VCP	Hs00997642_m1	DEG

** DEG= differentially expressed genes from the discovery cohort*

2.6 Prediction models in pSS-associated lymphoma

To identify the most important genes that can predict the group membership of pSS-associated lymphoma, I have used modelling techniques. The gene expression data from qRT-PCR of the Discovery cohort was used to build these models, using multiple logistic regression techniques. A training sets were based on a random two-thirds training set, with the remaining one-third of the cases was retained for testing of the models. A stepwise regression analysis was performed with SAS JMP software. A number of tools were used to minimise over-fitting including cross-validation, evaluation of Akaike and Bayesian Information Criteria and inspection of the residual deviance. Robustness of the models was evaluated by inclusion and exclusion of alternative gene candidates. Models based on the training set were then evaluated on the testing set comparing predicted group membership with observed group membership. For the dataset from the independent validation cohort the models were tested by comparison of observed and predicted group membership and misclassification analysis and the results were represented by mosaic plots.

Chapter 3

Optimisation of the identification of a whole blood gene expression signature in primary Sjögren's Syndrome by globin mRNA depletion

3.1. Introduction

Gene expression profiling in Primary Sjögren's Syndrome (pSS) using samples from a large patient cohort is very challenging for several reasons. Firstly, given the relatively low prevalence of pSS, samples from pSS patients and healthy controls have to be recruited from different centres across the UK. Moreover, because of the involvement of a variety of centres we needed a robust sample collection method that reduced variability between the samples. This led to our decision of using whole blood samples, an approach that is frequently employed in such multi-centre studies. Using whole blood sample has many advantages. Whole blood samples can reduce the variability in sample processing during the isolation of different cell subsets, as it involves additional steps. There are additional challenges in ensuring the same equipment and reagents are used in each tissue collection centre. Furthermore, the time from sample collection to processing may have a greater impact on the quality and quantity of RNA extracted from the samples. In addition, RNA extraction from whole blood is less time-consuming and less expensive, and provides an opportunity to study all the white blood cells subsets together (Vartanian et al., 2009). On the other hand, there are also disadvantages. For instance, whole blood samples contain numerous cellular and non-cellular components that contain RNA materials which may affect gene expression studies.

A major concern regarding whole blood gene expression studies is the abundance of globin mRNA within the samples. It is known that the whole blood consists of red blood cells (RBCs; ~95%), platelets (~5%) and white blood cells (WBCs). WBCs typically constitute less than 1% of the cellular content (Mastrokolias et al., 2012). Therefore, globin transcripts represent about 70% of the total amount of mRNA isolated from whole blood samples (Field et al., 2007). The abundance of globin mRNA in whole blood samples may therefore affect the sensitivity of the microarray by reducing fluorescent label availability for other less-abundant transcripts and hence preventing them from detection by the microarray assay (Liu et al., 2006, Mastrokolias et al., 2012). Therefore, one approach to overcome this

problem is to perform blood fractionation, which removes RBCs from the samples for gene expression study, leaving a more homogenous cell population. However, the blood fractionation procedure may affect gene expression by WBCs as well as introduce additional sample variability due to the additional experimental procedures (Fan and Hegde, 2005).

Globin mRNAs are stable mRNAs in RBCs that is important for the synthesis of globin proteins after RBC enucleation. There are 3 members of the globin gene family: α-globin genes, which are located in chromosome 16 (Goh et al., 2005); β-globin genes, which are located in chromosome 11 and are highly expressed in erythrocytes; and γ-globin genes, which are normally expressed only in foetal liver (Levings and Bungert, 2002). Being the dominant mRNA species in whole blood samples, α-globin and β-globin transcripts may interfere with gene expression profiling using these samples, reducing the sensitivity of microarray signals. Several studies have been carried out to investigate globin reduction for blood-based gene expression studies, with the majority of these studies concluding that globin reduction is preferable (Winn et al., 2011). This preference results from that the reduction of globin mRNA from whole blood samples produced microarray data similar to those obtained from PBMC samples (Raghavachari et al., 2009, Wright et al., 2008). Furthermore, some investigators suggest that the interference of globin mRNA with whole genome gene expression studies is more prominent if the genes of interest are not highly expressed. However, globin mRNA depletion involves additional procedures that may introduce sample variability and damage to mRNAs.

Various methods have been developed to eliminate globin mRNA from blood samples. One of these methods is the GLOBINclear^{TM} Kit. This kit uses a method that benefits from the strong binding between biotin and streptavidin molecules, nucleic acid hybridisation specificity and the use of magnetic beads to separate globin mRNAs from the remaining RNA species. This kit removes both α- and β-globin mRNA (Ambion, 2007). Another method of globin reduction involves blocking of the α- and β-globin mRNA by peptide nucleic acids (PNAs) during the cDNA synthesis. The globin-reduction PNA oligomers are a set of four oligomers complimentary to globin mRNA (Affymetrix, 2004). This method works as an immobiliser that masks globin mRNA during reverse transcription, making non-globin mRNA more available for detection in microarray analysis.

Up to date, a few gene expression studies using pSS peripheral blood samples have also been carried out. One of these studies revealed that genes of the IFN pathways are among the most overexpressed genes involved in pSS pathogenesis, with ten of the top 20 up-regulated genes in pSS being IFN-α inducible genes (Kimoto et al., 2011). The up-regulation of IFN-α inducible genes was also detected in PBMCs from pSS (Emamian et al., 2009). This activated IFN signature in pSS was present regardless of the different versions of gene chips used. Interestingly, the IFN signature correlated with high levels of anti-Ro/SSA and anti-La/SSB autoantibodies. Peripheral blood gene expression profiling also revealed differentially expressed genes that are related to the inflammatory and other immune-related pathways. These pathways include B and T cell receptors, IGF-1, GM-CSF, PPARα/RXRα, and PI3/AKT signalling. Moreover, the authors observed that the abundance of globin mRNA in peripheral blood may have reduced the sensitivity of detection of differentially expressed genes (DEGs) because the number of DEGs identified was greater when PBMC samples were used (Emamian et al., 2009). To my knowledge, no study has been reported regarding the impacts of globin mRNA on whole blood gene expression profiling in pSS.

In my first experiment, globin mRNA depletion was carried out using the GLOBINclear^{TM} Kit. The kit separates out globin mRNA, permitting use of non-3'-bias techniques including the chemical labelling of RNA. Although this method affects RNA quality due to the additional preparation steps, it results in an improved gene expression profile outcome compared with other methods (Liu et al., 2006). Different downstream steps were evaluated and compared with paired samples without globin depletion.

3.2 Aim and Design of the Experiment

The aim of this experiment is to evaluate the effect of globin mRNA depletion on the whole blood gene expression signature in pSS. The data will also be used to optimize the gene expression profiling protocol to identify the transcriptomic signature of pSS-associated lymphoma.

A total of 24 whole blood samples, collected in PAXgene RNA tubes, were selected from the UKPSSR biobank (pSS patients=12, healthy controls=12). The RNA was extracted, cleaned and concentrated (if required). Each sample was split into 2 aliquots, one aliquot

was subjected to the globin mRNA depletion protocol (G-depleted group), the other aliquot was kept without further processing (G non-depleted group). All forty-eight samples (G-depleted=24 and G non-depleted=24) were sent to Cambridge Genomic Services (CGS) at the University of Cambridge for whole genome microarray analysis. The design of the experiment is shown in Figure 3.1.

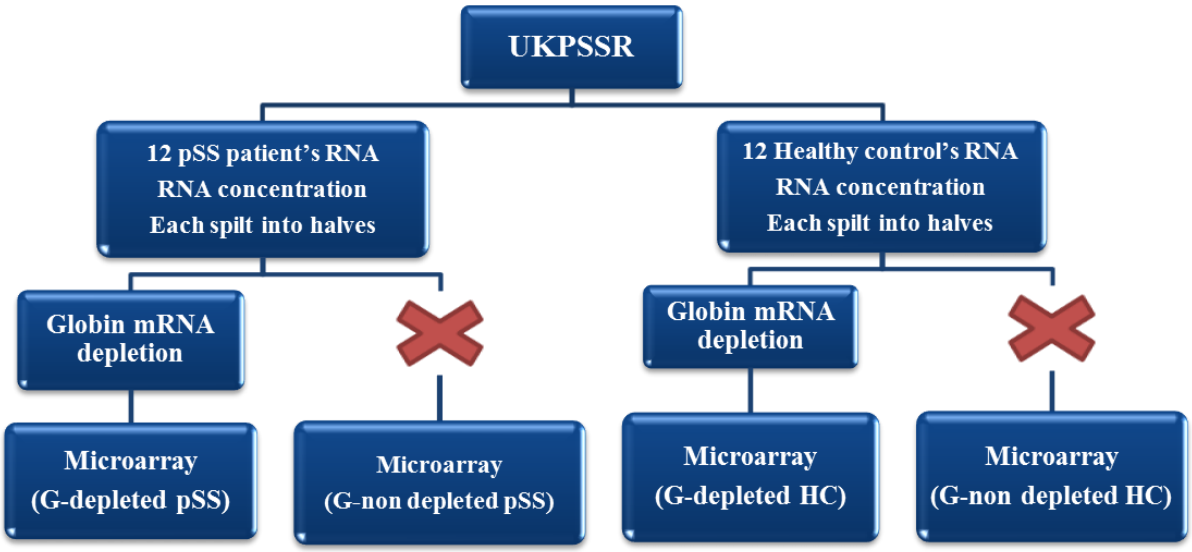


Figure 3.1 Globin mRNA depletion experimental design. *pSS= primary Sjögren's syndrome, HC= Healthy Controls*

3.4 Results:

3.4.1 Demography of study subjects

The pSS patients and healthy controls were selected from the UKPSSR (pSS patients=12, healthy controls=12). All the patients fulfilled the AECG classification criteria and were females with an average age of 52 yrs. (25–76 yrs). The presence of paraproteinemia, lymphoma or other types of cancer was not reported in the chosen pSS patients. The aged-matched healthy controls included two males (females=10, male=2) with an average of 49 yrs. (25–72 yrs). The clinical characteristics and demographics of the patients and healthy controls are shown in Table 3.1

Table 3.1 Clinical data of pSS patients selected for the globin mRNA depletion experiment

Clinical criteria	pSS patients (mean ± S.E.M)	Healthy control
Age (years)	52 ± 4.49	49 ± 3.00
Unstimulated oral salivary flow (ml/15 mins)	1.34 ± 0.54	N/A
Schirmer's test (mm/5 mins)	7.70 ± 3.34	N/A
Anti-Ro/SSA positive (%)	75%	N/A
Anti-La/SSB positive (%)	66.6%	N/A
WCC (x 10^9/l)	5.41 ± 0.88	N/A
Neutrophil (x 10^9/l)	3.13 ± 0.40	N/A
Lymphocytes (x 10^9/l)	1.87 ± 0.36	N/A
ESR (mm/hr)	25.60 ± 6.08	N/A
IgG (g/L)	16.14 ± 1.99	N/A
C3 (mg/dl)	1.18 ± 0.16	N/A
C4 (mg/dl)	0.40 ± 0.20	N/A
CRP (mg/L)	4.5 ± 1.53	N/A
RF (IU)	62.5 ± 35.83	N/A
Total SSDDI	3.5 ± 0.33	N/A
ESSPRI (0-10)	6.47 ± 0.56	N/A
ESSDAI (0-123)	4.45 ± 1.00	N/A

WCC= White Cell Count, ESR= Erythrocyte Sedimentation Rate, C3= Complement Component 3, C4= Complement Component 4, CRP=C-reactive Protein, RF=Rheumatoid Factor, IU=International units, ESSDAI score=(0-4 no activity, 5-12 moderate activity and ≥ 13 high activity)

3.4.2 Assessment of globin mRNA depletion from RNA samples:

Before sending the RNA samples for microarray, the efficacy of globin mRNA depletion was assessed using qRT-PCR. Since the β-globin gene is the main globin gene expressed in adult erythrocytes (Antoniou et al., 1988, Johnstone et al., 2013), I chose to measure the amount of β-globin mRNAs in the samples to assess the efficacy of globin mRNA depletion. As the amount of RNA available from the biobanked samples was limited, the levels of β-globin transcripts from 11 paired samples (HC=9 and pSS patients=2) with or without globin mRNA depletion were measured. The level of expression was normalised to the housekeeping gene 18S. The results showed a significant reduction (average of 64-fold) in the amount of the β-globin mRNA in the G-depleted samples in comparison with their corresponding G non-depleted samples ($p < 0.0001$, Mann-Whitney U test) (Figure 3.2).

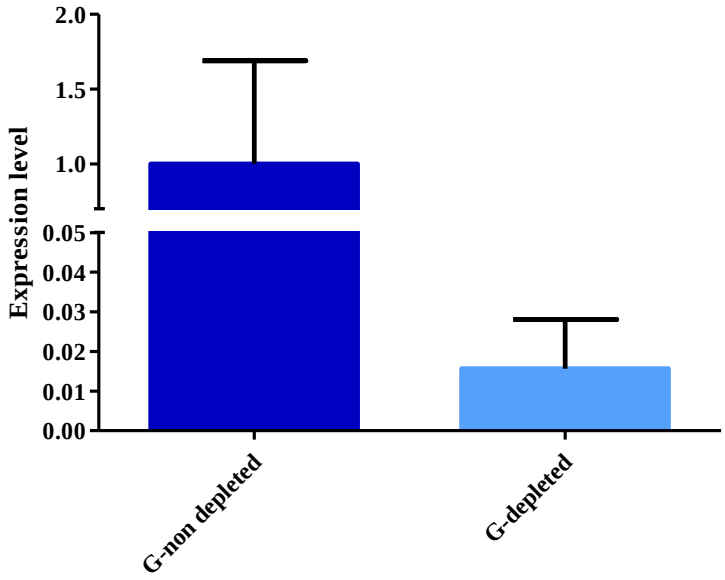


Figure 3.2 Relative gene expression levels of β-globin with and without globin mRNA depletion. *The amount of β-globin mRNA in paired samples, G-depleted and G non-depleted samples (n=11), measured using qRT-PCR. There was a significant reduction (average of 64-fold, p<0.0001) in the expression levels of β-globin mRNA in the G-depleted samples compared to the G non-depleted samples, the error bars show the standard error of the means.*

3.4.3 RNA quality for microarray analysis with or without globin mRNA depletion.

The Bioanalyzer measured the RNA quality, and the data were analysed using the RIN algorithm, which ranks different RNA features, including the 28S to 18S ratio and other factors to give a robust evaluation of RNA quality. The higher the RIN score, the better the RNA quality. Depletion of globin mRNA led to a slight reduction in the RNA quality as shown by a reduction of the RIN score. Nevertheless, the RNA quality of the G-depleted samples remained good (with RIN values being greater than 7), which is considered suitable for microarray experiments. As shown in Table 3.2, the RIN score for some samples could not be determined by the Bioanalyzer (shown in Table 3.2 as N/A). This does not necessarily mean that the RNA quality of the samples was low. An "N/A" result may be returned due to different factors, for instance, unusual ribosomal ratio and background noise. The quality of such samples was therefore judged by visual inspection of the electropherograms and the gel-like images of them. The electropherograms of the amplified cRNA from both pSS and healthy controls from the G-depleted group demonstrate the lack of a sharp peak of about 700 bp, which represents the Globin mRNA in the G-depleted group and which give the curve a bell shape similar to that obtained from PBMCs as demonstrated in Figure 3.3. All samples were judged by CGS staff to be of good quality for the microarray experiment. The electropherograms of the amplified cRNA found in supplementary figures SF1, SF2, SF3 and SF4.

Table 3.2 The RIN scores of RNA samples with or without globin mRNA depletion. *RIN scores were calculated by the algorithms provided by the Agilent Bioanalyzer 2100. Samples IDs with a suffix of "0" represented the healthy controls; whereas a suffix of "1" represented pSS patients*

Sample ID	G-depleted	G-non depleted	Sample ID	G-depleted	G-non depleted
BAS-017-0	7.4	8	NCL-011-1	8	N/A
LEE-059-0	7.5	8.3	DER-006-1	8.1	10
GAT-027-0	8.5	9.6	BIR-051-1	7.9	N/A
BAT-023-0	8.4	9.6	NCL-055-1	7.6	N/A
NCL-136-0	7.7	8.9	BIR-029-1	7.8	10
NCL-091-0	7.8	9.8	NCL-052-1	8	9.9
NCL-123-0	7	N/A	NCI-083-1	7.6	9.6
CAM-006-0	7.5	9.4	NCL-084-1	7.5	N/A
NCL-113-0	7.3	9.2	NCL-024-1	7.5	9.8
NCL-117-0	8.1	N/A	NCL-060-1	7.7	9.8
NCL-130-0	7.9	9.2	BIR-005-1	8.3	N/A
LEE-049-0	7.5	N/A	SUN-009-1	8.2	9.2

**N/A- RIN score not calculable by the Bioanalyzer software*

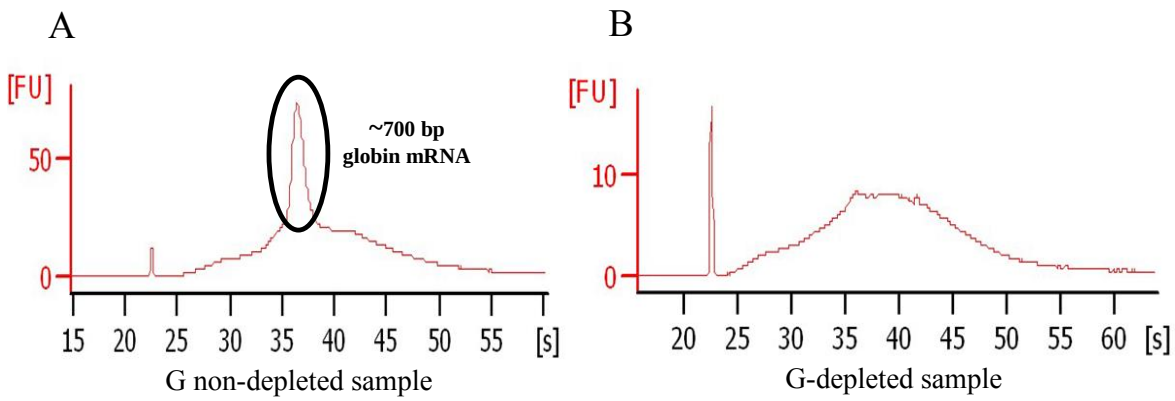


Figure 3.3 Representative electropherograms of amplified cRNA of a paired sample with or without globin mRNA depletion. *A) "G non-depleted" sample showing a sharp peak of about 700 base pair representing the globin mRNA, B) "G-depleted" sample shows a bell-shaped curve and the absence of the globin mRNA's sharp peak.*

3.3.4 The effect of globin mRNA depletion on microarray signal intensity and samples variability

The single signal intensity for each probe in the microarray was calculated by GenomeStudio software, after the Illumina BeadChips was scanned (Johnstone et al., 2013). The boxplots in Figure 3.4 illustrate how the probe signal intensity was distributed in the G-depleted and G non-depleted datasets. With regard to the raw data, the boxplots in the G non-depleted groups showed a more disproportionate spread and the mean of each array had a greater intra-group variability in comparison to the G-depleted groups. The normalised data (RSN method) showed a similar trend, with the G non-depleted samples showing a higher variability between individual arrays while the G-depleted group had a more homogeneous signal intensity across individual arrays. Furthermore, principle component analysis (PCA) of the normalised data of both groups revealed that the G-depleted samples clustered closer together comparing to the G-non depleted samples (Figure 3.5). When the samples were further sub-classified according to their disease status (i.e., pSS patients or healthy controls), again, G-depleted samples clustered closer together than the G non-depleted samples (Figure 3.6). Taken together, these observations suggest that the microarray data generated using G-depleted samples were of a better quality.

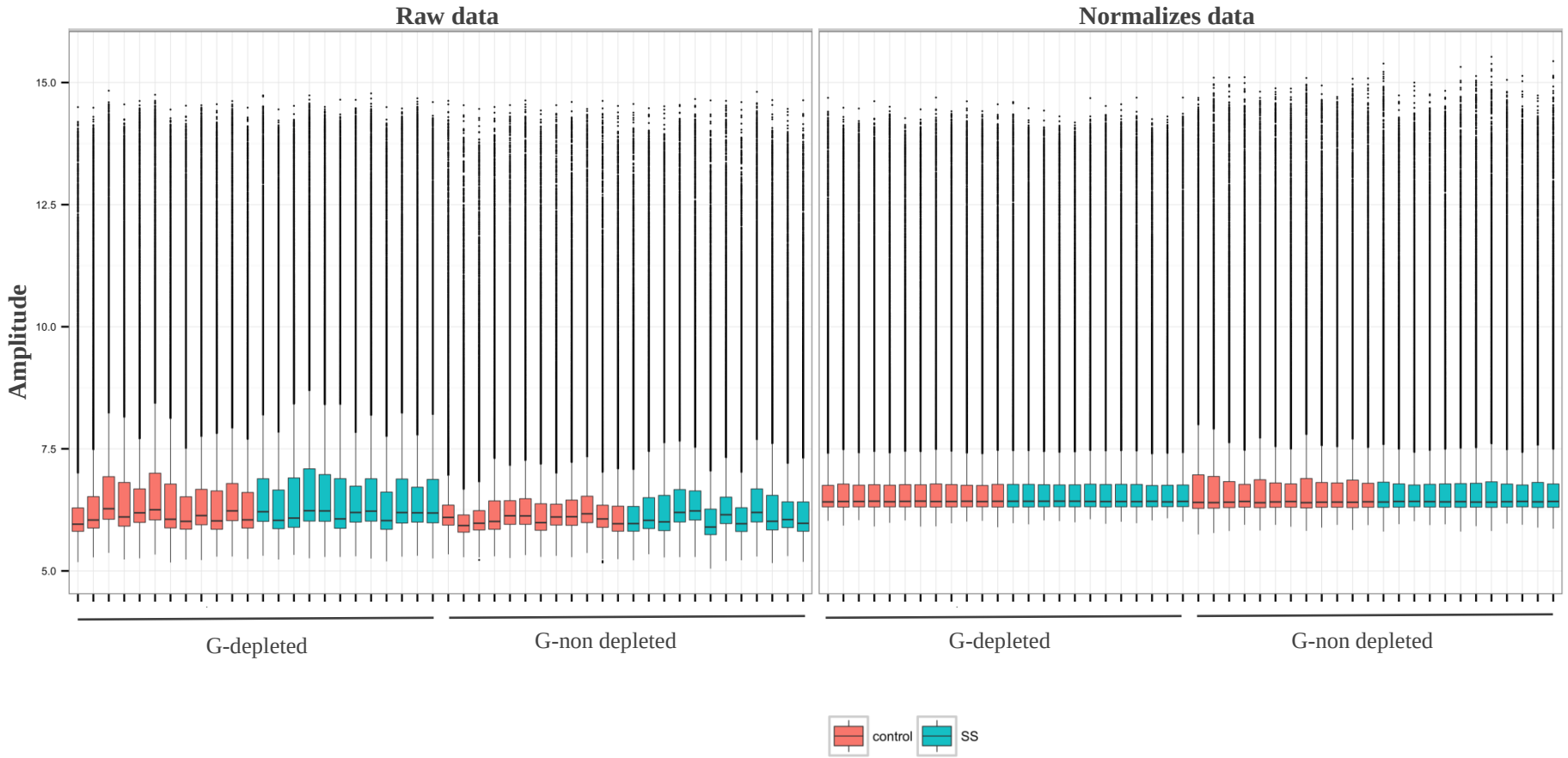


Figure 3.4 Boxplots of the microarray signal intensity of the paired samples with or without globin mRNA depletion. *Left: raw data, showing a great spread of the boxplots medians among the G-non depleted group for both pSS patients and HC in comparison to the G-depleted group. Right: normalised data (RSN method), the G-non depleted samples showed a slight variability unlike the G-depleted group, which revealed a more even signal intensity across individual microarrays.*

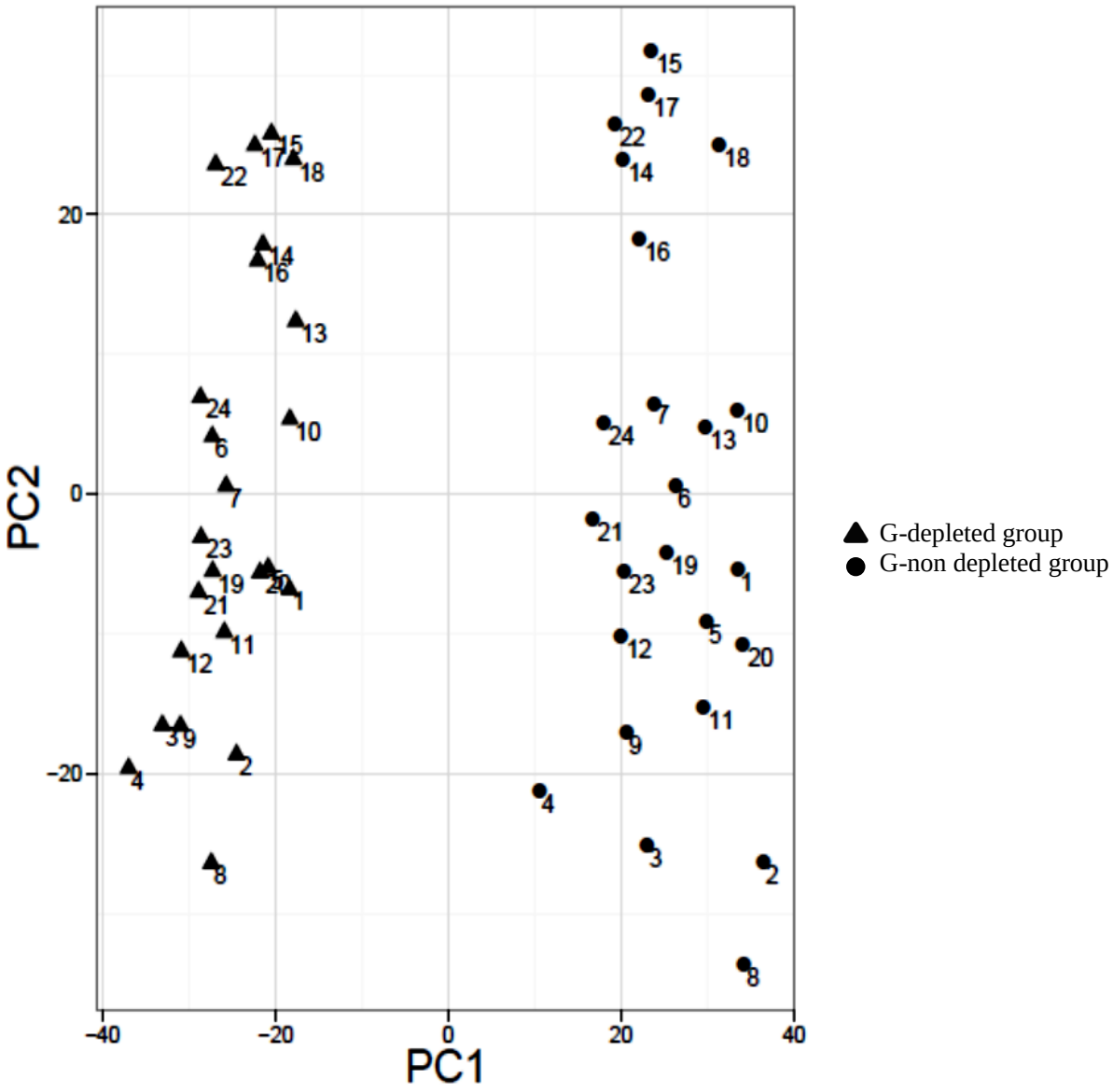


Figure 3.5 PCA of normalised microarray data with or without globin mRNA depletion. *The triangles represent the G depleted group whereas the G non-depleted group is represented by circles. Each paired samples have the same corresponding number showing its clustering location in their respective groups*

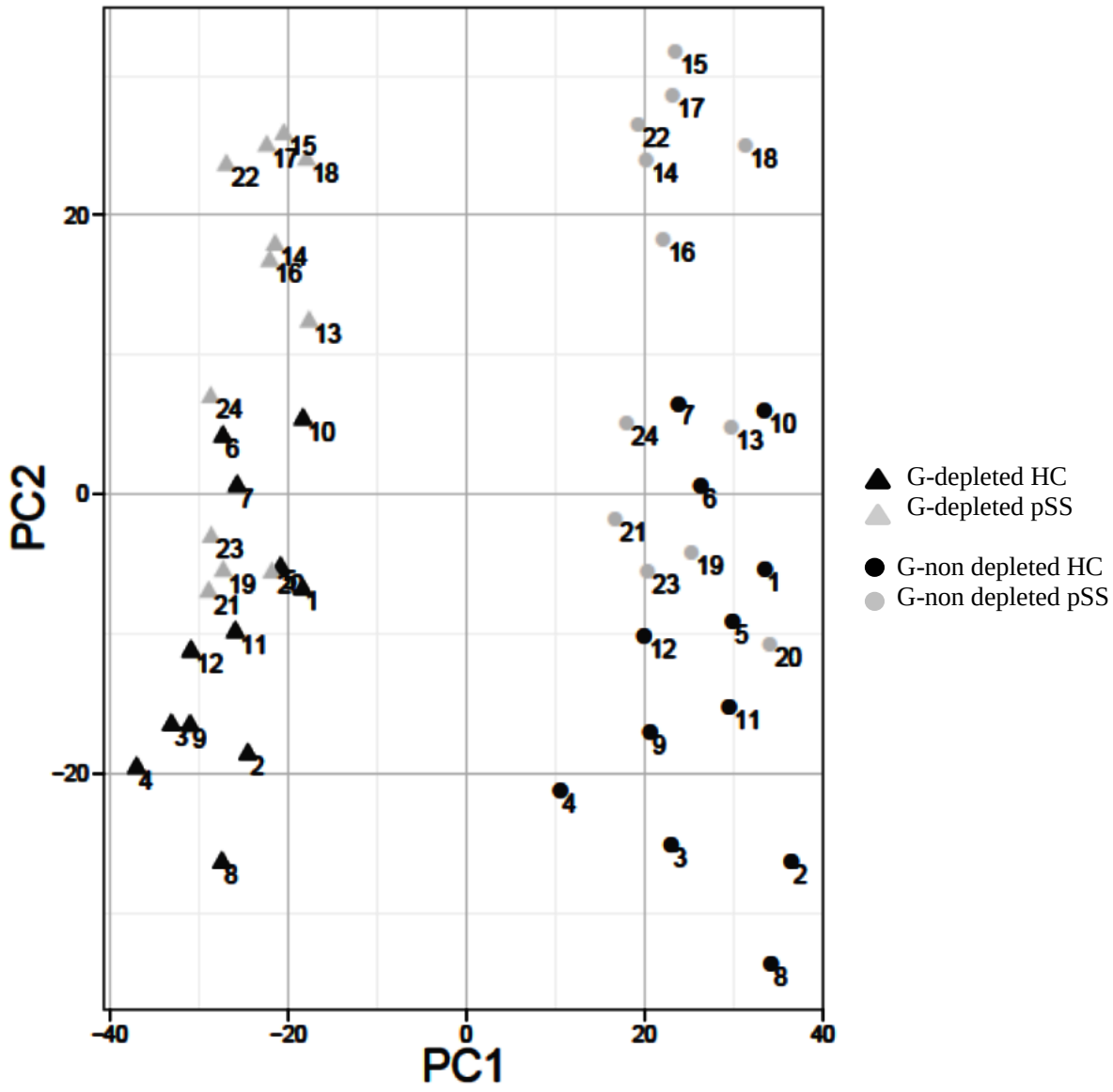


Figure 3.6 PCA of normalised microarray data of pSS patients and healthy controls with or without globin mRNA depletion . *Each paired sample has the same corresponding number showing its clustering location in their respective groups.*

3.3.5 Differentially expressed genes in pSS with or without globin mRNA depletion:

The gene expression profile in pSS patients was investigated. Analysis of the microarray data (Illumina HT-12v4) revealed a total of 24642 detected probes. Moreover, when the analysis was performed separately for the G-depleted and G non-depleted groups, there were 21929 and 17851 detected transcripts/entities respectively. These numbers represented the number of probes that had passed the basic filter step with a detection p-value threshold <0.01. A total of 17211 transcripts/entities overlapped between the two groups. There were 640 distinctive transcripts/entities in the G non-depleted group and 4718 distinctive transcripts/entities in the G-depleted group (Figure 3.7(A)).

There were also differences in the number of differentially-expressed genes (DEGs) (fold-change ≥ 1.2, adjusted $p < 0.05$) detected for pSS using G-depleted or G-non depleted samples. When using the G-depleted samples, a total of 1070 DEGs (733 up-regulated and 337 down-regulated DEGs) between pSS and HC were identified. When using the G non-depleted samples, 942 DEGs (670 up-regulated and 272 down-regulated DEGs) between pSS and HC were identified. Among the identified DEGs using these two groups of samples, 559 DEGs were in common (Figure 3.7(B)). The volcano plot in Figure 3.8 shows the DEGs using the G-depleted and G non-depleted samples. Among the up-regulated DEGs identified using the two groups of samples, 80% of the DEGs in the top 20 and top 50 of up-regulated genes for pSS were identical. However only 64% of the DEGs were the same in the top 100 up-regulated genes for the G-depleted and G-non-depleted groups. The IFN-inducible genes dominated the top 20 up-regulated DEGs regardless of whether G-depleted and G non-depleted samples were used, consistent with the published literature (Emamian et al., 2009). For the down-regulated genes between pSS and HC, however, there were more differences in the DEGs identified using the two groups of samples. Only 30%, 28%, 32% of the DEGs were the same in the top 20, top 50, and top 100 down-regulated genes respectively , The top 20 up-regulated genes are shown in Table 3.3 and the top 20 down-regulated genes are shown in Table 3.4.

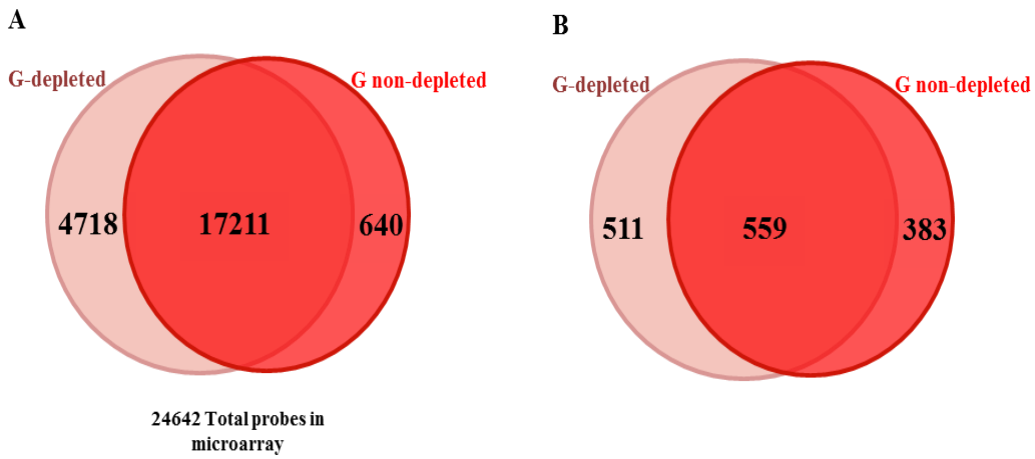


Figure 3.7 Venn diagram of the number of detectable transcripts/entities and DEGs identified using samples with or without globin mRNA depletion. A) *Venn diagram of the number of detectable transcripts/entities that had passed the detection threshold in the microarray analysis in the G-depleted and G non-depleted groups.* ***B)*** *Venn diagram of the number of differentially expressed genes in pSS identified using in G-depleted and G non-depleted groups samples*

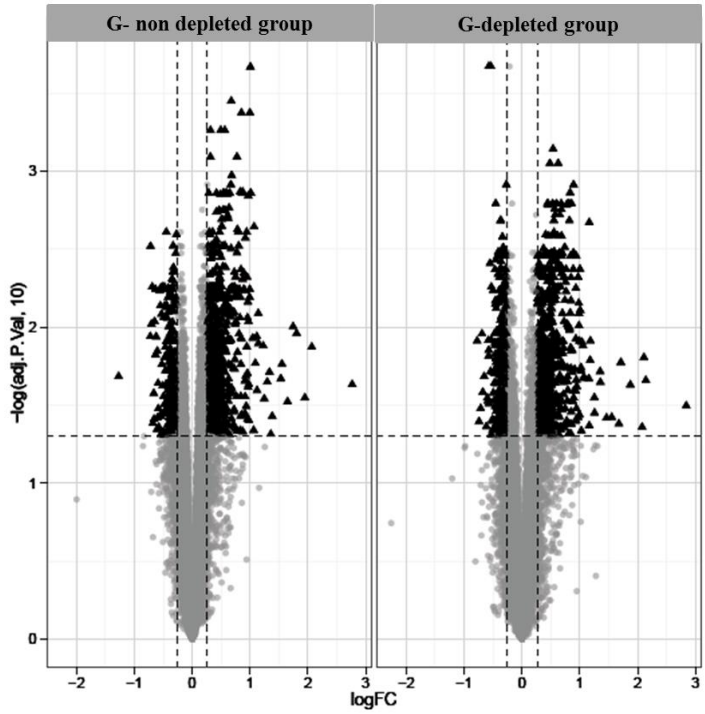


Figure 3.8 Volcano plots of the differentially expressed genes in pSS using samples with or without globin mRNA depletion.*The G-depleted (right) and G non-depleted (left) groups. The x axis represents the log2 of the fold change (FC) and the y axis represents the–log10 of the p-value; the dots in black represent the differentially expressed genes using the threshold values of FC ≥ 1.2, and adjusted p value < 0.05.*

Table 3.3 Top 20 up-regulated DEGs in pSS in the G-depleted and G non-depleted groups, p<0.05, FC≥1.2

G- depleted group			G non-depleted group		
Gene symbol	**adjusted P Value**	**Fold change**	**Gene symbol**	**adjusted P Value**	**Fold change**
IFI27	0.0319	7.170	*IFI27*	0.0233	6.802
IFIT1	0.0219	4.418	*IFIT1*	0.0134	4.192
IFI44	0.0156	4.317	*IFI44L*	0.0284	3.859
IFI44L	0.0437	4.224	*IFI44*	0.0110	3.496
RSAD2	0.0234	3.669	*EPSTI1*	0.0099	3.353
EPSTI1	0.0168	3.274	*ISG15*	0.0300	3.137
ISG15	0.0416	3.199	*RSAD2*	0.0172	2.932
OAS3	0.0378	2.936	*IFIT3*	0.0214	2.894
IFIT3	0.0382	2.757	*HERC5*	0.0373	2.610
IFIT3	0.0226	2.563	*IFITM3*	0.0486	2.567
IFIT2	0.0192	2.554	*IFIT2*	0.0194	2.525
OAS1	0.0355	2.389	*OAS3*	0.0225	2.459
IFIT3	0.0283	2.384	*IFI6*	0.0288	2.373
RPL26	0.0163	2.265	*S100A8*	0.0132	2.342
CHMP5	0.0021	2.243	*OAS1*	0.0127	2.214
OAS1	0.0428	2.242	*IFIT3*	0.0082	2.208
XAF1	0.0213	2.232	*XAF1*	0.0254	2.175
S100A8	0.0210	2.109	*RPL26*	0.0115	2.170
GBP1	0.0125	2.054	*OAS1*	0.0171	2.132
RPL31	0.0461	2.040	*TRIM22*	0.0023	2.101

Table 3.4 Top 20 down-regulated DEGs on pSS in the G-depleted and G non-depleted groups, p<0.05, FC≥1.2

G- depleted group			G non-depleted group		
Gene symbol	**adjusted P Value**	**Fold change**	**Gene symbol**	**adjusted P Value**	**Fold change**
GPR162	0.0123	1.710	*HLA-H*	0.0207	2.413
CLIC3	0.0403	1.671	*PABPC1*	0.0030	1.647
PTGDS	0.0332	1.638	*IMPA2*	0.0096	1.629
IMPA2	0.0110	1.612	*RPS28*	0.0055	1.615
NINJ2	0.0261	1.597	*RNA18S5*	0.0414	1.601
MATK	0.0207	1.571	*NINJ2*	0.0266	1.584
MPZL1	0.0122	1.495	*MARCH2*	0.0390	1.581
RYBP	0.0153	1.489	*RPL14*	0.0057	1.553
ZBTB16	0.0140	1.488	*GPR162*	0.0115	1.540
CABIN1	0.0002	1.487	*HLA-G*	0.0452	1.522
ZNF467	0.0309	1.484	*SIPA1L3*	0.0380	1.513
NCR3	0.0039	1.466	*MATK*	0.0437	1.492
CCNY	0.0002	1.449	*TUBA4A*	0.0056	1.455
DDIT4	0.0120	1.444	*SH3GLB2*	0.0082	1.444
VPS37C	0.0055	1.442	*CD7*	0.0356	1.443
MAPK8IP1	0.0490	1.442	*RPL23AP64*	0.0173	1.438
CD7	0.0399	1.427	*AMY1C*	0.0055	1.432
MARCH6	0.0390	1.423	*ABTB1*	0.0278	1.409
ARRB1	0.0058	1.411	*RPLP1*	0.0087	1.406
TBL1X	0.0264	1.408	*FTH1P2*	0.0361	1.405

3.3.6 Pathway analysis

I then examined whether the biological pathways identified using G-depleted and G non-depleted samples were similar. Using the KEGG pathway analysis, two pathways "Spliceosome" and "Oxidative phosphorylation" were identified to be differentially activated in pSS patients in both the G-depleted and G non-depleted datasets. Using the gene ontology analysis, 319 and 493 Gene Ontology (GO) terms from the G-depleted and G non-depleted datasets were identified to be associated with pSS. A total of 230 of these GO terms were identical between the two datasets. For the top 10 GO terms, 6 (60%) were in common between the G-depleted and G non-depleted datasets (Table 3.5). The top three GO terms for both datasets were "Antigen processing and presentation of exogenous peptide antigen via MHC class Ib," "Antigen processing and presentation of exogenous protein antigen via MHC class Ib, TAP-dependent" and "Detection of virus"

Table 3.5 Top 10 GO terms in pSS using samples from the G-depleted and G-non depleted groups

G- depleted group			G- non depleted group		
GO ID	**P value**	**Term**	**GO ID**	**P value**	**Term**
GO:0002477	0.00498	antigen processing and presentation of exogenous peptide antigen via MHC class Ib	GO:0002477	0.0046242	antigen processing and presentation of exogenous peptide antigen via MHC class Ib
GO:0002481	0.00498	antigen processing and presentation of exogenous protein antigen via MHC class Ib, TAP-dependent	GO:0002481	0.0046242	antigen processing and presentation of exogenous protein antigen via MHC class Ib, TAP-dependent
GO:0009597	0.00498	detection of virus	GO:0009597	0.0046242	detection of virus
GO:0033364	0.00498	mast cell secretory granule organization	GO:0032439	0.0046242	endosome localization
GO:0034343	0.00498	type III interferon production	GO:0033364	0.0046242	mast cell secretory granule organization
GO:0034344	0.00498	regulation of type III interferon production	GO:0034343	0.0046242	type III interferon production
GO:0036337	0.00498	Fas signaling pathway	GO:0034344	0.0046242	regulation of type III interferon production
GO:0044565	0.00498	dendritic cell proliferation	GO:0035616	0.0046242	histone H2B conserved C-terminal lysine deubiquitination
GO:1902044	0.00498	regulation of Fas signaling pathway	GO:0035726	0.0046242	common myeloid progenitor cell proliferation
GO:0002428	0.01423	antigen processing and presentation of peptide antigen via MHC class Ib	GO:0036257	0.004624	multivesicular body organization

3.4 Discussion

The aim of this experiment was to evaluate the effect of globin mRNA depletion on whole blood gene expression profiling in pSS. The data generated helped me to determine the approach that I should use to identify the whole blood gene expression signature in pSS-associated lymphoma. The reasons for choosing pSS gene expression profiling instead of pSS-associated lymphoma gene expression profiling was two-fold. First, whole blood gene expression data on pSS are available in the literature as reference data for my experiment, while there is no such data available for pSS-associated lymphoma. Second, only a limited amount of whole blood RNA was available from the pSS-associated lymphoma group.

Whole blood RNA samples

The use of whole blood samples might increase noise and decrease the responsiveness in microarray experiments (Feezor et al., 2004). The use of PAXgene blood RNA tubes and their stabilising reagents helps to stabilise the transcriptome once the specimen has been taken, and minimises RNA degradation for a long period of time (Rainen et al., 2002). Therefore, this blood collecting system is suitable for gene expression studies using RT-PCR or microarrays (Stordeur et al., 2003, Thach et al., 2003). In this study, most of the RNA samples extracted from PAXgene tubes required additional "clean up" and "concentration" steps before globin mRNA depletion with the GLOBINclear kit. The need for these extra preparatory steps has also been reported (Liu et al., 2006) and was thought to be related to the different incubation times of the PAXgene tubes resulting in different RNA yields (Wang et al., 2004).

RNA quality

The RIN scores (thus the RNA quality) of my samples decreased after globin mRNA depletion. This observation is not unexpected and is in agreement with previous reports (Mastrokolias et al., 2012, Choi et al., 2014). The electropherogram of the cRNAs demonstrating the lack of a globin mRNA peak in my globin-depleted samples is also consistent with previous studies (Liu et al., 2006, Vartanian et al., 2009). The efficiency of the novel hybridisation capture technology used in the GLOBINclear kit as a globin reduction protocol has previously been evaluated for next generation sequencing (Mastrokolias et al., 2012). It is believed that the design of these oligonucleotides is a

crucial factor for the effectiveness in globin removal (Affymetrix, 2003). Regardless, the efficient depletion of β-globin mRNA in my G-depleted samples was satisfying.

Microarray data

The improvement of the microarray signal intensity and the reduction of signal intensity variability among the samples following globin mRNA depletion is consistent with previous reports using samples from humans, mice and rats (Liu et al., 2006, Whitley P, 2007, Whitley P, 2005, Winn et al., 2010). In addition, the principle component analysis (PCA) illustrated a reduction in the overall variability within the respective subject groups (pSS patients and healthy controls) in the 'G-depleted' dataset. This observation suggests that globin mRNA reduction improved sample homogeneity. Furthermore, the number of detectable transcripts/entities (present call) were higher in the G-depleted dataset, which also has been reported previously (Affymetrix, 2003).

With regard to the ability to detect DEGs in pSS, more DEGs were identified using the G-depleted samples than using G non-depleted samples. However, it may not be a simple case of more DEGs being detected in the G-depleted datasets, since only approximately half of the DEGs identified were in common using both G-depleted and G non-depleted samples. The difference was more apparent for down-regulated genes. One possible explanation is that globin depletion may be more efficient in unmasking less-abundant transcripts (Liu et al., 2006). Nevertheless, for both G-depleted and G non-depleted datasets, the up-regulated genes showed a dominant pattern of IFN-inducible genes in the top 20 up-regulated genes, consistent with the hypothesis that an activated IFN signature is a hallmark of pSS (Brkic and Versnel, 2014).

Pathway analysis

Pathway analysis showed that similar pathways were enriched in both G-depleted and G non-depleted datasets and 60% of the top 10 GO terms was in common. Interestingly, the top three GO terms were related to immune response to viruses; this finding is consistent with the hypothesis that viruses (such as EBV, CMV, retroviruses) may play a role in the pathogenesis of pSS (Venables and Rigby, 1997, Willoughby et al., 2002).

In summary, globin mRNA depletion improved microarray signal intensity, and reduced sample variability. Regarding its impact on gene expression profiling, globin mRNA depletion resulted in the detection of more DEGs, but had relatively little impact on the DEGs with the highest fold changes. Globin mRNA depletion also appears to have a small impact on the outcome of pathway analysis of the microarray data. Based on the data from this study, I believe that globin mRNA depletion may offer a slight advantage in my subsequent whole blood gene expression study of pSS-associated lymphoma.

Chapter 4

Identification of Whole Blood Gene Expression Signature in primary Sjögren's Syndrome -associated lymphoma

4.1 Introduction

I. Primary Sjögren's Syndrome (pSS)-associated lymphoma

Lymphoma is one of the most serious extraglandular manifestations of pSS. It was first reported in pSS patients in 1963 (Talal and Bunim, 1964). The prevalence of lymphoma in pSS patients is approximately 5%. It has been reported that pSS patients have a 44-fold increased risk of developing non-Hodgkin lymphoma (NHL) (Kassan et al., 1978), although another study has reported a 16-fold increased risk to develop NHL (Theander et al., 2006). A meta-analysis investigating the association of NHL with autoimmune diseases showed a smaller SIR of 4.9 in pSS patients (Fallah et al., 2014). A summary of the studies regarding lymphoma prevalence in pSS has also been provided in the introduction chapter (Table 1.4). It has been documented that the most common type of pSS-associated lymphoma is the mucosa-associated lymphoid tissue (MALT) lymphoma followed by the nodal marginal zone lymphomas (NMZLs), and then the diffuse large B-cell lymphomas (DLBCLs), based on a study of 53 pSS patients with lymphoma from a cohort of 584 pSS patients (Voulgarelis et al., 2012). As lymphoma is a potentially fatal manifestation of pSS, the determination of lymphoma's risk factors provides a useful background to understand its evolution. Several risk factors have been linked to the increased risk of lymphoma development among pSS patients. These risk factors include: low C4 levels in the blood, cryoglobulinaemia (De Vita et al., 2012), paraproteinemia, persistent salivary gland swelling, and palpable purpura (Skopouli et al., 2000). Reduced ratio of $CD4^+/CD8^+$ T cells and a low count of $CD4^+$ T cells in peripheral blood are also associated with an increased risk; in this case, it has been speculated that the infiltration of $CD4^+$ into the glandular tissue results in reduced $CD4^+$ T cell number and reduced $CD4^+/CD8^+$ ratio in the blood (Pillemer, 2006). A recent paper from France reports that rheumatoid factor is also a risk factor (Nocturne et al., 2015c).

Few studies have been carried out investigating gene expression in pSS-associated lymphoma. These studies were limited by small sample size. As mentioned in the introduction chapter, one of the studies focused on the gene expression of MALT lymphoma in the parotid gland of pSS patients has identified eight candidate genes (*GRB2, ARHGDIB, CD40, PSMB9, ALDOA, PRDXS, PARC* and *PPIA)* that can differentiate pSS-associated MALT lymphoma from pSS (Hu et al., 2009). Another study analysed the differentially expressed genes in the whole blood of pSS-associated lymphoma (marginal zone B cell lymphoma) patients using cDNA microarray; it showed that the ribosomal protein genes S29 and S27 are upregulated in pSS patients with lymphoma, while the IFN-inducible genes, which are extensively overexpressed in pSS, were not. Interestingly, subsequently to chemotherapy and rituximab, the ribosomal protein genes S29 and S27 were downregulated; this observation suggests that S29 and S27 may serve as biomarkers of pSS-related lymphoma (Kimoto et al., 2011, Ogawa, 2010). However, the molecular mechanisms that are important in the development of lymphoma remain poorly understood.

II. Gene expression signature by microarray

Recently, microarray technology has become a common tool to investigate gene expression. A microarray measures the expression of thousands of genes, giving the opportunity to use this huge source of information in different fields. This technology contributes widely in cancer research (Branca, 2003, Jeruss et al., 2008), disease treatments, and the identification of new biomarkers. The bead chip technology used in the Illumina Human HT-12 v4 Expression BeadChip utilises 3-micron silica beads that aggregate on either a planar silica slides or a fiber optic bundle, as illustrated in Figure 4.1. The BeadChip array is a more desirable technique as it requires less sample input and it is cheaper than other available platforms (Consortium, 2006). These beads are characterised by their ability of random self-assembly into microwells. This assemblage happens via van der Waals forces as well as the hydrostatic interaction of the beads and the well's wall. The beads are covered with hundreds of thousands of specific copies of oligonucleotides, or probes, that capture particular sequences in the assay (www.illumina.com). Because of the random aggregation of the beads, the probes will be randomly distributed in different numbers and locations in the array. The analysis process includes the identification of the

bead locations, bead types, and the microarray intensity. Subsequently, images were obtained from the scanner, and a pre-processing procedure was followed. The pre-processing method consists of three steps: registration of beads, insertion of the lasting beads, and pinpointing the grid above the array (Gunderson et al., 2004, Arteaga-Salas et al., 2008). Eventually, the probes on each bead were identified. The microarray data was corrected for background noise, which was produced by the hybridisation of nonspecific transcripts in the samples. As a result, the foreground, which is the mean of sharpened intensities from a 4 × 4 pixel square on the bead-centre, is separated from the background (Tarca et al., 2006, Wu, 2009, Smith et al., 2010).

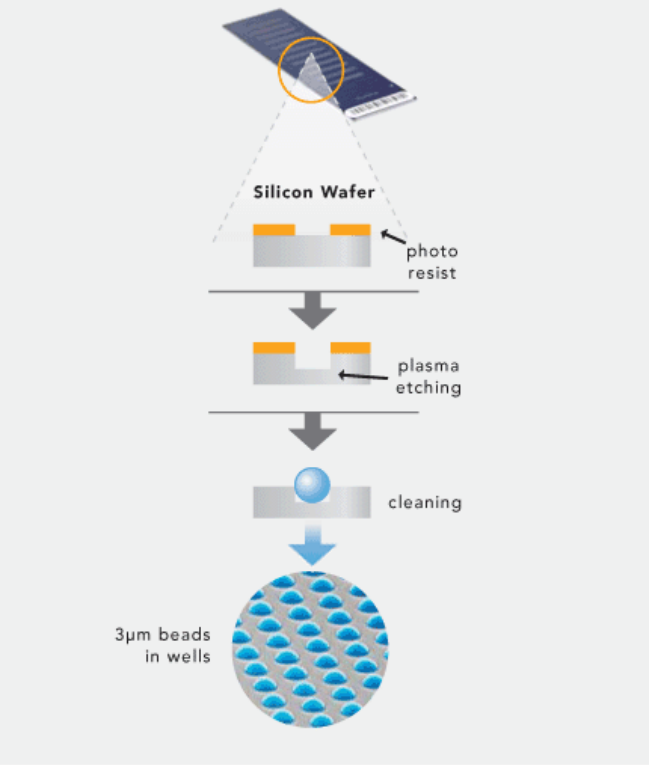


Figure 4.1 Illumina BeadChip technology. *This is the technology conducted by Illumina whole genome microarray.* (www.illumina.com)

The next step in microarray data analysis is transformation of the data into a log scale. This step is important as the data might be skewed and the transformation might bring it closer

to a normal distribution. The data is transformed either using Log_2 scale or variance-stabilising transformation (VST). This latter has been reported to be more efficient than the Log_2 transformation when using the Illumina platform (Lin et al., 2008). The selected normalisation method for the microarray data is crucial for further downstream analysis (Schmid et al., 2010). There are several normalisation methods that can be used to minimise the data variability that originates from noise and artifacts. The method that I used to analyse my data is the robust spline normalisation (RSN). This algorithm has the features of both the quantile and LOESS (Local regression) normalisation. In addition, this function, which is included in the "Lumi" package, will run the lumiT function if the data is not variance-stabilised, as it is designed to normalise VST-transformed data (Lin et al., 2008).

Conducting a large gene expression experiment involves many challenges. One such challenge is *batch effects*. Batch effects are derived from miniature differences in non-biological factors, such as the use of different batches of reagents in sample preparation or minor differences in sample handling by different researchers and many others. It has even been reported that ozone levels have an impact on the microarray data (Fare et al., 2003, Chen et al., 2011, Luo et al., 2010). Many methods were introduced in order to minimise batch effects in order to obtain an accurate gene expression dataset for downstream analysis. One of these methods is *ComBat* function. ComBat is an empirical Bayes framework that effectively eliminates batch effects. The method works by "borrowing information" from across the probes and experimental conditions in hope that the "borrowed information" will lead to better estimates or more stable inferences. (Johnson et al., 2007, Stein et al., 2015). In comparison with other batch effect removal tools, ComBat is more effective in eliminating batch effects and producing more accurate microarray data (Chen et al., 2011). The final step in microarray analysis was to identify a list of differentially expressed genes and/or biological pathways between the comparative groups. Such differentially expressed genes might be valuable in biomarker discovery and the identification of key genes or biological pathways in the pathogenesis of the disease being studied.

III. Validation of microarray data by Real-Time RT-PCR

High-throughput experiments, such as microarray, might be affected by the complexity of the assay. This complexity is reflected in the manufacturing of the array and the preparation procedures of the samples (Eisen and Brown, 1999). For this reason, it is important to use another method to confirm the findings of differentially expressed genes identified in microarray experiments. Real-Time RT-PCR is a standard technique in quantitative measurements of mRNA. RT-PCR has many advantages: it is inexpensive, not time-consuming, and requires only a small amount of mRNA (Rajeevan et al., 2001). In this project, TaqMan® real-time PCR primers and probes were used in the validation of my potential gene expression signature in pSS-associated lymphoma identified from microarray. The TaqMan® Gene Expression Assays kit includes: one pair of unlabeled PCR primers, TaqMan® probe that is labeled with FAM^{TM} or VIC$^{®}$ dye, and the third component is the minor groove binder (MGB) and non-fluorescent quencher (NFQ) on the 3' end. The MGB presence in the assay is very significant as it increases the probe's melting temperature (T_m), leading to the stabilisation of the probe-target complex. This advantage enables the probes to be shorter than the outdated probes, and it also allows more specific detection targets by using probe sequences that are more specific to the target mRNA (www.lifetechnologies.com, 2012). Figure 4.2 illustrates the TaqMan® gene expression assay technology.

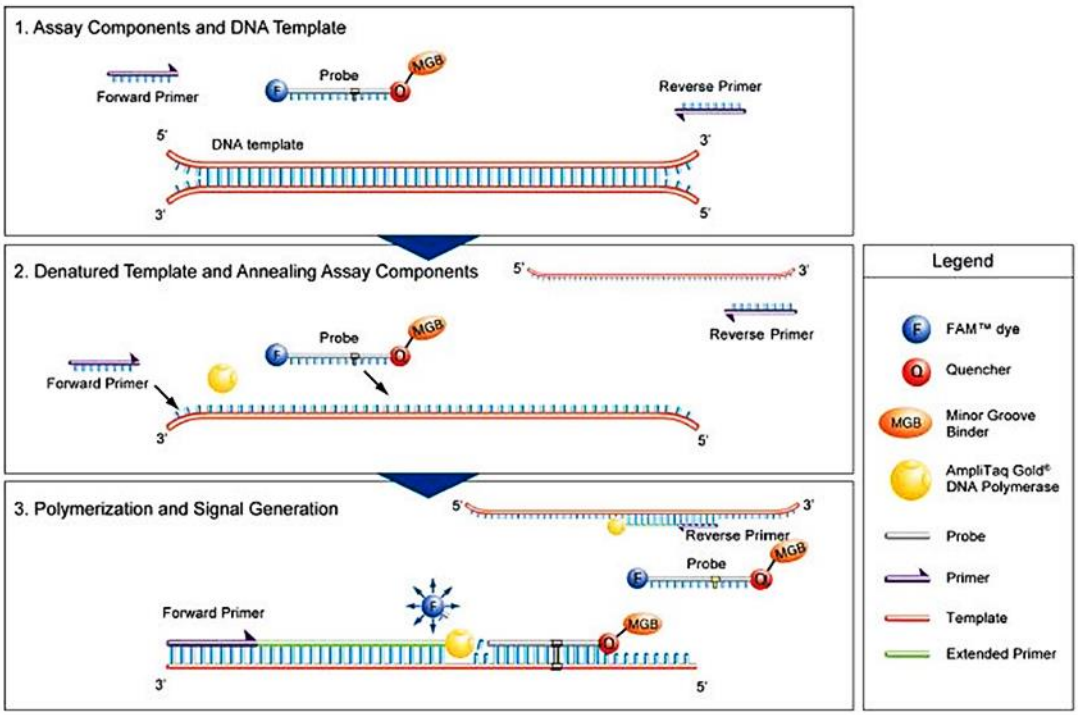


Figure 4.2 The TaqMan gene expression assay reaction steps.*(taken from www.lifetechnologies.com)*

TaqMan® Array 96-Well Fast plates were used to perform the validation experiment. The advantages of this method are speed and the requirement of only small amounts of mRNA. Therefore, this method is well suited for a large experiment with many samples and genes to validate, such as mine. Within each of the plates, the predesigned TaqMan® Gene Expression Assays were dried down into the wells of the plates. The assays work with the pre-prepared cDNA of the sample, and the expression data for a given gene is measured (www.lifetechnologies.com, 2011).

IV. Prediction models in pSS-associated lymphoma

After the identification and technical validation of the whole blood potential gene expression signature in pSS-associated lymphoma, it was useful to find out which genes are the most important in predicting the membership in pSS-associated lymphoma. A *prediction model* is a collection of mathematical techniques that aims to identify a mathematical relationship between the genes of interest and the phenotype. When using prediction models, it is important to keep in mind that they are usually imperfect; and

therefore it is important to include with the prediction a measurement of the percentage of uncertainty (Dickey, 2012).

4.2 Aim and experimental design

The ultimate aim of the project is to identify a whole blood gene expression signature in pSS-associated lymphoma. In order to achieve this goal, 144 samples were selected from the UKPSSR (Discovery cohort), and categorised into five groups. The group of interest is the pSS-associated lymphoma group (n=16) that will be used for comparison with the non-lymphoma pSS group (n=61), which represents the disease controls. Additional controls/comparison groups were also included. The pSS-paraproteinemia group (n=23) was included to represent an "at risk" group, as paraproteins play an important role in many malignant diseases (Cook and Macdonald, 2007) and are a risk factor for lymphoma development in pSS. The pSS-other cancers group (n=21) was included to explore whether any of the altered gene expression detected in the pSS-lymphoma group is associated with malignancies 'in general' or is specific to lymphoma. Finally, a healthy control group (n=23) was also included.

My first experiment is to perform a whole genome gene expression microarray. The microarray was performed by CGS after I extracted the RNA samples and depleted the globin mRNA. Data from the microarray were used to identify the differentially expressed genes between the lymphoma and the non-lymphoma groups. These differentially expressed genes were technically validated using qRT-PCR. Differentially expressed genes that have been confirmed using qRT-PCR were further validated using an independent set of samples (Validation cohort). In the independent cohort, I have also included a group of patients with pSS-associated lymphoma but have not yet received any treatment for their lymphoma (untreated pSS-associated lymphoma). Using the data from the RT-PCR of the Discovery cohort, I have tested the biological 'signature' based on these differentially expressed genes using a prediction modeling fit. Figure 4.3 summarises the experimental design.

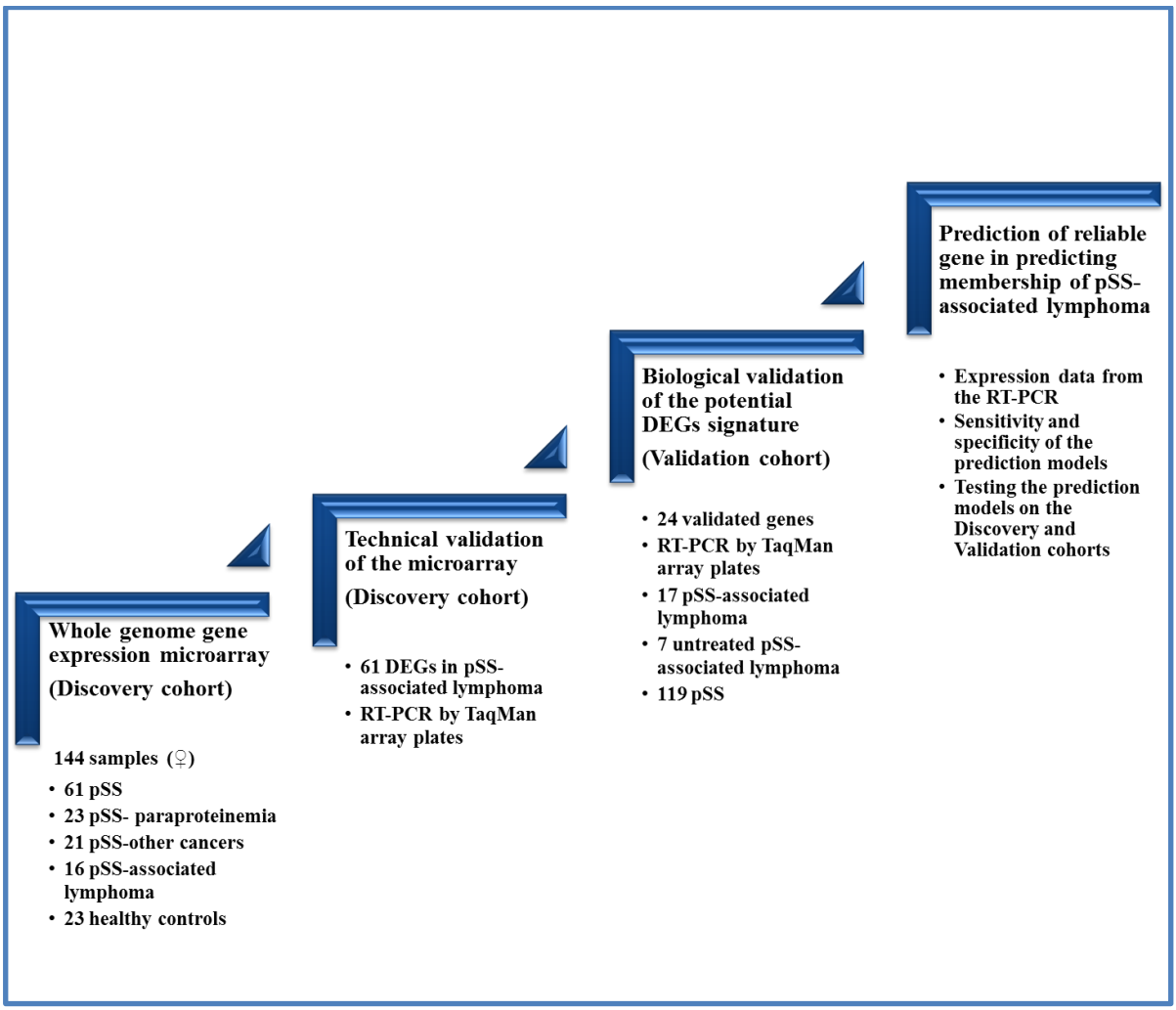


Figure 4.3 The identification of whole blood gene expression signature in pSS-associated lymphoma experimental design

4.4 Results

4.4.1 Demography of study subjects

All selected pSS patients used for identifying the gene expression signature in pSS-associated lymphoma fulfil the AECG classification criteria. The clinical parameters of the subjects in the Discovery cohort, Validation cohort, and untreated lymphoma are illustrated in Tables 4.1 and 4.2, respectively. In the Discovery cohort, after applying the Mann-Whitney test on the data, the results showed that the white cell count was significantly lower in the "pSS-associated lymphoma" group in comparison with the "pSS" group (p value = 0.040). In addition, a significantly lower level of C4 in the "pSS-associated

lymphoma" group was observed in comparison with the "pSS" and "pSS-other cancers" groups (p value = 0.0029 and 0.0027, respectively). The titre of IgG was significantly higher in the "pSS-associated lymphoma' group in comparison with the "pSS-other cancer" group (p value = 0.038). SSDDI scores were statistically different among all of the groups (p value = 0.000).

In the Validation cohort, after applying the Mann-Whitney test on the data (only from the samples from the UKPSSR biobank), the results showed a significantly lower C4 level in the "pSS-associated lymphoma" group in comparison with "pSS" (p = 0.0199). Significant differences were also showed in the SSDDI and ESSDAI between the "pSS-associated lymphoma" and the "pSS" groups with p values of 0.0026 and 0.0266 respectively.

The "pSS-associated lymphoma" group consists of patients who had different histological types of lymphomas. The type of lymphoma for each patient is listed in Table 4.3. In addition, all patients in the lymphoma group from the Discovery cohort had received treatment for their lymphoma at the time of blood sampling.

Table 4.1 Clinical data of pSS subgroups subjects in the Discovery cohort

Clinical criteria	**Mean ± S.E.M**				
	pSS	**pSS-paraproteinemia**	**pSS-other cancers**	**pSS-associated lymphoma**	**Healthy control**
Age (years)	60 ± 1.43	64 ± 2.73	66 ± 1.5	62 ± 3.33	57 ± 2.9
Unstimulated oral salivary flow (ml/15 mins)	0.38 ± 0.08	0.63 ± 0.25	0.45 ± 0.14	0.18 ± 0.08	N/A
Schirmer's test (mm/5 mins)	3.07 ± 0.85	4.39 ± 1.17	4.76 ± 1.31	3.53 ± 1.43	N/A
Anti-Ro/SSA positive (%)	93.4%	86.9%	80.9%	93.7%	N/A
Anti-La/SSB positive (%)	83.6%	73.9%	57.1%	81.2%	N/A
WCC (x10^9/l)	5.79 ± 0.29	5.27 ± 0.40	5.49 ± 0.39	**4.37 ± 0.43**	N/A
Neutrophil (x10^9/l)	3.65 ± 0.22	3.21 ± 0.28	3.50 ± 0.29	3.07 ± 0.35	N/A
Lymphocytes (x10^9/l)	1.55 ± 0.08	2.31 ± 0.80	1.38 ± 0.11	1.23 ± 0.16	N/A
ESR (mm/hr)	32.11 ± 3.30	36.30 ± 6.20	23.81 ± 5.53	39.63 ± 7.08	N/A
IgG (g/L)	17.65 ± 0.98	15.64 ± 1.59	13.88 ± 1.75	**17.39 ± 1.96**	N/A
C3 (mg/dl)	1.17 ± 0.04	1.17 ± 0.04	1.22 ± 0.10	1.10 ± 0.09	N/A
C4 (mg/dl)	0.23 ± 0.02	0.16 ± 0.02	0.22 ± 0.02	**0.14 ± 0.02**	N/A
CRP (mg/L)	3.87 ± 0.56	3.60 ± 0.62	6.39 ± 2.22	3.78 ± 0.90	N/A
RF (IU)	56.45 ± 7.64	63.17 ± 14.13	42.38 ± 9.67	133.87 ± 72.05	N/A
Total SSDDI	4.21 ± 0.20	4.61 ± 0.42	4.24 ± 0.41	**9.06 ± 0.41**	N/A
ESSPRI (0-10)	4.86 ± 0.28	5.41 ± 0.46	5.48 ± 0.52	5.92 ± 0.79	N/A
ESSDAI (0-123)	6.16 ± 0.92	4.61 ± 0.95	5.43 ± 1.02	6.06 ± 1.89	N/A

WCC= White Cell Count, ESR= Erythrocyte Sedimentation Rate, C3= Complement Component 3, C4= Complement Component 4, CRP=C-reactive Protein, RF=Rheumatoid Factor, IU= International Units, ESSDAI score=(0-4 low activity, 5-12 moderate activity and ≥ 13 high activity), parameters in bold depict those values that were statistically significantly different to other subject groups when tested using Mann-Whitney test.

Table 4.2 Clinical data of pSS and pSS-associated lymphoma subjects in the Validation cohort

Clinical criteria	**Mean ± S.E.M**	
	pSS (n = 119)	**pSS-associated lymphoma (n = 14)**
Age (years)	57 ± 1.11	60 ± 4.65
Unstimulated oral salivary flow (ml/15 mins)	0.78 ± 0.15	0.59 ± 0.25
Schirmer's test (mm/5 mins)	6.05 ± 0.69	7.14 ± 3.36
Anti-Ro/SSA positive (%)	88.2%	86.6%
Anti-La/SSB positive (%)	64.7%	64.2%
WCC (x10^9/l)	5.33 ± 0.17	5.54 ± 0.35
Neutrophil (x10^9/l)	3.73 ± 0.51	3.55 ± 0.35
Lymphocytes (x10^9/l)	1.73 ± 0.28	1.43 ± 0.18
ESR (mm/hr)	25.39 ± 1.99	23.50 ± 6.76
IgG (g/L)	18.18 ± 1.85	13.96 ± 1.32
C3 (mg/dl)*	19.33 ± 12.97	1.16 ± 0.13
C4 (mg/dl)	0.29 ± 0.07	**0.16 ± 0.02**
CRP (mg/L)*	3.15 ± 0.30	4.87 ± 1.43
RF (IU)*	64.15 ± 11.95	68.71 ± 23.07
Total SSDDI*	3.32 ± 0.17	**6.43 ± 1.09**
ESSPRI (0-10)*	5.02 ± 0.25	6.57 ± 0.52
ESSDAI (0-123)	4.10 ± 0.44	**6.79 ± 1.32**

** number of pSS-associated lymphoma samples = 7, WCC= White Cell Count, ESR= Erythrocyte Sedimentation Rate, C3= Complement Component 3, C4= Complement Component 4, CRP=C-reactive Protein, RF=Rheumatoid Factor, IU= International Units, ESSDAI score=(0-4 low activity, 5-12 moderate activity and ≥ 13 high activity), parameters in bold are those values that were statistically significantly different between groups when tested using the Mann-Whitney test.*

Table 4.3 Histological types of the subjects in the two pSS-associated lymphoma cohorts

Type of lymphoma	Discovery cohort	Validation cohort	
		treated pSS-associated lymphoma	Untreated pSS-associated lymphoma
Non- Hodgkin's lymphoma			
DLBCL	-	1	-
Low grade follicular lymphoma	-	-	1
Low grade MALT lymphoma	-	1	-
MALT lymphoma	7	13	5
MZL	2	-	-
Not specific	3	1	-
Parotid lymphoma	1	-	1
Periorbital lymphoma	1	-	-
T cell/ histiocyte rich b cell lymphoma	-	1	-
Hodgkin's lymphoma			
Hodgkin's lymphoma	2	-	-
Total number of samples	**16**	**17**	**7**

DLBCL = Diffused Large B-cell Lymphoma, MALT = mucosa-associated lymphoid tissue, MZL = Marginal Zone lymphoma

4.4.2 RNA quality

Total RNA from patients and healthy controls were extracted from the PAXgene tubes, and the concentration of each sample was measured initially by Nano-drop. The quality and the concentration of the samples were also checked using an Agilent 2100 Bioanalyzer at the CGS. Generally, samples with RIN above 7 are considered achieving the quality standards for microarray experiments. A total of 117 RNA samples out of 144 were of good quality, with an RIN score that ranged from 7 to 8.8. Three RNA samples from the pSS group have N/A RIN values, but the samples were subsequently inspected visually by the staff at CGS and were considered as being of good quality (by comparing them with the profiling of a good-quality sample). Twenty-four samples have an RIN score < 7, of which 18 were considered being of 'medium' quality, with RIN scores ranging between 5 and 7. These 18 samples were from 2 healthy controls, 5 pSS, 3 pSS-paraproteinemia, 6 pSS-other cancers and 2 pSS-associated lymphoma. The remaining 6 samples were of low quality with RIN

score < 5; these 6 samples were from 1 healthy control, 1 pSS-other cancers, and 4 pSS patients. The distribution of the samples according to their RIN scores is shown in Figure 4.4. The amplification cRNA for all the samples showed the lack of a sharp peak of ~700 bp that represents the globin mRNA, indicating that globin mRNA had been effectively removed.

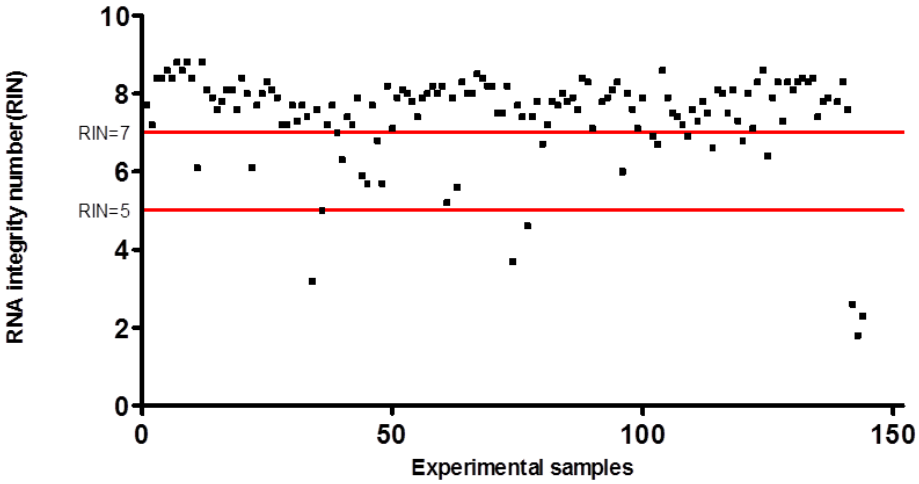


Figure 4.4 RNA integrity number (RIN) of the Discovery cohort. *The black squares represent each sample, and the horizontal red lines mark out RIN score =5 and RIN score =7 cutoffs.*

4.4.3 Whole genome gene expression Bead Chip of pSS-associated lymphoma

1. Microarray data analysis

The microarray data analysis was performed using R packages. The data were transformed and normalised using the RSN method, followed by quality control analysis. These include removal of technical outliers that were detected during the analysis by Principal Component Analysis (PCA). The reason behind removing the technical outliers is that they might affect the downstream analysis of the data because they have dissimilar behaviours from the other samples in the data. The RNA quality of the samples was also taken into consideration.

Since the sample size of the pSS-associated lymphoma group is the smallest among the five subject groups, all samples from the four different pSS subgroups with RIN score < 7 were

removed, but the two pSS-associated lymphoma samples with RIN score < 7 were kept during the analysis. The reason for keeping these two lymphoma samples is that their RIN score were more than 5, which is still considered acceptable for microarray experiments. Furthermore, excluding those 2 samples would have substantially reduced the power of the study due to the resultant smaller sample size of the lymphoma group. All technical outliers were removed from the analysis (including a pSS-associated lymphoma sample). This analysis will be referred to as Analysis A later in the chapter, and all of the figures and tables presented are generated using this analysis unless stated otherwise.

2. Batch effect

A large experiment such as this one might be affected by what is known as *batch effects*. PCAs were used to investigate these potential batch effects. PCA prior to batch correction showed batch effects that separate the samples into 3 groups, where each group represents a number of samples that were processed and scanned at a specific time, which differs from the time that the other two groups were processed and scanned. These differences created a non-biological variation leading to the batch effects as shown in Figure 4.5. Following batch correction, which was achieved by applying the ComBat correction method, the PCA showed that all of the samples were distributed more evenly as one group, indicating effective elimination of batch effects (Figure 4.6).

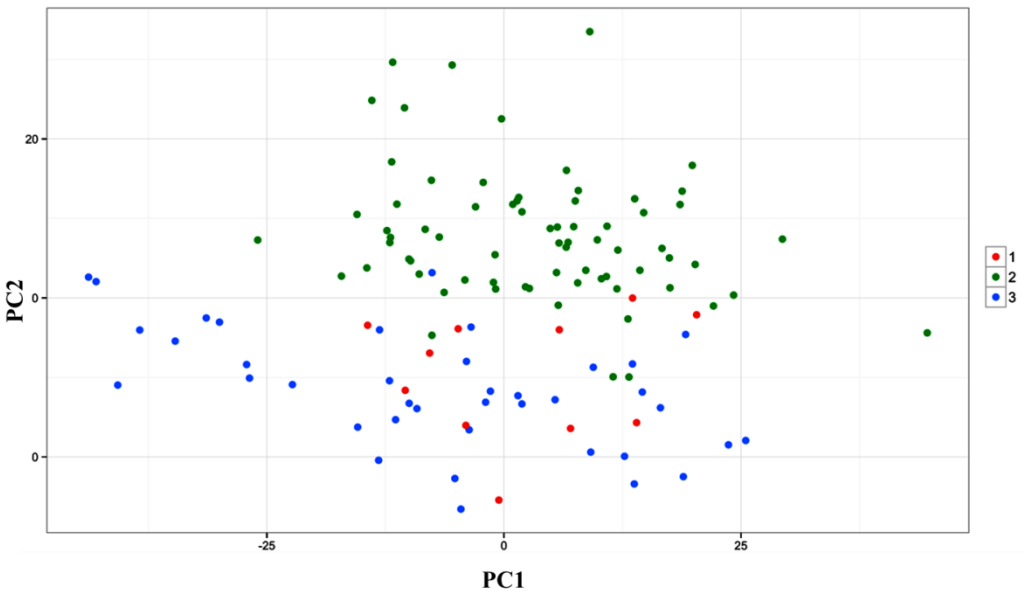


Figure 4.5 Batch effects of the microarray data of the Discovery cohort. *Principal component analysis (PCA) showed that batch effects were present in the normalised gene expression microarray data. Three color coded groups of samples (red (1), green (2) and blue(3)) were clustered separately due to non-biological variation caused by differences in the time of sample processing and scanning. Each group is represented by a unique color to distinguish the clustering.*

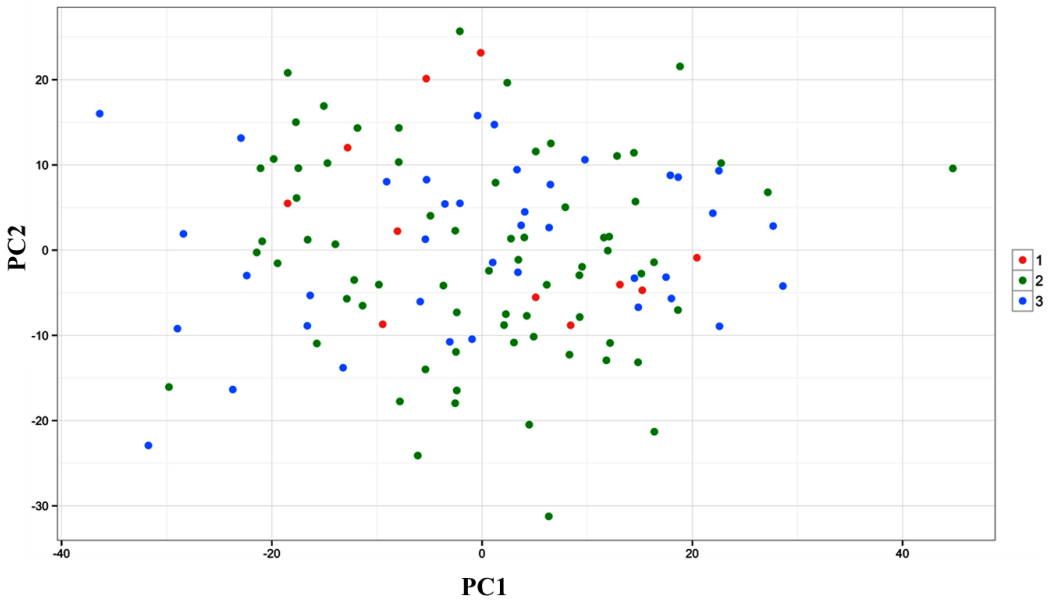


Figure 4.6 Batch correction of the microarray data of the Discovery cohort. *Principal component analysis (PCA) after batch correction of the normalised gene expression microarray data. All samples from the three color coded groups of samples (red (1), green (2) and blue(3)) were now more evenly distributed as a single group.*

3. Differentially expressed genes in pSS-associated lymphoma

As mentioned earlier, the pSS-associated lymphoma group has the smallest sample size (n= 16). In the main analysis (Analysis A), for the lymphoma group, I have removed the one technical outlier in the analysis but included the two pSS-associated lymphoma samples with RIN scores 5.6 and 5.2; thus, the sample size was reduced from 16 to 15 samples. For the pSS-non-lymphoma group, the sample size, after removing samples based on technical outliers and RIN score (< 7), was 52 samples. The list of differentially expressed genes was generated from the comparison between the "pSS" and "pSS-associated lymphoma" groups. A negative log fold change (Log (FC)) means a downregulation of that gene in the pSS (non-lymphoma) group (or upregulation in the pSS-associated lymphoma group). Similarly, a positive Log (FC) represents an upregulation of the gene in the pSS group (i.e. downregulated in pSS-associated lymphoma group). Throughout the chapter, I will use the pSS-associated lymphoma group as the reference point to describe the level of gene expression (i.e. the direction of regulation) of the differentially expressed genes (DEGs). In the main analysis (Analysis A), a total of 57 DEGs were identified: 11 genes were upregulated in the pSS-associated lymphoma group, and 40 genes were downregulated, 6 probes were identified as not applicable (NA), theses NAs represent a potential gene which has not been officially identified or validated. The NAs were excluded from the presented data shown in Table 4.4. The DEGs were also shown using a volcano plot (Figure 4.7). The heat map and the hierarchical clustering of the samples according to the DEGs in Analysis A are shown in Figure 4.8.

Table 4.4 Differentially expressed genes in pSS-associated lymphoma from Analysis A of the microarray data of the Discovery cohort.
Data analysis performed in R, Cutoff values: adjusted $p < 0.05$, and fold change ≥ 1.2

Gene symbol	Name of gene	Adjusted P. value	Regulation in pSS-associated lymphoma	Fold change
LGALS1	lectin, galactoside-binding, soluble, 1	0.0199	upregulation	1.36
KCTD12	potassium channel tetramerisation domain containing 12	0.0329	upregulation	1.28
DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)	0.0280	upregulation	1.27
RBP7	retinol binding protein 7, cellular	0.0272	upregulation	1.26
MGST3	microsomal glutathione S-transferase 3	0.0423	upregulation	1.23
NUDT14	nudix (nucleoside diphosphate linked moiety X)-type motif 14	0.0058	upregulation	1.22
UBXN11	UBX domain protein 11	0.0081	upregulation	1.21
ATG12	autophagy related 12	0.0199	upregulation	1.21
C10orf32(BORCS7)	chromosome 10 open reading frame 32	0.0224	upregulation	1.21
DYNLL1	dynein, light chain, LC8-type 1	0.0195	upregulation	1.2
RNF7	ring finger protein 7	0.0294	upregulation	1.2
RNA28S5	RNA, 28S ribosomal 5	0.0030	downregulation	1.6
LEF1	lymphoid enhancer-binding factor 1	0.0390	downregulation	1.42
SPOCK2	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2	0.0195	downregulation	1.4
ETS1	v-ets avian erythroblastosis virus E26 oncogene homolog 1	0.0292	downregulation	1.39
POM121C	POM121 transmembrane nucleoporin C	0.0062	downregulation	1.35
MYC	v-myc avian myelocytomatosis viral oncogene homolog	0.0199	downregulation	1.32
SGK223	homolog of rat pragma of Rnd2	0.0442	downregulation	1.32
HSP90B1	heat shock protein 90kDa beta (Grp94), member 1	0.0071	downregulation	1.31
RPL23AP5	ribosomal protein L23a pseudogene 5	0.0224	downregulation	1.31
BCL11B	B-cell CLL/lymphoma 11B (zinc finger protein)	0.0353	downregulation	1.3
DDB1	damage-specific DNA binding protein 1, 127kDa	0.0058	downregulation	1.28
CYFIP2	cytoplasmic FMR1 interacting protein 2	0.0498	downregulation	1.27
WAC	WW domain containing adaptor with coiled-coil	0.0058	downregulation	1.27

ESYT1	extended synaptotagmin-like protein 1	0.0442	downregulation	1.26
MAGED1	melanoma antigen family D, 1	0.0442	downregulation	1.26
PRPF8	pre-mRNA processing factor 8	0.0390	downregulation	1.26
SUN2	Sad1 and UNC84 domain containing 2	0.0417	downregulation	1.26
HNRNPUL1	heterogeneous nuclear ribonucleoprotein U-like 1	0.0390	downregulation	1.25
CD96	CD96 molecule	0.0498	downregulation	1.24
CDR2	cerebellar degeneration-related protein 2, 62kDa	0.0206	downregulation	1.24
NAT10	N-acetyltransferase 10 (GCN5-related)	0.0330	downregulation	1.24
VCP	valosin containing protein	0.0195	downregulation	1.24
ALDH9A1	aldehyde dehydrogenase 9 family, member A1	0.0212	downregulation	1.23
BAG3	BCL2-associated athanogene 3	0.0235	downregulation	1.23
LRIG1	leucine-rich repeats and immunoglobulin-like domains 1	0.0159	downregulation	1.23
NCSTN	Nicastrin	0.0465	downregulation	1.23
PAF1	Paf1, RNA polymerase II associated factor, homolog (S. cerevisiae)	0.0199	downregulation	1.23
SF3A1	splicing factor 3a, subunit 1, 120kDa	0.0442	downregulation	1.23
ATP1A1	ATPase, Na+/K+ transporting, alpha 1 polypeptide	0.0058	downregulation	1.22
KHDRBS1	KH domain containing, RNA binding, signal transduction associated 1	0.0405	downregulation	1.22
LRFN3	leucine rich repeat and fibronectin type III domain containing 3	0.0452	downregulation	1.22
RBL2	retinoblastoma-like 2	0.0224	downregulation	1.22
RRN3	RRN3 RNA polymerase I transcription factor homolog (S. cerevisiae)	0.0153	downregulation	1.22
SDHA	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	0.0390	downregulation	1.22
CBLL1	Cbl proto-oncogene-like 1, E3 ubiquitin protein ligase	0.0244	downregulation	1.21
CDV3	CDV3 homolog (mouse)	0.0475	downregulation	1.21
FOXJ3	forkhead box J3	0.0153	downregulation	1.21
SMARCA2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	0.0299	downregulation	1.21
HNRNPDL	heterogeneous nuclear ribonucleoprotein D-like	0.0350	downregulation	1.2
HSPA9	heat shock 70kDa protein 9 (mortalin)	0.0202	downregulation	1.2

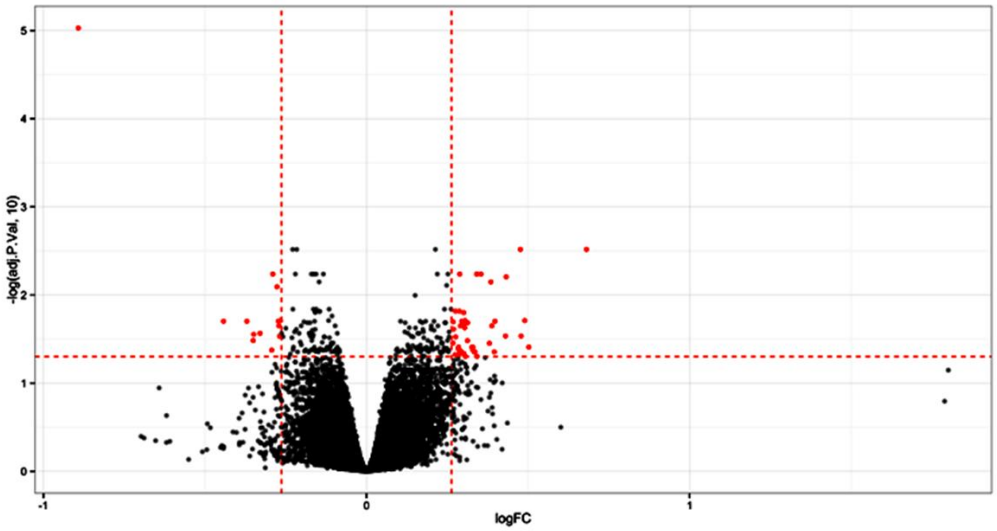


Figure 4.7 Volcano plot of the differentially expressed genes in pSS-associated lymphoma in the Discovery cohort. *The x axis represents log2 of the fold change and the y axis represents the –log10 of the adjusted p value. The dots in red represent the differentially expressed genes, and the red horizontal and vertical lines indicate the cut-off values of the adjusted p value and the fold changes.*

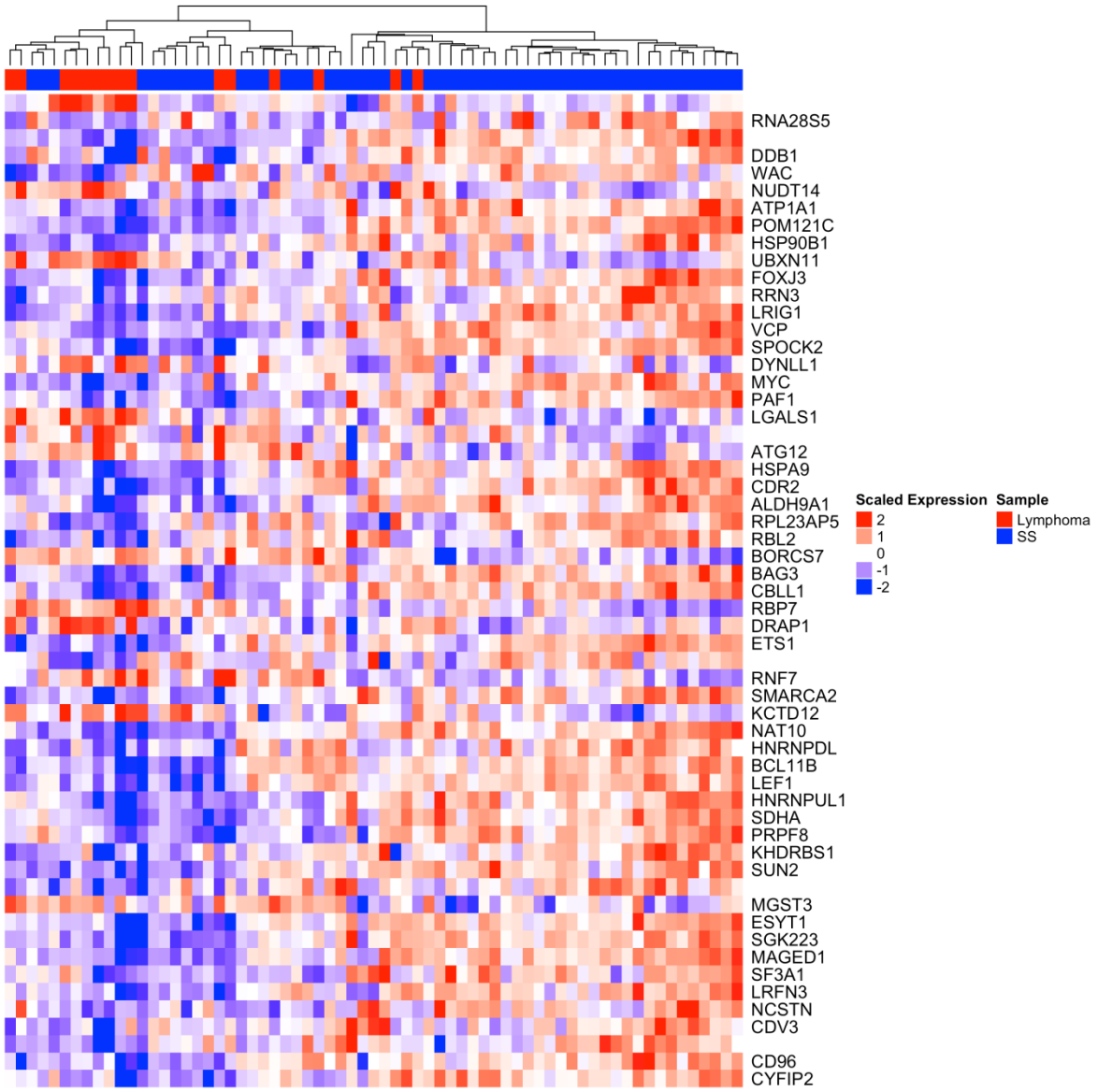


Figure 4.8 Heat map and hierarchical clustering of samples in the Discovery cohort based on the Analysis A of the microarray data. *The empty places represent the probes that were identified as NAs in the analysis.*

Since the main objective of the Discovery cohort was to identify a list of candidate genes that distinguish the lymphoma and the non-lymphoma groups, I have performed additional analyses using different RIN cut-off values (Analysis B, C, and D). Although these analyses have included lymphoma samples that did not meet the 'standard' quality requirements for microarrays, I considered this approach to be acceptable for the Discovery cohort for two reasons. Firstly, the candidate genes identified in the Discovery cohort will

be validated technically and with an independent cohort; therefore, it will reduce the risk of identifying any false DEGs revealed using these additional analyses. Secondly, the lymphoma group has the smallest sample size and was therefore the key factor limiting the power of the study. Increasing the sample size of the lymphoma group could enable the detection of DEGs that are important but did not reach the statistical significance values due to a small sample size. The description of these additional analyses is provided in Table 4.5. Each of these different analyses has yielded a set of deferentially expressed genes. For Analysis B (14 pSS-associated lymphoma samples) has generated 20 DEGs including 4 NAs, Analysis C (16 pSS-associated lymphoma samples) has generated 45 DEGs including 9 NAs, while Analysis D (13 pSS-associated lymphoma samples), generates 19 DEGs including 3 NAs. Each set of DEGs from these four analyses contains genes that are in common with one or more analyses, but there are also DEGs that were unique to each analysis (see Venn diagram in Figure 4.9). The lists of DEGs with the exclusion of the NAs of Analyses B, C, and D are shown in Tables 4.6, 4.7, and 4.8, respectively. The final list of candidate DEGs in pSS-associated lymphoma was generated by combining the lists of DEGs from all 4 analyses. The combined genes list contains 68 DEGs from the microarray experiments is referred to as the "**68-DEGs-Mi**". The workflow of the Discovery cohort and the genes selection for validation is shown schematically in Figure 4.10.

Table 4.5 The analytic criteria used in the four microarray data analyses of pSS-associated lymphoma gene expression signature in the Discovery cohort.

	pSS group		**pSS-associated lymphoma group**		
	RIN<7 removed	**Technical outliers removed**	**RIN < 7 removed**	**Technical outliers removed**	**No. of pSS-associated lymphoma samples**
Analysis A	Yes	Yes	No	Yes	15
Analysis B	Yes	Yes	Yes	No	14
Analysis C	Yes	Yes	No	No	16
Analysis D	Yes	Yes	Yes	Yes	13

Table 4.6 Differentially expressed genes in pSS-associated lymphoma from Analysis B of the microarray data of the Discovery cohort.
Data analysis performed in R, cutoff values: adjusted $p < 0.05$, fold change ≥ 1.2

Gene symbol	Gene name	Adjusted p value	Regulation in pSS-associated lymphoma	Fold change
DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)	0.0418	upregulated	1.29
DYNLL1	dynein, light chain, LC8-type 1	0.0155	upregulated	1.23
SRP14	signal recognition particle 14kDa (homologous Alu RNA binding protein)	0.0419	upregulated	1.23
PSMC1	proteasome (prosome, macropain) 26S subunit, ATPase, 1	0.0408	upregulated	1.23
UBXN11	UBX domain protein 11	0.0167	upregulated	1.22
OAF	OAF homolog (Drosophila)	0.0418	upregulated	1.22
NUDT14	nudix (nucleoside diphosphate linked moiety X)-type motif 14	0.0167	upregulated	1.22
SEC61G	Sec61 gamma subunit	0.0363	upregulated	1.21
CNPY3	canopy FGF signaling regulator 3	0.0226	upregulated	1.21
C10orf32(BORCS7)	chromosome 10 open reading frame 32	0.0418	upregulated	1.21
RNA28S5	RNA, 28S ribosomal 5	0.0015	downregulated	1.67
RPL23AP5	ribosomal protein L23a pseudogene 5	0.0398	downregulated	1.31
WAC	WW domain containing adaptor with coiled-coil	0.0145	downregulated	1.27
HSP90B1	heat shock protein 90kDa beta (Grp94), member 1	0.0408	downregulated	1.26
DDB1	damage-specific DNA binding protein 1, 127kDa	0.0392	downregulated	1.23
RPA2	replication protein A2, 32kDa	0.0225	downregulated	1.20

Table 4.7 Differentially expressed genes in pSS-associated lymphoma from Analysis C of the microarray data of the Discovery cohort.
Data analysis performed in R, cutoff values: adjusted $p < 0.05$, fold change ≥ 1.2

Gene symbol	Gene name	Adjusted p value	Regulation in pSS-associated lymphoma	Fold change
LGALS1	lectin, galactoside-binding, soluble, 1	0.0166	upregulated	1.34
DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)	0.0096	upregulated	1.32
UBXN11	UBX domain protein 11	0.0019	upregulated	1.25
NUDT14	nudix (nucleoside diphosphate linked moiety X)-type motif 14	0.0019	upregulated	1.24
RBP7	retinol binding protein 7, cellular	0.0390	upregulated	1.23
ATG12	autophagy related 12	0.0102	upregulated	1.22
RAB37	RAB37, member RAS oncogene family	0.0127	upregulated	1.21
HCFC1R1	host cell factor C1 regulator 1 (XPO1 dependent)	0.0096	upregulated	1.21
EHBP1L1	EH domain binding protein 1-like 1	0.0397	upregulated	1.20
HLA-DRB1	major histocompatibility complex, class II, DR beta 1	0.0486	downregulated	3.59
RNA28S5	RNA, 28S ribosomal 5	0.0015	downregulated	1.63
LEF1	lymphoid enhancer-binding factor 1	0.0399	downregulated	1.39
ETS1	v-ets avian erythroblastosis virus E26 oncogene homolog 1	0.0245	downregulated	1.38
SPOCK2	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2	0.0381	downregulated	1.34
RPL23AP5	ribosomal protein L23a pseudogene 5	0.0116	downregulated	1.32
POM121C	POM121 transmembrane nucleoporin C	0.0161	downregulated	1.30
HSP90B1	heat shock protein 90kDa beta (Grp94), member 1	0.0071	downregulated	1.30
RASGRP1	RAS guanyl releasing protein 1 (calcium and DAG-regulated)	0.0486	downregulated	1.30
MYC	v-myc avian myelocytomatosis viral oncogene homolog	0.0190	downregulated	1.30
BCL11B	B-cell CLL/lymphoma 11B (zinc finger protein)	0.0315	downregulated	1.29
WAC	WW domain containing adaptor with coiled-coil	0.0015	downregulated	1.29
ITK	IL2-inducible T-cell kinase	0.0346	downregulated	1.29
DDB1	damage-specific DNA binding protein 1, 127kDa	0.0112	downregulated	1.25

CD96	CD96 molecule	0.0382	downregulated	1.24
RBL2	retinoblastoma-like 2	0.0114	downregulated	1.23
ALDH9A1	aldehyde dehydrogenase 9 family, member A1	0.0137	downregulated	1.23
CDR2	cerebellar degeneration-related protein 2, 62kDa	0.0184	downregulated	1.23
RRN3	RRN3 RNA polymerase I transcription factor homolog (S. cerevisiae)	0.0078	downregulated	1.23
HNRNPA1P10	heterogeneous nuclear ribonucleoprotein A1 pseudogene 10	0.0441	downregulated	1.23
LRIG1	leucine-rich repeats and immunoglobulin-like domains 1	0.0137	downregulated	1.22
KHDRBS1	KH domain containing, RNA binding, signal transduction associated 1	0.0292	downregulated	1.22
CDV3	CDV3 homolog (mouse)	0.0346	downregulated	1.22
SMARCA2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	0.0255	downregulated	1.20
VCP	valosin containing protein	0.0346	downregulated	1.20
HSPA9	heat shock 70kDa protein 9 (mortalin)	0.0137	downregulated	1.20
PRKCQ	protein kinase C, theta	0.0479	downregulated	1.20

Table 4.8 Differentially expressed genes in pSS-associated lymphoma from Analysis D of the microarray data of the Discovery cohort.
Data analysis performed in R, cutoff values: adjusted $p < 0.05$, fold change ≥ 1.2

Gene symbol	Gene name	Adjusted p value	Regulation in pSS-associated lymphoma	Fold change
DYNLL1	dynein, light chain, LC8-type 1	0.0102	upregulated	1.25
C10orf32(BORCS7)	chromosome 10 open reading frame 32	0.0249	upregulated	1.23
HNMT	histamine N-methyltransferase	0.0470	upregulated	1.23
OAF	OAF homolog (Drosophila)	0.0470	upregulated	1.22
SEC61G	Sec61 gamma subunit	0.0470	upregulated	1.21
RNA28S5	RNA, 28S ribosomal 5	0.0073	downregulated	1.64
SPOCK2	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2	0.0450	downregulated	1.41
POM121C	POM121 transmembrane nucleoporin C	0.0249	downregulated	1.34
CYFIP2	cytoplasmic FMR1 interacting protein 2	0.0470	downregulated	1.32
DDB1	damage-specific DNA binding protein 1, 127kDa	0.0185	downregulated	1.27
HSP90B1	heat shock protein 90kDa beta (Grp94), member 1	0.0481	downregulated	1.26
WAC	WW domain containing adaptor with coiled-coil	0.0327	downregulated	1.24
ATP1A1	ATPase, Na+/K+ transporting, alpha 1 polypeptide	0.0124	downregulated	1.23
PAF1	Paf1, RNA polymerase II associated factor, homolog (S. cerevisiae)	0.0470	downregulated	1.23
BMS1	BMS1 ribosome biogenesis factor	0.0260	downregulated	1.20
BTBD11	BTB (POZ) domain containing 11	0.0278	downregulated	1.20

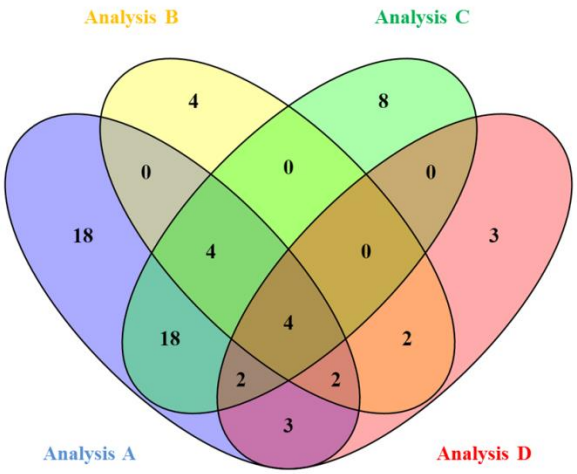


Figure 4.9 Venn diagram for the DEGs that are in common and unique among the four different microarray data analyses (A, B, C, and D) in the Discovery cohort

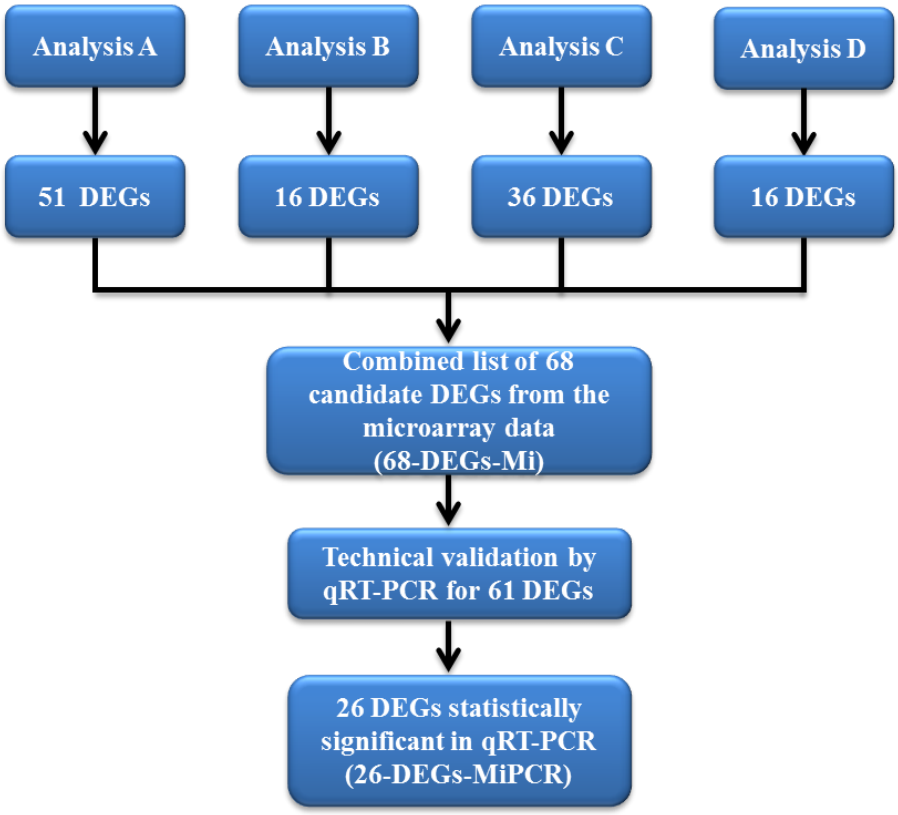


Figure 4.10 Schematic representation of the microarray data analytic approach of the Discovery cohort. *The NAs were excluded from the total number of the DEGs in each analysis. DEGs= Differentially expressed genes, IPA= Ingenuity Pathway analysis*

4.4.4 Technical validation of the differentially expressed genes in pSS-associated lymphoma with qRT-PCR

1. Housekeeping genes selection

An important step in validating the DEGs associated with pSS-associated lymphoma is to select the most stable housekeeping genes for comparison. I used the NormFinder algorithm (Andersen et al., 2004) to calculate the stability of the housekeeping genes between the different subject groups. This algorithm revealed that *UBC* being the housekeeping gene with the most stable level of expression, with a stability value of 0.003 (Figure 4.11). Moreover, taking into consideration of both the intra- and inter-group variations, the two genes with the most stable levels of expression are *UBC* and *YWHAZ*, with stability values of 0.003. In addition, I have included the *ACTB* gene, which came third in the stability value ranking based on NormFinder. Finally, I have also included the *18S* gene, as it has been frequently used in other studies and has been used in a previous study in our group to investigate the miRNA expression in pSS-associated lymphoma.

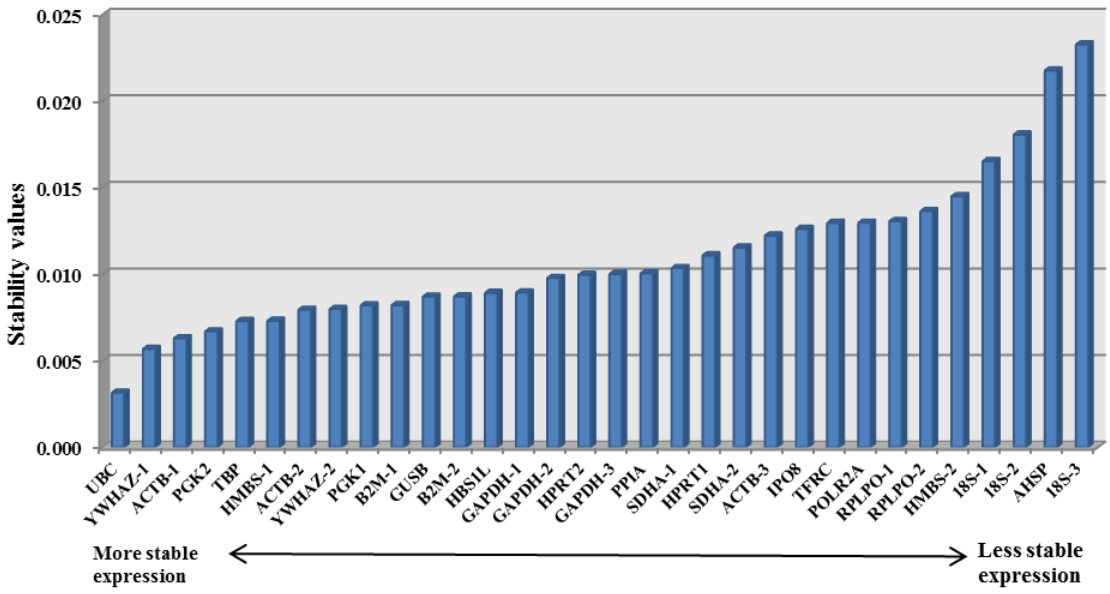


Figure 4.11 The expression stability values for the housekeeping genes calculated by NormFinder. *The bar chart shows the most stable housekeeping gene on the left and the most variable ones on the right. As the microarray contains more than one probe for some housekeeping genes, these probes were distinguished by adding a suffix of 1, 2, 3, etc. after the gene name.*

Furthermore, the four selected housekeeping genes were also measured by qRT-PCR to check their levels of expression in the 5 subject groups. The qRT-PCR threshold cycle (Ct) value for each housekeeping gene was calculated and then compared among the 5 subject groups. The expression levels of *YWHAZ* and *ACTB* were stable across the subject groups. There was no statistically significant variation in the level of expression of these genes among the five groups (ANOVA p values of 0.819 and 0.892 for *YWHAZ* and *ACTB*, respectively). Furthermore, there were also no statistically significant differences when comparing the subject groups individually to each other. Finally, the standard deviations of expression levels for both genes across all samples were small (0.5532 and 0.7951, respectively). With regard to the *UBC* gene, there was no statistically significant variation between the five subject groups when analysed using ANOVA (p=0.123); however, there was a significant difference in the expression level when comparing the pSS-associated lymphoma versus pSS (p=0.009) groups. Interestingly, the *18S* gene showed the highest variability of the 4 housekeeping genes that I have selected. Still, there was no significant variation between the 5 subject groups (ANOVA p=0.062). However, when comparing the pSS-associated lymphoma groups with all 4 other groups individually, there were statistically significant differences in the level of expression of 18S. I have therefore decided to exclude this housekeeping gene as well as *UBC* from my analysis. The details of these statistical analyses are provided in Table 4.9, and the comparisons of the Cts values are shown in Figure 4.12. Since the expression levels of *YWHAZ* are stable across all subject groups and have the lowest standard deviation, the data were normalised to this housekeeping gene in subsequent analysis. Due to limited RNA availability, the stability testing of the expression levels of the four housekeeping genes were performed in only the first two batches, while for the latter two batches I have used only *YWHAZ* and *ACTB*. The same strategy was employed with the second batch of qRT-PCR array plates, and the results were consistent with the results obtained from the first batch (Table 4.10).

Table 4.9 Statistical analysis of the expression levels stability of the 4 housekeeping genes across the 5 subject groups in the first batch of qRT-PCR plates. ** = Statistically Significant ($p < 0.05$)*

Parameters		Housekeeping genes			
		ACTB	***YWHAZ***	***UBC***	***18S***
	C_T mean , SD	22.92 (0.7951)	29.43 (0.5532)	25.67 (0.6636)	12.70 (1.523)
p-values	**pSS vs pSS- associated lymphoma**	0.81	0.26	0.009*	0.001*
	pSS- associated lymphoma vs pSS- other cancers	0.32	0.55	0.12	0.006*
	pSS- associated lymphoma vs pSS- paraproteinemia	0.45	0.83	0.1	0.006*
	pSS vs pSS- other cancers	0.4	0.93	0.19	0.7
	pSS vs pSS-paraproteinemia	0.7	0.53	0.19	0.54
	pSS- other cancers vs pSS-paraproteinemia	0.87	0.67	0.91	0.91
	pSS vs Healthy controls	0.51	1	0.35	0.94
	pSS- associated lymphoma vs Healthy controls	0.37	0.35	0.21	0.01*
	pSS- other cancers vs Healthy controls	0.75	0.66	0.88	0.78
	pSS-paraproteinemia vs Healthy controls	0.92	0.98	1	0.76
	ANOVA	0.892	0.819	0.123	0.062

Table 4.10 Statistical analysis of the expression level stability of the 4 housekeeping in the second batch of qRT-PCR plates. ** = Statistically Significant ($p < 0.05$)*

Parameters		Housekeeping genes			
		ACTB	***YWHAZ***	***UBC***	***18S***
	C_T mean , SD	23.47 (0.7677)	29.66 (0.5675)	25.79 (0.6184)	12.76 (1.441)
p-values	**pSS vs pSS- associated lymphoma**	0.82	0.24	0.03*	0.001*
	pSS- associated lymphoma vs pSS- other cancers	0.39	0.52	0.19	0.009*
	pSS- associated lymphoma vs pSS- paraproteinemia	0.26	0.53	0.19	0.003*
	pSS vs pSS- other cancers	0.39	0.79	0.3	0.66
	pSS vs pSS-paraproteinemia	0.48	0.93	0.51	0.85
	pSS- other cancers vs pSS-paraproteinemia	0.98	0.87	0.77	0.87
	pSS vs Healthy controls	0.21	0.92	0.58	0.74
	pSS- associated lymphoma vs Healthy controls	0.54	0.44	0.22	0.004*
	pSS- other cancers vs Healthy controls	0.95	1	0.75	0.66
	pSS-paraproteinemia vs Healthy controls	0.89	0.89	0.82	0.82
	ANOVA	0.8	0.94	0.38	0.07

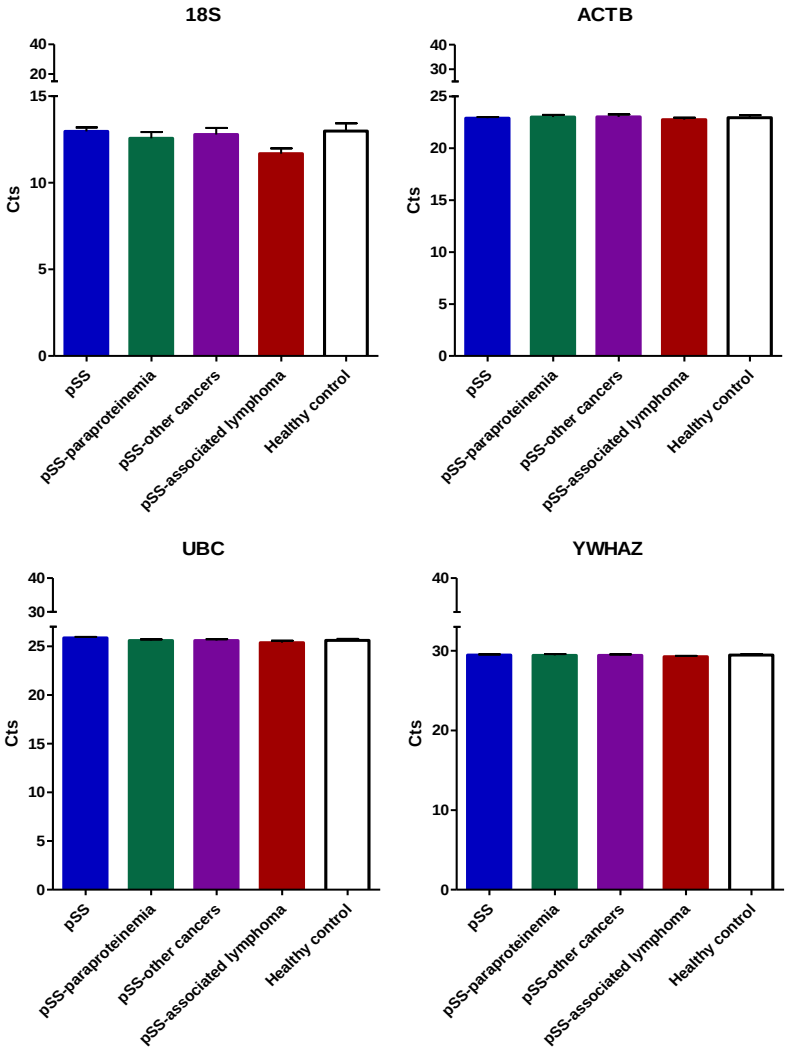


Figure 4.12 Bar charts of the Cts values of the 4 housekeeping genes measured with qRT-PCR for the 5 subject groups in the Discovery cohort. *YWHAZ and ACTB had the most stable levels of expression. There was a significant difference in the expression level of UBC when comparing the pSS-associated lymphoma vs pSS groups. The expression level of 18S was the most variable among the samples.*

2. Technical validation of the whole blood gene expression signature in pSS-associated lymphoma

As the availability of RNA samples was limited, I was not able to technically validate differential expression of all of the candidate genes. In order to select the best candidates for technical validation, I used the following criteria. Firstly, the final list of DEGs (68-DEGs-Mi) is ranked according to the fold change and the p value. Secondly, DEGs that were identified by all four analyses were selected. A total of 61 out of the 68 DEGs were selected, the remaining 7 genes included 2 pseudogenes and 5 genes with small fold changes. The real-time RT-PCR experiment confirmed differential expression of 26 DEGs (Mann-Whitney test under $P < 0.05$) out of the 61 genes tested. These 26 DEGs included both upregulated and downregulated genes in pSS-associated lymphoma. The downregulated genes were *BMS1, BTBD11, CBLL1, CYFIP2, ESYT1, HNRNPUL1, LEF1, MAGED1, POM121C, PRPF8, SF3A1, SGK223,* and *VCP,* while the upregulated genes were *C10orf32(BORCS7), CNPY3, DRAP1, DYNLL1, HNMT, LGALS1, MGST3, NUDT14, OAF, RBP7, SEG61G, SRP14,* and *UBXN11*. Twenty of these 26 genes were identified from the main analysis (Analysis A), and the remaining 6 genes were identified from the other analyses (Table 4.11). These 26 genes constituted the 26 differentially expressed genes from microarray and PCR (referred to as "**26-DEGs-MiPCR")**.

Apart from these 26 DEGs, the expression level of *RNA28S5* was also significantly different between the lymphoma and non-lymphoma groups, but the direction of regulation (i.e. up/down-regulation) was different in the qRT-PCR experiment and in the microarray. It is also important to mention that for *HLA-DRB1,* the expression levels could not be determined by the qRT-PCR experiment, as the amplification curves were jagged. A heat map representing the sample clustering according to the 26-DEGs-MiPCR is shown in Figure 4.13.

Table 4.11 The 26-DEGs-MiPCR of pSS-associated lymphoma from the Discovery cohort. *For each DEG, the microarray analysis in which it was identified as candidate, the p value (measured by Mann-Whitney U test), the fold change and the regulation direction, Fc was calculated according to the housekeeping gene normalisation*

Gene symbol	Gene name	Analysis A	Analysis B	Analysis C	Analysis D	p value	Fc & regulation
NUDT14	Nudix (nucleoside diphosphate linked moiety X)-type motif 14	+	+	+	-	0.0000	2.09 ↑
OAF	OAF homolog (Drosophila)	-	+	-	+	0.0001	2.05 ↑
C10orf32(BORCS7)	Chromosome 10 open reading frame 32	+	+	-	+	0.0000	1.91 ↑
RBP7	Retinol binding protein 7, cellular	+	-	+	-	0.0005	1.88 ↑
LGALS1	Lectin, galactoside-binding, soluble, 1	+	-	+	-	0.0001	1.74 ↑
DYNLL1	Dynein, light chain, LC8-type 1	+	+	-	+	0.0000	1.65 ↑
HNMT	histamine N-methyltransferase	-	-	-	+	0.0231	1.55 ↑
DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)	+	+	+	-	0.0000	1.50 ↑
SRP14	Signal recognition particle 14kDa (homologous Alu RNA binding protein)	-	+	-	-	0.0344	1.39 ↑
UBXN11	UBX domain protein 11	+	+	+	-	0.0017	1.38 ↑
SEC61G	Sec61 gamma subunit	-	+	-	+	0.0005	1.37 ↑
CNPY3	Canopy FGF signaling regulator 3	-	+	-	-	0.0004	1.32 ↑
MGST3	Microsomal glutathione S-transferase 3	+	-	-	-	0.0079	1.07 ↑
VCP	Valosin containing protein	+	-	+	-	0.0057	2.45 ↓
HNRNPUL1	Heterogeneous nuclear ribonucleoprotein U-like 1	+	-	-	-	0.0238	2.35 ↓
ESYT1	Extended synaptotagmin-like protein 1	+	-	-	-	0.0118	1.86 ↓
SGK223	Homolog of rat pragma of Rnd2	+	-	-	-	0.0495	1.67 ↓
BTBD11	BTB (POZ) domain containing 11	-	-	-	+	0.0338	1.66 ↓
BMS1	BMS1 ribosome biogenesis factor	-	-	-	+	0.0012	1.50 ↓
PRPF8	Pre-mRNA processing factor 8	+	-	-	-	0.0238	1.41 ↓
MAGED1	Melanoma antigen family D, 1	+	-	-	-	0.0333	1.40 ↓
SF3A1	Splicing factor 3a, subunit 1, 120kDa	+	-	-	-	0.0106	1.32 ↓
LEF1	Lymphoid enhancer-binding factor 1	+	-	+	-	0.0289	1.30 ↓
CYFIP2	Cytoplasmic FMR1 interacting protein 2	+	-	-	+	0.0314	1.26 ↓
POM121C	POM121 transmembrane nucleoporin C	+	-	+	+	0.0375	1.23 ↓
CBLL1	Cbl proto-oncogene-like 1, E3 ubiquitin protein ligase	+	-	-	-	0.0421	1.14 ↓

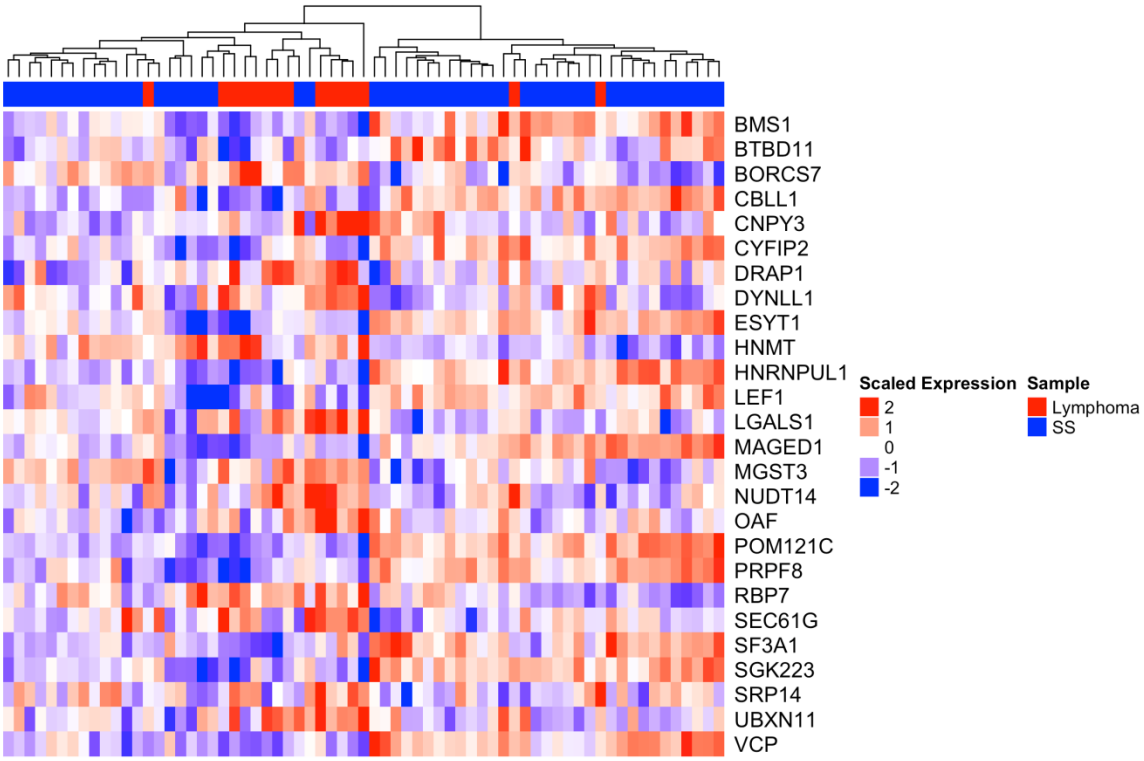


Figure 4.13 Heat map and hierarchical clustering of samples of the Discovery cohort according to the 26-DEGs-MiPCR

4.4.5 Biological validation of the potential whole blood gene expression signature of pSS-associated lymphoma

A second independent set of samples (referred to as the Validation cohort) was used to validate the potential gene expression signature of pSS-associated lymphoma identified using the Discovery cohort. The Validation cohort consisted of only two groups, the pSS-associated lymphoma group with 17 samples, and the pSS group with 119 samples. The clinical data for the samples are shown in Table 4.2. The reason for not including the additional pSS subgroups as in the Discovery cohort was that comparisons between the lymphoma group and the other pSS subgroups did not contribute to the identification of DEGs in the Discovery cohort (see chapter 6 for more details). Twenty-four genes were included in this biological validation experiment, of which only 3 genes were found to be differentially expressed between the lymphoma and the non-lymphoma groups. The qRT-PCR data were normalised using *YWHAZ* as a housekeeping gene, the expression levels of which was also stable across the pSS-associated lymphoma and pSS groups in this cohort.

These 3 validated DEGs were *BMS1 (*downregulated gene in pSS-associated lymphoma), *NUDT14* and *MGST3* (upregulated). These genes were significantly differentially expressed with *p* values of <0.0000, 0.0137, and 0.0209 respectively, and fold changes of 2, 1.40, and 1.28, respectively (Figure 4.14). Thus, these 3 genes (*BMS1*, *NUDT14*, and *MGST3*) were found to be the gene expression signature for pSS-associated lymphoma for this study (referred to as the **"3-gene biosignature of pSS-associated lymphoma"** in the thesis). For the remaining 21 genes, I have also examined in more detail those genes that satisfy the following criteria: (a) those with p values between 0.05 and 0.25 and (b) a consistent direction of regulation as observed in the Discovery cohort. Three genes satisfied these criteria: *LEF1*, *OAF*, and *DRAP1*. In addition, another five genes have a p value > 0.25 but the same direction of regulation of the Discovery cohort; these genes were *LGALS1*, *CBLL1*, *C10orf32 (BORCS7)*, *DYNLL1*, and *SGK223* (Table 4.12). The p values, fold changes, and the direction of regulation of the remaining 13 non-validated genes are provided in supplementary table S3.

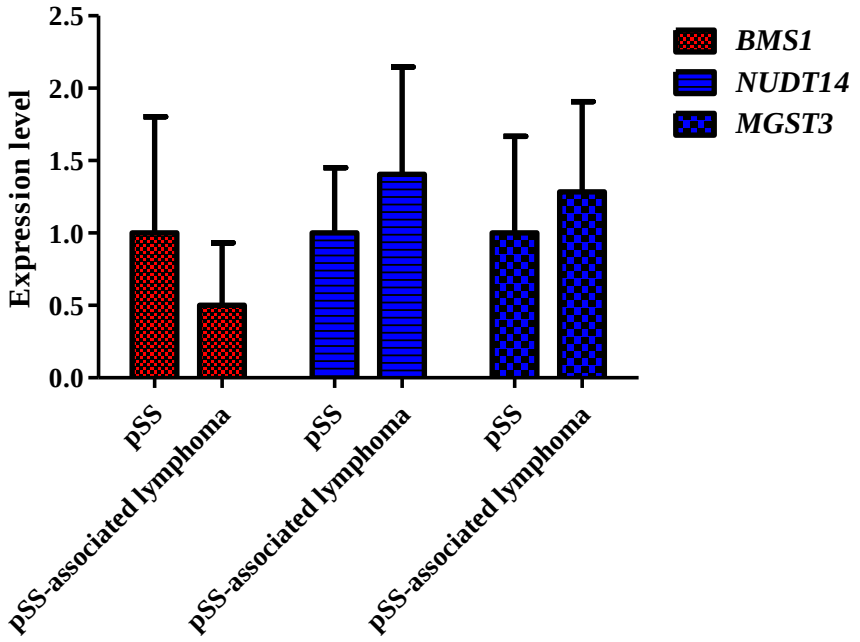


Figure 4.14 The relative gene expression levels of the 3-gene biosignature of pSS-associated lymphoma. *The expression levels of the genes were expressed as the ratio to the expression level of the housekeeping gene YWHAZ. The downregulated BMS1 gene is in red (p<0.0000, FC=2); the upregulated NUDT14 and MGST3 genes are in blue (p=0.0137, 0.0209 ;FC=1.40,1.28 respectively). The total number of samples used in the comparison were pSS = 119 and pSS-associated lymphoma = 17.*

Table 4.12 The 3-gene biosignature of pSS-associated lymphoma and other potential candidate genes identified in the Validation cohort *(i.e. those genes with expression levels that were not statistically significantly different between subject groups but have the same direction of regulation as the Discovery cohort). Fc was calculated according to the normalisation against the pSS group.*

Gene symbol	Gene name	p value	Fold change and regulation
3-gene biosignature of pSS-associated lymphoma (sig. p values & consistent regulation with the Discovery cohort)			
BMS1	BMS1 ribosome biogenesis factor	0.0000	2.00↓
NUDT14	Nudix (nucleoside diphosphate linked moiety X)-type motif 14	0.0137	1.40 ↑
MGST3	Microsomal glutathione S-transferase 3	0.0209	1.28 ↑
Genes with 0.05 < p values > 0.25 consistent direction of regulation as the Discovery cohort			
LEF1	Lymphoid enhancer-binding factor 1	0.0607	1.25 ↓
OAF	OAF homolog (Drosophila)	0.1370	1.33↑
DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)	0.1220	1.19 ↑
Genes with p values > 0.25 consistent direction of regulation as the Discovery cohort			
LGALS1	Lectin, galactoside-binding, soluble, 1	0.3109	1.12 ↑
CBLL1	Cbl proto-oncogene-like 1, E3 ubiquitin protein ligase	0.5063	1.11 ↓
C10orf32 (BORCS7)	chromosome 10 open reading frame 32	0.4221	1.07 ↑
DYNLL1	Dynein, light chain, LC8-type 1	0.6263	1.07 ↑
SGK223	homolog of rat pragma of Rnd2	0.6124	1.01 ↓

4.4.6 Testing the whole blood gene expression signature in untreated pSS-associated lymphoma patients

A set of untreated pSS-associated lymphoma samples was also included within the Validation cohort. These samples were collected from 7 patients with untreated lymphoma at the time of sampling (4 samples were obtained from our collaborators in Sweden, while the remaining 3 were from the UKPSSR). Six genes (*NUDT14*, *DRAP1*, *DYNLL1*, *RBP7*, *SF3A1*, and *VCP)* were differentially expressed between "untreated lymphoma" and "non-lymphoma" groups with $p < 0.05$ (Table 4.13). Among these 6 DEGs, *NUDT14* was a DEG for both treated and untreated pSS-associated lymphoma groups. *DRAP1* and *DYNLL1* were not validated in the Validation cohort, but they have the same direction of regulation in both treated lymphoma cohorts (i.e. Discovery and Validation cohorts). *RBP7* was also not validated in the Validation cohort with p value > 0.05, and having an opposite direction of regulation between the Discovery and the Validation cohorts. Similarly, *SF3A1* and *VCP* were downregulated in the "pSS-associated lymphoma" group in the Discovery cohort but

were 'upregulated' in the Validation cohort, albeit both with $p < 0.05$. Therefore, I considered that *NUDT14*, *DRAP1*, and *DYNLL1* as genes that could be potential biomarkers for untreated pSS-associated lymphoma (Figure 4.15). The p values, fold changes, and the direction of regulation of the remaining genes are provided in supplementary table S4.

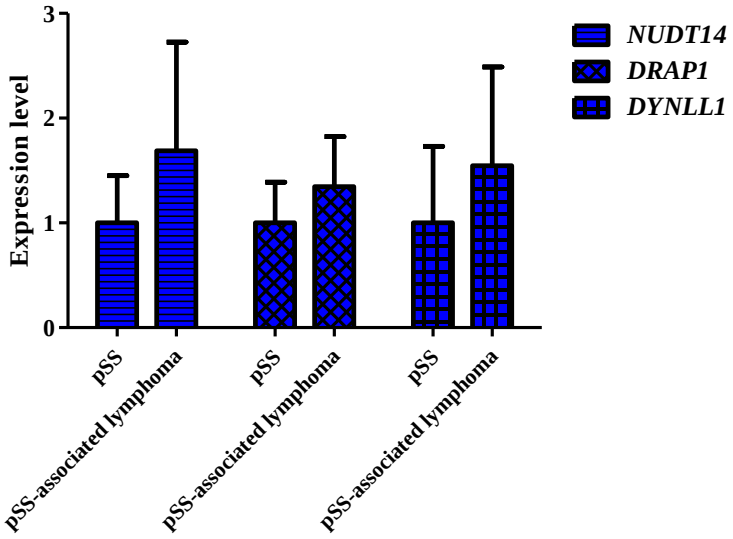


Figure 4.15 The relative gene expression levels of the 3 DEGs in untreated pSS-associated lymphoma. *All 3 genes (NUDT14, DRAP1 and, DYNLL1) were upregulated in the untreated lymphoma group with p=0.0097, 0.0323 and 0.0106; FC=1.68, 1.34 and 1.54 respectively. The total number of samples used in the comparison were pSS = 119 and untreated pSS-associated lymphoma = 7*

Table 4.13 The significant DEGs in untreated pSS-associated lymphoma*. Fc was calculated according to the normalisation against the control group.*

Gene symbol	Gene name	p value	Fold change and regulation
3-gene biosignature in pSS-associated lymphoma			
NUDT14	Nudix (nucleoside diphosphate linked moiety X)-type motif 14	0.0097	1.68 ↑
DEGs in untreated pSS-associated lymphoma (significant p values and consistent regulation direction with the treated lymphoma groups in both Discovery and Validation cohort)			
DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)	0.0323	1.34 ↑
DYNLL1	Dynein, light chain, LC8-type 1	0.0106	1.54 ↑
DEGs in untreated pSS-associated lymphoma (but not consistent regulation direction within treated lymphoma)			
SF3A1	splicing factor 3a, subunit 1, 120kDa	0.0123	1.79 ↑
RBP7	Retinol binding protein 7, cellular	0.043	1.29 ↑
VCP	Valosin containing protein	0.0306	1.27 ↑

4.4.7 Prediction models for transcriptomic biomarkers of pSS-associated lymphoma

The gene expression data obtained from the qRT-PCR experiment of the Discovery cohort (60 DEGs) were used to build prediction models for pSS-associated lymphoma using logistic regression. Inspection of simple logistic regression plots for each of the individual candidates, followed by a stepwise logistic regression procedure, led to the identification of two candidate genes for subsequent modelling with one third of the data retained for model cross-validation, then the Two-Gene model.

1. Two-Gene model

In this model, a cross-validation method was used first to test the model within the data in the Discovery cohort, where the data is divided into thirds. Two thirds were used to build the model and the third part used to test it. Next, the Two-Gene model was applied to the whole dataset of the Discovery cohort.

a. Cross-validation of the Two-Gene model

The dataset set was randomly divided into thirds. Two-thirds of which were used as a "training set" to build the model and the remaining third was used for validation (named as the "test set"). *NUDT14* and *UBXN11* were identified as key candidates. Both genes were retained in the model. The correlation between the expression level of *UBXN11* and the probability of having lymphoma in pSS patients was poor (p = 0.7743), while the correlation between *NUDT14* expression levels and having lymphoma was statistically significant (p=0.0210) and robust to the inclusion of *UBXN11*, as in Table 4.14A. I also applied the "likelihood ratio test", a statistical test used to compare the goodness of fit of the model to the null hypothesis. The test determines how many times more likely (i.e. the likelihood ratio) the data fit one model over the null model with the corresponding p value. In this model, the likelihood ratio test showed a significant p value for *NUDT14* of 0.0027 (Table 4.14 B), suggesting that *NUDT14* may be a better predictor of pSS-associated lymphoma than *UBXN11*.

The receiver operating characteristic (ROC) curve for the cross-validation of the Two-Gene model is shown in Figure 4.16 and the area under the curve (AUC) values for the training and testing sets were 0.875 and 0.944 respectively, indicating good overall performance. In order to further evaluate the performance of the model, we performed a misclassification analysis comparing the observed and expected diagnoses according to the model in both the

training and test sets. The off-diagonal values in the confusion matrix in Table 4.15 (A and B) highlight these misclassification errors. Overall, the model correctly predicted 37 of the 42 samples (88.09% accuracy) in the training set and correctly predicted 14 of the 18 samples (77.7% accuracy) in the test set. The model did not misclassify any of the 32 non-lymphoma pSS controls, but 4 of the 9 pSS-associated lymphoma patients were misclassified in the training set. In the test set, the model did not misclassify any of the 12 non-lymphoma pSS controls, but 4 of the 6 pSS-associated lymphoma patients were misclassified. Note that this model could be further refined by incorporating the prevalence of lymphoma in pSS patients as prior probabilities.

Table 4.14 Statistical analysis of the Two-Gene model using cross-validation approach. *A. Parameter Estimates. B. Effect Likelihood Ratio tests. The results were generated by JMP SAS.*

A

Parameter Estimates				
Term	**Estimate**	**Std Error**	**ChiSquare**	**Prob>ChiSq**
intercept	-5.3179632	2.5135689	4.48	0.0344*
NUDT14	27.7794556	12.038592	5.32	0.0210*
UBXN11	0.37521799	1.3085782	0.08	0.7743

B

Effect Likelihood Ratio Tests				
Source	**Nparm**	**DF**	**ChiSquare**	**Prob>ChiSq**
NUDT14	1	1	8.98441796	0.0027
UBXN11	1	1	0.08133799	0.7755

** = Staistically signficant*

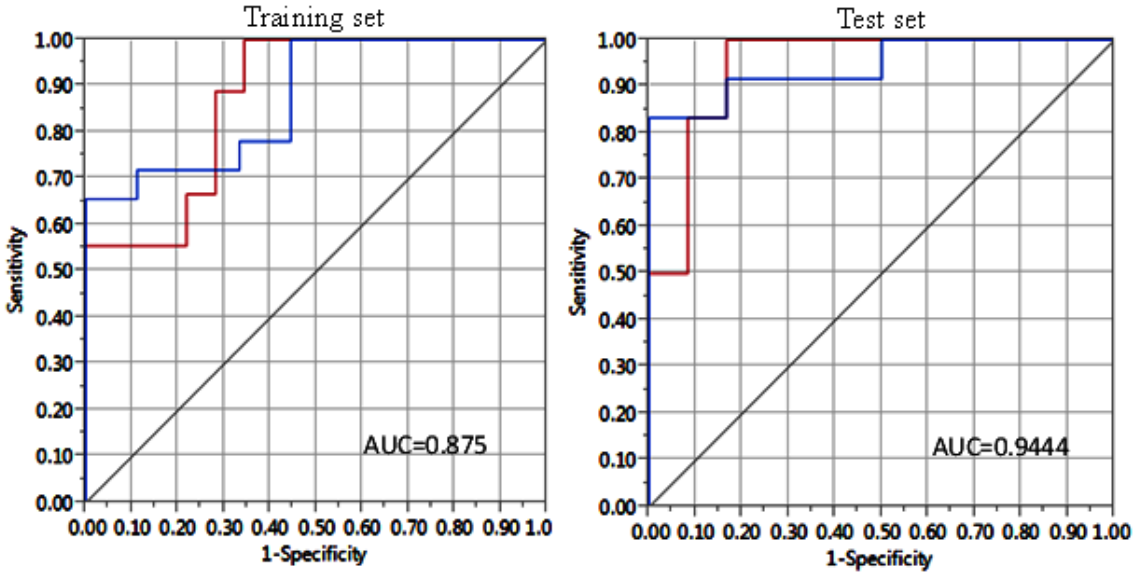


Figure 4.16 Receiver operating characteristic (ROC) curves of the Two-Gene model using cross-validation approach*. The specificity and sensitivity of the Two-Gene model to classify pSS patients with or without lymphoma using the cross-validation approach. The left figure is the ROC of the training set with AUC=0.875, and the right figure is the ROC of the test set with AUC=0.944.*

Table 4.15 Confusion matrix of the Two-Gene model using cross-validation approach. *Showing actual and predicted classifications in training set (A) and test set (B). The results were generated by JMP SAS*

A

Actual classification		**Predicted**
Training set	**pSS-associated lymphoma**	**pSS**
pSS-associated lymphoma	**5**	**4**
pSS	**0**	**32**

B

Actual classification		**Predicted**
Test set	**pSS-associated lymphoma**	**pSS**
pSS-associated lymphoma	**2**	**4**
pSS	**0**	**12**

b. Building the Two-Gene model

In this model, a stepwise ordinal logistic regression method was used to identify the best classifier for pSS-associated lymphoma. For this model, I used the 60 DEGs (i.e. training and test sets were pooled together). *NUDT14* and *UBXN11* were again identified as key candidates. The correlation between *UBXN11* expression and the probability of having lymphoma was not significant (p=0.0606), while that between *NUDT14* expression levels and lymphoma development was statistically significant (p=0.0095) (Table 4.16A).

I also applied the likelihood ratio test for this model, which showed a significant p value for *NUDT14* of 0.0013 (Table 4.16B), suggesting that *NUDT14* was a better predictor of pSS-associated lymphoma than *UBXN11*. Similarly, the probability of the Lack of Fit test is not statistically significant, with a value of 0.927, suggesting that the goodness of fit of the model was good (Table 4.16 C).

The ROC curve for this Two-Gene model has an area AUC value of 0.886, indicating good overall performance (Figure 4.17). In order to further evaluate the performance of the model, I performed a misclassification analysis comparing the observed and expected diagnoses according to the model. Overall, this model correctly predicted 50 of the 59 samples (84.7% accuracy). The model misclassified just 2 of the 44 non-lymphoma pSS patients, but 7 of the 15 pSS associated lymphoma patients were misclassified. This model could be further refined by incorporating the prevalence of lymphoma in pSS patients as prior probabilities.

Table 4.16 Statistical analysis of the Two-Gene model. *A. Parameter Estimates; B. Effect likelihood ratio tests; and C. the lack of fit. The results were generated by JMP SAS.*

A

Parameter Estimates				
Term	**Estimate**	**Std Error**	**ChiSquare**	**Prob>ChiSq**
Intercept	-7.0777384	2.0191171	12.29	0.0005*
NUDT14	25.4739019	9.8198523	6.73	0.0095*
UBXN11	1.50683684	0.80304	3.52	0.0606

B

Effect Likelihood Ratio Tests				
Source	**Nparm**	**DF**	**ChiSquare**	**Prob>ChiSq**
NUDT14	1	1	10.3598995	0.0013*
UBXN11	1	1	3.60432543	0.0576

C

Lack of Fit			
Source	**DF**	**-LogLikelihood**	**ChiSquare**
Lack of fit	56	20.714426	41.42885
Saturated	58	0.000000	Prob>ChiSq
Fitted	2	20.714426	0.927

** = Staistically signficant*

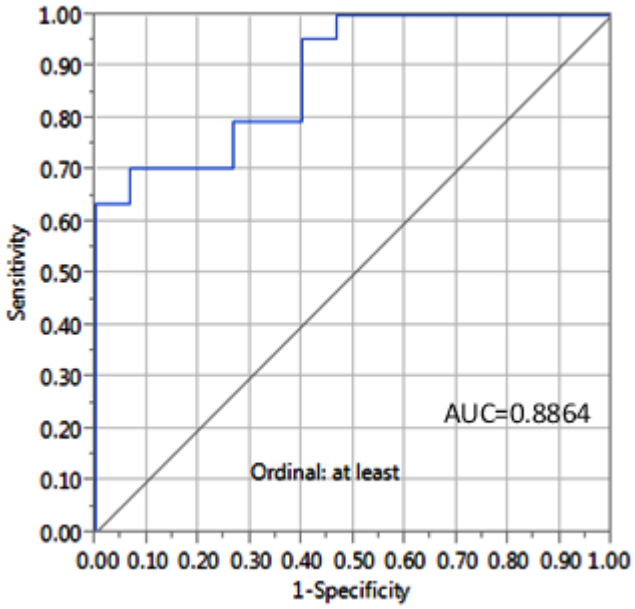


Figure 4.17 Receiver operating characteristic (ROC) curve of the Two-Gene model. *This represents the specificity and sensitivity of the Two Genes model to classify pSS patients with or without lymphoma; the AUC=0.8864.*

Table 4.17 Confusion Matrix of the Two-Gene model. *Showing actual and predicted classifications in the Discovery cohort. The results were generated by JMP SAS*

Actual classification		**Predicted**
Training	**pSS-associated lymphoma**	**pSS**
pSS-associated lymphoma	**8**	**7**
pSS	**2**	**42**

2. The Single Gene model

In the Two-Gene model, the *UBXN11* gene achieves a small improvement on predictions, though it was only marginally significant. Since *UBXN11* was not making a statistically significant contribution in the Two-Gene model, I have chosen to investigate whether a model that uses only the *NUDT14* gene will also perform well in predicting membership of pSS-associated lymphoma.

a. Building the Single Gene model

The logistic fit for *NUDT14* is shown in Figure 4.18. It appears that *NUDT14* plays an important role in distinguishing pSS patients with lymphoma from those without lymphoma. Thus, there was a statistically significant (p = 0.0011) correlation between the expression level of *NUDT14* and the probability of having lymphoma in pSS patients as in Table 4.18A.

I then applied the likelihood ratio test, which showed a significant p value of 0.0001 (Table 4.18 B). Similarly, the lack of fit test (p = 0.874) suggested that the goodness of fit of the model was good (Table 4.18 C).

The ROC curve for the Single Gene model is shown in Figure 4.19 and the AUC was 0.859, indicating good overall performance. I then performed a misclassification analysis (Table 4.19) Overall, the model correctly predicted 48 of the 59 samples (81.3% accuracy). The model misclassified just 3 of the 44 non-lymphoma pSS patients, but 8 of the 15 pSS-associated lymphoma patients were misclassified. This model could be further refined by incorporating the prevalence of lymphoma in pSS patients as prior probabilities.

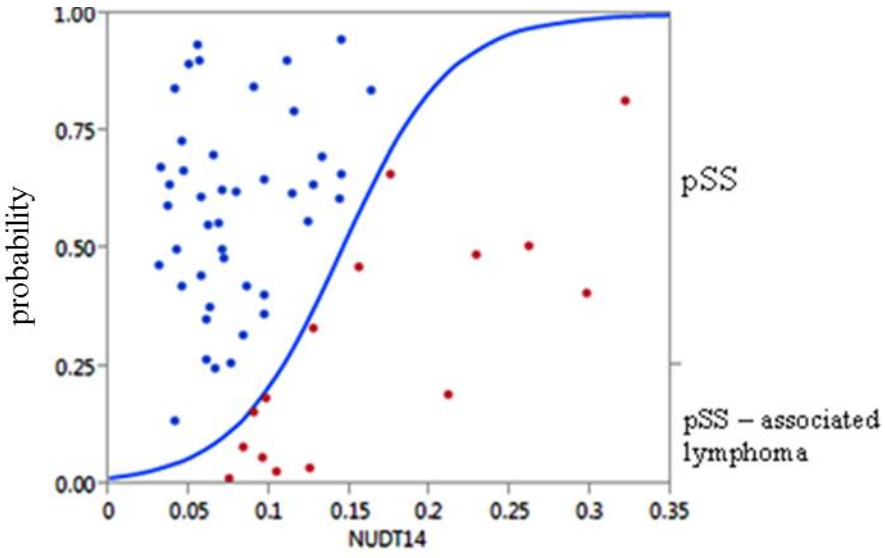


Figure 4.18 Logistic fit of NUDT14 in the Single Gene model. *The red dots represent the pSS-associated lymphoma cases, and the blue dots represent the pSS without lymphoma cases.*

Table 4.18 Statistical analysis of the Single Gene model. *A. Parameter Estimates; B. Effect Likelihood ratio tests; and C. the Lack of Fit. The results were generated by JMP SAS.*

A

Parameter Estimates				
Term	**Estimate**	**Std Error**	**ChiSquare**	**Prob>ChiSq**
Intercept	-4.2933101	1.0582869	16.46	<0.0001*
NUDT14	29.6484091	9.0728471	10.68	0.0011*

B

Effect Likelihood Ratio Tests				
Source	**Nparm**	**DF**	**ChiSquare**	**Prob>ChiSq**
NUDT14	1	1	21.8660478	<0.0001*

C

Lack of fit			
Source	**DF**	**-LogLikelihood**	**ChiSquare**
Lack of fit	57	22.516588	45.03318
Saturated	58	0.000000	Prob>ChiSq
Fitted	1	22.516588	0.8743

** = Staistically signficant*

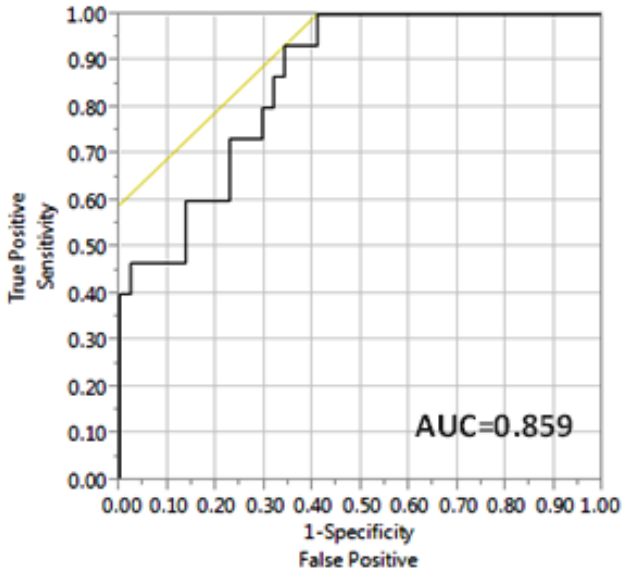


Figure 4.19 Receiver operating characteristic (ROC) curve of the Single Gene model. *This represents the sensitivity and specificity of the Single Gene model to separate pSS patients with or without lymphoma; the AUC=0.859.*

Table 4.19 Confusion matrix of the Single Gene model. *Showing actual and predicted classifications in the Discovery cohort. The results were generated by JMP SAS*

Actual classification		**Predicted**
Training	**pSS-associated lymphoma**	**pSS**
pSS-associated lymphoma	**7**	**8**
pSS	**3**	**41**

3. Testing the Two-Gene and the Single Gene models using the Discovery and Validation cohorts

Two-Gene model

The Two-Gene model was applied to the Discovery and the Validation cohorts. The prediction was visualised with a mosaic plot that shows the predictions of non-lymphoma pSS and pSS-associated lymphoma groups. The x-axis represents the subject groups, while the y-axis represents the probability of patients belonging to the lymphoma group. The probability of predicting lymphoma membership for the lymphoma group was 80% in the Discovery cohort (Figure 4.20) but decreased to 38.5% in the Validation cohort (Figure 4.21).

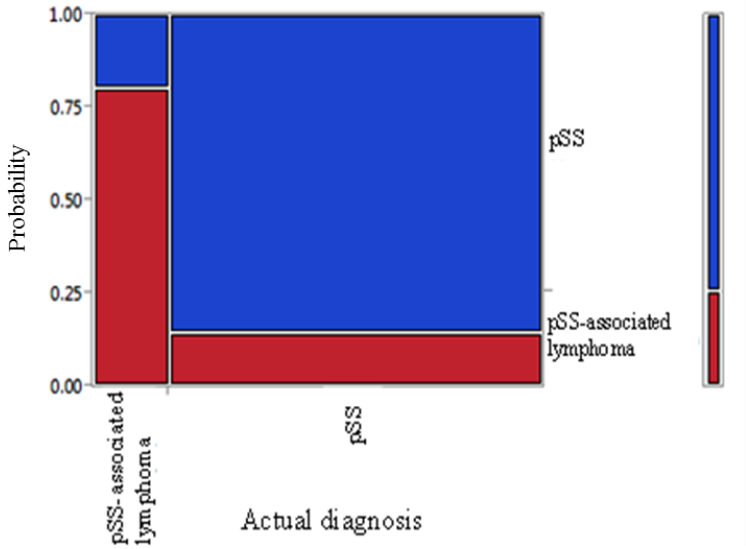


Figure 4.20 Mosaic plot of the prediction probability of the Two-Gene model in the Discovery cohort. *The red rectangles represent the proportion of patients predicted to have lymphoma. The blue rectangles represent the proportion of samples predicted not to have lymphoma. The actual diagnosis is shown on the horizontal axis. The vertical axis is the proportion of the prediction for a particular diagnosis.*

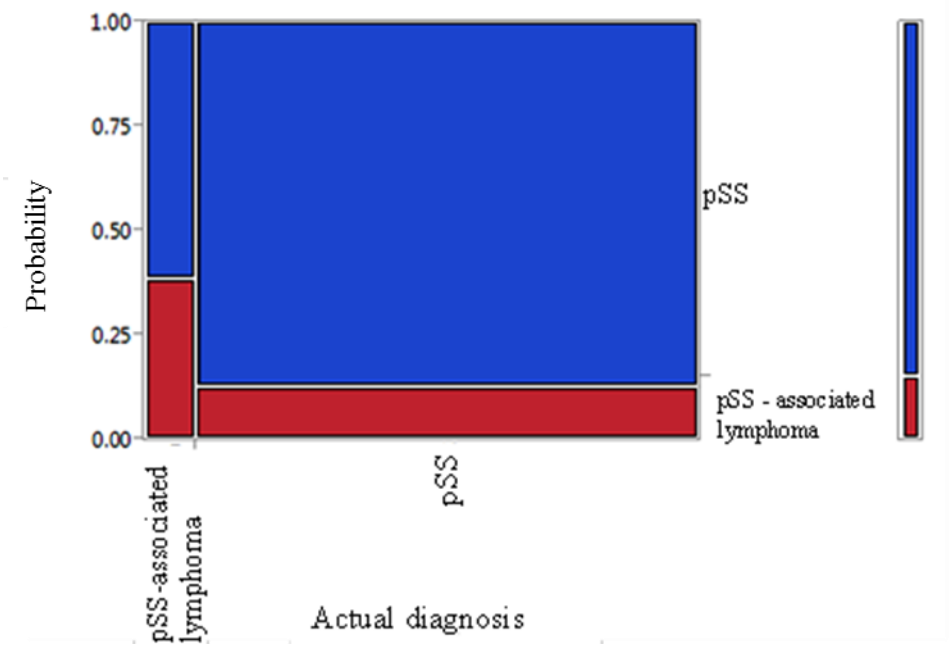


Figure 4.21 Mosaic plot of the predictions probability of the Two-Gene model in the Validation cohort. *The red rectangles represent the proportion of patients predicted to have lymphoma. The blue rectangles represent the proportion of samples predicted not to have lymphoma. The actual diagnosis is shown on the horizontal axis. The vertical axis is the proportion of the prediction for a particular diagnosis.*

The Single Gene model

When the Single Gene model is applied to the datasets of the Discovery and Validation cohorts, the proportions of the prediction for each subject group are shown on the mosaic plots. The frequency of correctly predicting lymphoma cases in the lymphoma group was 70% in the Discovery cohort (Figure 4.22), which reduced to 41.7% in the Validation cohort (Figure 4.23).

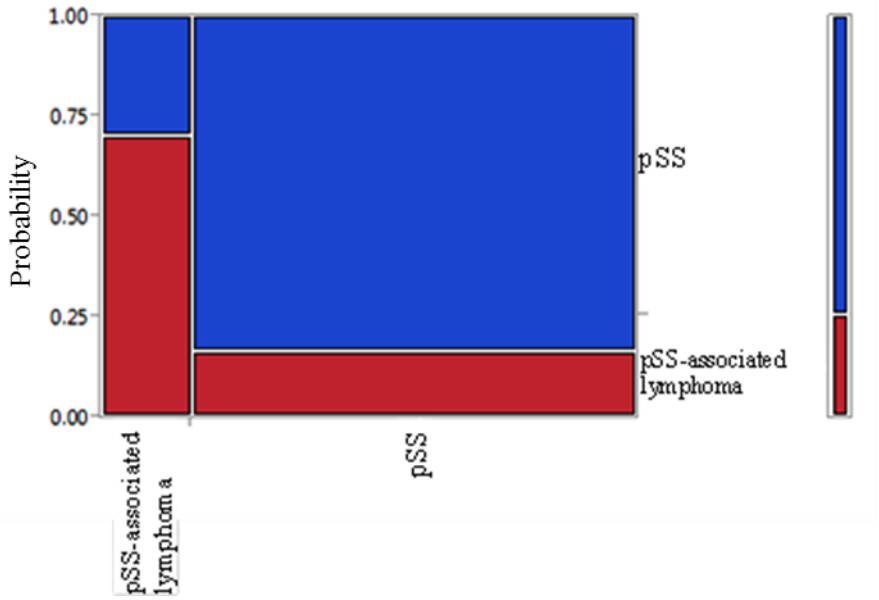


Figure 4.22 Mosaic plot of the prediction probability of the Single Gene model in the Discovery cohort. *The red rectangles represent the proportion of patients predicted to have lymphoma. The blue rectangles represent the proportion of samples predicted not to have lymphoma. The actual diagnosis is shown on the horizontal axis. The vertical axis is the proportion of the prediction for a particular diagnosis.*

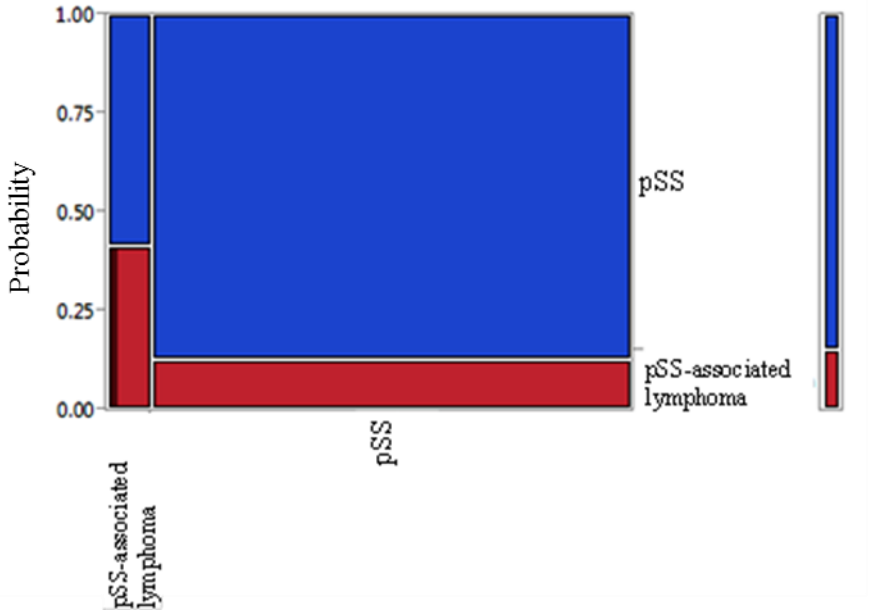


Figure 4.23 Mosaic plot of the predictions probability of the Single Gene model in the Validation cohort. *The red rectangles represent the proportion of patients predicted to have lymphoma. The blue rectangles represent the proportion of samples predicted not to have lymphoma. The actual diagnosis is shown on the horizontal axis. The vertical axis is the proportion of the prediction for a particular diagnosis.*

4.5 Discussion

The goal of this project is to identify a whole blood gene expression signature in pSS-associated lymphoma. For this goal, 144 whole blood globin-depleted samples represent the Discovery cohort. Then validate the potential gene expression signature identified in this cohort in an independent cohort. I also test the potential signature in a set of untreated pSS-associated lymphoma and I attempted to use a machine learning approach to investigate which gene might be a candidate to predict lymphoma development in pSS patients.

RNA quality

The globin mRNA depletion is a multistep process, and adding more preparation steps might result in RNA degradation, as RNA is a molecule with thermodynamic stability. RNA degradation will have adverse impact on downstream analysis of RNA expression (Auer et al., 2003). In the Discovery cohort, the RIN score of 24 samples were below 7. The sub-optimal RIN scores might be caused by the globin depletion process, as it has been discussed in Chapter 3 and also in the literature (Shin et al., 2014, Choi et al., 2014, Mastrokolias et al., 2012). On the other hand, all of the amplified cRNA samples from the Discovery cohort lacked the ~700 bp peak that represents globin mRNA in the electropherograms (Vartanian et al., 2009), indicating efficient globin mRNA depletion.

Microarray data analysis

The Illumina Human HT-12 v4 BeadChip microarray was used for the screening of the potential candidate gene expression signature of pSS-associated lymphoma. In general, avoiding bias is crucial in such experiments. For this reason, samples from the five subject groups were selected randomly to be run on the same microarray chip (each chip can measure up to 12 samples). This approach will avoid bias and systematic error being introduced (Pannucci and Wilkins, 2010). Because the sample size of the Discovery cohort is relatively large, it was therefore unavoidable to perform the microarray at different times and in different batches, which in turn will introduce batch effects. Batch effect detection and correction were therefore necessary when I analyse the microarray data. The correction of batch effects during microarray data analysis is important, and so is the choice of method to achieve that, in order to yield a robust data set (Larsen et al., 2014). Many methods have been developed for batch effect correction. The ComBat method has been suggested to have better performance than other batch effect correction methods (Chen et al., 2011). The

ComBat method collects information across genes and it minimises the parameter of the batch effects in the direction of the overall mean of the batch across the genes (Johnson et al., 2007, Larsen et al., 2014).

In this project, I have identified many candidate genes from the Discovery cohort as a potential whole blood gene expression signature for pSS-associated lymphoma (68-DEGs-Mi). The list of candidate genes is related to many different biological functions. For example, some of these genes have been linked to lymphoma development (such as *BCL11B* and *MYC)*. Whether these genes are actually involved in lymphoma development in pSS is still under investigation.

Technical validation of the microarray data

I used real-time RT-PCR to technically validate the 68-DEGs-Mi. The results shown that DEGs were confirmed in 26 / 60 DEGs by qRT-PCR (26-DEGs-MiPCR). The are several reasons why differential expression was not confirmed with RT-PCR for the remaining 34 genes. It has been suggested that filtering of the microarray data by the fold change and p values is important to gain robust results (Morey et al., 2006). The fold changes of the DEGs from the microarray data ranged from 1.2-1.8, which is relatively small. It is therefore unsurprising that not all of the DEGs were validated with qRT-PCR. A previous report has indicated that the correlation between microarray data and qRT-PCR decreases when the fold changes were less than 1.5 (Dallas et al., 2005). Furthermore, the biological and technical differences between the two techniques must be taken into consideration as a source of variation (Wurmbach et al., 2003, Chuaqui et al., 2002). One of the genes, *HLA-DRB1,* failed to generate a normal amplification plot during the qRT-PCR. The failure in detecting this gene with RT-PCR might be due to the RNA concentration of the samples. Moreover, another reason of this failure is due to the polymorphism of *HLA-DRB1*. The polymorphism of *HLA-DRB1* led to the use of alternative PCR techniques such as sequence-specific primers PCR to detect its expression (Song et al., 2012, Gersuk and Nepom, 2009). Certainly, the *HLA-DRB1* gene plays an important role in the pathogenesis of pSS and other autoimmune diseases (Guggenbuhl et al., 1998, Doherty et al., 1998). Therefore, it warrants further investigation in the future.

Twenty of these 26 genes were identified from the main analysis (Analysis A), and the remaining 6 genes were identified from the other 3 sub-analyses. The validation of these 6 additional genes justifies the inclusion of these analyses.

Biological validation of whole blood gene expression signature in pSS-associated lymphoma

Biological validation was carried out using an independent cohort. Differential expression of three genes (*BMS1*, *NUDT14* and *MGST3*) was confirmed in this cohort. These three genes are referred to as the **3-gene biosignature of pSS-associated lymphoma**. The downregulated *BMS1* or ribosome biogenesis factor, which is also known as *ACC* and *BMS1L*, is located in the nucleus of the cell. *BMS1* encodes a protein called ribosome biogenesis factor, which plays a role in ribosomal assembly and is critical for the 40S ribosome formation. Limited data are available concerning the role of *BMS1*. Nonetheless, it has been shown that the decrease in *BMS1* level leads to late maturation of 18S rRNA. Furthermore, *BMS1* mutation is linked to aplasia cutis congenital (ACC), which is caused by defects in cell cycle, reducing cell proliferation and affecting skin morphology (Marneros, 2013). No information linking this gene to any type of malignancy has been reported in the literature.

NUDT14 or nudix (nucleoside diphosphate linked moiety X) – type motif 14, which was upregulated in pSS-associated lymphoma, is also known as NUDIX MOTIF 14, uridine diphosphate glucose pyrophosphatase (UGPP), UDPG pyrophosphatase, and UGPPase. The protein that this gene encodes is uridine diphosphate glucose pyrophosphatase, and it has a role in producing glucose 1-phosphate and UMP by hydrolysing UDPG (Yagi et al., 2003). *NUDT14* plays a role in the pyrophosphorylation of substrates that contain nucleosides (Yagi et al., 2003). In mice, the role of *NUDT14* appears to be more related to ADP-ribose hydrolysis than UDP-glucose hydrolysis (Heyen et al., 2009). *NUDT14* is located on chromosome 14q32.3. Deletion of *NUDT14* has been described in the 14q32 deletion syndrome, which causes mild facial dysmorphisms and intellectual disability (Holder et al., 2012).

Regarding the involvement of *NUDT14* in malignancies, Choi and co-workers (Choi et al., 2011) reported an upregulation of *NUDT14* in a rectal carcinoma cell line. However, the role of this gene in malignancy development is still unknown. Recently, modulation role of *NUDT14* in viral infection, in particular, the human cytomegalovirus (HCMV), was reported. Wang and colleagues (Wang et al., 2016) found that the HCMV RL13 protein, which is encoded by the HCMV RL 13 gene, is co-localised with the *NUDT14* protein in the cell membrane and the cytoplasm of human embryonic kidney HEK293 cells, and that they interact with each other. Furthermore, when the *NUDT14* expression decreased, the number of the viral DNA copies in the infected cells increased. However, the overexpression of *NUDT14* had no effect on the number of the viral DNA copies. *NUDT14* is differentially expressed in pSS-associated lymphoma cases regardless of their treatment status. Therefore, undertaking more studies is worthwhile to investigate the possible role of this gene in lymphoma development.

MGST3 (microsomal glutathione S-transferase 3) is also known as GST III and GST-3. This gene encodes a protein named microsomal glutathione S-transferase 3. The protein is located mostly in the endoplasmic reticulum membrane, commonly expressed in skeletal muscles, adrenal cortex, and heart. The main function of this protein is to serve as a glutathione peroxidase (Jakobsson et al., 1997). The microsomal glutathione S-transferase 3 (*MGST3*) is not evolutionarily related to the glutathione S-transferase (GST) gene family, but it encodes membrane-bound enzymes that possess GST-like activity (Nebert and Vasiliou, 2004). Young and Woodside (2001) have reported that the glutathione peroxidases contribute to the process of detoxification from both endogenous and exogenous toxins by eliminating lipid peroxidase (Young and Woodside, 2001). Furthermore, *MGST3* is highly abundant in the liver as it is involved in drug metabolism (Morgenstern et al., 2011, Uno et al., 2013). Along with its main function in glutathione-dependent peroxidase activity, *MGST3* also contributes to the synthesis of the leukotriene C4, which leads to the production of inflammatory and hypersensitivity mediators (Ford-Hutchinson, 1990).

In the validation cohort, *OAF, LEF1,* and *DRAP1* had adjusted p values between 0.05 and 0.25 with the same direction of regulation to the Discovery cohort. Five additional genes have a p value more than 0.25 but the same direction of regulation of the Discovery cohort.

these genes are *CBLL1*, *C10orf32 (BORCS7)*, *DYNLL1*, *LGALS1*, and *SGK223*. I believe that these genes are also worthy of further investigation as potential gene expression signatures for pSS-associated lymphoma in the future. This is because the probability values report the likelihood of an effect being detected by chance but not the size of the effect. The p values are often affected by the sample size (Sullivan and Feinn, 2012).

Identify possible whole blood gene expression signature for untreated pSS-associated lymphoma

In this project, I had the opportunity to test whether the potential signature for pSS-associated lymphoma was also present in untreated lymphoma samples. When compared to non-lymphoma pSS samples, 6 genes were found to be differentially expressed. Interestingly, *NUDT14*, which was upregulated in the treated lymphoma cases in both the Discovery and Validation cohorts, was also upregulated in untreated lymphoma cases. *DRAP1* and *DYNLL1* may also be of interest. Differential expression of these two genes was confirmed in the treated lymphoma samples from the Discovery cohort. These two genes also have the same direction of regulation in the treated lymphoma cases in the Validation cohort albeit with a p values of greater than 0.05. *DRAP1* or DR1-associated protein 1 (also known as negative cofactor 2 alpha (NC2-alpha)) is a transcriptional repressor. In order to initiate a transcription, the assembly of RNA polymerase II is required as well as general transcriptional factors (GTFs) such as TFIIA, TFIIB, and TFIID. *DRAP1* interacting with TATA-binding protein (TBP) of TFIID precludeing TBP-DNA complex formation (Schluesche et al., 2007). Furthermore, evidence showed that in hypoxic conditions, the activity of NC2 increases suggesting that a *DRAP1* may play an important role in gene regulation in hypoxic conditions (Denko et al., 2003). In cancer settings, upregulation of *DRAP1* has been reported to be an accurate predictor of radioresistance in prostate cancer patients undergoing radiotherapy, although the data were derived from a small dataset and therefore needs to be replicated (Valdagni et al., 2009).

DYNLL1 or dynein light chain encodes the protein LC8-type 1, which is an enzyme complex that has a role in intracellular transport and motility. Evidence has revealed that the binding of the *DYNLL1* encoded protein to pilin of the pathogen *Pseudomonas aeruginosa* stimulates an inflammatory response in the host, which might be harnessed for treatment purposes (Kausar et al., 2013). LC8 contributes to microtubule stability as

overexpression of LC8 increases microtubule acetylation and reduces microtubule susceptibility to cold- and nocodazole-induced depolymerisation, while decrease of LC8 leads to disruption of bipolar spindle assembly, suggesting a novel microtubule-associated protein-like function for LC8 (Asthana et al., 2012). LC8 also regulates the activity of the proapoptotic Bcl-2 family member BH3-only protein Bim, through impounding Bim to the cellular microtubules in the state of healthy cells (Puthalakath et al., 1999). Therefore, the regulatory role of LC8 on Bim might have a role in malignancy development and resistance to cancer treatments (Izidoro-Toledo et al., 2013). Recently, *DYNLL1* was reported to interact with human immunodeficiency virus type 1 integrase (HIV-1 integrase (IN)), leading to the reverse transcription and the multiplication of the virus (Jayappa et al., 2015). More experiments are needed to understand the possible role of *DYNLL1* in pSS-associated lymphoma.

Identifying the most important genes in predicting the group's membership in pSS-associated lymphoma

To identify the key constituent genes of the pSS-associated lymphoma gene signature, I have examined two prediction models using the qRT-PCR data from the Discovery cohort based on the 60 DEGs. Both models indicated that *NUDT14* is the best predictor or classifier of pSS-associated lymphoma membership. The data generated from these models were also consistent with my experimental findings, as *NUDT14* has been shown to be differentially expressed in the two treated lymphoma cohorts and the untreated lymphoma cohort. Although there is no data in the literature that implicate *NUDT14* in lymphoma development, my data strongly suggest that further investigation into the role of *NUDT14* in the pathogenesis of lymphoma in pSS is worthwhile.

In conclusion, the data from the microarray and qRT-PCR from the Discovery and Validation cohorts have identified several genes that warrant further investigations in larger cohorts of treated and untreated lymphoma as well as in longitudinal studies. Many of these candidate genes have not been linked directly to lymphoma development. Therefore, if the roles of these signature genes for pSS-associated lymphoma are confirmed, it will unravel novel mechanisms of lymphomagenesis in pSS.

Chapter 5
Biological pathway analysis in primary Sjögren's Syndrome - associated lymphoma

5.1 Introduction:

Biological pathways in pSS, lymphoma and pSS-associated lymphoma

A biological pathway refers to a series of molecular interactions within the cell that lead to the production of a certain molecule or a biological change within the cell. A valuable approach to the analysis of microarray gene expression data is to study whether there are indications of changes in biological pathways within the datasets. This linkage between the gene expression data and biological processes will help to gain a better understanding of a given disease.

For instance, in pSS, analysis of microarray gene expression data have led to the discovery that the type I interferon pathway plays a major role in pSS pathology (Gottenberg et al., 2006, Emamian et al., 2009). This observation is also consistent with the hypothesis that viruses being involved in the aetiology of pSS, as viral infections are potent inducers of IFN-alpha (Bave et al., 2005). The 'Toll-like receptor (TLR) pathway' is another pathway that has been shown to be activated in pSS. TLRs appear to be linked to salivary-gland epithelial cell pathology in pSS; epithelial cells from salivary glands of pSS patients have a higher level of TLR expression compared to those from healthy controls (Spachidou et al., 2007).

Many pathways have also been shown to be dysregulated in lymphoma. For example, it has been shown that the NF-κB signalling pathway is activated in lymphoma (Troppan et al., 2015). Moreover, the NF-κB signalling pathway can be activated by the stimulation of B-cells receptor in Diffused Large B Cells lymphoma (DLBCL). NF-κB signalling pathway also has an activating effect on the JAK2/STAT3 signalling pathway by stimulating TLR and interleukin receptors (IL-1 and IL-18), which leads to the progress of DLBCL lymphoma (Ngo et al., 2011, Turturro, 2015). Recently, the discovery of therapeutic drug that act as an inhibition of antigen receptor signalling pathway, including B-cell receptors, T-cells receptors, TLR and BAFF might be a useful treatment for B cell lymphoma (Zhang

et al., 2015). Interestingly, the NF-κB signalling pathway has also been implicated in pSS pathogenesis (G'Sell et al., 2015). It is therefore possible that exploring the changes in biological pathways in pSS-associated lymphoma may improve our understanding of lymphomagenesis in pSS, reveal the biological meaning of the gene-expression signature for pSS-associated lymphoma and help to establish novel therapies.

To date, there has been no published research on the use of microarray data to explore the biological processes that may be involved in pSS-associated lymphoma. Therefore, this study is the first attempt to investigate such relationships.

Ingenuity Pathway Analysis (IPA):

Many programs have been developed to analyse biological pathways for microarray data (Table 5.1). MetaCore by GeneGo, GenMapp and Ingenuity Pathway Analysis (IPA) (Bogner et al., 2011) are examples of such software. A key difference between these different tools is how up-to-date their databases are. In this regard, the knowledge database of Ingenuity Pathway Analysis is manually curated, literature supported, and frequently updated. IPA is a web-based software program that enables researchers to discover the biological significance of their high-throughput biological data (Kramer et al., 2014). With IPA, specific pathways and networks that might be relevant to a certain disease (for example, pSS-associated lymphoma) can be identified. IPA uses a p-value calculated by the Fisher's exact test, right-tailed and an activation z-score to present the outcome (Ingenuitysystems, IPA). A brief definition of some key terms used in the IPA is provided below:

Focus genes: Probe identifiers of significantly differentially expressed genes (DEGs) uploaded from my experiment that are represented in the suggested network or a pathway.

P-value: reflects the significance metric representative of the probability that the molecules or the genes are involved in a given pathway or network not by random chance. The lower the p-value below a given threshold (0.01 or 0.05) is the more significant.

Z-score: reflects the statistical state of the direction of a relationship that is expected of the genes in a given experiment and the regulation direction that derived from the literature in IPA; normally, a z-score > 2 or z-score < -2 is considered significant.

Score: is the number of genes in the network, including the genes of the uploaded DEGs list from a given experiment and the genes that are added from the indirect interactions by IPA.

Overlap p-value: is used in the upstream regulator analysis. It represents whether there is any statistically significant in common between the DEGs in a given experiment and their regulators.

Ratio: the ratio is used in the canonical pathway analysis. It represents the ratio of the genes in the uploaded DEGs list from my experiment and the total number of genes that make up each pathway. Each indirect association is supported by literature.

IPA maps a list of genes (commonly the list of DEGs) in a certain experiment to the Ingenuity Knowledge Base (IKB). The IKB is a database with detailed information on known biological interactions and functional annotations. The information was built from existing individual relationships between different molecules such as genes, proteins, cells, chemicals and diseases that can be found in the literature and is updated on a regular basis (http://www.ingenuity.com, 2015). Using IPA has many advantages: it provides biological function as well as sub-functions for each gene. Moreover, the bibliography in which the relationship is defined is provided through a direct link to the resource. Furthermore, the network analysis shows the interactions of the genes based on current knowledge (Jiménez-Marín et al., 2009).

Table 5.1 Software available for microarray data analyses(Jin et al., 2014b)

Software	Feature	Annotations	URL
PathMAPA	A tool for displaying gene expression and performing statistical tests on metabolic pathways at multiple levels for Arabidopsis, based on expression data	Local databases	http://bioinformatics.med.yale.edu/pathmapa.htm
MetaCore	Based on a high-quality, manually curated database, MetaCore is an integrated software suite for functional analysis of microarray, metabolic, SAGE, proteomics, NGS, copy number variation, siRNA, microRNA and screening data	MetaRodent, MetaLink, MetaSearch	http://www.genego.com/
Ingenuity Pathway Analysis (IPA)	A comprehensive software/database search tool for finding functions and pathways for specific biological states. Manually curated in house	GO, KEGG, BIND	https://www.ingenuity.com/
ePath3D	An easy-to-use and powerful software for creating and managing illustrated 3D pathways for publications and presentations	eProtein, ePathway	http://www.proteinlounge.com/epath3d/

Pathway Builder	An online pathway drawing tool which is the fastest and easiest method of creating signal transduction pathways, enabling the users to design their own project or use pre-made pathway templates to help get them started	GenBank, Uniprot/Swiss-Prot, TrEMBL, KEGG, ENZYME, *etc*	http://www.pathwaybuilder.com/
Interactive Pathways Explorer (iPath)	A web-based tool for the visualization, analysis and customization of various pathways maps from KEGG. The recently released version 2 could deal with metabolic pathway, regulatory pathway and biosynthesis of secondary metabolites	KEGG	https://pathwayexplorer.genome.tugraz.at
GSEA-P & R-GSEA	GSEA-P is a desktop application for Gene Set Enrichment Analysis, with a friendly graphic interface. R-GSEA is provided as a stand-alone R program.	MSigDB, Gene Set Cards, GEO	http://www.broadinstitute.org/gsea/
DAVID	A tool for augmenting and integrating functional annotations from other databases	KEGG,GO	http://david.abcc.ncifcrf.gov/
MetaCyc	Applications include serving as an encyclopaedia of metabolism, providing a reference data set for the computational prediction of metabolic pathways in sequenced organisms, supporting metabolic engineering and helping to compare biochemical networks	KEGG, BioCyc, EcoCyc	http://metacyc.org/
Reactome	Intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modelling, systems biology and education	KEGG	http://www.reactome.org/
GenMAPP	Designed to visualize gene expression and other genomic data on maps representing biological pathways and groupings of genes	GenMAPP, GO	http://www.genmapp.org
FunCluster	An integrative tool for analysing gene co-expression networks from microarray expression data; the analytic model implemented in the library involves two abstraction layers: transcriptional and functional (biological roles)	GO, KEGG	http://corneliu.henegar.info/FunCluster.htm
Graphite web	A novel web tool for pathway analyses, consisting of topological-based analysis and network visualization for gene expression data of both microarray and RNA-sequencing experiments	KEGG, Reactome	http://graphiteweb.bio.unipd.it/

5.2 Aim and experimental design:

In this chapter, I will describe the biological pathway analysis in pSS-associated lymphoma, using the 68 DEGs from the microarray data generated in the Discovery cohort (68-DEGs-Mi) in order to gain a better understanding regarding the biology of pSS-associated lymphoma. I used IPA for this analysis. In Chapter 4, I described the list of DEGs in pSS-associated lymphoma from the microarray experiment (68-DEGs-Mi), which is a combined list of DEGs that were generated from the four analyses of the microarray data using different criteria regarding sample inclusion.

The reason behind choosing the 68-DEGs-Mi is that the list contains the largest number of DEGs in comparison with the validated lists (26-DEGs-MiPCR and the 3-gene biosignature in pSS-associated lymphoma); the more genes the list contains the more associated pathways can be identified, the added power from 4 variations of the differential expression tests could yield more informative, disease relevant pathways. The uploaded DEGs list can be found in Table 5.2.

Table 5.2 The 68-DEGs-Mi list used in Ingenuity pathway analysis

Gene symbol	adjusted P-value	Fold change	regulation in pSS-associated lymphoma
LGALS1	0.019861522	1.36	upregulated
DRAP1	0.009554204	1.32	upregulated
KCTD12	0.032858824	1.28	upregulated
RBP7	0.027203463	1.26	upregulated
DYNLL1	0.010184961	1.25	upregulated
UBXN11	0.001878176	1.25	upregulated
NUDT14	0.001878176	1.24	upregulated
C10orf32	0.024934079	1.23	upregulated
HNMT	0.046978282	1.23	upregulated
SRP14	0.041910977	1.23	upregulated
PSMC1	0.040796915	1.23	upregulated
MGST3	0.042260308	1.23	upregulated
OAF	0.046978282	1.22	upregulated
ATG12	0.010241707	1.22	upregulated
SEC61G	0.046978282	1.21	upregulated
RAB37	0.012720244	1.21	upregulated
HCFC1R1	0.009554204	1.21	upregulated
CNPY3	0.022617674	1.21	upregulated
RNF7	0.029377902	1.20	upregulated
EHBP1L1	0.039664006	1.20	upregulated
HLA-DRB1	0.048625399	3.59	downregulated
RNA28S5	0.001454575	1.67	downregulated
LEF1	0.038999435	1.42	downregulated
SPOCK2	0.045033431	1.41	downregulated
ETS1	0.029209021	1.39	downregulated
POM121C	0.00623275	1.35	downregulated
RPL23AP5	0.011557085	1.32	downregulated
MYC	0.019861522	1.32	downregulated
CYFIP2	0.046978282	1.32	downregulated
SGK223	0.044226587	1.32	downregulated
HSP90B1	0.007110164	1.31	downregulated

BCL11B	0.03531743	1.30	downregulated
RASGRP1	0.048625399	1.30	downregulated
WAC	0.001538691	1.29	downregulated
ITK	0.03462919	1.29	downregulated
DDB1	0.00579247	1.28	downregulated
MAGED1	0.044226587	1.26	downregulated
ESYT1	0.044226587	1.26	downregulated
PRPF8	0.039035069	1.26	downregulated
SUN2	0.041735128	1.26	downregulated
HNRNPUL1	0.038999435	1.25	downregulated
CDR2	0.020642125	1.24	downregulated
NAT10	0.033024722	1.24	downregulated
CD96	0.038246765	1.24	downregulated
VCP	0.019508447	1.24	downregulated
RBL2	0.011367948	1.23	downregulated
BAG3	0.023506309	1.23	downregulated
ATP1A1	0.012439918	1.23	downregulated
NCSTN	0.046450529	1.23	downregulated
ALDH9A1	0.013677032	1.23	downregulated
LRIG1	0.015856148	1.23	downregulated
RRN3	0.007770882	1.23	downregulated
HNRNPA1P10	0.044075354	1.23	downregulated
PAF1	0.019861522	1.23	downregulated
SF3A1	0.044226587	1.23	downregulated
LRFN3	0.045230731	1.22	downregulated
SDHA	0.038999435	1.22	downregulated
KHDRBS1	0.040527676	1.22	downregulated
CDV3	0.03462919	1.22	downregulated
FOXJ3	0.015259771	1.21	downregulated
SMARCA2	0.029889934	1.21	downregulated
CBLL1	0.024405233	1.21	downregulated
BMS1	0.026000737	1.20	downregulated
HSPA9	0.020187775	1.20	downregulated
RPA2	0.022508992	1.20	downregulated
HNRNPDL	0.035037482	1.20	downregulated
PRKCQ	0.047925118	1.20	downregulated
BTBD11	0.027803403	1.20	downregulated

5.3 Results:

5.3.1 The Canonical pathways of pSS-associated lymphoma

A total of 188 canonical pathways have been identified as being over-representated in pSS-associated lymphoma ($p < 0.05$, Fischer's exact test, right-tailed). Moreover, the number of genes in common between 68-DEGs-Mi and the total number of the genes included in each pathway was calculated by IPA and were shown as "ratio". The top 5 most significant (i.e., with the lowest p-values,) canonical pathways that were enriched in pSS-associated lymphoma are shown in Table 5.3. 'Aryl Hydrocarbon Receptor (AHR) signalling' was the statistically most significant pathway in pSS-associated lymphoma. The other top 4 pathways were 'histamine degradation,' 'unfolded protein response,' 'Neuregulin signaling' and 'T-cell receptor signalling.' Supplementary table S5 shows the complete list of the 188 canonical pathways identified in this analysis.

Of note, one of the five genes that was in common between the genes in the 'AHR signalling' pathway and the 68-DEGs-Mi is *MGST3,* which was also one of the three genes in the 3-gene biosignature of pSS-associated lymphoma. Indeed, when I focussed on the three genes in the 3-genes biosignature of pSS-associated lymphoma (*NUDT14*, *MGST3* and *BMS1*) and the canonical pathways that these genes might be involved in (Table 5.4), interesting observations emerged. *NUDT14* was involved in six canonical pathways. *MGST3* was involved in five other canonical pathways, in addition to the 'AHR signalling' pathway. Both genes appeared to be involved in metabolic pathways. The third gene, the down-regulated *BMS1,* was not associated with any of the canonical pathways identified in pSS-associated lymphoma.

Furthermore, *DYNLL1,* the gene that I found to be significantly differentiated in the untreated pSS-associated lymphoma was involved in one canonical pathway named 'Phagosome maturation' with a –log (p-value) of 1.20E+00. While *DRAP1* did not show any associated with any of the canonical pathways identified in pSS-associated lymphoma.

The IPA, however, was unable to predict the directionality (i.e. whether the pathways were activated or suppressed) of these enriched canonical pathways (the z-score was 0 or near 0 in all pathways). Figure 5.1 shows the directionality analysis of the top 10 canonical pathways.

Table 5.3 Top 5 canonical pathways in pSS-associated lymphoma identified by IPA

Ingenuity Canonical Pathways	**-log (p-value)****	**No. of genes in common between the canonical pathway and genes of interest**	**Genes in common between the canonical pathway and genes of interest**
Aryl Hydrocarbon Receptor Signalling	3.99E+00	5/135 (3.7 %)	*MYC, ALDH9A1, RBL2, HSP90B1,* ***MGST3****
Histamine Degradation	3.12E+00	2/12 (16.7 %)	*HNMT, ALDH9A1*
Unfolded Protein Response	3.09E+00	3/53 (5.7 %)	*HSPA9, HSP90B1, VCP*
Neuregulin Signalling	2.50E+00	3/85 (3.5 %)	*MYC, HSP90B1, PRKCQ*
T-Cell Receptor Signalling	2.38E+00	3/94 (3.2 %)	*PRKCQ, ITK, RASGRP1*

** The gene in bold represents the gene from the three-gene biosignature of pSS-associated lymphoma*

*** The p values are presented in a –log scale, therefore, the higher the value on –log scale, the higher is the statistical significance (i.e. lower p-value).*

Table 5.4 The canonical pathways involvement of the three genes in the 3-genes biosignature in pSS-associated lymphoma

Ingenuity Canonical Pathways	**-log(p-value)**	**Ratio** **68-DEGs-Mi / genes in pathway**
NUDT14		
D-myo-inositol (1,4,5,6)-Tetrakisphosphate Biosynthesis	1.18E+00	1.64E-02
D-myo-inositol (3,4,5,6)-tetrakisphosphate Biosynthesis	1.18E+00	1.64E-02
D-myo-inositol-5-phosphate Metabolism	1.08E+00	1.45E-02
3-phosphoinositide Degradation	1.08E+00	1.44E-02
3-phosphoinositide Biosynthesis	1.03E+00	1.34E-02
Super-pathway of Inositol Phosphate Compounds	8.69E-01	1.08E-02
MGST3		
Aryl Hydrocarbon Receptor Signalling	3.99E+00	3.70E-02
Xenobiotic Metabolism Signalling	1.93E+00	1.56E-02
NRF2-mediated Oxidative Stress Response	1.63E+00	1.69E-02
Glutathione Redox Reactions I	1.22E+00	5.56E-02
Glutathione-mediated Detoxification	1.10E+00	4.17E-02
LPS/IL-1 Mediated Inhibition of RXR Function	7.92E-01	9.62E-03
BMS1		
No significant pathway identified	N/A	N/A

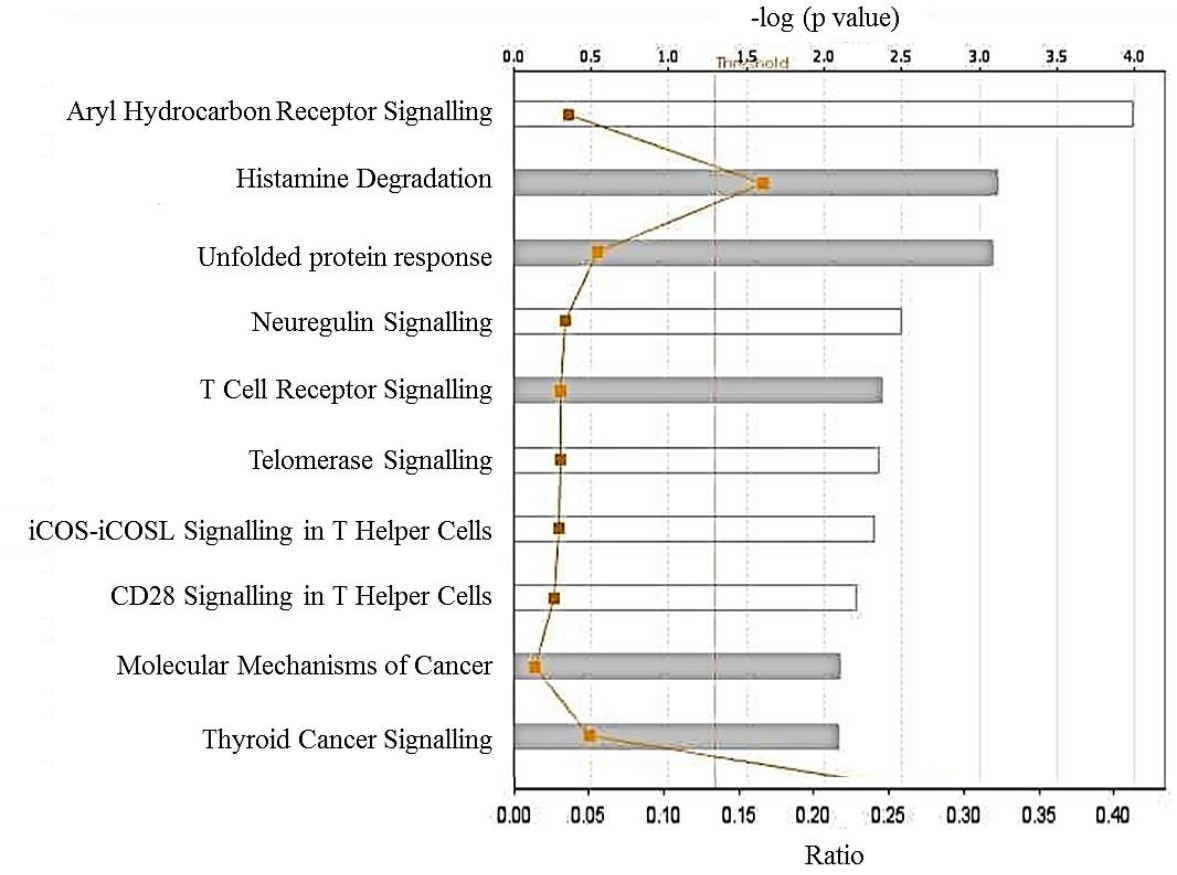


Figure 5.1 Bar chart of the top 10 canonical pathways in pSS-associated lymphoma identified by IPA. *Grey bars represent that no prediction of directionality can be made. White bars represent the canonical pathways with a z-score near 0. The vertical threshold line in yellow represents the p-value cutoff of statistical significance. The yellow dots represent the ratio of the genes in common between the 68-DEGs-Mi and the total number of genes that made up each pathway.*

5.3.2 The Downstream Effects Analysis of pSS-associated lymphoma:

Another approach to biological pathway analysis is to study the downstream effects of the genes in pSS-associated lymphoma (68-DEGs-Mi) and their connection with known biological functions or pathological conditions. For example, gene A may be reported to be associated with cancer development or gene B with apoptosis. Such knowledge may provide additional insight apart from the biological pathways of the pSS-associated lymphoma. In this analysis 'disease and function' of the 68-DEGs-Mi were analysed by

IPA. The directionality of the prediction was made based on casual effects derived from the literature. Three possible predictions can be made according to the direction of change as follow:

1. If the direction of change is the same as reported in the literature then the function is considered increased in pSS-associated lymphoma
2. If the direction of change was opposite to that reported in the literature then the function is decreased in pSS-associated lymphoma
3. If the direction of change does not have a clear relationship with that reported in the literature then IPA will not make a prediction.

Ten "diseases and functions" were identified (Figure 5.2). The ten diseases and functions were analyzed using the z-score algorithm, which avoids creating random significant predictions in 68-DEGs-Mi (Ingenuitysystems, IPA). Nine of these functions were increased and one was decreased. Three of these diseases and functions were biased, which suggest that their regulations are skewed to particular direction (Table 5.5).

Furthermore, 490 other "diseases and functions" were identified by IPA to be associated with the genes from the 68-DEGs-Mi list. These "diseases and function" were all statistically significant but IPA was unable to make prediction of the directionality of the expected relationship according to the z-score (i.e., $-2 < \text{z-score} < 2$).

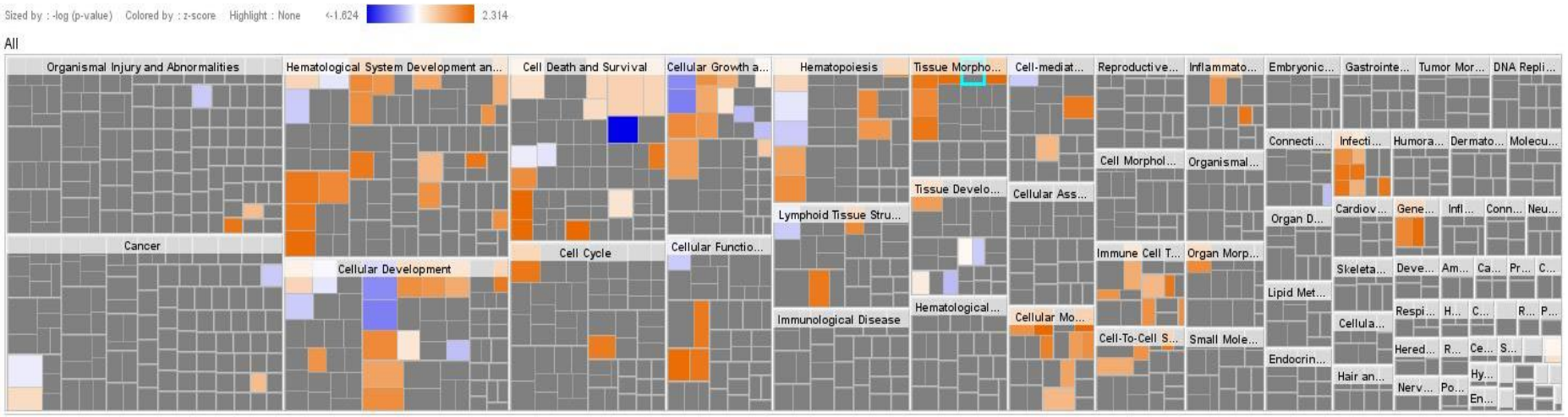


Figure 5.2 Heatmap of the downstream effects analysis of the pSS-associated lymphoma 68-DEGs-Mi by IPA. *The square size corresponds to the statistical significance of the p value (i.e., larger square = more significant (lower p-value)). The colours of the squares reflect the direction of change (activated or inhibited). Orange: IPA predicts that the biological process or function is increased in pSS-associated lymphoma with a positive z-score (z-score ≥ 2). Blue: IPA predicts that the biological process or function is decreased in pSS-associated lymphoma with a negative z-score (z-score ≤ -2) Gray: represent that no prediction can be made with regard to directionality. White: represent the z-score is near 0. The strength of the prediction is represented by the intensity of the colour.*

Table 5.5 The Downstream Effects analysis of pSS-associated lymphoma 68-DEG-Mi by IPA. *The ten statistically significant biological processes and functions and the genes involved*

Categories	Diseases or Functions Annotation	p-Value	Predicted Activation State	Activation z-score	Notes	Genes	No. of genes
Infectious Diseases, Organismal Injury and Abnormalities	infection of embryonic cell lines	2.51E-02	Increased	2	bias	*HNRNPDL,KHDRBS1,PRPF8,SF3A1*	4
Haematological System Development and Function, Tissue Morphology	quantity of T lymphocytes	5.52E-04	Increased	2.211		*ETS1,HLA-DRB1,HSP90B1,ITK,LGALS1, PRKCQ,RASGRP1*	7
Cellular Function and Maintenance	cellular homeostasis	5.42E-04	Increased	2.228		*ATG12,ATP1A1,BAG3,BCL11B,ETS1,HSP90B1, ITK,LEF1,LGALS1,MYC,NCSTN,PRKCQ, RASGRP1,VCP,WAC*	15
Cell Death and Survival	cell viability	2.04E-03	Increased	2.253	bias	*ATG12,BAG3,HNRNPUL1,HSP90B1,LEF1, LRIG1,MYC,NCSTN,PRKCQ,PRPF8,SF3A1, SMARCA2,VCP*	13
Cell Death and Survival	cell survival	1.05E-03	Increased	2.415		*ATG12,BAG3,HNRNPUL1,HSP90B1,LEF1,LRIG , MYC,NCSTN,PRKCQ,PRPF8,RRN3,SF3A1, SMARCA2,VCP*	14
Cellular Movement	cell movement	8.82E-03	Increased	2.539		*ATG12,BCL11B,CBLL1,ETS1,ITK,KHDRBS1, LEF1,LGALS1,LRIG1,MYC,PAF1,PRKCQ,RASG ,RP1,RNF7,SUN2,VCP*	16
Cell Death and Survival	cell viability of tumour cell lines	9.35E-05	Increased	2.557	bias	*BAG3,HNRNPUL1,HSP90B1,LEF1,LRIG1,MYC, NCSTN,PRKCQ,PRPF8,SF3A1,SMARCA2,VCP*	12
Tissue Morphology	quantity of cells	1.40E-03	Increased	2.602		*DYNLL1,ETS1,HLA-DRB1,HSP90B1,HSPA9, ITK,KHDRBS1,LEF1,LGALS1,MYC,NAT10, PRKCQ,RASGRP1,RBL2*	14
Cellular Movement	migration of cells	6.32E-03	Increased	2.893		*ATG12,BCL11B,CBLL1,ETS1,ITK,KHDRBS1, LEF1,LGALS1,MYC,PAF1,PRKCQ,RASGRP1, RNF7,SUN2,VCP*	15
Cell Death and Survival	cell death of melanoma cell lines	3.96E-04	Decreased	-2.166		*BAG3,CYFIP2,ETS1,MYC,PAF1,SF3A1*	6

5.3.3 The Molecular Networks analysis of the 68-DEGs-Mi of pSS-associated lymphoma:

Molecular Networks analysis provide information regarding the visualization of the relations of the genes of interest and the directionality of their regulation. Molecular networks analysis are related to the downstream effects analysis. In the Molecular Network analysis, two terms , 'Focus Molecules' and the 'Score.' are used in the report deserve further explanation. The 'Focus molecules' of a network represent the number of the genes in the uploaded DEGs list from my experiment that are represented in the network, while the 'Score' is the total number of genes in the network, including the genes of the 68-DEGs-Mi and the genes that were added from the indirect interactions predicted by IPA. The networks analysis by IPA revealed nine potentially significant molecular networks in pSS-associated lymphoma. The top four networks showed the highest Scores (between 21 and 32) while the other five showed much lower Scores (2). Therefore, only the top four networks are discussed in this chapter.

The top network (Network 1) has a Score of 32 with Focus Molecules of 17. The top diseases and functions involved in this network, shown in Figure 5.3, include 'Cell Death and Survival,' 'Cell-mediated Immune Response' and 'Cellular Development.' Figure 5.4 shows the second network (Network 2), involving 'Cell Cycle,' 'DNA Replication,' 'Recombination and Repair' and 'Gene Expression,' and has a Score of 23 and Focus Molecules of 13. Network 3 (Figure 5.5) included the diseases and functions of 'Cell-mediated Immune Response,' 'Cellular Development' and 'Cellular Function and Maintenance'; this network has a Score and Focus Molecules of 21 and 12 respectively. Network 4 has the same Score and Focus Molecules as Network 3. The most important diseases and biological functions within the fourth network were 'Cell Death and Survival,' 'DNA Replication,' 'Recombination and Repair' and 'Cancer' (Figure 5.6).

Interestingly, two of these networks (Networks 2 and 4) included two genes in the three-gene biosignature in pSS-associated lymphoma (*NUDT14* and *MGST3*). The other two networks (Networks 1 and 3) included the two genes that were significantly associated with untreated pSS-associated lymphoma (*DYNLL1* and *DRAP1*). The significant gene associated with untreated pSS-associated lymphoma, *DYNLL1,* was included in network 1 and has an indirect interaction with complexes ERK and Jnk. In network 2, *MGST3* was

included and was regulated by *PPARA* in a direct interaction. *DRAP1*, which was also significant in untreated pSS-associated lymphoma, has a direct interaction with *myc* in network 3. Network 4 contains *NUDT14*, the most consistently differentially expressed gene throughout all the cohorts. The interactions in Network 4 are most likely related to cancer biology, as it includes *TP53,* which binds to *WT1*, which in turn has a direct relationship with *TCOF1*. *TCOF1* has an indirect interaction with *NUDT14*.

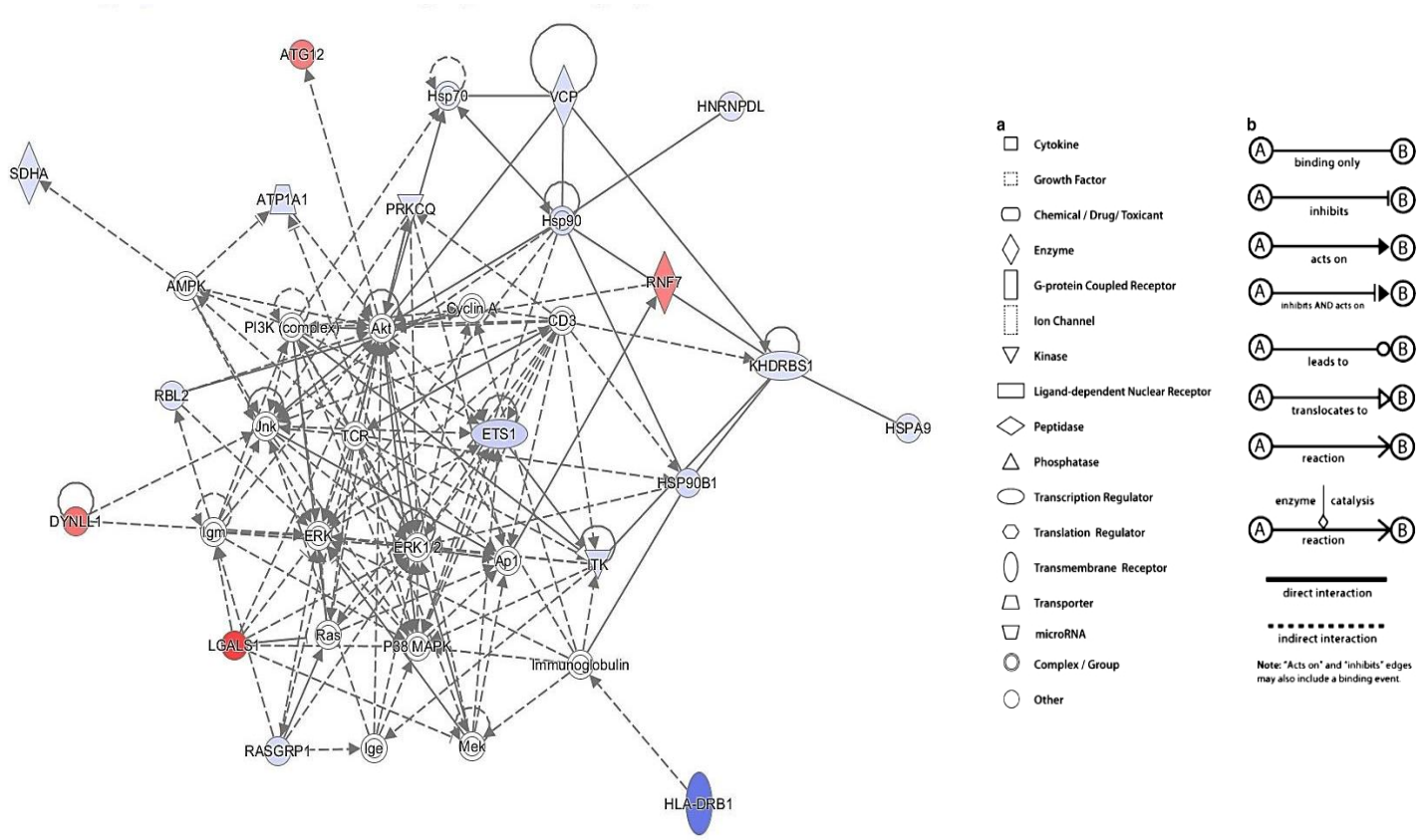


Figure 5.3 Network 1 of pSS-associated lymphoma identified by IPA. *A total of 17 genes from the pSS-associated lymphoma gene expression signature are involved in 'Cell Death and Survival' ,'Cell-mediated Immune Response' and 'Cellular Development'. Genes in red represent up-regulated genes in pSS-associated lymphoma. Genes in blue represent down-regulated genes in pSS-associated lymphoma. Genes in white represent genes in the global network but not included in the putative pSS-associated lymphoma gene-expression signature. DYNLL1, which was differentially expressed in untreated pSS-associated lymphoma, was included in this network.*

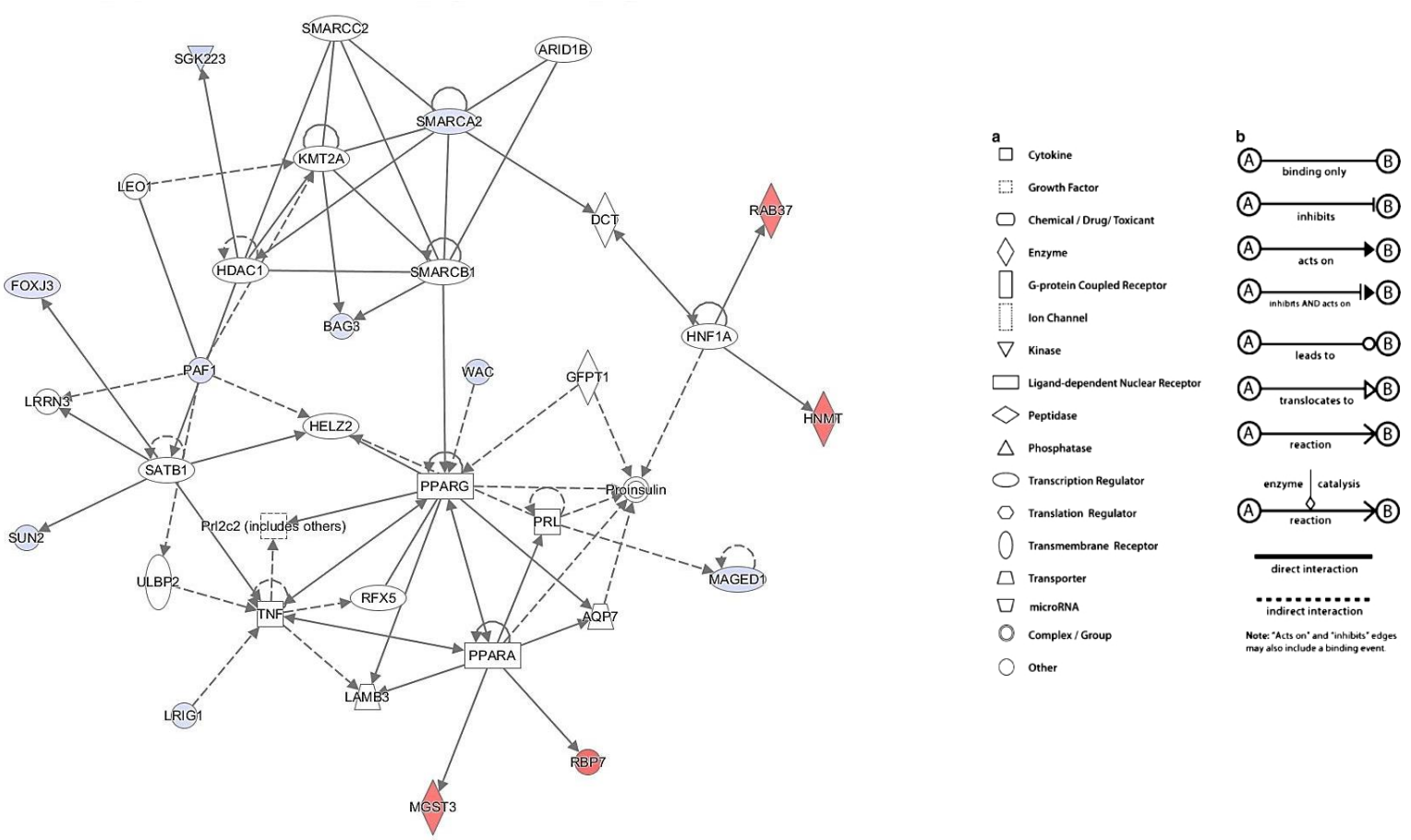


Figure 5.4 Network 2 of pSS-associated lymphoma identified by IPA. *A total of 13 genes from the pSS-associated lymphoma gene expression signature are involved in 'Cell Cycle', 'DNA Replication', 'Recombination and Repair' and 'Gene Expression'. Genes in red represent up-regulated genes in pSS-associated lymphoma. Genes in blue represent down-regulated genes in pSS-associated lymphoma. Genes in white represent genes in the global network not included in the putative pSS-associated lymphoma gene-expression signature. MGST3, which was differentially expressed in pSS-associated lymphoma (3-gene biosignature), was included in this network.*

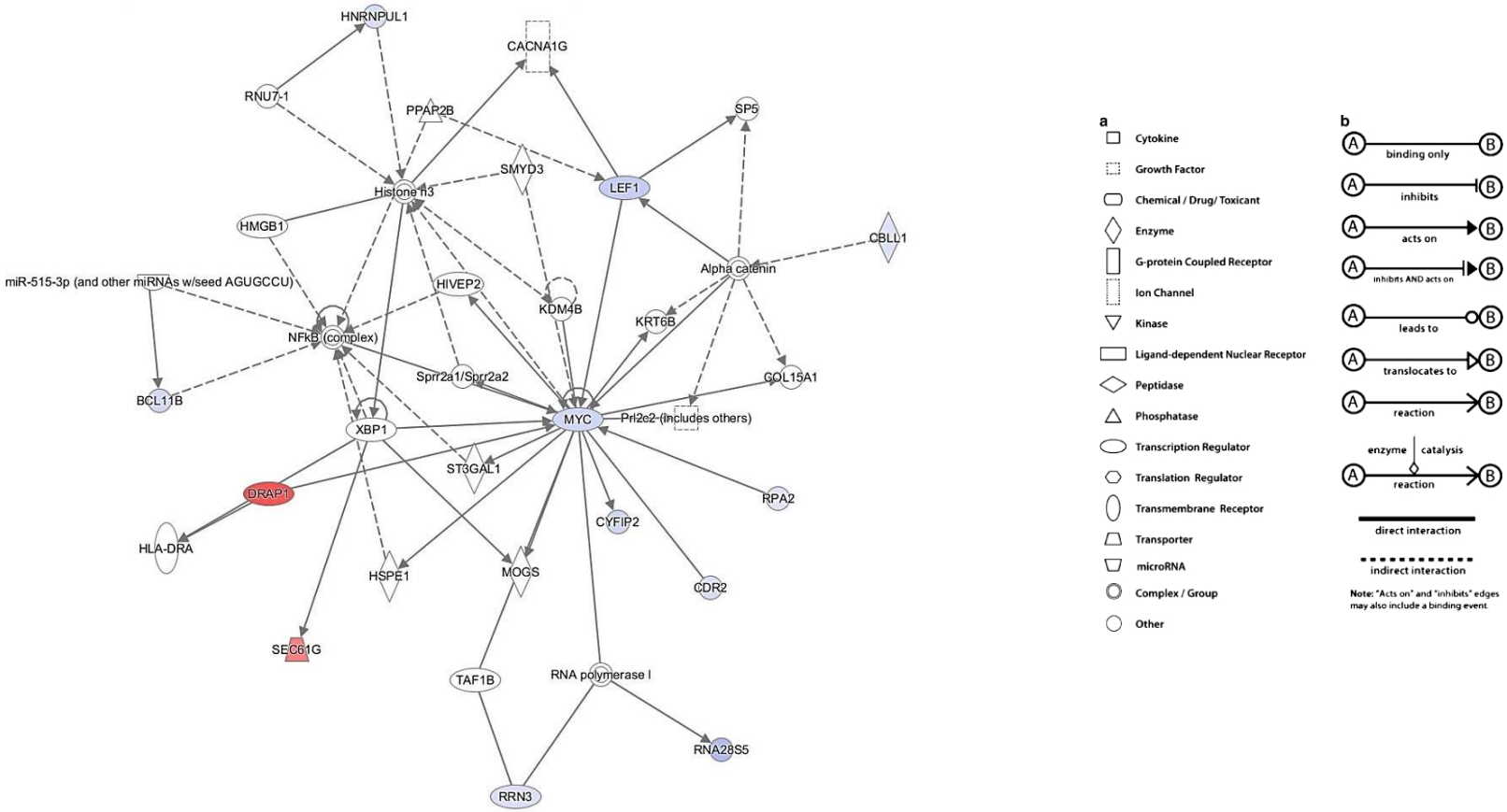


Figure 5.5 Network 3 of pSS-associated lymphoma identified by IPA. *A total of 12 genes from the pSS-associated lymphoma gene expression signature are involved in 'Cell-mediated Immune Response', 'Cellular Development' and 'Cellular Function and Maintenance'. Genes in red represent up-regulated genes in pSS-associated lymphoma. Genes in blue represents the down-regulated genes in pSS-associated lymphoma. Genes in white represent genes in the global network not included in the pSS-associated lymphoma gene-expression signature. DRAP1, which was differentially expressed in untreated pSS-associated lymphoma, was included in this network.*

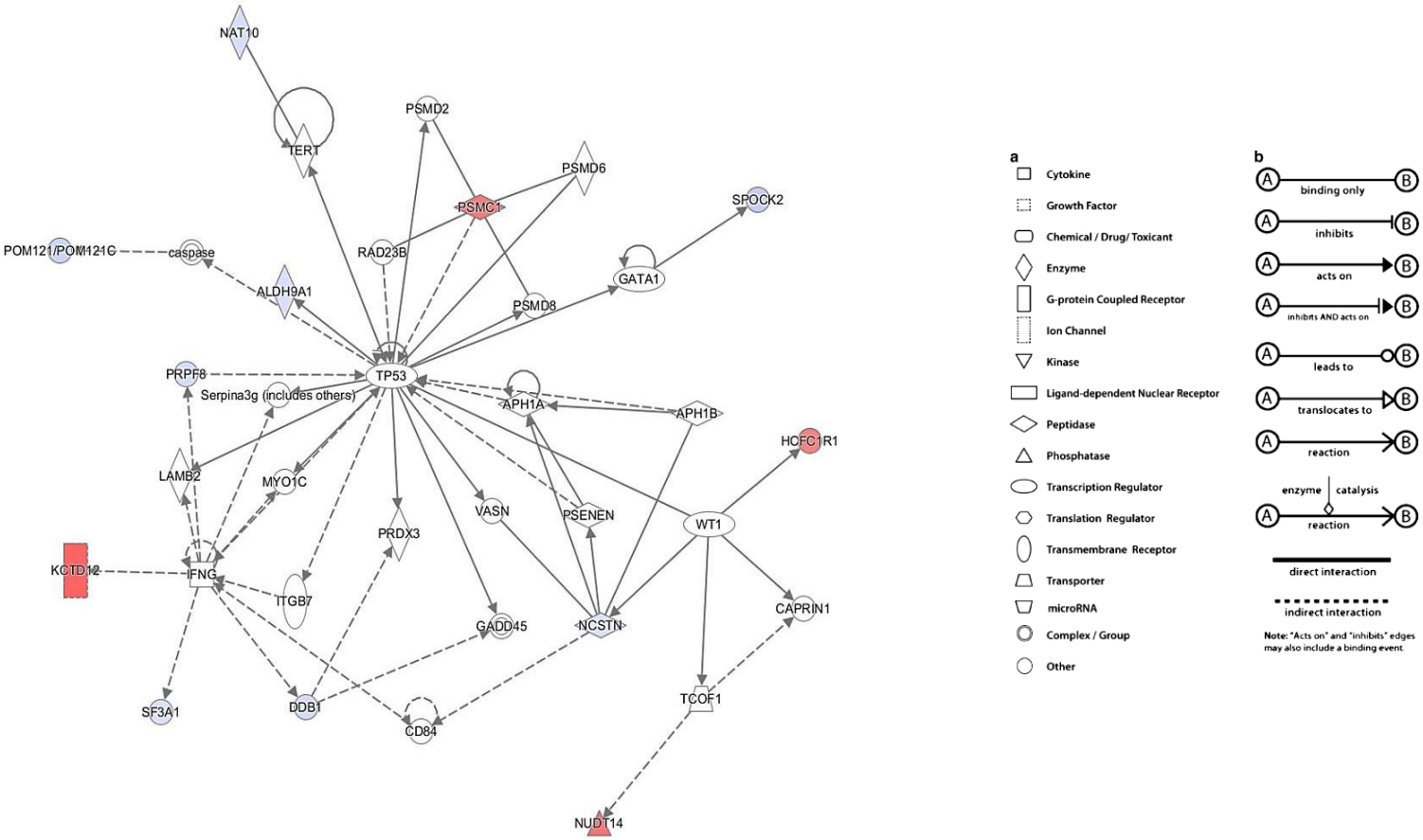


Figure 5.6 Network 4 of pSS-associated lymphoma identified by IPA. *A total of 12 genes from the pSS-associated lymphoma gene expression signature are involved in 'Cell Death and Survival', 'DNA Replication', 'Recombination and Repair', and 'Cancer'. Genes in red represent up-regulated genes in pSS-associated lymphoma. Genes in blue represent down-regulated genes in pSS-associated lymphoma. Genes in white represent genes in the global network not included in the putative pSS-associated lymphoma gene-expression signature. NUDT14, which was differentially expressed in pSS-associated lymphoma (3-gene biosignature), is included in this network.*

5.3.4 The Upstream Regulators Analysis of pSS-associated lymphoma:

The Upstream Regulator Analysis by IPA allows researchers to identify the cascade of the upstream transcriptional factors that control the genes of interest. The interactions between certain genes and the upstream regulators may explain changes in the expression of these genes and the effects of these changes on a certain biological function. As in other analyses in IPA, it also predicts whether the regulators that are involved in genes of interest are activated or inhibited by calculating the z-score.

Applying Upstream Regulator Analysis to my dataset (68-DEGs-Mi), I have identified 219 regulators that were significantly linked to pSS-associated lymphoma ($p < 0.05$, Fisher's Exact test). Focusing on my 3-gene biosignature of pSS-associated lymphoma, IPA predicted upstream regulators for *NUDT14* and *MGST3* but not *BMS1*. One regulator (*TCOF1*) was predicted to regulate *NUDT14*. Two regulators (*NFE2L2* and *PPARA*) were predicted to be involved in the regulation of *MGST3*.

Furthermore, I searched for upstream regulators that regulate the differentially expressed genes in untreated lymphoma. Five regulators (*NFE2L2*, *PPARA*, *SLC6A2*, *BARX2* and *SLC18A2*) were predicted to regulate *DYNLL1*. However, *SLC6A2*, *BARX2* and *SLC18A2* seem to be less important regulators as they only regulate one gene (*DYNLL1*). No upstream regulators were predicted for *DRAP1*. These six regulators identified included transcription regulators, transporters and ligand-dependent nuclear receptors (Table 5.6). Despite the statistically significant association, IPA was unable to make prediction to whether these regulators were activated or inhibited in pSS-associated lymphoma as the calculated absolute z-scores were not significant.

Table 5.6 The upstream regulators of pSS-associated lymphoma identified by IPA. *Six regulators were predicted to be involved in the regulation of the genes in the 3-genes biosignature and the differentially expressed genes in untreated pSS-associated lymphoma. Z-score is not significant for all the upstream regulators.*

Upstream Regulator	Molecule Type	p-value	Target molecules
NFE2L2	Transcription regulator	5.66E-04	*ATP1A1, DYNLL1, HSP90B1, MGST3, PSMC1, VCP*
PPARA	Ligand-dependent nuclear receptor	2.52E-02	*DYNLL1, MGST3, MYC, RBP7*
TCOF1	Transporter	2.36E-02	*HNRNPDL, NUDT14*
BARX2	Transcription regulator	2.72E-02	*DYNLL1*
SLC18A2	Transporter	3.06E-02	*DYNLL1*
SLC6A2	Transporter	1.71E-02	*DYNLL1*

5.4 Discussion:

In this chapter I have focused on the biological processes/pathways that might be associated with pSS-associated lymphoma. The identified biological pathways will provide additional insight in particularly to understand the interactions of the genetic factors that are important in the pathogenesis of lymphoma in pSS.

The Canonical pathways in pSS-associated lymphoma

I applied different analytic methods provided by IPA to my dataset (68-DEGs-Mi). The first analysis in the IPA core analysis is the 'Canonical Pathway Analysis'. The 'Canonical Pathway Analysis' identified a total of 188 pathways, most of these were related to cancer biology. It is important to keep on mind that many of these pathways were generated using cancer as research model, and therefore may introduce potential bias. The canonical pathway with the lowest p-value was the 'AHR signalling' pathway. Despite the highly significant p-value, no prediction of the direction of change (i.e., activation or inhibition) for the pathway could be made. One possible explanation is that the small number of genes in 68-DEGs-Mi may make it difficult to link the genes of interest with the pathway information in the Ingenuity Knowledge Base (IKB). AHR is a ligand-activation transcriptional factor in the Pern-Arnt-Sim (PAS) family (Burbach et al., 1992). This receptor is also considered to be a xenobiotic receptor. Originally, the role of AHR was believed to be limited to xenobiotic metabolism, until it was discovered that AHR has the

ability to mediate toxic responses. For this reason, AHR was heavily investigated in toxicology studies (Barouki et al., 2007). The link between AHR and immune responses and other endogenous functions was only recently discovered. For instance, AHR has a role in regulating cell shape, cell adhesion and cell migration (Carvajal-Gonzalez et al., 2009, Ikuta and Kawajiri, 2006). In addition, Aryl Hydrocarbon Receptor nuclear translocators (ARNT) are crucial for haematopoietic stem cell (HSC) viability. Furthermore, it has been found that the HSCs with ARNT deficiency experienced programmed death process (apoptosis) (Krock et al., 2015). Many studies have also been conducted to investigate the role of AHR in different types of lymphoma and cancer. Ding and co-workers recently reported that the AHR/ARNT complex regulates *MEF2B*, a transcription factor that regulates the expression of B-cell lymphoma 6 (*BCL6*). The regulation of *BCL6* by both AHR/ARNT and *MEF2B* leads to the expression of the germinal centre markers in diffused B-cell lymphomas (Ding et al., 2015). Furthermore, AHR has been suggested to play a role in other malignancies such as pleomorphic adenoma, which is a benign mixed tumour of the parotid gland (Drozdzik et al., 2015).

On the other hand, the linkage between the environmental and endogenous ligand by AHR can be harnessed as an adjunct to the treatment of cancers. For example, activation of AHR in the presence of chemo-preventive agents (Chrysin) in cancer cell lines, leads to the induction of Chrysin-induced apoptosis. This induction of apoptosis is due to the activation of TNF-α and TNF-β, which is dependent on AHR serving as a ligand to Chrysin (Ronnekleiv-Kelly et al., 2015). In another study, the presence of TNF-α plays a role in modulating the AHR function as an activator of apoptosis. These modulation effects were investigated by using 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), which acts as a ligand to AHR. In the presence of TNF-α and TCDD, lymphocyte apoptosis is reduced (Ghatrehsamani et al., 2015). Thus, it is noteworthy that TNF-α is over-expressed in pSS (Kang et al., 2011), which might suggest similar mechanisms lead to the development of lymphoma in pSS by reducing lymphocyte apoptosis.

Emerging data indicate that AHR acts as a co-factor in autoimmune disease development. Veldhoen et al. (2008) have shown that AHR activation through a high-affinity ligand leads to the development of T_H17 cells. Notably, this activation must occur during the development of this cell subset. Eventually, the increased percentage of T_H17 cells leads to

an increase in T_H17 cytokines, which include IL-22 (Veldhoen et al., 2008). Moreover, AHR also regulates the function of regulatory T cells (T_{reg}), an important mechanism of self-tolerance (Quintana et al., 2008). It is known that IL-17 and IL-22 are elevated in pSS (Miletić et al., 2012, Lavoie et al., 2011). Therefore, it is plausible that similar mechanisms might be involved in the pathology of pSS and pSS-associated lymphoma. The role of AHR and the linkage between environmental and endogenous ligands is illustrated in Figure 5.7.

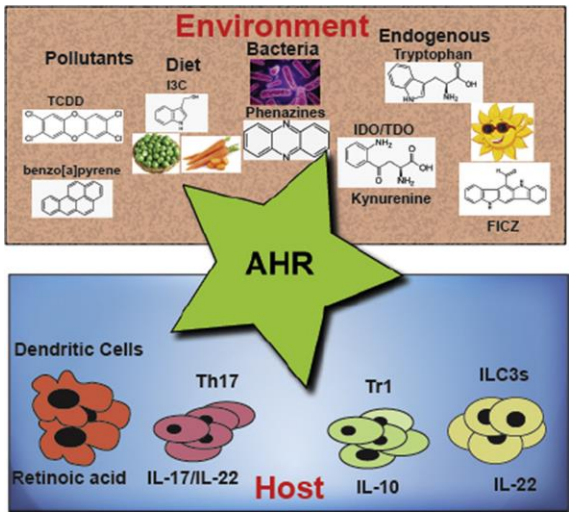


Figure 5.7 The Aryl Hydrocarbon Receptor (AHR) integrates responses from environmental and endogenous ligands to mount appropriate immune responses at barrier organs *(Cella and Colonna, 2015)*

It is of interest that one of the genes included within the AHR pathway is *MGST3*, one of the genes in the 3-gene biosignature in pSS-associated lymphoma. The genes connected to AHR pathways are summarized below:

MGST3: Since AHR is involved in response to xenobiotics such as chemicals and drugs, many researchers have studied the role of AHR in drug metabolism. The microsomal *GST3 (MGST3)* gene is known for its involvement in metabolic reactions. Given the close connection between the AHR pathway and the role of *MGST3*, it is not surprising that studies have confirmed the correlation (upregulation/activation) of *MGST3* and AHR in drug metabolism in mice. Similar studies involving the gene *ALDH9A*1, which controls the metabolism of aldehydes, have demonstrated their correlation to AHR in drug metabolism in the liver of mice. Furthermore, *MGST3* was reported to correlate with transcriptional

factors such as *NFE2L2* and *PPARA* (Aleksunes and Klaassen, 2012, Fu and Klaassen, 2014). Interestingly, the same transcriptional factors (*NFE2L2* and *PPARA*) were discovered to regulate *MGST3* in my dataset.

MYC: Yang and colleagues in 2005 demonstrated that the activation of AHR in Hs578T cancer cells could inhibit the expression of the oncogene *MYC*. Down-regulation of *MYC* in our Discovery cohort microarray was consistent with this observation. Furthermore, AHR-mediated regulation could affect apoptosis (increase or decrease) depending on the cell line used (Yang et al., 2005).

HSP90: It is known that there is a link between AHR and heat shock proteins (HSP). Thus, *HSP90* associates with AHR, leading to the activation of the receptor in response to xenobiotics (Tsuji et al., 2014).

RBL2: It has been shown that the expression of AHR is associated with the retinoblastoma-like 2 (*RBL2*) gene (also known as Rb2 and p130), and such association affected the AHR-mediated cell cycle in the presence of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) in 5L hepatoma cells (Ge and Elferink, 1998). Overexpression of the *rbl2* gene acted as a tumour suppressor in nude mice. It has been suggested that *RBL2* modulates angiogenetic balance, which is essential for tumour formation (Claudio et al., 2001). Consistently, *RBL2* was down-regulated in my dataset.

To summarise, 'AHR signalling' may play an important role in pSS pathogenesis and may provide a link between environmental factors and the progression of pSS (Inoue et al., 2012). AHR signalling also has an important role in tumourigenesis, anti-cancer therapies, as well as pSS pathology. Further studies should be conducted to investigate the role of this pathway in pSS-associated lymphoma.

The Histamine Degradation pathway is another canonical pathway in pSS-associated lymphoma. Moreover, histamine, in particular H4 histamine receptor (H4R), is important in the pathology of pSS. Studies have shown that dendritic cells and T-cells synthesize histamine (Oda et al., 2000). The role of H4R is important to sustain the health of the tubuloacinar epithelium (Stegaev et al., 2012). Furthermore, H4R activation inhibits TNF-α/IMD-0354-induced apoptosis in salivary gland cells in pSS (Stegajev et al., 2014).

In my dataset, histamine N-methyltransferase (*HNMT*) was one of the genes that was involved in the histamine degradation pathway and was present in both the 68-DEGs-Mi and 26-DEGs-MiPCR signatures. Due to the fact that limited RNA samples were available, the expression of the *HNMT* gene was not tested during biological validation. *HNMT* has been linked to many malignancies. For example, *HNMT* has been identified as a key gene that predicted the prognosis of paediatric acute lymphoblastic leukaemia (ALL) (Gao et al., 2015). Another study also reported that the *HNMT* gene was expressed in the bone marrow of breast cancer patients (Del Valle et al., 2014). Histamine degradation is also linked to the expression of *ALDH9A1*. Of note, *ALDH9A1* gene was also differentially expressed in the microarray dataset in the Discovery cohort. Thus, the role of the 'Histamine degradation pathway' in pSS-associated lymphoma also warrants more investigation.

Another canonical pathway associated with pSS-associated lymphoma is the 'Unfolded protein response (UPR)'. This pathway is responsible for endoplasmic reticulum (ER) homeostasis. The key role of the ER is the organization of the biosynthesis and the secretion of proteins. The UPR consists of three 'sensors': first, the inositol-requiring transmembrane kinase/endoribonuclease 1 (IRE1), second, the double-stranded RNA (PKR)–activated protein kinase-like eukaryotic initiation factor 2α kinase (PERK), and third, the activating transcription factor-6 (ATF6). Figure 5.8 illustrates the relationships between the UPR pathway and infectious inflammatory diseases and cancer.

The genes from my dataset that overlapped with genes that make up the UPR pathway were two heat-shock proteins (*HSPA9, HSP90B1*) and valosin-containing protein (*VCP*). Both heat shock protein genes were down-regulated in pSS-associated lymphoma. Data have shown that inhibition of the heat shock protein *HSP90* will activate UPR in a myeloma cell line (Davenport et al., 2007). Furthermore, *HSP90B1* inhibition has a role in the chaperoning of integrin and Toll-like receptor (TLR) in B cells, which is important in cancer and lymphoma pathogenesis (Liu and Li, 2008). *VCP* is associated with both heat shock protein and the activation of UPR (Abisambra et al., 2013). *VCP* has also been linked to the development of malignancies such as hepatocellular carcinoma (HCC) and has been considered a potential therapeutic target (Yi et al., 2012).

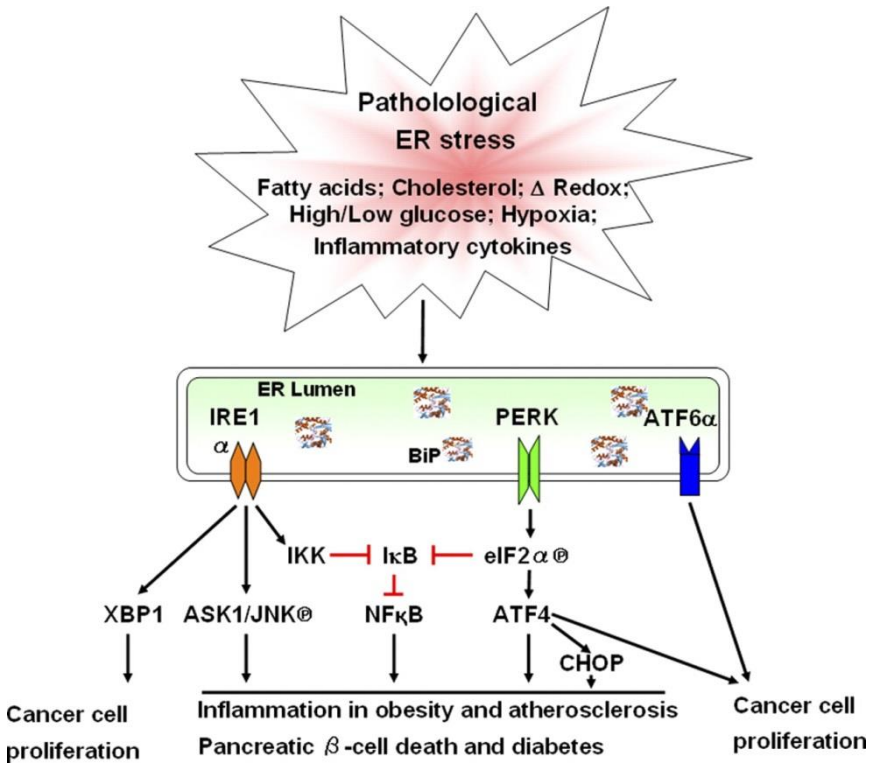


Figure 5.8 The Unfolded Protein Receptor (UPR) signalling in diseases*. The three arms of the UPR pathway are IRE1α-XBP1s, PERK-eIF2α phosphorylation-ATF4 and ATF6, which are all important for tumour cell survival and growth under hypoxic conditions. IRE1α and PERK can trigger c-JUN N-terminal kinase (JNK) and NFκB to stimulate inflammation and apoptosis that can contribute to inflammation in the pancreatic β-cell death and obesity in diabetes. In addition, CHOP production in the PERK pathway exacerbates oxidative stress in diabetes and atherosclerosis, aggravating these diseases (Wang and Kaufman, 2012).*

The fourth pSS lymphoma-associated pathway is 'Neuregulin Signalling.' This pathway has been implicated in psychiatric disorders such as schizophrenia (Buonanno, 2010, Hatzimanolis et al., 2013). The neuregulins (NRGs) are members of the growth factor family, which has multiple functions regarding important organs such as the heart and the nervous system. It has also been linked to diseases such as cancer. For instance, overexpression of NRG was found in many ovarian carcinoma cell lines (Gilmour et al., 2002). The neuregulin family consists of four members: NRG1, NRG2, NRG3 and NRG4. NRG binds to the receptor tyrosine kinases (the human epidermal growth factor receptor HER) ErbB3 or ErbB4, leading to the formation of homo or heterodimer ErbB2, which is also known as *HER2*. Previous studies have demonstrated that NRG4 and *HER4* are

predominantly expressed in MALT lymphoma clinical samples from patients with gastrointestinal lymphoma. The recombinant NRG4 stimulates *HER4* tyrosine phosphorylation leading to proliferation in a lymphoma cell line (Ebi et al., 2011). Stimulation of heat-shock protein 70 inhibits the NRG1-induced demyelination through increased proteasomal degradation of c-Jun (Li et al., 2012). These biological mechanisms might provide an explanation to the association of 'Neuregulin Signalling' pathway in pSS-associated lymphoma, as the heat-shock protein gene (*HSP90B1*) was down-regulated in the lymphoma group in my dataset. Investigating the role of NGRs and HER in pSS-associated lymphoma may provide a better understanding of the pathogenesis of lymphoma in pSS.

The fifth canonical pathway associated with pSS-lymphoma is the 'T-cell receptor signalling' pathway. T-cell receptor signalling, which involves many steps including the activation of NF-κB, has an important role in the development of lymphoma in haemophagocytosis (An et al., 2011).

While the Canonical Pathway analysis has demonstrated that these pathways were enriched in pSS-associated lymphoma, IPA could not make a prediction of whether these canonical pathways were activated or inhibited in pSS-associated lymphoma. There are several reasons why no prediction on the directionality of these canonical pathways in pSS-associated lymphoma could be made. First, there was insufficient information in the literature and/or the IPA knowledge base. Second, there were only 68 differentially expressed genes included in the analysis. Interestingly, among the three genes in the three-gene biosignature in pSS-associated lymphoma, *BMS1* does not appear to be associated with any of the 188 lymphoma-associated canonical pathways. In contrast, both *NUDT14* and *MGST3* were associated with six different canonical pathways. The canonical pathways that were associated with *NUDT14* were all metabolic pathways. Similarly, *MGST3* is involved in metabolic pathways as well as the xenobiotic pathway (AHR signalling pathway).

The Downstream Effects and Networks analysis in pSS-associated lymphoma

With regard to the top 'diseases and biological function' that might be implicated in pSS-associated lymphoma, IPA predicted that most of these functions were related to cell

function, cell viability, cell survival and cell movement. These functions are known to be associated with different types of malignancies and with lymphoma, in particularly, non-Hodgkin lymphoma (Pon et al., 2015).

The network analysis illustrated the involvement of the DEGs in pSS-associated lymphoma and their interactions through nine networks. Knowledge of such gene-gene interactions provides information of how my genes of interest may interact with each other and with other genes from the global molecular network at the molecular level. These data may improve the understanding of the pathogenesis of pSS-associated lymphoma.

The Upstream Regulators analysis in pSS-associated lymphoma

The Upstream Regulator analysis identified upstream regulators for the 2 genes of the 3-gene biosignature of pSS-associated lymphoma and 1 gene of the significant genes in the untreated pSS-associated lymphoma (See Table 5.6). Nuclear factor (erythroid-derived 2)-like 2 (*NFE2L2*) is known for its role in the regulation of the genes with antioxidant function *MGST3*. Both *NFE2L2* and *MGST3* were functioning within the same pathway in weakening the lung function in smokers. More interestingly, both genes were involved in the 'AHR signalling' pathway (Curjuric et al., 2012), reinforcing the association between pSS-associated lymphoma development and AHR signalling.

Another upstream regulator of interest is peroxisome proliferator-activated receptor alpha (*PPARA*). *PPARA* regulates *MGST3* and *DYNLL1* and is a vital upstream regulator of genes that involve in cell metabolism (Blavy et al., 2014).

The third identified upstream regulator is Treacher Collins-Franceschetti syndrome 1 (*TCOF1*), which regulates *NUDT14,* involved in nucleotide sugar catabolism (Yagi et al., 2003). *TCOF1* encodes a protein called 'treacle' which is important for ribosomal RNA (rRNA) transcription. The gene was investigated in mouse Neuroblastoma cell line N1E-115 in order to identify its role in the regulation of proliferation and differentiation of cells. In this study, genes with changes in expression level that were concordant with the *TCOF1* expression level were involved in promoting cell proliferation, whereas genes with changes in expression level discordant with *TCOF1* were either involved in proliferation repression or cell-death stimulation. Interestingly, *NUDT14* was one of the genes that correlated

negatively with *TCOF1* (Mogass et al., 2004). Consistent with my dataset, *NUDT14* is thought to have an indirect interaction with *TCOF1*.

The upstream regulator analysis in pSS-associated lymphoma revealed the importance of metabolic genes in the development of lymphoma. Regarding the upstream regulators analysis of the genes (*DYNLL1*) that were implicated in untreated pSS-associated lymphoma, BARX homobox 2 (*BARX2*) is a transcriptional factor that controls several genes through its homeodomain binding sites. A homeodomain is a 60 amino acid helix-turn-helix DNA-binding domain. The DNA sequence that encodes the homeodomain is called the 'homeobox' and homeobox-containing genes are known as 'hox genes'. *BARX2* abnormality has been linked to ovarian cancer cells through its role as a suppressor of cell adhesion, migration and invasion of these cells (Sellar et al., 2001). Another study has reported that inhibition of *BARX2* affects the cell growth of a human breast cancer cell line (MCF7 breast cancer cell line) (Stevens et al., 2004).

It would be of interest to investigate whether BARX2 may also affect pSS-associated lymphoma cell growth via its regulation of *DYNLL1*. The solute carrier family 18 (vesicular monoamine transporter), member 2 (*SLC18A2*) and the solute carrier family 6 (neurotransmitter transporter) member 2 (*SLC6A2*) are upstream regulators of *DYNLL1*. *SLC18A2* acts as a vesicular monoamine transporter; abnormalities in this gene have been associated with neuropsychiatric disorders through its role in regulating monoamine neurotransmitters. The only type of cancer that has been reported in the literature linked to *SLC18A2* is prostate cancer. Thus, DNA hyper-methylation that causes silencing of *SLC18A2* in prostate cancer, make this gene a novel predictor to the response to treatment (Sorensen et al., 2009). *SLC6A2* is a norepinephrine transporter (NET). Studies have demonstrated that polymorphism of this gene contributes to Major Depression (MD) susceptibility (Wang et al., 2015). These associations might be of interest as depression and fatigue are common in pSS.

In conclusion, the identification of the canonical pathways provides an improved understanding of the possible mechanisms that might be involved in pSS-associated lymphoma. Additionally, network analysis and upstream regulators analyses revealed molecules that interact with or regulate the genes of interest in pSS-associated lymphoma. As it is likely that many biological processes and mechanisms are involved in lymphoma

development in pSS, IPA has provided me with clues for further investigations into such mechanisms.

Chapter 6

Gene expression profiling and pathway analysis of different pSS subgroups in the Discovery cohort

6.1 Introduction

The association between pSS and cancers, especially haematological tumours, is widely acknowledged. However, the associations with other types of cancer have also been reported, but without conclusive overall evidence. A meta-analysis of 14 cohort studies consisting of a total of over 14,523 patients reported a significant increase in the malignancy development including all forms of cancer. The authors also recommended the conduction of more studies to examine the role of NHL in increasing the risk of overall malignancy (Liang et al., 2014). The determination of the risk factors for other types of tumour in pSS has been poorly understood. However, in Sweden it has been reported that 10 excess tumour cases were found in pSS patients in addition to the NHL cases (Theander et al., 2006). Lazarus and colleagues reported an increased risk of developing more than one type of cancer in pSS (Lazarus et al., 2006). One example is the lung adenocarcinoma association with pSS as pulmonary involvement is a known pSS systemic manifestation (Takabatake et al., 1999). Another example was a case report suggested hepatocellular carcinoma as a possible outcome of pSS (Yan et al., 2013). Breast cancer is another type of non-haematological tumours that might be linked to pSS, as evidence has suggested that systemic inflammation might affect neoplasia of the breast epithelia, especially in elderly women (Gadalla et al., 2009). The prevalence of different types of cancers other than lymphoma was also recorded in the UKPSSR cohort.

Paraproteinemia or monoclonal gammopathy (MG) is a recognised extra-glandular manifestation of pSS. Paraproteinemia is defined as the presence of monoclonal immunoglobulins in the blood or immunoglobulin light chains (Bence Jones protein) in the urine. Paraproteinemia can be detected in two clinical conditions: monoclonal gammopathy of undetermined significance (MGUS), and multiple myeloma. Additionally, paraproteinemia is associated with many other disorders, including lymphoproliferative disorders such as chronic lymphocytic leukemia (CLL) and lymphoplasmacytic lymphoma/Waldenstrom macroglobulinaemia (LPL/WM) (reviewed in (Cook and

Macdonald, 2007). The association between paraproteinemia and autoimmune diseases has also been documented . Moreover, paraproteinemia indicates an increased risk of malignant progression in patients with these autoimmune disorders (Kelly et al., 1991). In pSS, the presence of MG was reported to be associated with a high risk for the development of multiple myeloma (MM), more so than for lymphoma development (Tomi et al., 2015). However, the presence of IgM and IgG paraproteinemia was observed in patients with low-grade B-cell lymphoma (Iwatani et al., 2014). Recently, lymphoplasmacytic lymphoma, defined as a type of B-cell non-Hodgkin's lymphoma, was linked to the presence of IgA paraproteinemia. More studies were needed to define more accurately the lymphoma type in this study (Guan et al., 2015). MG was correlated with high systemic disease activity in pSS (as measured using ESSDAI), which was in turn associated with a higher risk of death in those patients (Brito-Zeron et al., 2014). The UKPSSR database includes information on whether paraproteinemia was present in the pSS patients.

Although identifying a gene expression signature in pSS-associated lymphoma is my main goal, given the documented association between paraproteinemia and the development of haematological malignancies, comparing the gene expression profiles of pSS patients with lymphoma and those with paraproteinemia is worthwhile. Similarly, comparing and contrasting the gene expression profiles of pSS-associated lymphoma and pSS patients with other cancers is also of interest. Because of the limited availability of biobanked RNA, I will focus my analysis on the DEGs from the microarray data and biological pathway analysis (using IPA).

6.2 Aim and experimental design

The aim of this chapter is to generate more evidence demonstrating that the whole blood gene expression signature of pSS-associated lymphoma that I identified in the Discovery cohort was "specific." For this reason, nine different comparisons were made. The first two comparisons were made between pSS-associated lymphoma and pSS-other cancers and pSS-associated lymphoma and pSS-paraproteinemia. Additionally, the pSS-other cancers, pSS-paraproteinemia and pSS (without lymphoma) groups were compared with each other. These additional comparative analyses provide assurance that the identified signature is specific to pSS-associated lymphoma. The final four comparisons examined pSS, pSS-

associated lymphoma, pSS-other cancers and pSS-paraproteinemia versus healthy controls. To sum up, the new comparisons I have performed are as follows:

1. pSS-associated lymphoma vs pSS-other cancers
2. pSS-associated lymphoma vs pSS-paraproteinemia
3. pSS vs pSS-other cancers
4. pSS vs pSS-paraproteinemia
5. pSS-other cancers vs pSS-paraproteinemia
6. pSS vs HC
7. pSS-associated lymphoma vs HC
8. pSS-other cancers vs HC
9. pSS-paraproteinemia vs HC

As mentioned in Chapter 4, in the Discovery cohort, 144 globin mRNA-depleted samples were used, and the samples were categorized into five groups: pSS-associated lymphoma (n=16); pSS (n=61); pSS-paraproteinemia (n=23); pSS-other cancers (n=21) and healthy controls (n=23). Data from the whole genome gene expression microarray were analysed to generate lists of differentially expressed genes (DEGs) between groups (Analysis A) as described in Chapter 4. Each list of DEGs was then uploaded onto the IPA platform for pathway analysis, as described in Chapter 5.

Patient demography can be found in Chapter 4, Table 4.1. The pSS-other cancer group includes a variety of different cancers, such as breast cancer, bowel cancer, cervical cancer, thyroid papillary carcinoma, benign meningioma, endometrial cancer, uterine cancer and renal cancer.

6.3 Results

6.3.1 RNA quality and RIN score

The RNA quality of the samples has been described in detail in Section 4.4.2.

6.3.2 Whole genome gene expression Bead Chip of pSS subgroups

1. Microarray data analysis

The microarray data analysis was performed using R Packages, as previously described in chapter 4, section 4.4.3. All samples with RIN < 7 were excluded from the analysis for all the pSS subgroups. An exception was made for the pSS-associated lymphoma, where two samples with RIN < 7 were included. In addition, the technical outliers were removed. The data were transformed and normalised using the RSN method. Finally, the quality control was performed tested. Finally the DEGs were generated using a p-value cut-off of $p < 0.05$ or the fold change cut-off of 1.2.

2. Batch effects on the microarray data

The ComBat function was applied to remove the batch effects from the microarray data, as described in Chapter 4 (Figures 4.5 and 4.6).

3. Differentially expressed genes from the microarray data analysis

The analyses to generate the list of differentially expressed genes (DEGs) for each of the nine comparisons. Not all the comparisons have generated a DEGs list, as the differences in gene expression levels between the comparison groups did not reach either the adjusted p-value cut-off of $p < 0.05$ or the fold change cut-off of 1.2. The summary of the DEGs analysis is as follows:

1. pSS vs HC ⟶ Generation of a DEGs list
2. pSS vs pSS-other cancers ⟶ No DEGs
3. pSS vs pSS-paraproteinemia ⟶ No DEGs
4. pSS-associated lymphoma vs HC ⟶ Generation of a DEGs list
5. pSS-associated lymphoma vs pSS-other cancers ⟶ Only one DEG (*RNA28S5*)
6. pSS-associated lymphoma vs pSS-paraproteinemia ⟶ No DEGs
7. pSS-other cancers vs HC ⟶ Generation of a DEGs list
8. pSS-other cancers vs pSS-paraproteinemia ⟶ No DEGs
9. pSS-paraproteinemia vs HC ⟶ Generation of a DEGs list

DEGs were identified only in the comparisons between pSS subgroups and healthy controls and the number of DEGs identified varied for each comparison. A total of 278 DEGs were identified in the comparison between pSS and healthy controls (221 upregulated genes, 57

downregulated genes). For pSS-paraproteinemia versus healthy controls, a list of 178 DEGs (148 upregulated genes, 30 downregulated genes) was generated, while pSS-other cancers versus healthy controls generated a list of 123 DEGs (113 upregulated genes, 10 downregulated genes). The comparison of pSS-associated lymphoma against healthy controls generates the highest number of DEGs, with 557 (301 upregulated genes, 256 downregulated genes). The comparison of pSS-other cancers and pSS-associated lymphoma identified only one DEG (*RNA28S5*). It is noticeable that the interferon-inducible genes dominated the DEGs lists in these comparisons. The top 10 up- and down-regulated genes in each comparison group are listed in Table 6.1. The DEGs from the four comparison groups are visualized using volcano plots in Figure 6.1. The complete DEGs lists (NAs excluded) for the four comparisons can be found in the supplementary tables S6, S7, S8 and S9.

Focusing on comparing the 68-DEGs-Mi signature identified in the Discovery cohort to the DEGs list from the comparison of pSS-associated lymphoma versus healthy control, there were some DEGs that were in common between the two lists. Firstly, 54 (79.4%) of the 68-DEGs-Mi signature were in common with the DEGs in the comparison of pSS-associated lymphoma versus healthy controls. Thus, 14 DEGs (20.6%) were specific for the comparison between pSS and pSS-associated lymphoma. The three genes of the 3-gene biosignature in pSS-associated lymphoma (*NUDT14*, *MGST3* and *BMS1*) were found within the 54 DEGs that are in common. The rest of the 503 genes were differentially expressed only in pSS-associated lymphoma versus healthy controls. Furthermore, comparing the 26-DEGs-MiPCR signature identified in the Discovery cohort to the DEGs list from the comparison of pSS-associated lymphoma versus healthy controls, 20 DEGs (76.9%) were in common, with 6 DEGs (23.1%) unique to the 26-DEGs-MiPCR signature and 537 DEGs differentially expressed only in pSS-associated lymphoma versus healthy controls.

Focusing on comparing the 68-DEGs-Mi signature identified in the Discovery cohort to the DEGs lists from the other comparisons of pSS subgroups versus healthy control. First, two genes, *DRAP1* and *SMARCA2,* were in common with the 68-DEGs-Mi signature, in the differentially expressed genes of pSS-other cancers group versus healthy controls. Second, three DEGs (*DRAP1, SMARCA2, BCL11B)* were in common with the 68-DEGs-Mi

signature in the comparison between pSS-paraproteinemia and healthy controls. *DRAP1* was included in the 26-DEGs-MiPCR but *DRAP1, SMARCA2* and *BCL11B* were not validated in the Validation cohort; *DRAP1* was however differentially expressed in the untreated pSS-associated lymphoma group. Third, only *DRAP1* was in common between the 68-DEGs-Mi signature and pSS versus healthy controls. There were no DEGs in common between the 3-gene biosignature and the DEGs lists for different pSS subgroups and healthy controls. Overall, these comparisons showed that *DRAP1* was differentially expressed in all pSS subgroups as well as in pSS patients with untreated lymphoma.

Table 6.1 Top 10 up and down regulated DEGs in microarray analysis among different pSS subgroups versus healthy controls. *A. Upregulated genes. B. Downregulated genes.*

A

pSS			pSS-paraproteinemia			pSS-other cancers			pSS-associated lymphoma		
Gene symbol	**Adjusted P value**	**Fold change**	**Gene symbol**	**Adjusted P value**	**Fold change**	**Gene symbol**	**Adjusted P value**	**Fold change**	**Gene symbol**	**Adjusted P value**	**Fold change**
IFI27	6.12E-07	11.51	*IFI27*	0.00078165	9.26	*IFI44L*	8.9011E-05	7.13	*IFI27*	1.5037E-05	16.86
IFI44L	9.852E-11	8.75	*IFI44L*	1.58534E-06	7.62	*IFI27*	0.03815265	6.60	*IFI44L*	7.4968E-09	12.44
ISG15	2.151E-10	5.63	*IFIT1*	1.58534E-06	5.09	*IFIT1*	8.9011E-05	4.90	*IFIT1*	7.4968E-09	7.66
IFIT1	9.852E-11	5.60	*ISG15*	4.80787E-06	4.88	*ISG15*	8.9011E-05	4.79	*RSAD2*	9.0914E-09	7.33
RSAD2	1.618E-10	5.39	*RSAD2*	4.80787E-06	4.61	*RSAD2*	0.000102181	4.43	*ISG15*	2.5047E-08	7.23
IFI44	1E-09	5.23	*IFI44*	6.83017E-05	4.05	*IFIT3*	5.16374E-05	4.00	*IFI44*	1.0884E-07	6.51
OAS3	8.28E-10	4.35	*IFIT3*	1.58534E-06	3.82	*IFI44*	0.00085592	3.95	*LY6E*	1.2523E-07	5.43
EPSTI1	9.852E-11	4.29	*IFITM3*	1.80416E-05	3.79	*OAS3*	0.000166919	3.77	*EPSTI1*	7.4968E-09	5.39
IFIT3	9.852E-11	4.28	*OAS3*	1.50298E-05	3.76	*EPSTI1*	8.9011E-05	3.62	*IFIT3*	7.4968E-09	5.34
IFITM3	6.055E-09	4.09	*HERC5*	4.44311E-06	3.66	*LY6E*	0.000653176	3.56	*OAS3*	1.2032E-07	5.19

B

pSS			pSS-paraproteinemia			pSS-other cancers			pSS-associated lymphoma		
Gene symbol	**Adjusted P value**	**Fold change**	**Gene symbol**	**Adjusted P value**	**Fold change**	**Gene symbol**	**Adjusted P value**	**Fold change**	**Gene symbol**	**Adjusted P value**	**Fold change**
HLA-DRB4	0.02191445	2.24	*NELL2*	0.0016615	1.51	*NELL2*	0.01363656	1.49	*HLA-DRB1*	0.01957886	4.75
MYOM2	0.00632069	1.84	*TXNDC12*	0.04806856	1.45	*ANXA7*	0.00506038	1.25	*MYH9*	0.00154487	1.59
TXNDC12	0.00356056	1.46	*BCL11B*	0.03711457	1.32	*ESYT2*	0.04281725	1.25	*SPOCK2*	0.00077385	1.55
IMPA2	0.00518347	1.39	*EIF3L*	0.00087021	1.31	*SMARCA2*	0.04281725	1.24	*SGK223*	0.00041848	1.55
GPR162	0.02652862	1.37	*EIF4B*	0.00364144	1.27	*MID2*	8.9011E-05	1.24	*LEF1*	0.00844086	1.52
PYGL	0.03946276	1.35	*TBC1D14*	0.02947428	1.27	*USP9X*	0.00895127	1.24	*TXNDC12*	0.01253812	1.52
FAM212B	0.00969497	1.32	*EEF2*	0.01107357	1.27	*HNRNPA0*	0.048907	1.22	*NELL2*	0.00229604	1.51
PPM1F	0.0006649	1.31	*SERTAD2*	0.03359516	1.26	*NR3C2*	0.00206934	1.21	*ABLIM1*	0.00995283	1.49
RPL3	0.00202947	1.30	*FEZ1*	0.03100429	1.26	*DSC1*	0.03419917	1.20	*BCL11B*	0.00050068	1.48
MIR181A2HG	0.03875588	1.29	*ALDOC*	0.02060631	1.26				*PIK3IP1*	0.00202645	1.48

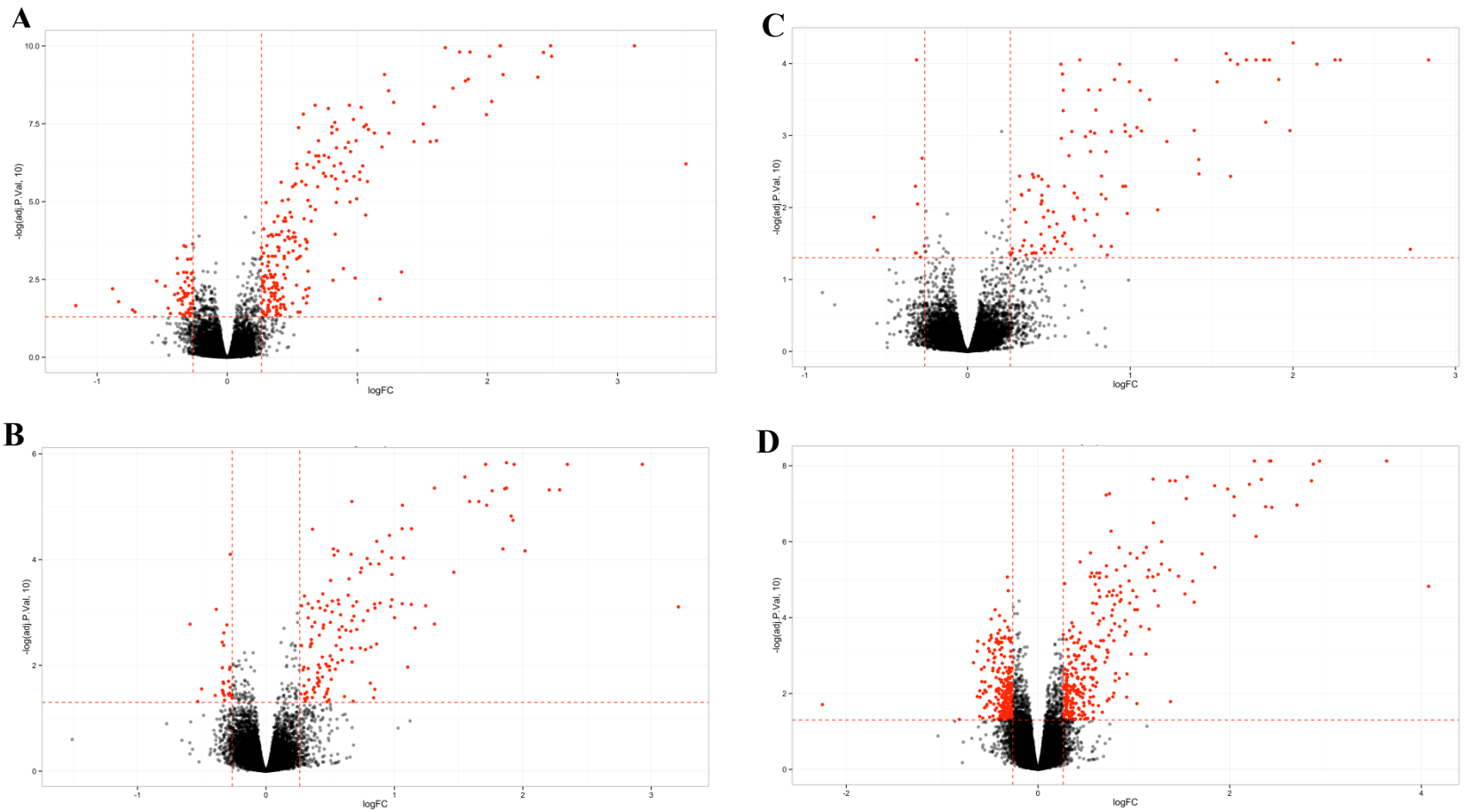


Figure 6.1 Volcano plots of the differentially expressed genes in the four pSS subgroups. A. *pSS vs HC.* ***B.*** *pSS-paraproteinemia vs HC.* ***C.*** *pSS-other cancers vs HC.* ***D.*** *pSS-associated lymphoma vs HC. The x-axes represent log2 of the fold change and the y-axes represent the –log10 of the adjusted p value. The red dots represent the differentially expressed genes and the red lines indicate the cut-offs for the adjusted p value and fold change.*

4. Validation of *DRAP1* in pSS subgroups by qRT-PCR

DRAP1 was the only gene from the 68-DEGs-Mi and 26-DEGs-MiPCR signatures of the Discovery cohort that was also differentially expressed in the pSS subgroups versus healthy control comparison. Furthermore, although *DRAP1* was not included in the 3-gene biosignature in pSS-associated lymphoma, it was differentially expressed in the untreated pSS-associated lymphoma group. The gene was included in the qRT-PCR plates that I ran to detect the stability of the housekeeping genes. For this reason, qRT-PCR expression data from all the pSS groups was available for this gene. *DRAP1* was significantly upregulated in all the pSS subgroups compared to healthy controls. The p values were < 0.0000, 0.0007, 0.0005 and < 0.0000, with fold change values of 1.54, 1.54, 1.55 and 2.37 for pSS, pSS-paraproteinemia, pSS-other cancers and pSS-associated lymphoma, respectively (Figure 6.2).

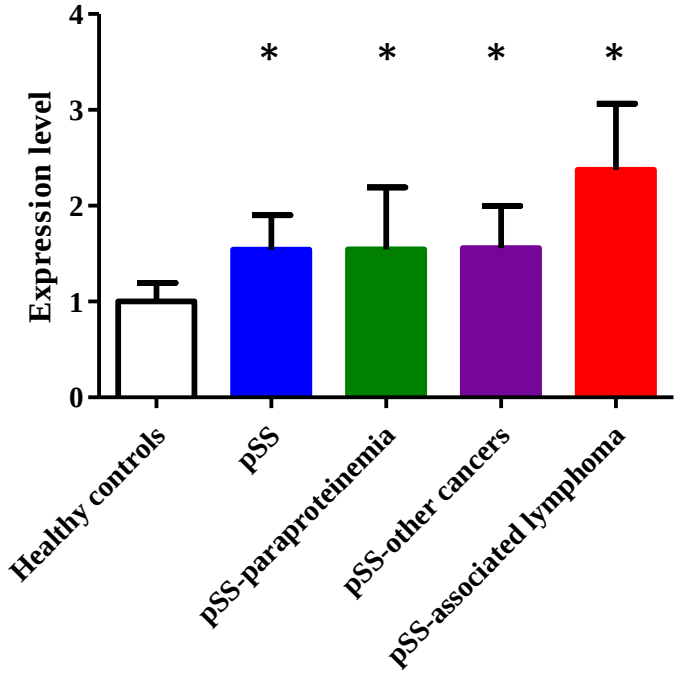


Figure 6.2 Relative expression levels of DRAP1 in pSS subgroups evaluated by qRT-PCR*. DRAP1 was differentially expressed in all pSS subgroups and the levels were significantly higher compared to healthy controls(n = 16). The p values were < 0.0000, 0.0007, 0.0005 and < 0.0000 and fold change values of 1.54, 1.54, 1.55 and 2.37 for pSS (n = 44), pSS-paraproteinemia (n = 15), pSS-other cancers (n = 14) and pSS-associated lymphoma (n = 15), respectively.*

6.3.3 Pathways analysis of pSS subgroups

1. Canonical pathways in different pSS subgroups

For this canonical pathway analysis, the 4 lists of DEGs from comparing the 4 pSS subgroups with healthy controls (HC) were uploaded onto the IPA platform. Each comparison using different pSS subgroups revealed a different set of canonical pathways. In total, 178 pathways were identified in the pSS vs HC comparison: 216 pathways in the pSS-paraproteinemia vs. HC comparison and 74 pathways in the pSS-other cancers vs. HC comparison. The pSS-associated lymphoma vs HC comparison revealed the highest number of canonical pathways (351 pathways). The Venn diagram in Figure 6.3A showed that sixty-seven pathways were in common between all four comparisons. The p value was calculated using the Fischer's exact right-tailed test. $P < 0.05$ was considered statistically significant. Moreover, the number of genes in the uploaded pSS subgroups' DEGs lists from my experiment that were in common with the total number of the genes included in each pathway was calculated. The top 5 canonical pathways (based on levels of statistical significance) in each comparison are shown in Table 6.2 and Figures 6.4 and 6.5. The "Interferon Signalling pathway" was activated in all four pSS subgroups and was the most statistically significant pathway. Interestingly, 178 out of 188 canonical pathways (94.9%) identified from the pSS-associated lymphoma vs. pSS comparison were in common with the pathways identified in the pSS-associated lymphoma versus HC comparisons. The ten canonical pathways that are only identified in pSS versus pSS-associated lymphoma are: Retinoate Biosynthesis II, Mitochondrial Dysfunction, Amyloid Processing, Serotonin Degradation, GABA Receptor Signalling, Basal Cell Carcinoma Signalling, Role of Wnt/GSK-3β Signalling in the Pathogenesis of Influenza, Oxidative Phosphorylation, Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis and G-Protein Coupled Receptor Signalling. The canonical pathways identified in pSS-associated lymphoma versus healthy controls that have a significant z-score can be found in supplementary table S10.

The number of pathways identified in the pSS versus pSS-associated lymphoma comparison from chapter 5 that were also identified in other group comparisons were as follows: 37 in the pSS-other cancers subgroup (all included in the list of the 178 canonical pathways that are in common when comparing pSS and/or healthy controls to pSS-

associated lymphoma), 78 in the pSS group and 113 pSS-paraproteinemia group (for both groups 77 and 112 canonical pathways were included in the 178 canonical pathways that are in common when comparing pSS and/or healthy controls to pSS-associated lymphoma) Figure 6.3B.

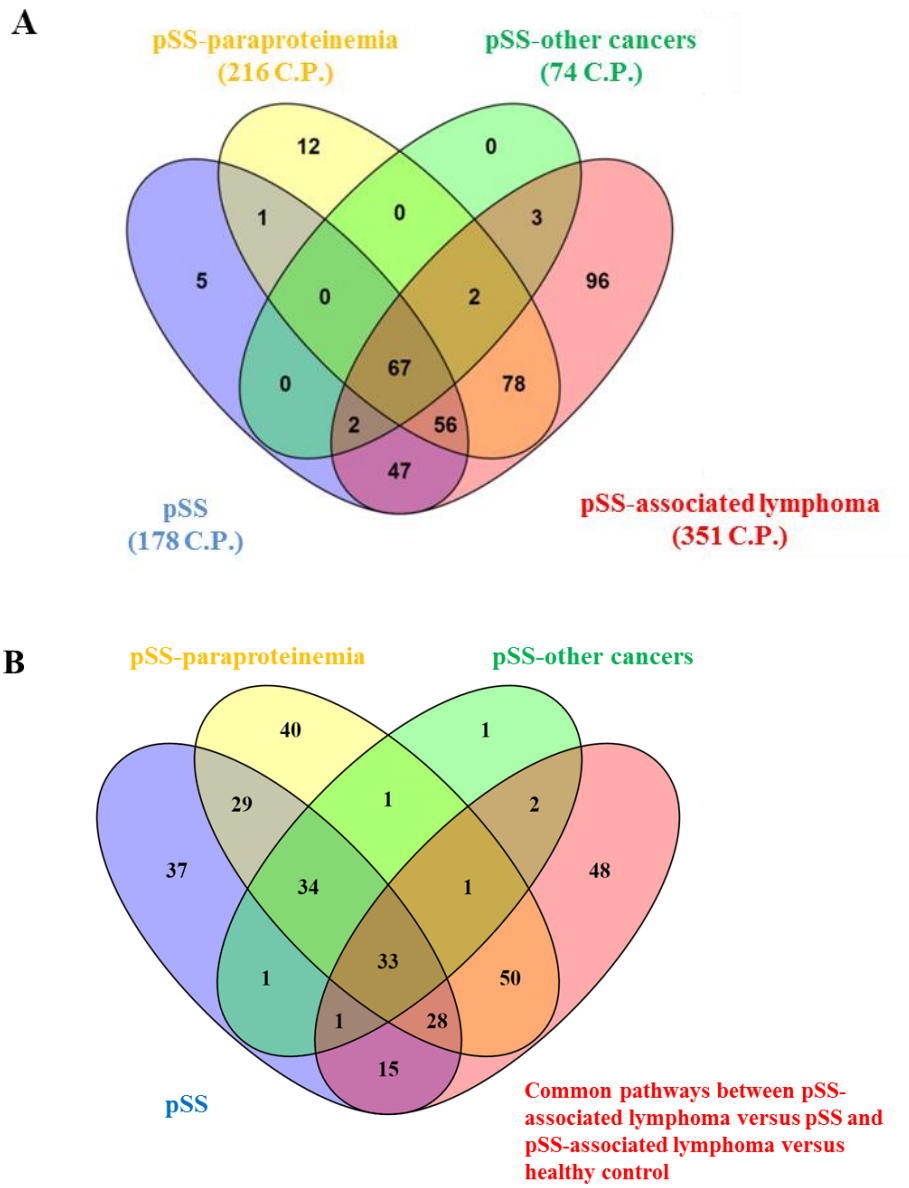


Figure 6.3 Venn diagram of the canonical pathways identified in the four pSS subgroup comparisons with healthy controls (HC). A. *Sixty-seven canonical pathways were in common between the four pSS subgroups comparisons.* ***B.*** *The common canonical pathways between pSS subgroups and the common pathway identified between pSS-associated lymphoma vs. pSS and pSS-associated lymphoma vs. healthy control. C.P.= Canonical Pathways.*

Table 6.2 Top 5 canonical pathways analysis in the four pSS subgroups (compared to healthy controls)

Ingenuity Canonical Pathways	-log (p-value)	Ratio	Genes
pSS			
Interferon Signaling	1.89E+01	3.89E-01	*OAS1,IRF9,IFITM1,IFIT3,STAT2,IFI6,IFITM3, TAP1,IFIT1,STAT1,ISG15,MX1,IFI35,SOCS1*
Activation of IRF by Cytosolic Pattern Recognition Receptors	1.07E+01	1.75E-01	*DHX58,IRF9,IRF7,STAT1,IFIH1,STAT2,ISG15, ADAR,DDX58,ZBP1,IFIT2*
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	7.40E+00	8.73E-02	*EIF2AK2,OAS1,IRF7,MYD88,IFIH1,OAS3, C1QB,OAS2,CASP1,DDX58,C3AR1*
Antigen Presentation Pathway	7.25E+00	1.89E-01	*HLA-A,TAP2,HLA-DRB4,TAP1,PSMB9, HLA-F,HLA-DRA*
Retinoic acid Mediated Apoptosis Signaling	5.66E+00	1.13E-01	*PARP10,PARP12,ZC3HAV1,TNFSF10,RXRA, PARP9,PARP14*
pSS-paraproteinemia			
Interferon Signaling	1.74E+01	3.33E-01	*OAS1,IRF9,IFIT1,IFITM1,STAT1,IFIT3,STAT2, ISG15,MX1,IFI35,IFI6,IFITM3*
Activation of IRF by Cytosolic Pattern Recognition Receptors	1.26E+01	1.75E-01	*DHX58,IRF9,IRF7,STAT1,IFIH1,STAT2,ISG15, ADAR,DDX58,ZBP1,IFIT2*
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	5.78E+00	6.35E-02	*EIF2AK2,OAS1,IRF7,IFIH1,OAS3,OAS2, CASP1,DDX58*
UVA-Induced MAPK Signaling	5.77E+00	7.95E-02	*PARP10,PARP12,ZC3HAV1,STAT1,PARP9, PARP14,RRAS*
Retinoic acid Mediated Apoptosis Signaling	5.52E+00	9.68E-02	*PARP10,PARP12,ZC3HAV1,TNFSF10,PARP9, PARP14*
pSS-other cancers			
Interferon Signaling	1.94E+01	3.33E-01	*OAS1,IRF9,IFIT1,IFITM1,STAT1,IFIT3,STAT2, ISG15,MX1,IFI35,IFI6,IFITM3*
Activation of IRF by Cytosolic Pattern Recognition Receptors	1.27E+01	1.59E-01	*DHX58,IRF9,IRF7,STAT1,STAT2,ISG15,ADAR, DDX58,ZBP1,IFIT2*
Retinoic acid Mediated Apoptosis Signaling	5.06E+00	8.06E-02	*PARP10,PARP12,TNFSF10,PARP9,PARP14*
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	4.66E+00	4.76E-02	*EIF2AK2,OAS1,IRF7,OAS3,OAS2,DDX58*
UVA-Induced MAPK Signaling	4.32E+00	5.68E-02	*PARP10,PARP12,STAT1,PARP9,PARP14*
pSS-associated lymphoma			
Interferon Signaling	1.74E+01	4.44E-01	*OAS1,IRF9,IFITM1,IFIT3,STAT2,JAK1,IFI6,IF ITM3,TAP1,IFIT1,STAT1,ISG15,MX1,IFI35, SOCS1,PSMB8*
Activation of IRF by Cytosolic Pattern Recognition Receptors	7.25E+00	1.75E-01	*DHX58,IRF9,IRF7,STAT1,IFIH1,STAT2,ISG15, ADAR,DDX58,ZBP1,IFIT2*
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	5.70E+00	1.03E-01	*EIF2AK2,OAS1,MYD88,CREB1,PRKCQ,OAS2, CASP1,DDX58,C3AR1,IRF7,OAS3,IFIH1, PRKCH*
mTOR Signaling	5.11E+00	8.02E-02	*PRKCQ,EIF4B,EIF3L,EIF3B,RPS3,ULK1, RPS6KA5,EIF3D,PDPK1,RPS4X,PRKCH, EIF4A3,VEGFB,RRAS,FNBP1*
Huntington's Disease Signaling	4.65E+00	6.99E-02	*GNB4,CREB1,PRKCQ,NAPA,CASP1,PSME1, PSME2,HSPA6,TAF4,HSPA9,PDPK1,GNG5, SIN3A,PRKCH,HDAC1,GLS*

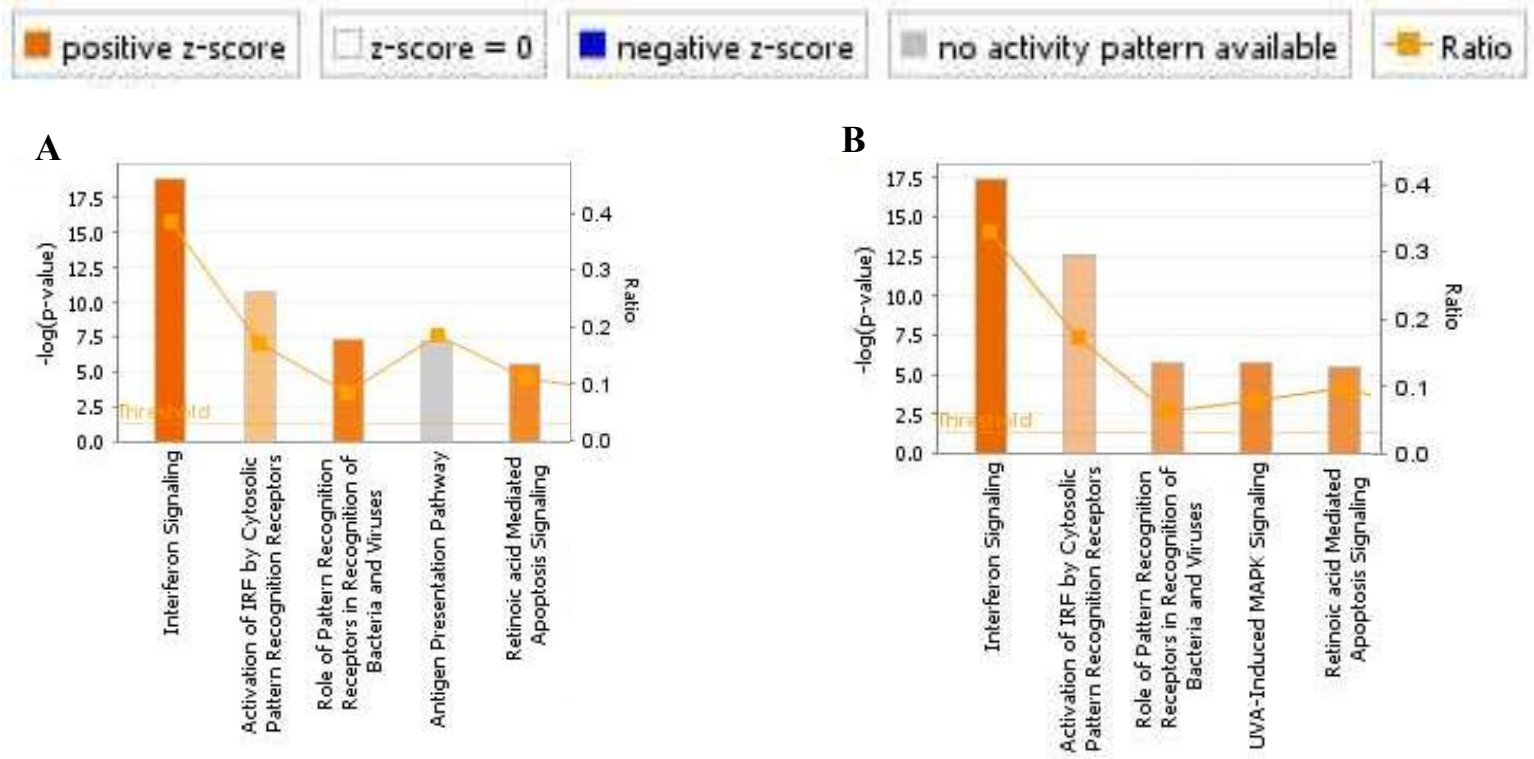


Figure 6.4 Top 5 canonical pathways analysis in pSS and pSS-paraproteinemia.*Grey bars indicate that no prediction of activation or inhibition can be made. Yellow bars represent pathways that are enriched and activated, with a significant positive z-score. Blue bars represent pathway that are enriched and inhibited, with a significant negative z-score. White bars represent the canonical pathways with a z-score near 0. The threshold line in yellow represents the statistically significant p value (0.05). The yellow dots represent the ratio of the number of genes that were present in my dataset to the total number of genes in each pathway. A. pSS, B. pSS-paraproteinemia.*

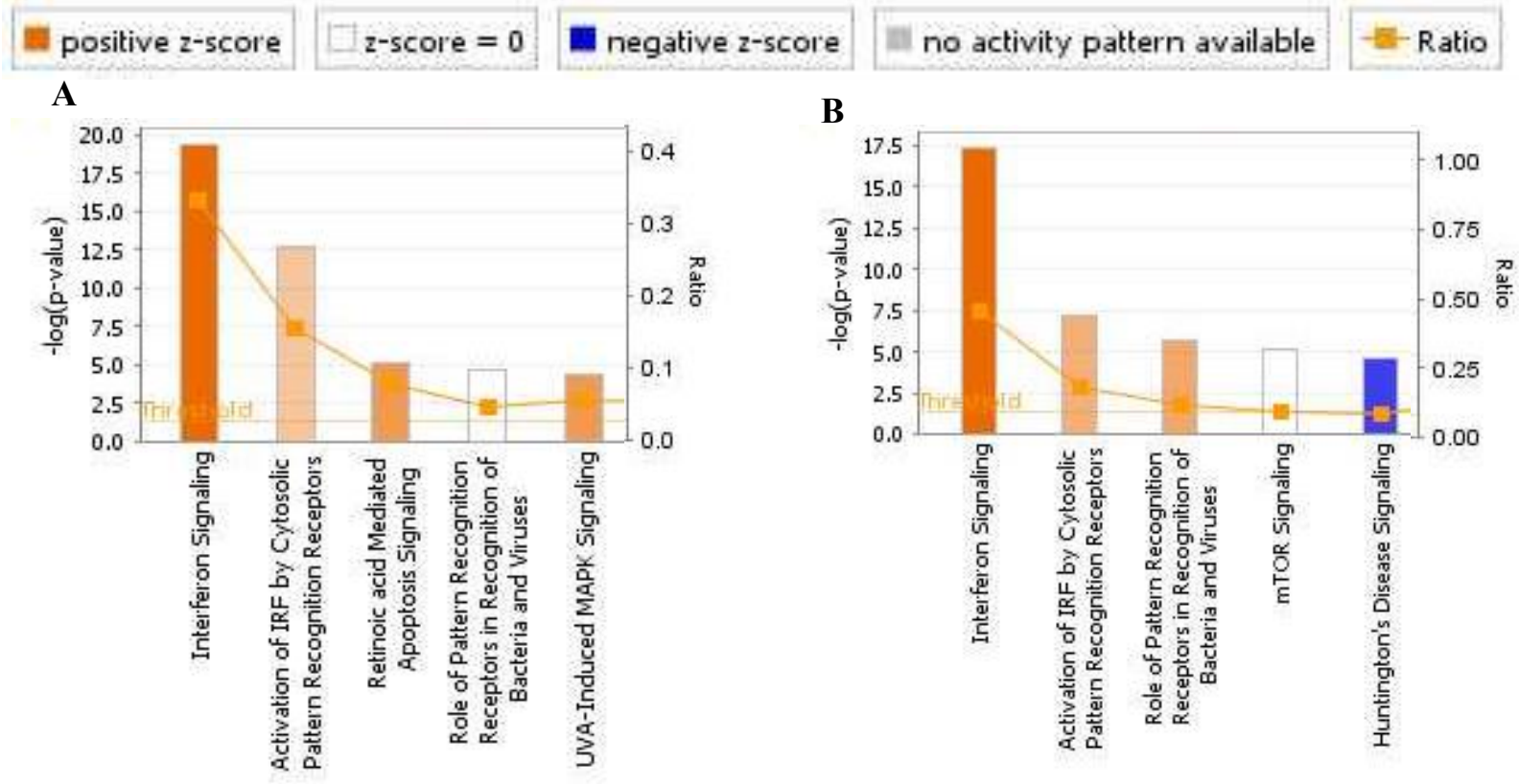


Figure 6.5 Top 5 canonical pathways analysis in pSS-other cancers and pSS-associated lymphoma *created by IPA Grey bars indicate that no prediction of activation or inhibition can be made. Yellow bars represent the pathways that are enriched and activated, with a significant positive z-score. Blue bars represent the pathways that are enriched and inhibited, with a significant negative z-score. White bars represent canonical pathways with a z-score near 0. The threshold line in yellow represents the statistically significant p value (0.05). The yellow dots represent the ratio of the number of genes present in my dataset to the total number of genes in each pathway. A. pSS-other cancers, B. pSS-associated lymphoma.*

2. The Downstream Effects analysis in pSS subgroups

The downstream effects analysis including the identification of diseases and functions in the pSS subgroups revealed the top diseases and functions associated with each pSS subgroup. For the pSS, pSS-paraproteinemia and pSS-other cancers groups, the most significant "disease and function" was "infectious diseases," as shown in Figures 6.6, 6.7 and 6.8, respectively. The most significant function for the pSS-associated lymphoma was "cell death and survival," as shown in Figure 6.9. Notably, the "infectious diseases" function was inhibited in all pSS subgroups. In contrast, the "cell death and survival" function was activated in all subgroups, and was the second most significant function in both the pSS-paraproteinemia and pSS-other cancers subgroups and the fourth most significant function in the pSS groups. "Inflammatory responses" was also increased in all 4 subgroups. "Haematological system development" and "immune cell trafficking" were increased in pSS subgroup, but were much less prominent in the other 3 subgroups. The Downstream Effects analysis with the diseases and functions identified in pSS-associated lymphoma versus healthy controls that have a significant z-score can be found in supplementary table S11.

Figure 6.6 Heatmap of the Downstream Effects analysis in the pSS group. *The size of the squares is a graphic representation of the statistical significance of the p value (larger square = more significant). The colours of the squares reflect the direction of change. Orange: IPA predicts that the biological process or function is increased, with a positive z-score (z-score ≥ 2). Blue: IPA predicts that the biological process or function is decreased, with a negative z-score (z-score ≤ -2). Grey: no prediction can be made in the current situation. White: the canonical pathways with a z-score near 0. The strength of the prediction is indicated by the intensity of the colour.*

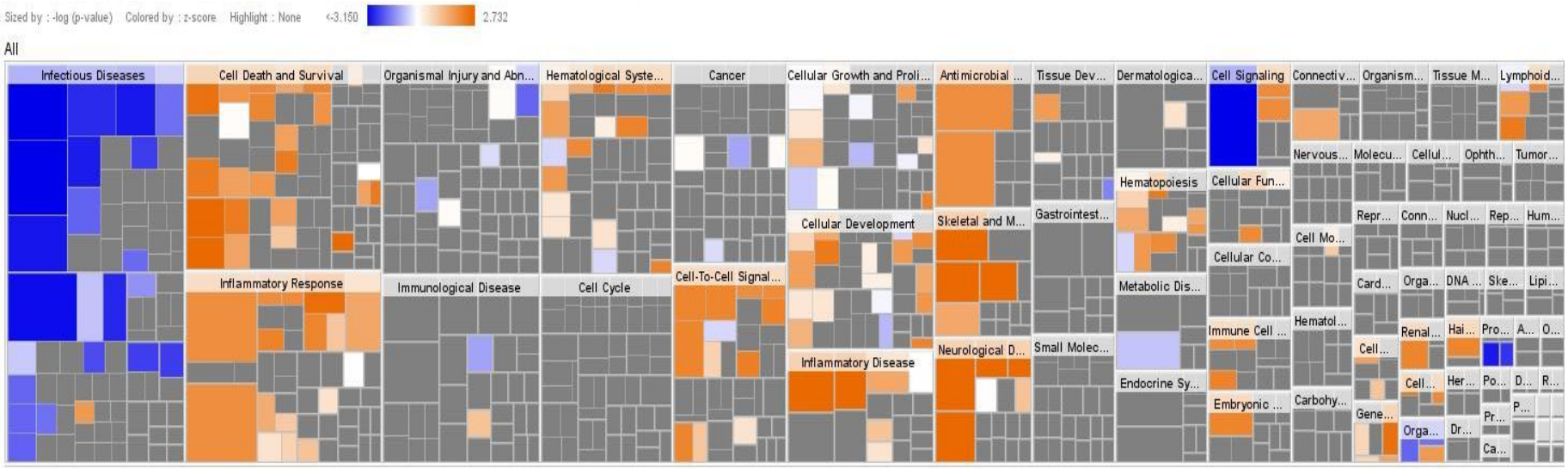


Figure 6.7 Heatmap of the Downstream Effects analysis in the pSS-paraproteinemia group. *The size of the squares is a graphic representation of the statistical significance of the p value (larger square = more significant). The colours of the squares reflect the direction of change. Orange: IPA predicts that the biological process or function is increased, with a positive z-score (z-score ≥ 2). Blue: IPA predicts that the biological process or function is decreased, with a negative z-score (z-score ≤ -2). Grey: no prediction can be made in the current situation. White: the canonical pathways with a z-score near 0. The strength of the prediction is indicated by the intensity of the colour.*

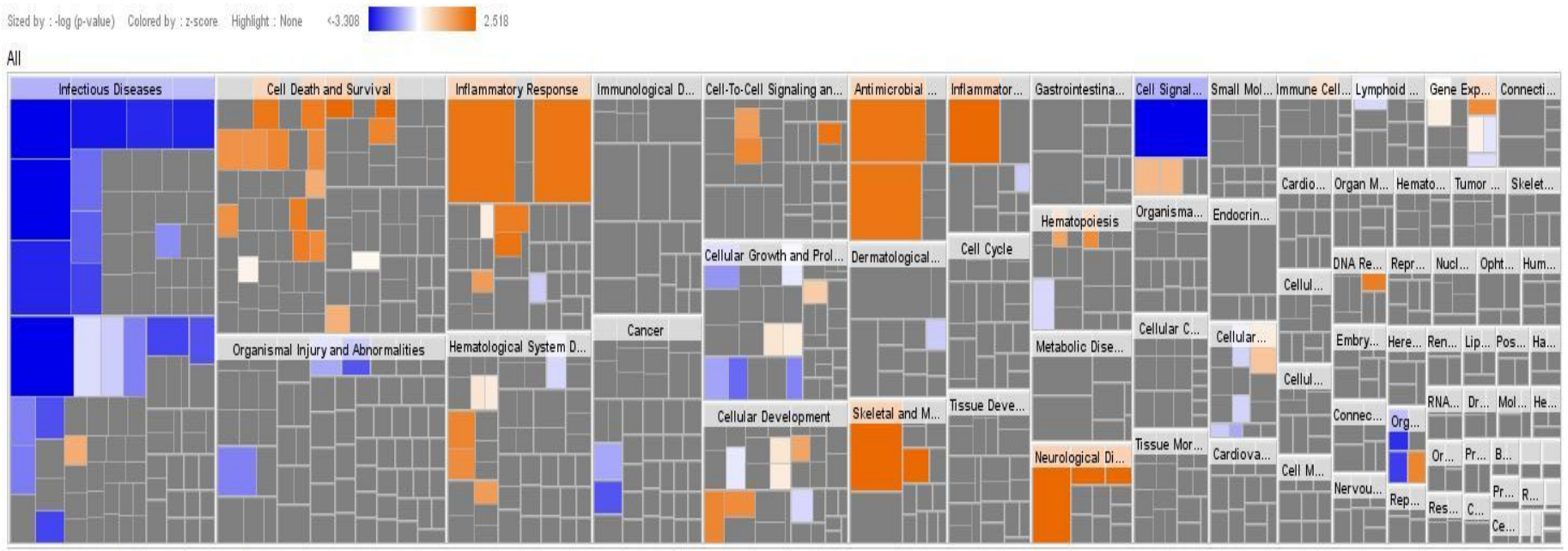


Figure 6.8 Heatmap of the Downstream Effects analysis in the pSS-other cancers group. *The size of the squares is a graphic representation of the statistical significance of the p value (larger square = more significant). The colours of the squares reflect the direction of change. Orange: IPA predicts that the biological process or function is increased, with a positive z-score (z-score ≥ 2). Blue: IPA predicts that the biological process or function is decreased with a negative z-score (z-score ≤ -2). Grey: no prediction can be made in the current situation. White: the canonical pathways with a z-score near 0. The strength of the prediction is indicated by the intensity of the colour.*

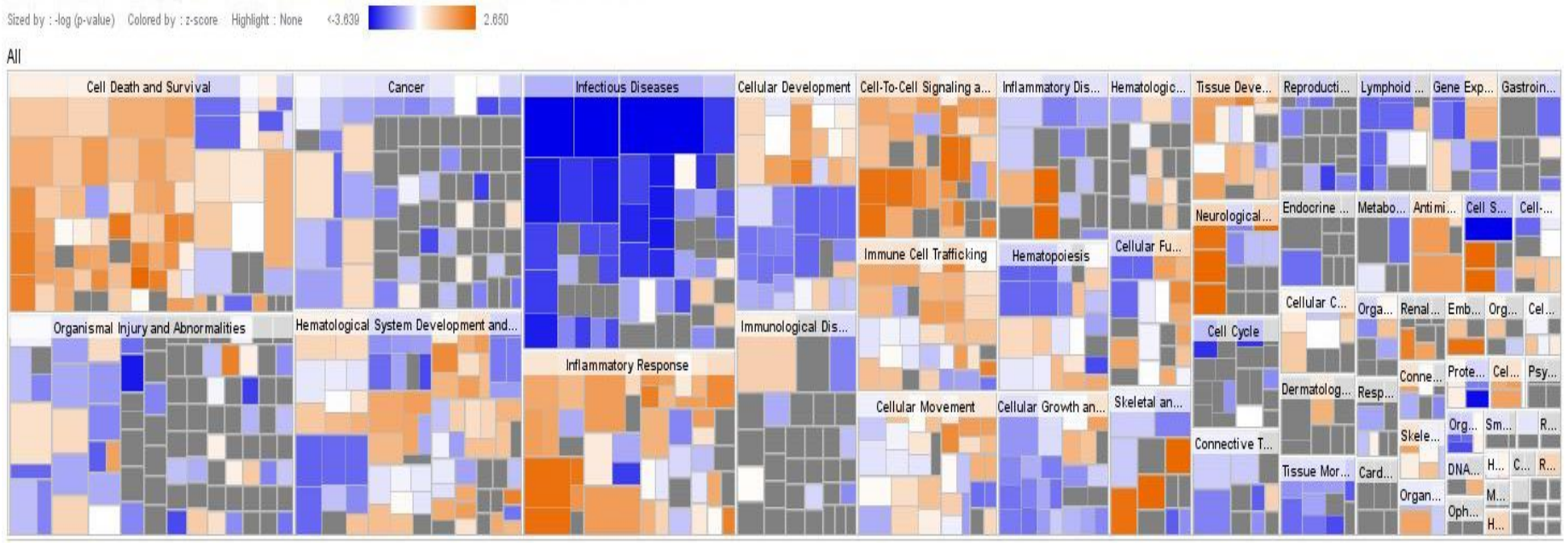


Figure 6.9 Heatmap of the Downstream Effects analysis in the pSS-associated lymphoma group. *The size of the squares is a graphic representation of the statistical significance of the p value (larger square = more significant). The colours of the squares reflect the direction of change. Orange: IPA predicts that the biological process or function is increased, with a positive z-score (z-score ≥ 2). Blue: IPA predicts that the biological process or function is decreased, with a negative z-score (z-score ≤ -2). Grey: no prediction can be made in the current situation. White: the canonical pathways with a z-score near 0. The strength of the prediction is indicated by the intensity of the colour.*

3. Molecular Networks analysis in pSS subgroups:

Many molecular networks were identified in each pSS subgroup. In the pSS group, IPA reported 13 networks. In the pSS-paraproteinemia and pSS-other cancers groups, IPA identified 10 and 8 networks respectively, whereas in the pSS-associated lymphoma group, 25 different networks were identified. In the network analysis, the term "Focus molecules" refers to the number of genes in the uploaded DEG list that were also represented in the network, while the "Score" is the number of genes in the network (including genes from the uploaded DEG list and the genes that were added from the indirect interactions by IPA). I first focus on *DRAP1*, which was the only DEG in common for all pSS subgroups compared to HC. In the pSS subgroup, the network that involves *DRAP1* includes biological functions such as "Antimicrobial Response, Inflammatory Response and Cell Signaling" (Score = 27, Focus molecules = 19) with *DRAP1* interacted directly with *HLA-DRA* in this network. Importantly, In the pSS-paraproteinemia and the pSS-other cancers groups, *DRAP1* was involved in the network including biological functions of "Cellular Development, Cell Death and Survival and Tissue Morphology" (Score = 22, Focus molecules = 15) and "Cellular Development, Haematological System Development and Function and Haematopoiesis" (Score = 22, Focus molecules = 14) respectively. In both networks, *DRAP1* interacted directly with *MYC*. The three networks are shown in Figures 6.10, 6.11 and 6.12.

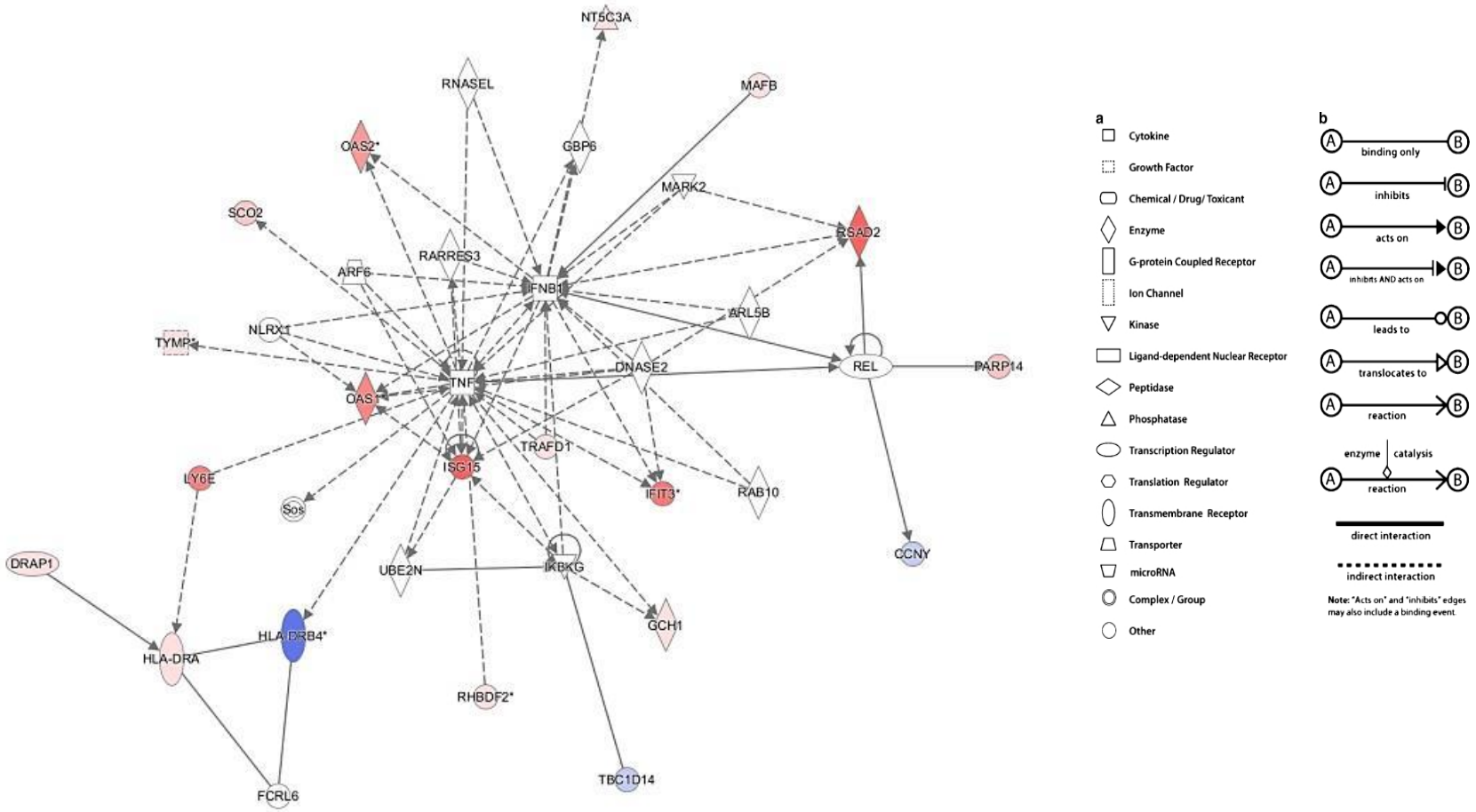


Figure 6.10 Network analysis of the pSS group. *The networks show the interactions of DRAP1.*

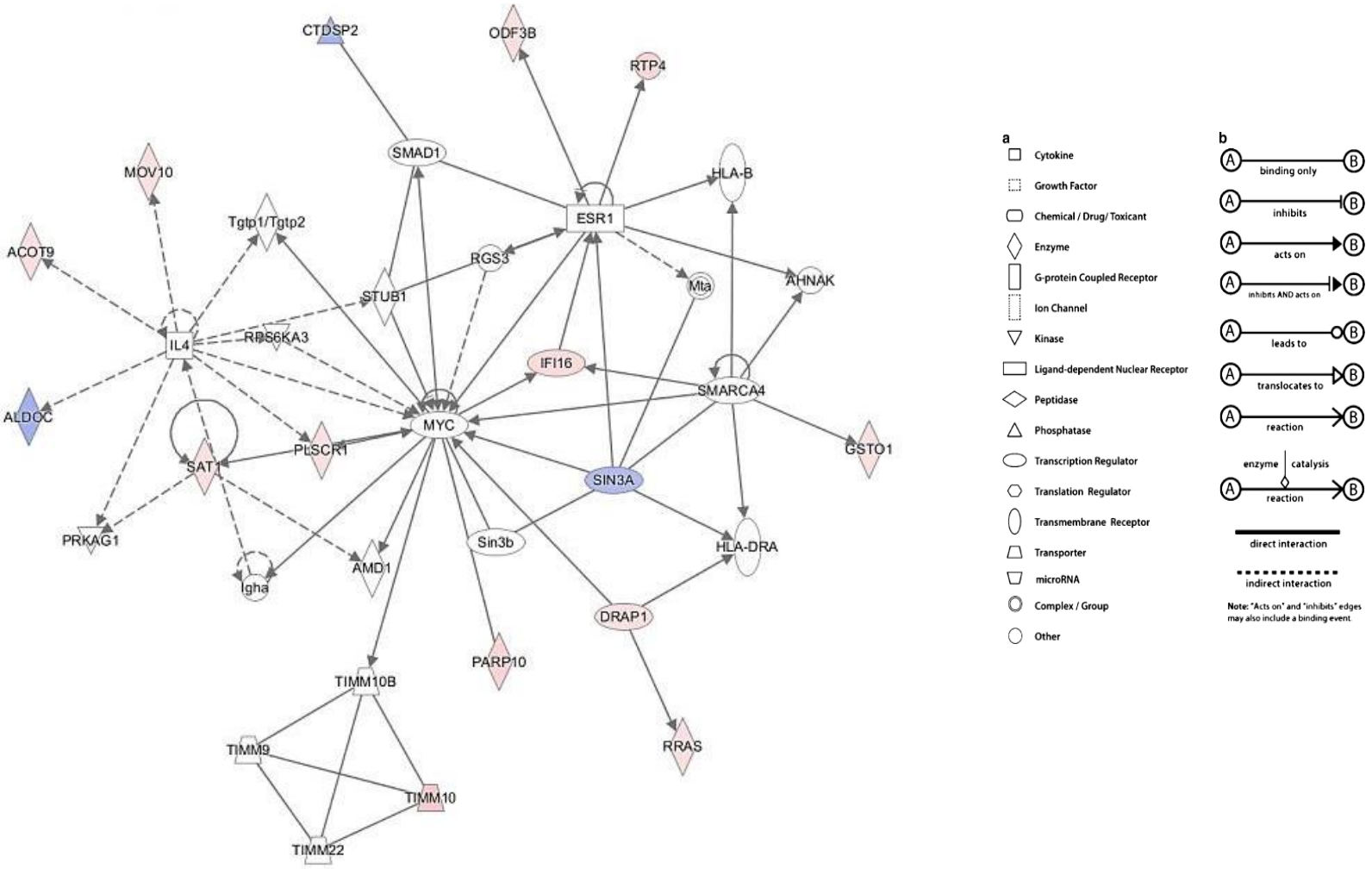


Figure 6.11 Network analysis of the pSS-paraproteinemia group. *The networks show the interactions of DRAP1.*

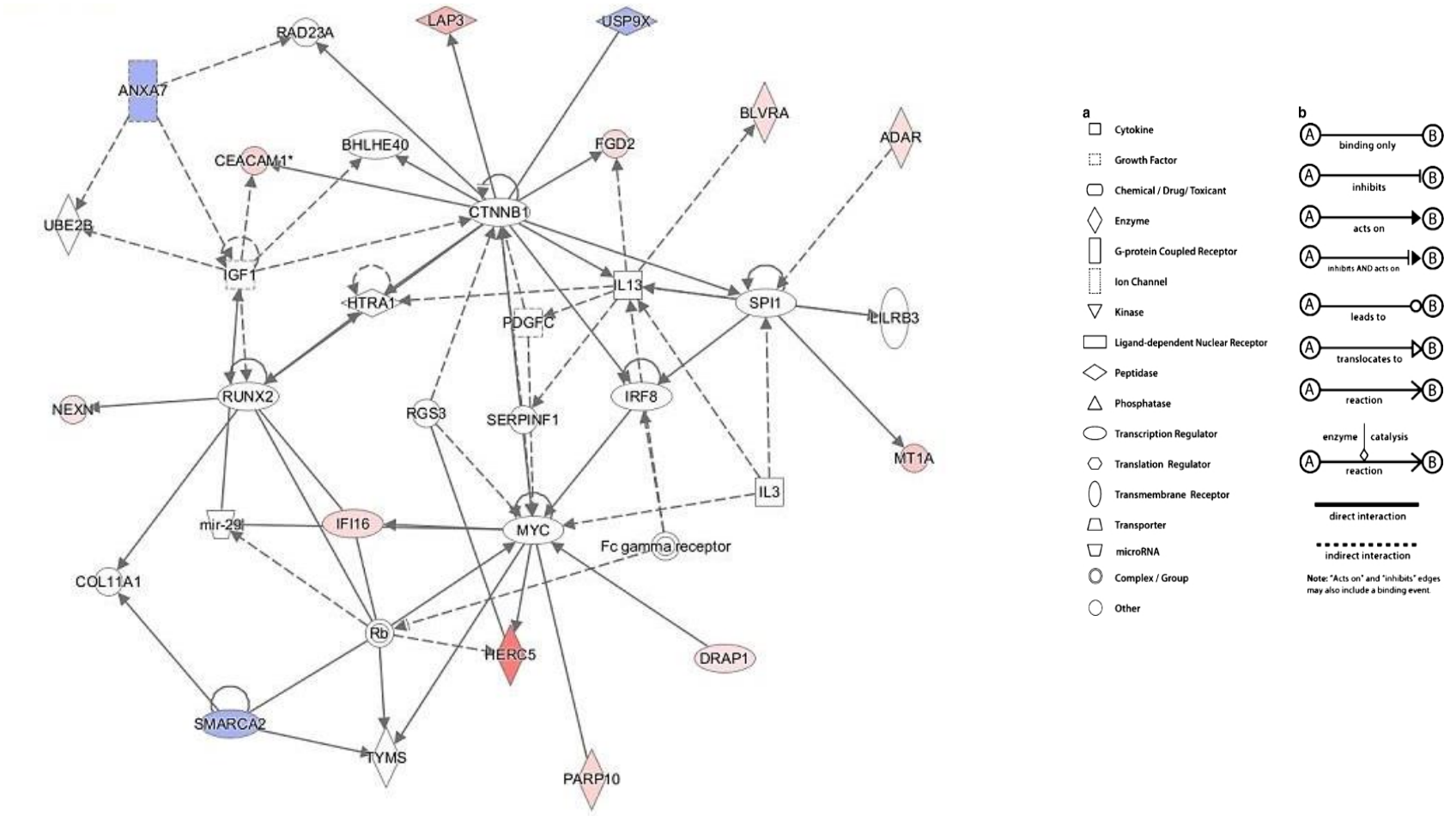


Figure 6.12 Network analysis of the pSS-other cancers group. *The networks show the interactions of DRAP1.*

In the comparison of pSS-associated lymphoma versus healthy controls, 25 networks were identified. When investigating the networks that involves the genes from the 3-gene biosignature in pSS-associated lymphoma, only two of the genes (*NUDT14* and *MGST3*) appeared in two networks. There were two other networks involving the two differentially expressed genes (*DRAP1* and *DYNLL1*) in untreated pSS-associated lymphoma. *NUDT14* was involved in the "Molecular Transport, RNA Trafficking, RNA Post-Transcriptional Modification" network (Score = 6, Focus molecules = 7). This gene interacted indirectly with *TCOF1* in the network. *TCOF1* had a direct interaction with *WT1*, which binds to *TP53* (Figure 6.13). *MGST3* was involved in the "Inflammatory Disease, Respiratory Disease, Antigen Presentation" network (Score = 15, Focus molecules = 15) and was regulated by *PPARA* (Figure 6.14). *DRAP1* was included in the network "Lipid Metabolism, Molecular Transport, Nucleic Acid Metabolism" (Score = 43, Focus molecules = 30), in which *DRAP1* directly interacted with *MYC* (Figure 6.15). Interestingly, the network containing *DRAP1* was the most significant in the pSS-associated lymphoma versus healthy control comparison. *DYNLL1* was included in the "Infectious Diseases, Antimicrobial Response, Inflammatory Response" network (Score = 32, Focus molecules = 25), with an indirect interaction with the *NF-κB* complex (Figure 6.16).

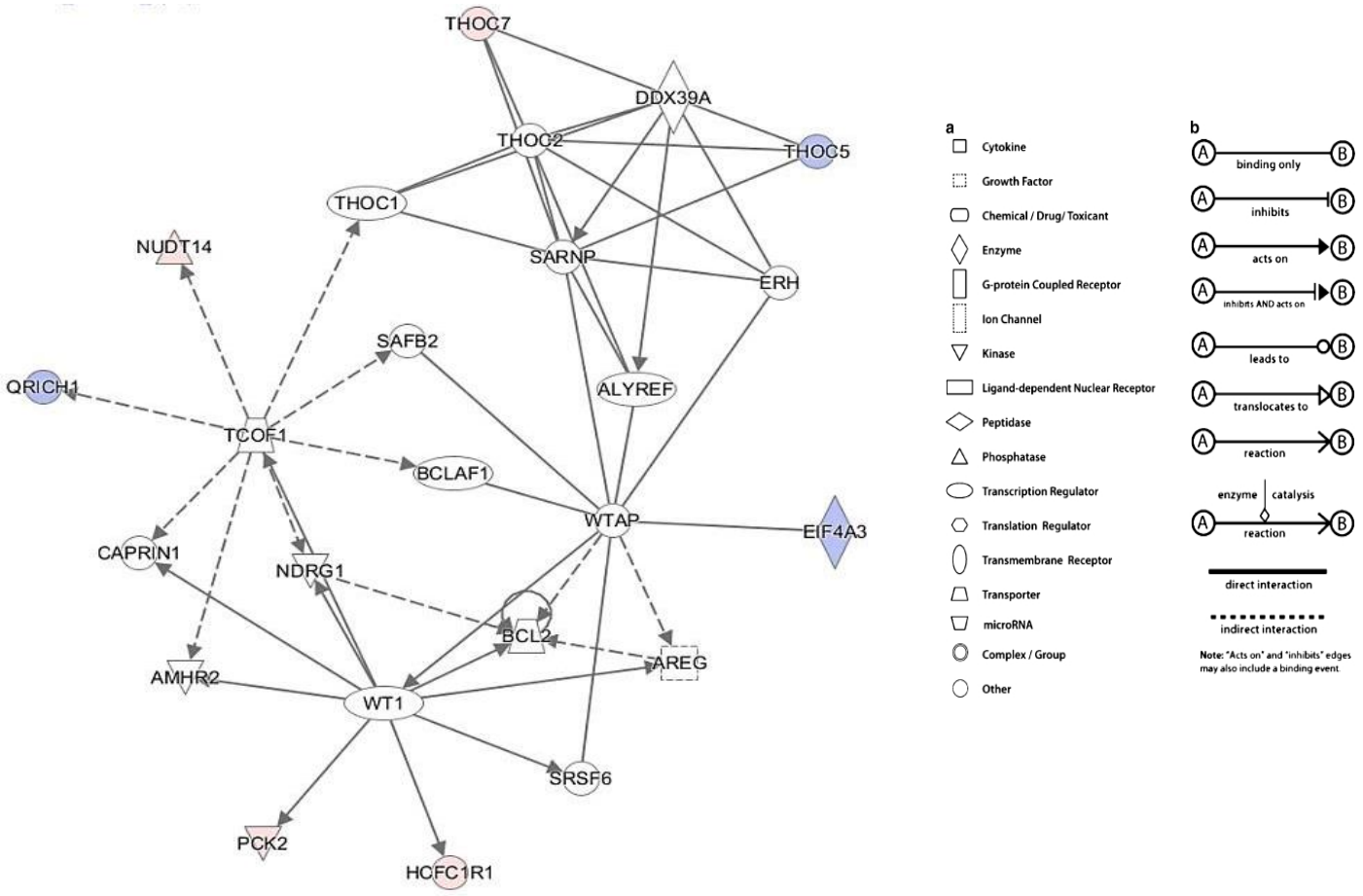


Figure 6.13 The interactions of NUDT14 in the network analysis of pSS-associated lymphoma versus healthy controls.

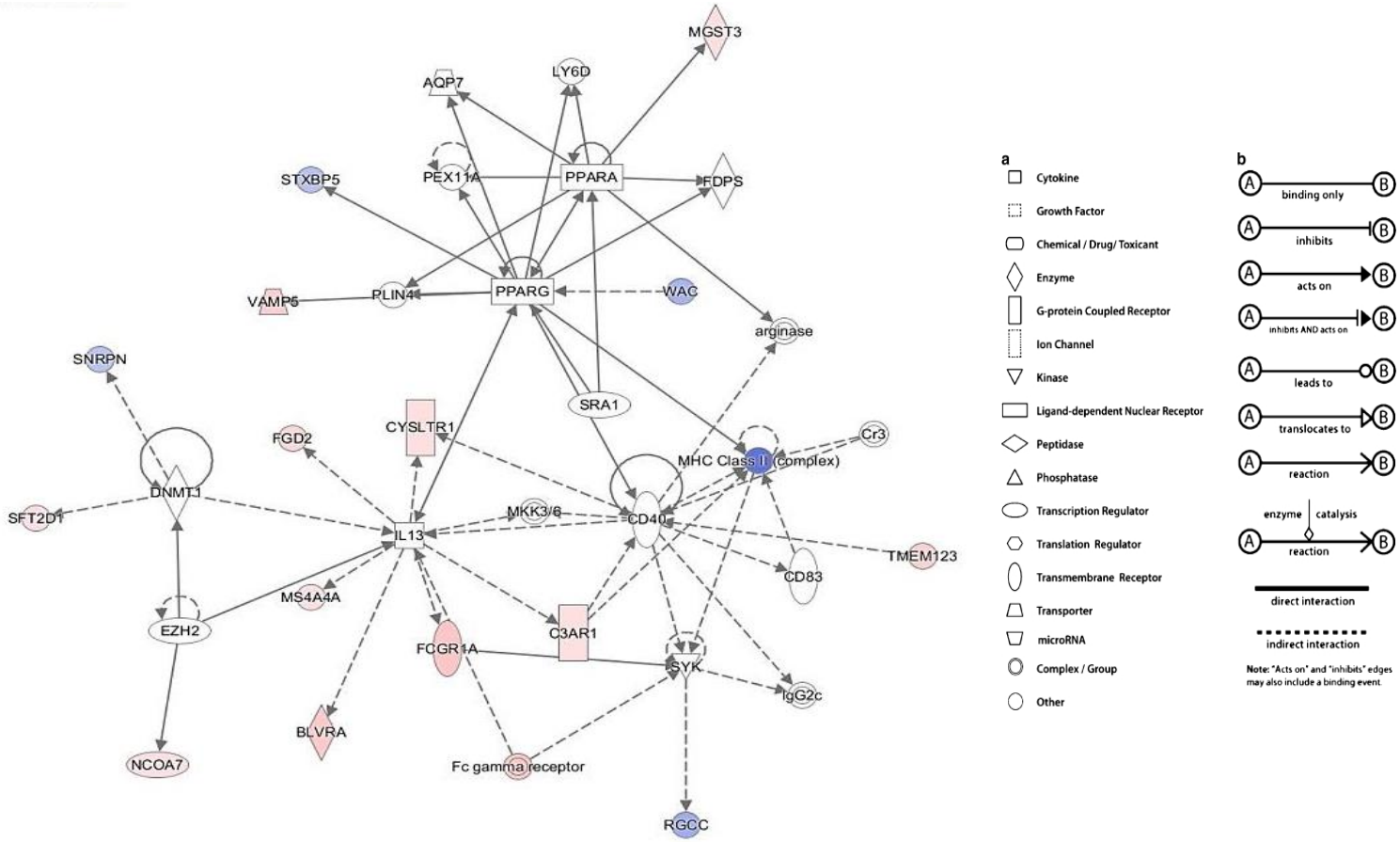


Figure 6.14 The interactions of MGST3 in the network analysis of pSS-associated lymphoma versus healthy controls.

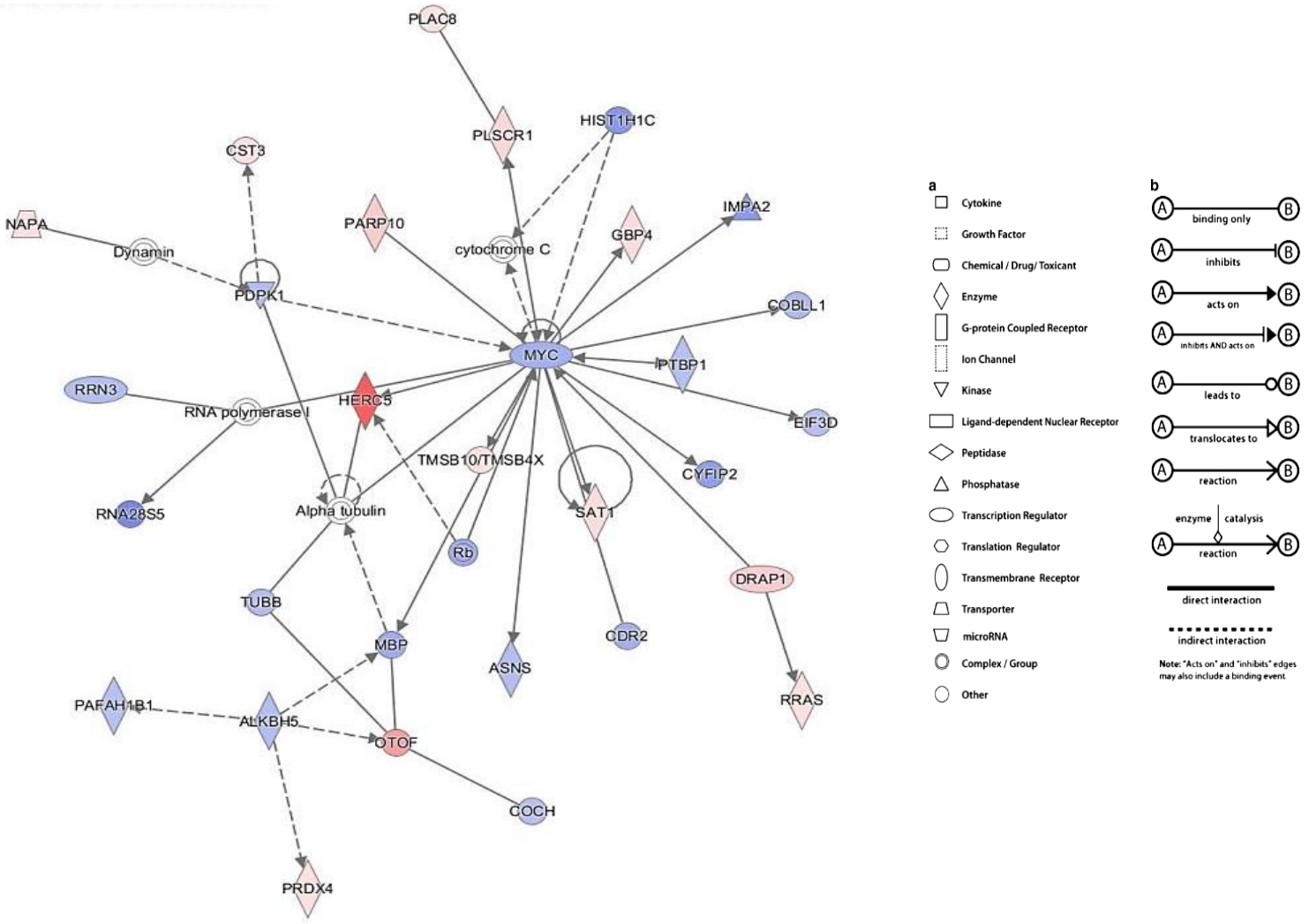


Figure 6.15 The interactions of DRAP1 in the network analysis of pSS-associated lymphoma versus healthy controls.

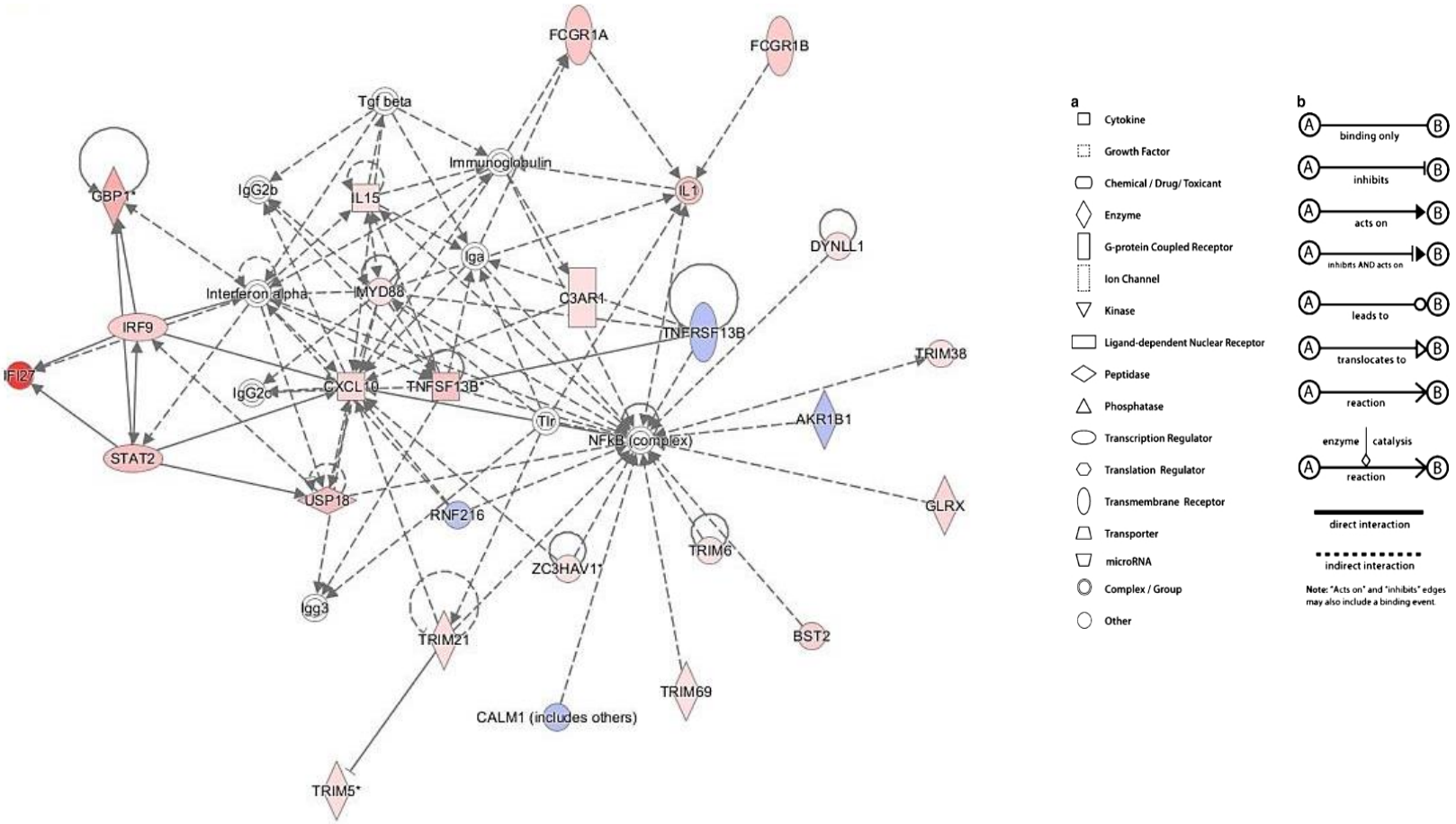


Figure 6.16 The interactions of DNYLL1 in the network analysis of pSS-associated lymphoma versus healthy controls.

4. Upstream regulators:

Analysis of the upstream regulators revealed that a total of 401 upstream regulators were identified in pSS, 370 in pSS-paraproteinemia and 255 in pSS-other cancers. For pSS-associated lymphoma 495 regulators were identified. Among these 495 regulators, 87 with absolute z-scores ≥ 2 or z-scores ≤ -2 (58 regulators were activated and 29 regulators were inhibited). The top 10 activated upstream regulators showed a predominance of IFN regulators, as shown in Table 6.3. When investigating the upstream regulators of the genes constituting the 3-gene biosignature of pSS-associated lymphoma and the 2 DEGs in untreated pSS-associated lymphoma, two of the three upstream regulators identified were the same as the upstream regulators identified previously in the "pSS versus pSS-associated lymphoma" comparison (chapter 5). These two regulators were *NFE2L2* and *TCOF1*. The other upstream regulator was *NR3C1*. Two regulators involved in the regulation of *NUDT14* and *MGST3* and one upstream regulator controls both *MGST3* and *DYNLL1* (Table 6.4). The Upstream regulators analysis identified in pSS-associated lymphoma versus healthy controls that have a significant z-score can be found in supplementary data S12.

Table 6.3 Top 10 activated upstream regulators in pSS-associated lymphoma versus healthy control

Upstream Regulator	Molecule Type	Predicted Activation State	Activation z-score	p-value
IFNL1	cytokine	Activated	6.664	9.18E-66
IFNA2	cytokine	Activated	6.539	6.77E-55
PRL	cytokine	Activated	6.291	2.15E-54
IRF7	transcription regulator	Activated	5.798	4.99E-40
IRF3	transcription regulator	Activated	5.944	1.46E-38
Ifnar	group	Activated	5.728	5.59E-37
IFNG	cytokine	Activated	7.442	1.91E-35
STAT1	transcription regulator	Activated	5.378	4.25E-31
MAVS	other	Activated	4.516	6.85E-26
IFNB1	cytokine	Activated	5.093	2.56E-25

Table 6.4 The upstream regulators of the genes constituting the 3-gene biosignature and DEGs in untreated pSS-associated lymphoma.

Upstream regulator	Molecule	p-value of overlap	Target molecules in uploaded DEG list from my experiment
NR3C1	Ligand-dependent nuclear receptor	3.99E-03	*ADCK3,AKTIP,ARRB1,BAG3,BTG1,CCL, CXCL10,**DYNLL1**,FOXO1,GADD45B, HSPA6,IFI6,IFIH1,IFIT2,IL15,ISG15, MYC,MYD88,OASL,SERTAD2,TNFAIP6*
NFE2L2	Transcription regulator	9.47E-03	*ACTG1,ARF1,ATP1A1,CXCL10,**DYNLL1**, EPHB4,GSTO1,HSP90B1,IL1RN,LY6E, **MGST3**,PAFAH1B1,VCP*
TCOF1	Transporter	4.49E-02	*AKR1B1,HNRNPDL,**NUDT14**,QRICH1*

** The genes from the 3-gene biosignature and those that were differentially expressed in untreated pSS-associated lymphoma are highlighted in bold*

6.4 Discussion

Microarray gene expression data in pSS subgroups

The aim of this chapter was to investigate whether the identified whole blood gene expression signature of pSS-associated lymphoma was also present in other pSS subgroups. The microarray gene expression data from the Discovery cohort revealed DEGs only in the comparisons between the pSS subgroups and healthy controls. In contrast, there were no DEGs identified in all other comparisons, with the exception of 1 gene (*RNA28S5)* that was differentially expressed between the pSS-associated lymphoma and pSS-other cancers groups but this gene was not validated. Importantly, the DEGs identified between the 3 non-lymphoma pSS subgroups and healthy controls did not include the genes of the 3-gene biosignature of pSS-associated lymphoma. These observations indicate that the gene expression signature I have identified is unique to pSS-associated lymphoma. On the other hand, a dominant interferon signature was apparent in all 4 pSS subgroups. It should however be noted that IFN activation is not only well-documented in pSS, but also one of the key shared signatures with other inflammatory systematic diseases such as SLE and RA (Emamian et al., 2009, Toro-Dominguez et al., 2014). At an individual gene level, the gene *DRAP1* may be of particular interest, as the differential expression of this gene was validated in all 4 pSS subgroups by qRT-PCR. *DRAP1* is a transcriptional factor that has a role in either inhibition or activation of transcription (Cang and Prelich, 2002, Castaño et al., 2000, Creton et al., 2002, White et al., 1994).

The Canonical pathways in pSS subgroups

Consistent with the microarray data, "Interferon signaling pathway" was a canonical pathway being identified in all 4 pSS subgroups. Type I interferon signaling activation in pSS has already been reported (Yao et al., 2013). Type I interferon signaling is crucial for pSS pathogenesis, as knockout mice (B6.*Aec1Aec2Ifnar1*$^{-/-}$) that lack interferon alpha receptor 1 (*Ifnar1*) exhibit less salivary gland dysfunction (Szczerba et al., 2013). Moreover, it has been reported that type I interferon, together with other factors, are associated with both pSS and pSS-associated lymphoma (reviewed recently by (Nezos and Mavragani, 2015). Interestingly, in the canonical pathway analysis, almost all of the canonical pathways identified in pSS versus healthy controls were included in the list of canonical pathways identified in pSS-associated lymphoma versus healthy controls, with the exception of only 6 canonical pathways. Interestingly, when comparing the canonical pathways identified in the comparison between pSS-associated lymphoma and pSS (identified in Chapter 5), the top 5 canonical pathways were not included in the top 40 canonical pathways identified from the comparison between pSS-associated lymphoma and healthy controls. These data suggest that the canonical pathways identified in Chapter 5 comparing pSS and pSS-Lymphoma represent the important and unique pathways for pSS-associated lymphoma.

The Downstream Effects and Networks analyses of pSS subgroups

Another interesting observation from the IPA was the Downstream Effects analysis. The pSS-associated lymphoma clearly displayed a different pattern from the other three pSS subgroups, which showed a more similar pattern with one other. Furthermore, the relative importance of the different "diseases and functions" also differed between the pSS-associated lymphoma and the other pSS subgroups. Notably, "cell death and survival" was the most important downstream effects in the lymphoma group. This is perhaps unsurprising given the link between cell death and survival is well known in malignancies (Labi and Erlacher, 2015).

The network analysis from different pSS subgroups revealed the interactions between the genes within the uploaded DEGs list from my experiment and with other genes in the pathways found in the literature. The networks analysis of pSS-associated lymphoma

versus healthy controls showed that two of the genes in the 3-gene biosignature in pSS-associated lymphoma (with the exception of *BMS1*) were included in two networks (out of a total of 25 networks) identified by IPA. Furthermore, the two genes that were differentially expressed in untreated pSS-associated lymphoma were also included in two networks. All these genes had interactions with similar genes to those identified and discussed in Chapter 5.

The Upstream regulators analysis in pSS subgroups

Focusing on the upstream regulators of the genes in the 3-gene biosignature, IPA revealed two upstream regulators *NFE2L2* and *TCOF1*. Both regulators have been also identified in when using the DEG list from the comparison between pSS and pSS-associated lymphoma (see chapter 5). The roles of *NFE2L2* and *TCOF1*, have been discussed in chapter 5 with the former regulator linked to antioxidant functions and the latter related to cell proliferation. The analysis also identified *NR3C1* as an upstream regulator of *DYNLL1*, which was differentially expressed in untreated the pSS-associated lymphoma. *NR3C1* (nuclear receptor subfamily 3 group C member 1), is a glucocorticoid receptor as well as a transcriptional factor. *NR3C1* has recently been linked to tumorigenesis of adult acute lymphoblastic leukaemia (ALL) (Safavi et al., 2015) as well as being identified as a candidate gene related to the development of lung adenocarcinoma (Zhao et al., 2015). Therefore, further investigation of *NR3C1* regulation of *DYNLL1* might be important in understanding the pathogenesis of pSS-associated lymphoma.

To summarise, no DEGs were identified in the comparison between the 3 non-lymphoma pSS subgroups. In the Downstream Effects analysis, all 3 non-lymphoma pSS subgroups exhibited a similar pattern with only minor differences, but were substantially different from the pSS-lymphoma group. This supports the notion that distinct biological mechanisms are involved in lymphoma development and pSS pathogenesis. Finally, further investigation of the role of the transcriptional factor *DRAP1* in pSS pathogenesis and lymphoma development might be worthwhile.

Chapter 7
Conclusion and future work

7.1 Conclusion:

During my PhD study, I have focused on the identification of a whole-blood gene expression signature in pSS-associated lymphoma. Several gene expression studies have been carried out regarding pSS, focusing on gene-expression profiling in the most affected organs (salivary and lacrimal glands). Only one study has used whole-blood samples from pSS patients (Emamian et al., 2009). No data was found in the literature, however, regarding whole-blood gene expression profiling in pSS-associated lymphoma.

By identifying a whole-blood gene expression signature in pSS-associated lymphoma, it will help to understand the molecular mechanisms underlying the development of lymphoma in pSS patients and to discover genetic biomarkers that might help to predict, diagnose or monitor lymphoma development in pSS. Such biomarkers will have substantial translational potentials in the clinic.

In contrast to salivary gland biopsy, which is an invasive procedure which can lead to permanent paraesthesia, obtaining peripheral blood samples is simple and widely acceptable to patients. In addition, collecting blood specimens can be easily repeated at regular intervals. One major problem with analysing gene expression using whole blood samples is the presence of globin mRNA. This abundance of globin mRNA can interfere with gene expression profiling studies. For this reason in my project, I first examined the impact of globin mRNA on gene expression study in pSS. This optimisation step is important as it has been recommended that globin mRNA depletion should be optimised for different experiments and microarray platforms (Dumeaux et al., 2008). In this experiment, I used pSS whole-blood samples and not pSS-associated lymphoma samples because peripheral blood gene expression data in pSS are available in the literature providing a robust set of data for comparison with my work. As described in chapter 3, I used paired samples with or without globin mRNA depletion from 12 pSS patients and 12 healthy controls. Globin mRNA depletion resulted in an increase in the microarray signal intensity, more transcripts being detected and a higher number of differentially expressed genes being identified. The increased number of detectable transcripts indicated that the depletion

of globin mRNA may improve the detection of low-abundance genes (Kam et al., 2012). However, the overall gene expression profiles for pSS using samples with or without globin mRNA depletion were comparable. Important, these gene expression profiles were consistent with the published data in the literature. The concordance between the globin-depleted and non-globin depleted datasets was more noticeable among the up-regulated genes, characterised by the presence of an activated type I IFN-inducible gene signature.

Unlike the clear differences in gene expression profiles between pSS patients and healthy controls, I reasoned that the differences in gene expression between pSS and pSS-associated lymphoma may be more subtle for two reasons. Firstly, lymphoma development is likely to be a multi-step process from pre-lymphoma stage to lymphoma, therefore, gene expression changes relevant to lymphoma development may have occurred in some of the non-lymphoma pSS patients. Furthermore, the majority of the samples that I will be using for the experiment were from pSS patients who have already received treatment for their lymphoma, which may also blunted the differences in gene expression profiles between lymphoma and non-lymphoma cases. Therefore, I have decided to deplete globin mRNA from my samples for my microarray experiments as it may increase the sensitivity of detecting differentially expressed genes between the lymphoma and non-lymphoma groups.

In the discovery experiment, 144 globin mRNA-depleted samples were used. There were five subject groups: first, the target group (pSS-associated lymphoma), and second, the disease control groups (pSS), in which the comparison will be made to identify the gene expression signature for pSS-associated lymphoma. The other three groups included two subgroups of pSS (pSS with paraproteinemia, pSS with other cancers) and a healthy control group. The inclusion of these groups was to test the "specificity" of the identified gene expression signature for pSS-associated lymphoma. Sixty-eight DEGs were identified in pSS-associated lymphoma from the microarray data of the Discovery cohort (68-DEGs-Mi). Due to limited availability of RNA samples, the expression levels of only 61 genes were measured with qRT-PCR, which validated the differential expression of 26 genes (26-DEGs-MiPCR). These 26-DEGs-MiPCR was selected as candidate whole-blood gene expression signature for pSS-associated lymphoma.

Due to limited availability of RNA samples, the expression levels of 24 out of the 26 genes were measured in a second independent cohort (Validation cohort) using qRT-PCR. The

Validation cohort consisted of only pSS patients without lymphoma and pSS-associated lymphoma. This cohort has included patients from our collaborators in Sweden and Norway as well as the UKPSSR samples. Three genes (*BMS1*, *NUDT14* and *MGST3)* were validated and they were referred to as the '3-gene biosignature of pSS-associated lymphoma'.

During the study, I was able to obtain a few samples from the UKPSSR and our collaborators in Sweden of pSS patients with lymphoma before treatment was initiated. The expression levels of the same 24 genes were also tested for these samples. Six genes (*NUDT14*, *DRAP1*, *DYNLL1*, *RBP7*, *SF3A1* and *VCP*) were differentially expressed between the untreated lymphoma and the non-lymphoma pSS groups. The differential expression for the latest 3 genes (*RBP7*, *SF3A1* and *VCP*) was not confirmed in the Validation cohort as the genes have opposite regulation direction, therefore, for my thesis I just considered the first 3 genes (*NUDT14*, *DRAP1* and *DYNLL1*) for discussion. Interestingly *NUDT14* is also one of the constituent genes of the 3-gene biosignature of treated pSS-associated lymphoma. Furthermore, *DRAP1* and *DYNLL1* were also candidate DEGs for treated pSS-associated lymphoma even though differential expression was not confirmed in the Validation cohort with $p > 0.05$ but they have the same regulation direction. The summary findings of the DEGs identified in each stage of the experiment are shown in Table 7.1.

Table 7.1 Summary of the DEGs in pSS-associated lymphoma at all the project's stages

Discovery cohort				**Validation cohort**		**6 -DEGs in untreated pSS-associated lymphoma**	
68-DEGs-Mi		**26-DEGs-MiPCR**		**3-gene biosignature of pSS-associated lymphoma**			
Gene symbol	**regulation in pSS-associated lymphoma**	**Gene symbol**	**regulation in pSS-associated lymphoma**	**Gene symbol**	**regulation in pSS-associated lymphoma**	**Gene symbol**	**regulation in pSS-associated lymphoma**
ATG12	upregulated	*C10orf32(BORCS7)*	upregulated	*MGST3*	upregulated	*DRAP1*	upregulated
C10orf32(BORCS7)	upregulated	*CNPY3*	upregulated	*NUDT14*	upregulated	*DYNLL1*	upregulated
CNPY3	upregulated	*DRAP1*	upregulated	*BMS1*	downregulated	*NUDT14*	upregulated
DRAP1	upregulated	*DYNLL1*	upregulated			*RBP7*	upregulated
DYNLL1	upregulated	*HNMT*	upregulated			*SF3A1*	upregulated
EHBP1L1	upregulated	*LGALS1*	upregulated			*VCP*	upregulated
HCFC1R1	upregulated	*MGST3*	upregulated				
HNMT	upregulated	*NUDT14*	upregulated				
KCTD12	upregulated	*OAF*	upregulated				
LGALS1	upregulated	*RBP7*	upregulated				
MGST3	upregulated	*SEC61G*	upregulated				
NUDT14	upregulated	*SRP14*	upregulated				
OAF	upregulated	*UBXN11*	upregulated				
PSMC1	upregulated	*BMS1*	downregulated				
RAB37	upregulated	*BTBD11*	downregulated				
RBP7	upregulated	*CBLL1*	downregulated				
RNF7	upregulated	*CYFIP2*	downregulated				
SEC61G	upregulated	*ESYT1*	downregulated				
SRP14	upregulated	*HNRNPUL1*	downregulated				

UBXN11	upregulated	*LEF1*	downregulated				
ALDH9A1	downregulated	*MAGED1*	downregulated				
ATP1A1	downregulated	*POM121C*	downregulated				
BAG3	downregulated	*PRPF8*	downregulated				
BCL11B	downregulated	*SF3A1*	downregulated				
BMS1	downregulated	*SGK223*	downregulated				
BTBD11	downregulated	*VCP*	downregulated				
CBLL1	downregulated						
CD96	downregulated						
CDR2	downregulated						
CDV3	downregulated						
CYFIP2	downregulated						
DDB1	downregulated						
ESYT1	downregulated						
ETS1	downregulated						
FOXJ3	downregulated						
HLA-DRB1	downregulated						
HNRNPA1P10	downregulated						
HNRNPDL	downregulated						
HNRNPUL1	downregulated						
HSP90B1	downregulated						
HSPA9	downregulated						
ITK	downregulated						
KHDRBS1	downregulated						
LEF1	downregulated						
LRFN3	downregulated						

LRIG1	downregulated						
MAGED1	downregulated						
MYC	downregulated						
NAT10	downregulated						
NCSTN	downregulated						
PAF1	downregulated						
POM121C	downregulated						
PRKCQ	downregulated						
PRPF8	downregulated						
RASGRP1	downregulated						
RBL2	downregulated						
RNA28S5	downregulated						
RPA2	downregulated						
RPL23AP5	downregulated						
RRN3	downregulated						
SDHA	downregulated						
SF3A1	downregulated						
SGK223	downregulated						
SMARCA2	downregulated						
SPOCK2	downregulated						
SUN2	downregulated						
VCP	downregulated						
WAC	downregulated						

Another approach that I used to identify the gene expression signature for pSS-associated lymphoma was to use machine-learning method. This method enable me to predict which genes among the 60 DEGs from the qRT-PCR data in the Discovery cohort were most important in predicting the group membership of pSS-associated lymphoma. This method has yielded two prediction models. Both models identified *NUDT14* being the best gene in distinguishing pSS patients with or without lymphoma.

To further explore my microarray data, pathway analysis was performed using the IPA. Due to the small number of validated DEGs between the lymphoma and non-lymphoma groups (i.e. only 3 genes), I chose to use the 68 DEGs identified from the microarray experiment in the Discovery cohort (68-DEGs-Mi) for the pathway analysis. The top 5 canonical pathways identified were "Aryl Hydrocarbon Receptor (AHR) signalling," "Histamine Degradation," "Unfolded protein response," "Neuregulin Signalling," and "T Cell Receptor Signalling." AHR signalling appeared to have a special and important role in the pathology of pSS. Moreover, *MGST3* is one of the genes in the AHR signalling pathway. Several other canonical pathways also included the genes of the 3-gene biosignature of pSS-associated lymphoma. Interestingly, all the pathways that contained these 3 genes related to metabolic functions. In addition to canonical pathway analysis, downstream effects and gene-gene interactions were explored through the molecular networks analysis. Additionally, important upstream regulators of the 3 biosignature genes include *NFE2L2*, *PPARA* and *TOCF1*. *NFE2L2* and *PPARA* are regulators of *MGST3* and also *DYNLL1* (which was differentially expressed in untreated pSS-associated lymphoma) whereas *TOCF1* regulates *NUDT14*.

In conjunction with the gene expression profiling and the pathway analysis of pSS-associated lymphoma, I have also investigated the gene expression profiling and performed pathway analyses of the other comparison groups from the Discovery cohort. There were no DEGs among these additional pSS subgroups comparisons (with the exception of one DEG (*RNA28S5*) between the "pSS-associated lymphoma" and "pSS-other cancers" groups). DEGs were however identified between each of these pSS subgroups and healthy controls. The DEGs between these pSS subgroups and healthy controls were used for pathway analyses using IPA. The canonical pathways showed 67 common pathways among all the pSS groups. Focusing on pSS-associated lymphoma versus healthy controls, 351

canonical pathways were identified, 94.9% of the canonical pathways in this comparison were in common with the canonical pathways identified when comparing pSS-associated lymphoma with pSS. The "Interferon Signaling pathway" was the top pathway for all pSS subgroups comparing with healthy controls. In addition, the non-lymphoma pSS subgroups (i.e. pSS, pSS-paraproteinemia and pSS-other cancers) showed similar patterns of "diseases and biological functions" in the downstream analysis, but were different from the "pSS-associated lymphoma" group. These results further support the biosignature of pSS-associated lymphoma identified was specific. Finally, *DRAP1*, a transcriptional factor, was found to be differentially expressed in all pSS subgroups compared to healthy controls. In addition, the molecular network analysis demonstrates *DRAP1* in all pSS subgroups. These results might provide a key to a deeper understanding and a direction for future studies to investigate the development of lymphoma in pSS patients.

7.2 Future work:

In this project, I have identified a whole-blood gene expression signature in pSS-associated lymphoma. To explore the clinical and biological significance of my findings, several future experiments can be pursued. Below is a list of such experiments:

- To investigate the presence or absence of the biosignature in different immune cell subsets. It will help to gain a better understanding of the role of theses genes in pSS-associated lymphoma. The expression of these genes could be measured using single-cell analysis technologies such as Mass Cytometry (CyTOF) or SmartFlare^{TM}.
- The determination of the type of cell responsible for the expression of each gene would facilitate knockout or knock-in studies in animal models of pSS. These kinds of studies help to understand the exact role of each gene in many biological functions, for instance, apoptosis and lymphoproliferation in pSS.
- Evaluation of the expression of the signature genes in salivary and lacrimal glands of pSS patients with or without lymphoma as well as the lymphoma tissues from lymphoma patients without pSS.
- Evaluation of the expression of the signature genes in the whole blood of lymphoma patients without pSS.

- A longitudinal study to determine the level of expression of the signature genes throughout different stages of pSS progression.
- A microarray experiment using whole blood samples from untreated pSS-associated lymphoma and pSS patients without lymphoma may reveal a new set of genes that are important for lymphoma development in pSS.

References

- ABDULAHAD, W. H., KROESE, F. G., VISSINK, A. & BOOTSMA, H. 2012. Immune regulation and B-cell depletion therapy in patients with primary Sjögren's syndrome. *J Autoimmun,* 39**,** 103-11.

- ABISAMBRA, J. F., JINWAL, U. K., BLAIR, L. J., O'LEARY, J. C., 3RD, LI, Q., BRADY, S., WANG, L., GUIDI, C. E., ZHANG, B., NORDHUES, B. A., COCKMAN, M., SUNTHARALINGHAM, A., LI, P., JIN, Y., ATKINS, C. A. & DICKEY, C. A. 2013. Tau accumulation activates the unfolded protein response by impairing endoplasmic reticulum-associated degradation. *J Neurosci,* 33**,** 9498-507.

- ADRIANTO, I., WEN, F., TEMPLETON, A., WILEY, G., KING, J. B., LESSARD, C. J., BATES, J. S., HU, Y., KELLY, J. A., KAUFMAN, K. M., GUTHRIDGE, J. M., ALARCON-RIQUELME, M. E., ANAYA, J. M., BAE, S. C., BANG, S. Y., BOACKLE, S. A., BROWN, E. E., PETRI, M. A., GALLANT, C., RAMSEY-GOLDMAN, R., REVEILLE, J. D., VILA, L. M., CRISWELL, L. A., EDBERG, J. C., FREEDMAN, B. I., GREGERSEN, P. K., GILKESON, G. S., JACOB, C. O., JAMES, J. A., KAMEN, D. L., KIMBERLY, R. P., MARTIN, J., MERRILL, J. T., NIEWOLD, T. B., PARK, S. Y., PONS-ESTEL, B. A., SCOFIELD, R. H., STEVENS, A. M., TSAO, B. P., VYSE, T. J., LANGEFELD, C. D., HARLEY, J. B., MOSER, K. L., WEBB, C. F., HUMPHREY, M. B., MONTGOMERY, C. G. & GAFFNEY, P. M. 2011. Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. *Nat Genet,* 43**,** 253-8.

- AFFYMETRIX 2003. Globin Reduction Protocol A Method for Processing Whole Blood RNA Samples for Improved Array Results. *www.affymetrix.com*.

- AFFYMETRIX 2004. GeneChip® Globin-Reduction Kit Handbook. *www.affymetrix.com*.

- AGMON-LEVIN, N., KIVITY, S., TZIOUFAS, A. G., LOPEZ HOYOS, M., ROZMAN, B., EFES, I., SHAPIRA, Y., SHAMIS, A., AMITAL, H., YOUINOU, P. & SHOENFELD, Y. 2012. Low levels of vitamin-D are associated with neuropathy and lymphoma among patients with Sjögren's syndrome. *J Autoimmun,* 39**,** 234-9.

- ALEKSUNES, L. M. & KLAASSEN, C. D. 2012. Coordinated regulation of hepatic phase I and II drug-metabolizing genes and transporters using AhR-, CAR-, PXR-, PPARalpha-, and Nrf2-null mice. *Drug Metab Dispos,* 40**,** 1366-79.

- ALEVIZOS, I., ALEXANDER, S., TURNER, R. J. & ILLEI, G. G. 2011. MicroRNA expression profiles as biomarkers of minor salivary gland inflammation and dysfunction in Sjögren's syndrome. *Arthritis Rheum,* 63**,** 535-44.

- ALUNNO, A., PETRILLO, M. G., NOCENTINI, G., BISTONI, O., BARTOLONI, E., CATERBI, S., BIANCHINI, R., BALDINI, C., NICOLETTI, I., RICCARDI, C. & GERLI, R. 2013. Characterization of a new regulatory CD4+ T cell subset in primary Sjögren's syndrome. *Rheumatology (Oxford),* 52**,** 1387-96.

- AMBION 2007. GLOBINclear™ Kit, (Cat #AM1980, AM1981): Instruction Manual. *https://www.lifetechnologies.com/order/catalog/product/AM1980*.

- AN, J., FUJIWARA, H., SUEMORI, K., NIIYA, T., AZUMA, T., TANIMOTO, K., OCHI, T., AKATSUKA, Y., MINENO, J., OZAWA, H., ISHIKAWA, F., KUZUSHIMA, K. & YASUKAWA, M. 2011. Activation of T-cell receptor signaling in peripheral T-cell lymphoma cells plays an important role in the development of lymphoma-associated hemophagocytosis. *Int J Hematol,* 93**,** 176-85.

- ANAND, A., KRISHNA, G. G., SIBLEY, R. K. & KAMBHAM, N. 2015. Sjogren Syndrome and Cryoglobulinemic Glomerulonephritis. *Am J Kidney Dis,* 66**,** 532-5.

- ANDERSEN, C. L., JENSEN, J. L. & ORNTOFT, T. F. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res,* 64**,** 5245-50.

- ANTONIOU, M., DEBOER, E., HABETS, G. & GROSVELD, F. 1988. The human beta-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. *Embo j,* 7**,** 377-84.

- AQRAWI, L. A., BROKSTAD, K. A., JAKOBSEN, K., JONSSON, R. & SKARSTEIN, K. 2012. Low number of memory B cells in the salivary glands of patients with primary Sjögren's syndrome. *Autoimmunity,* 45**,** 547-55.

- ARTEAGA-SALAS, J. M., ZUZAN, H., LANGDON, W. B., UPTON, G. J. G. & HARRISON, A. P. 2008. An overview of image-processing methods for Affymetrix GeneChips. *Briefings in Bioinformatics,* 9**,** 25-33.

- ASTHANA, J., KUCHIBHATLA, A., JANA, S. C., RAY, K. & PANDA, D. 2012. Dynein light chain 1 (LC8) association enhances microtubule stability and promotes microtubule bundling. *J Biol Chem,* 287**,** 40793-805.

- AUER, H., LYIANARACHCHI, S., NEWSOM, D., KLISOVIC, M. I., MARCUCCI, G. & KORNACKER, K. 2003. Chipping away at the chip bias: RNA degradation in microarray analysis. *Nat Genet,* 35**,** 292-3.

- BACMAN, S., STERIN-BORDA, L., CAMUSSO, J. J., ARANA, R., HUBSCHER, O. & BORDA, E. 1996. Circulating antibodies against rat parotid gland M3 muscarinic receptors in primary Sjögren's syndrome. *Clin Exp Immunol,* 104**,** 454-9.

- BAIMPA, E., DAHABREH, I. J., VOULGARELIS, M. & MOUTSOPOULOS, H. M. 2009. Hematologic manifestations and predictors of lymphoma development in primary Sjogren syndrome: clinical and pathophysiologic aspects. *Medicine (Baltimore),* 88**,** 284-93.

- BARCELOS, F., ABREU, I., PATTO, J. V., TRINDADE, H. & TEIXEIRA, A. 2009. Anti-cyclic citrullinated peptide antibodies and rheumatoid factor in Sjögren's syndrome. *Acta Reumatol Port,* 34**,** 608-12.

- BARONE, F., BOMBARDIERI, M., ROSADO, M. M., MORGAN, P. R., CHALLACOMBE, S. J., DE VITA, S., CARSETTI, R., SPENCER, J., VALESINI, G. & PITZALIS, C. 2008. CXCL13, CCL21, and CXCL12 expression in salivary glands of patients with Sjögren's syndrome and MALT lymphoma: association with reactive and malignant areas of lymphoid organization. *J Immunol,* 180**,** 5130-40.

- BAROUKI, R., COUMOUL, X. & FERNANDEZ-SALGUERO, P. M. 2007. The aryl hydrocarbon receptor, more than a xenobiotic-interacting protein. *FEBS Lett,* 581**,** 3608-15.

- BAVE, U., NORDMARK, G., LOVGREN, T., RONNELID, J., CAJANDER, S., ELORANTA, M. L., ALM, G. V. & RONNBLOM, L. 2005. Activation of the type I interferon system in primary Sjögren's syndrome: a possible etiopathogenic mechanism. *Arthritis Rheum,* 52**,** 1185-95.

- BLAVY, P., GONDRET, F., LAGARRIGUE, S., VAN MILGEN, J. & SIEGEL, A. 2014. Using a large-scale knowledge database on reactions and

regulations to propose key upstream regulators of various sets of molecules participating in cell metabolism. *BMC Syst Biol,* 8**,** 32.

- BLOCH, K. J., BUCHANAN, W. W., WOHL, M. J. & BUNIM, J. J. 1992. Sjögren's syndrome. A clinical, pathological, and serological study of sixty-two cases. 1965. *Medicine (Baltimore),* 71**,** 386-401; discussion 401-3.

- BOGNER, V., LEIDEL, B. A., KANZ, K. G., MUTSCHLER, W., NEUGEBAUER, E. A. & BIBERTHALER, P. 2011. Pathway analysis in microarray data: a comparison of two different pathway analysis devices in the same data set. *Shock,* 35**,** 245-51.

- BOLLAIN-Y-GOYTIA, J. J., AVALOS-DIAZ, E. & HERRERA-ESPARZA, R. 2000. Fas ligand and Bax gene transcription contributes to Ro60 ribonucleoprotein redistribution in UV-A irradiated human keratinocytes. *Joint Bone Spine,* 67**,** 283-9.

- BOWMAN, S. J., IBRAHIM, G. H., HOLMES, G., HAMBURGER, J. & AINSWORTH, J. R. 2004. Estimating the prevalence among Caucasian women of primary Sjögren's syndrome in two general practices in Birmingham, UK. *Scand J Rheumatol,* 33**,** 39-43.

- BRANCA, M. 2003. Genetics and medicine. Putting gene arrays to the test. *Science,* 300**,** 238.

- BRITO-ZERON, P., KOSTOV, B., SOLANS, R., FRAILE, G., SUAREZ-CUERVO, C., CASANOVAS, A., RASCON, F. J., QANNETA, R., PEREZ-ALVAREZ, R., RIPOLL, M., AKASBI, M., PINILLA, B., BOSCH, J. A., NAVA-MATEOS, J., DIAZ-LOPEZ, B., MORERA-MORALES, M. L., GHEITASI, H., RETAMOZO, S. & RAMOS-CASALS, M. 2014. Systemic activity and mortality in primary Sjogren syndrome: predicting survival using the EULAR-SS Disease Activity Index (ESSDAI) in 1045 patients. *Ann Rheum Dis,* 75**,** 348-55.

- BRKIC, Z., MARIA, N. I., VAN HELDEN-MEEUWSEN, C. G., VAN DE MERWE, J. P., VAN DAELE, P. L., DALM, V. A., WILDENBERG, M. E., BEUMER, W., DREXHAGE, H. A. & VERSNEL, M. A. 2013. Prevalence of interferon type I signature in CD14 monocytes of patients with Sjögren's syndrome and association with disease activity and BAFF gene expression. *Ann Rheum Dis,* 72**,** 728-35.

- BRKIC, Z. & VERSNEL, M. A. 2014. Type I IFN signature in primary Sjögren's syndrome patients. *Expert Rev Clin Immunol,* 10**,** 457-67.

- BUONANNO, A. 2010. The Neuregulin Signaling Pathway and Schizophrenia: From Genes to Synapses and Neural Circuits. *Brain research bulletin,* 83**,** 122-131.

- BURBACH, K. M., POLAND, A. & BRADFIELD, C. A. 1992. Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. *Proc Natl Acad Sci U S A,* 89**,** 8185-9.

- CANG, Y. & PRELICH, G. 2002. Direct stimulation of transcription by negative cofactor 2 (NC2) through TATA-binding protein (TBP). *Proc Natl Acad Sci U S A,* 99**,** 12727-32.

- CARAMASCHI, P., BIASI, D., CARLETTO, A., MANZO, T., RANDON, M., ZEMINIAN, S. & BAMBARA, L. M. 1997. Sjögren's syndrome with anticentromere antibodies. *Rev Rhum Engl Ed,* 64**,** 785-8.

- CARSONS, S. 2012. New Immunosuppressive Agents for the Treatment of Sjögren's Syndrome. *In:* RAMOS-CASALS, M., STONE, J. H. & MOUTSOPOULOS, H. (eds.) *Sjögren's Syndrome: Diagnosis and Therapeutics.* London: Springer-Verlag London.

- CARVAJAL-GONZALEZ, J. M., MULERO-NAVARRO, S., ROMAN, A. C., SAUZEAU, V., MERINO, J. M., BUSTELO, X. R. & FERNANDEZ-SALGUERO, P. M. 2009. The dioxin receptor regulates the constitutive expression of the vav3 proto-oncogene and modulates cell shape and adhesion. *Mol Biol Cell,* 20**,** 1715-27.

- CASTAñO, E., GROSS, P., WANG, Z., ROEDER, R. G. & OELGESCHLäGER, T. 2000. The C-terminal domain-phosphorylated IIO form of RNA polymerase II is associated with the transcription repressor NC2 (Dr1/DRAP1) and is required for transcription activation in human nuclear extracts. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 7184-7189.

- CELLA, M. & COLONNA, M. 2015. Aryl hydrocarbon receptor: Linking environment to immunity. *Semin Immunol,* 27**,** 310-4.

- CHARITAKI, E., LIAPIS, K., MOUTZOURIS, D. A., MARINOS, L., ADAMIDIS, K., MARGELLOS, V. & BAKIRI, M. 2011. Primary renal MALT lymphoma presenting with cryoglobulinaemia. *Nephrol Dial Transplant,* 26**,** 3819-21.

- CHEN, C., GRENNAN, K., BADNER, J., ZHANG, D., GERSHON, E., JIN, L. & LIU, C. 2011. Removing Batch Effects in Analysis of Expression Microarray Data: An Evaluation of Six Batch Adjustment Methods. *PLoS ONE,* 6**,** e17238.

- CHEN, J. Q., ZILAHI, E., PAPP, G., SIPKA, S. & ZEHER, M. 2015a. Simultaneously increased expression of microRNA-155 and suppressor of cytokine signaling 1 (SOCS1) gene in the peripheral blood mononuclear cells of patients with primary Sjögren's syndrome. *Int J Rheum Dis*. ahead of print.

- CHEN, W., CAO, H., LIN, J., OLSEN, N. & ZHENG, S. G. 2015b. Biomarkers for Primary Sjögren's Syndrome. *Genomics, Proteomics & Bioinformatics,* 13**,** 219-223.

- CHEN, Y. W., LEE, K. C., CHANG, I. W., CHANG, C. S., HSU, S. P. & KUO, H. C. 2013. Sjögren's syndrome with acute cerebellar ataxia and massive lymphadenopathy : a case report. *Acta Neurol Taiwan,* 22**,** 81-6.

- CHOI, I., BAO, H., KOMMADATH, A., HOSSEINI, A., SUN, X., MENG, Y., STOTHARD, P., PLASTOW, G. S., TUGGLE, C. K., REECY, J. M., FRITZ-WATERS, E., ABRAMS, S. M., LUNNEY, J. K. & GUAN, L. L. 2014. Increasing gene discovery and coverage using RNA-seq of globin RNA reduced porcine blood samples. *BMC Genomics,* 15**,** 954.

- CHOI, S. Y., JANG, J. H. & KIM, K. R. 2011. Analysis of differentially expressed genes in human rectal carcinoma using suppression subtractive hybridization. *Clin Exp Med,* 11**,** 219-26.

- CHRISTODOULOU, M. I., KAPSOGEORGOU, E. K. & MOUTSOPOULOS, H. M. 2010. Characteristics of the minor salivary gland infiltrates in Sjögren's syndrome. *J Autoimmun,* 34**,** 400-7.

- CHRISTODOULOU, M. I., KAPSOGEORGOU, E. K., MOUTSOPOULOS, N. M. & MOUTSOPOULOS, H. M. 2008. Foxp3+ T-regulatory cells in Sjögren's syndrome: correlation with the grade of the autoimmune lesion and certain adverse prognostic factors. *Am J Pathol,* 173**,** 1389-96.

- CHUAQUI, R. F., BONNER, R. F., BEST, C. J., GILLESPIE, J. W., FLAIG, M. J., HEWITT, S. M., PHILLIPS, J. L., KRIZMAN, D. B., TANGREA, M. A., AHRAM, M., LINEHAN, W. M., KNEZEVIC, V. & EMMERT-BUCK, M. R. 2002. Post-analysis follow-up and validation of microarray experiments. *Nat Genet,* 32 Suppl**,** 509-14.

- CICCIA, F., GUGGINO, G., RIZZO, A., BOMBARDIERI, M., RAIMONDO, S., CARUBBI, F., CANNIZZARO, A., SIRECI, G., DIELI, F., CAMPISI, G., GIACOMELLI, R., CIPRIANI, P., DE LEO, G., ALESSANDRO, R. & TRIOLO, G. 2015. Interleukin (IL)-22 receptor 1 is over-expressed in primary Sjögren's syndrome and Sjogren-associated non-Hodgkin lymphomas and is regulated by IL-18. *Clin Exp Immunol,* 181**,** 219-29.

- CLAUDIO, P. P., STIEGLER, P., HOWARD, C. M., BELLAN, C., MINIMO, C., TOSI, G. M., RAK, J., KOVATICH, A., DE FAZIO, P., MICHELI, P., CAPUTI, M., LEONCINI, L., KERBEL, R., GIORDANO, G. G. & GIORDANO, A. 2001. RB2/p130 gene-enhanced expression down-regulates vascular endothelial growth factor expression and inhibits angiogenesis in vivo. *Cancer Res,* 61**,** 462-8.

- COLL, J., RIVES, A., GRINO, M. C., SETOAIN, J., VIVANCOS, J. & BALCELLS, A. 1987. Prevalence of Sjögren's syndrome in autoimmune diseases. *Ann Rheum Dis,* 46**,** 286-9.

- CONSORTIUM, M. 2006. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology,* 24**,** 1151-1161.

- COOK, L. & MACDONALD, D. H. 2007. Management of paraproteinaemia. *Postgrad Med J,* 83**,** 217-23.

- CRETON, S., SVEJSTRUP, J. Q. & COLLART, M. A. 2002. The NC2 α and β subunits play different roles in vivo. *Genes & Development,* 16**,** 3265-3276.

- CURJURIC, I., IMBODEN, M., NADIF, R., KUMAR, A., SCHINDLER, C., HAUN, M., KRONENBERG, F., KUNZLI, N., PHULERIA, H., POSTMA, D. S., RUSSI, E. W., ROCHAT, T., DEMENAIS, F. & PROBST-HENSCH, N. M. 2012. Different genes interact with particulate matter and tobacco smoke exposure in affecting lung function decline in the general population. *PLoS One,* 7**,** e40175.

- DALLAS, P. B., GOTTARDO, N. G., FIRTH, M. J., BEESLEY, A. H., HOFFMANN, K., TERRY, P. A., FREITAS, J. R., BOAG, J. M., CUMMINGS, A. J. & KEES, U. R. 2005. Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR -- how well do they correlate? *BMC Genomics,* 6**,** 59.

- DAVENPORT, E. L., MOORE, H. E., DUNLOP, A. S., SHARP, S. Y., WORKMAN, P., MORGAN, G. J. & DAVIES, F. E. 2007. Heat shock protein inhibition is associated with activation of the unfolded protein response pathway in myeloma plasma cells. *Blood,* 110**,** 2641-9.

- DAVIDSON, B. K., KELLY, C. A. & GRIFFITHS, I. D. 1999. Primary Sjögren's syndrome in the North East of England: a long-term follow-up study. *Rheumatology (Oxford),* 38**,** 245-53.

- DE VITA, S., QUARTUCCIO, L., SALVIN, S., CORAZZA, L., ZABOTTI, A. & FABRIS, M. 2012. Cryoglobulinaemia related to Sjögren's syndrome or HCV infection: differences based on the pattern of bone marrow involvement, lymphoma evolution and laboratory tests after parotidectomy. *Rheumatology (Oxford),* 51**,** 627-33.

- DE VITA, S., QUARTUCCIO, L., SEROR, R., SALVIN, S., RAVAUD, P., FABRIS, M., NOCTURNE, G., GANDOLFO, S., ISOLA, M. & MARIETTE, X. 2015. Efficacy and safety of belimumab given for 12 months in primary Sjögren's syndrome: the BELISS open-label phase II study. *Rheumatology (Oxford),* 54**,** 2249-56.

- DEL VALLE, P. R., MILANI, C., BRENTANI, M. M., KATAYAMA, M. L., DE LYRA, E. C., CARRARO, D. M., BRENTANI, H., PUGA, R., LIMA, L. A., ROZENCHAN, P. B., NUNES BDOS, S., GOES, J. C. & AZEVEDO KOIKE FOLGUEIRA, M. A. 2014. Transcriptional profile of fibroblasts obtained from the primary site, lymph node and bone marrow of breast cancer patients. *Genet Mol Biol,* 37**,** 480-9.

- DELALEU, N., MYDEL, P., KWEE, I., BRUN, J. G., JONSSON, M. V. & JONSSON, R. 2015. High Fidelity Between Saliva Proteomics and the Biologic State of Salivary Glands Defines Biomarker Signatures for Primary Sjögren's Syndrome. *Arthritis & Rheumatology,* 67**,** 1084-1095.

- DENKO, N., WERNKE-DOLLRIES, K., JOHNSON, A. B., HAMMOND, E., CHIANG, C. M. & BARTON, M. C. 2003. Hypoxia actively represses

transcription by inducing negative cofactor 2 (Dr1/DrAP1) and blocking preinitiation complex assembly. *J Biol Chem,* 278**,** 5744-9.

- DEUTSCH, O., KRIEF, G., KONTTINEN, Y. T., ZAKS, B., WONG, D. T., AFRAMIAN, D. J. & PALMON, A. 2015. Identification of Sjögren's syndrome oral fluid biomarker candidates following high-abundance protein depletion. *Rheumatology (Oxford),* 54**,** 884-90.

- DEVAUCHELLE-PENSEC, V., MARIETTE, X., JOUSSE-JOULIN, S., BERTHELOT, J. M., PERDRIGER, A., PUECHAL, X., LE GUERN, V., SIBILIA, J., GOTTENBERG, J. E., CHICHE, L., HACHULLA, E., HATRON, P. Y., GOEB, V., HAYEM, G., MOREL, J., ZARNITSKY, C., DUBOST, J. J., PERS, J. O., NOWAK, E. & SARAUX, A. 2014. Treatment of primary Sjogren syndrome with rituximab: a randomized trial. *Ann Intern Med,* 160**,** 233-42.

- DICKEY, D. A., CAROLINA STATE N. U. AND RALEIGH, N.C. 2012. *Introduction to Predictive Modeling with Examples In: Statistics and Data Analysis,* USA, SAS Conference Proceedings: SAS Global Forum 2010.

- DINARELLO, C. A. 2007. Interleukin-18 and the pathogenesis of inflammatory diseases. *Semin Nephrol,* 27**,** 98-114.

- DING, J., DIRKS, W. G., EHRENTRAUT, S., GEFFERS, R., MACLEOD, R. A., NAGEL, S., POMMERENKE, C., ROMANI, J., SCHERR, M., VAAS, L. A., ZABORSKI, M., DREXLER, H. G. & QUENTMEIER, H. 2015. BCL6--regulated by AhR/ARNT and wild-type MEF2B--drives expression of germinal center markers MYBL1 and LMO2. *Haematologica,* 100**,** 801-9.

- DOHERTY, D. G., PENZOTTI, J. E., KOELLE, D. M., KWOK, W. W., LYBRAND, T. P., MASEWICZ, S. & NEPOM, G. T. 1998. Structural basis of specificity and degeneracy of T cell recognition: pluriallelic restriction of T cell responses to a peptide antigen involves both specific and promiscuous interactions between the T cell receptor, peptide, and HLA-DR. *J Immunol,* 161**,** 3527-35.

- DONG, L., CHEN, Y., MASAKI, Y., OKAZAKI, T. & UMEHARA, H. 2013. Possible Mechanisms of Lymphoma Development in Sjögren's Syndrome. *Curr Immunol Rev,* 9**,** 13-22.

- DROZDZIK, A., KOWALCZYK, R., LIPSKI, M., LAPCZUK, J., URASINSKA, E. & KURZAWSKI, M. 2015. The role of aryl hydrocarbon

receptor (AhR) in the pathology of pleomorphic adenoma in parotid gland. *Arch Oral Biol,* 61**,** 53-59.

- DU, M., PENG, H., SINGH, N., ISAACSON, P. G. & PAN, L. 1995. The accumulation of p53 abnormalities is associated with progression of mucosa-associated lymphoid tissue lymphoma. *Blood,* 86**,** 4587-93.

- DU, P., KIBBE, W. A. & LIN, S. M. 2008. lumi: a pipeline for processing Illumina microarray. *Bioinformatics,* 24**,** 1547-8.

- DUMEAUX, V., LUND, E. & BORRESEN-DALE, A. L. 2008. Comparison of globin RNA processing methods for genome-wide transcriptome analysis from whole blood. *Biomark Med,* 2**,** 11-21.

- EBI, M., KATAOKA, H., SHIMURA, T., HIRATA, Y., MIZUSHIMA, T., MIZOSHITA, T., TANAKA, M., TSUKAMOTO, H., OZEKI, K., TANIDA, S., KAMIYA, T., INAGAKI, H. & JOH, T. 2011. The role of neuregulin4 and HER4 in gastrointestinal malignant lymphoma. *Mol Med Rep,* 4**,** 1151-5.

- EISEN, M. B. & BROWN, P. O. 1999. DNA arrays for analysis of gene expression. *Methods Enzymol,* 303**,** 179-205.

- EK, M., POPOVIC, K., HARRIS, H. E., NAUCLER, C. S. & WAHREN-HERLENIUS, M. 2006. Increased extracellular levels of the novel proinflammatory cytokine high mobility group box chromosomal protein 1 in minor salivary glands of patients with Sjögren's syndrome. *Arthritis Rheum,* 54**,** 2289-94.

- EMAMIAN, E. S., LEON, J. M., LESSARD, C. J., GRANDITS, M., BAECHLER, E. C., GAFFNEY, P. M., SEGAL, B., RHODUS, N. L. & MOSER, K. L. 2009. Peripheral blood gene expression profiling in Sjögren's syndrome. *Genes Immun,* 10**,** 285-96.

- FALLAH, M., LIU, X., JI, J., FORSTI, A., SUNDQUIST, K. & HEMMINKI, K. 2014. Autoimmune diseases associated with non-Hodgkin lymphoma: a nationwide cohort study. *Ann Oncol,* 25**,** 2025-30.

- FAN, H. & HEGDE, P. S. 2005. The transcriptome in blood: challenges and solutions for robust expression profiling. *Curr Mol Med,* 5**,** 3-10.

- FARE, T. L., COFFEY, E. M., DAI, H., HE, Y. D., KESSLER, D. A., KILIAN, K. A., KOCH, J. E., LEPROUST, E., MARTON, M. J., MEYER, M. R., STOUGHTON, R. B., TOKIWA, G. Y. & WANG, Y. 2003. Effects of atmospheric ozone on microarray data quality. *Anal Chem,* 75**,** 4672-5.

- FAUCHAIS, A. L., MARTEL, C., GONDRAN, G., LAMBERT, M., LAUNAY, D., JAUBERTEAU, M. O., HACHULLA, E., VIDAL, E. & HATRON, P. Y. 2010. Immunological profile in primary Sjogren syndrome: clinical significance, prognosis and long-term evolution to other auto-immune disease. *Autoimmun Rev,* 9**,** 595-9.

- FAZAA, A., BOURCIER, T., CHATELUS, E., SORDET, C., THEULIN, A., SIBILIA, J. & GOTTENBERG, J. E. 2014. Classification criteria and treatment modalities in primary Sjögren's syndrome. *Expert Rev Clin Immunol,* 10**,** 543-51.

- FEEZOR, R. J., BAKER, H. V., MINDRINOS, M., HAYDEN, D., TANNAHILL, C. L., BROWNSTEIN, B. H., FAY, A., MACMILLAN, S., LARAMIE, J., XIAO, W., MOLDAWER, L. L., COBB, J. P., LAUDANSKI, K., MILLER-GRAZIANO, C. L., MAIER, R. V., SCHOENFELD, D., DAVIS, R. W. & TOMPKINS, R. G. 2004. Whole blood and leukocyte RNA isolation for gene expression analyses. *Physiol Genomics,* 19**,** 247-54.

- FERRER, G., HODGSON, K., MONTSERRAT, E. & MORENO, C. 2009. B cell activator factor and a proliferation-inducing ligand at the cross-road of chronic lymphocytic leukemia and autoimmunity. *Leuk Lymphoma,* 50**,** 1075-82.

- FIELD, L. A., JORDAN, R. M., HADIX, J. A., DUNN, M. A., SHRIVER, C. D., ELLSWORTH, R. E. & ELLSWORTH, D. L. 2007. Functional identity of genes detectable in expression profiling assays following globin mRNA reduction of peripheral blood samples. *Clin Biochem,* 40**,** 499-502.

- FIFE, R. S., CHASE, W. F., DORE, R. K., WIESENHUTTER, C. W., LOCKHART, P. B., TINDALL, E. & SUEN, J. Y. 2002. Cevimeline for the treatment of xerostomia in patients with Sjogren syndrome: a randomized trial. *Arch Intern Med,* 162**,** 1293-300.

- FORD-HUTCHINSON, A. W. 1990. Leukotriene B4 in inflammation. *Crit Rev Immunol,* 10**,** 1-12.

- FOX, R. I., KONTTINEN, Y. & FISHER, A. 2001. Use of muscarinic agonists in the treatment of Sjögren's syndrome. *Clin Immunol,* 101**,** 249-63.

- FU, Z. D. & KLAASSEN, C. D. 2014. Short-term calorie restriction feminizes the mRNA profiles of drug metabolizing enzymes and transporters in livers of mice. *Toxicol Appl Pharmacol,* 274**,** 137-46.

- FURUZAWA-CARBALLEDA, J., HERNANDEZ-MOLINA, G., LIMA, G., RIVERA-VICENCIO, Y., FEREZ-BLANDO, K. & LLORENTE, L. 2013. Peripheral regulatory cells immunophenotyping in primary Sjögren's syndrome: a cross-sectional study. *Arthritis Res Ther,* 15**,** R68.

- G'SELL, R. T., GAFFNEY, P. M. & POWELL, D. W. 2015. A20-Binding Inhibitor of NF-kappaB Activation 1 is a Physiologic Inhibitor of NF-kappaB: A Molecular Switch for Inflammation and Autoimmunity. *Arthritis Rheumatol,* 67**,** 2292-302.

- GADALLA, S. M., AMR, S., LANGENBERG, P., BAUMGARTEN, M., DAVIDSON, W. F., SCHAIRER, C., ENGELS, E. A., PFEIFFER, R. M. & GOEDERT, J. J. 2009. Breast cancer risk in elderly women with systemic autoimmune rheumatic diseases: a population-based case-control study. *Br J Cancer,* 100**,** 817-21.

- GAO, H. Y., LUO, X. G., CHEN, X. & WANG, J. H. 2015. Identification of key genes affecting disease free survival time of pediatric acute lymphoblastic leukemia based on bioinformatic analysis. *Blood Cells Mol Dis,* 54**,** 38-43.

- GARCIC-CARRASCO, M., FONT, J., FILELLA, X., CERVERA, R., RAMOS-CASALS, M., SISO, A., AYMAMI, A., BALLESTA, A. M. & INGELMO, M. 2001. Circulating levels of Th1/Th2 cytokines in patients with primary Sjögren's syndrome: correlation with clinical and immunological features. *Clin Exp Rheumatol,* 19**,** 411-5.

- GE, N. L. & ELFERINK, C. J. 1998. A direct interaction between the aryl hydrocarbon receptor and retinoblastoma protein. Linking dioxin signaling to the cell cycle. *J Biol Chem,* 273**,** 22708-13.

- GENTLEMAN, R. C., CAREY, V. J., BATES, D. M., BOLSTAD, B., DETTLING, M., DUDOIT, S., ELLIS, B., GAUTIER, L., GE, Y., GENTRY, J., HORNIK, K., HOTHORN, T., HUBER, W., IACUS, S., IRIZARRY, R., LEISCH, F., LI, C., MAECHLER, M., ROSSINI, A. J., SAWITZKI, G., SMITH, C., SMYTH, G., TIERNEY, L., YANG, J. Y. & ZHANG, J. 2004.

Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol,* 5**,** R80.

- GERSUK, V. H. & NEPOM, G. T. 2009. A real-time PCR assay for the rapid identification of the autoimmune disease-associated allele HLA-DQB1*0602. *Tissue antigens,* 73**,** 335-340.

- GHAFOOR, M. 2012. Sjögren's Before Sjogren: Did Henrik Sjogren (1899-1986) Really Discover Sjögren's Disease? *J Maxillofac Oral Surg,* 11**,** 373-4.

- GHATREHSAMANI, M., SOLEIMANI, M., ESFAHANI, B. A., SHIRZAD, H., HAKEMI, M. G., MOSSAHEBIMOHAMMADI, M., ESKANDARI, N. & ADIB, M. 2015. Tumor necrosis factor-alpha inhibits effects of aryl hydrocarbon receptor ligands on cell death in human lymphocytes. *Adv Biomed Res,* 4**,** 216.

- GILMOUR, L. M., MACLEOD, K. G., MCCAIG, A., SEWELL, J. M., GULLICK, W. J., SMYTH, J. F. & LANGDON, S. P. 2002. Neuregulin expression, function, and signaling in human ovarian cancer cells. *Clin Cancer Res,* 8**,** 3933-42.

- GOH, S. H., LEE, Y. T., BHANU, N. V., CAM, M. C., DESPER, R., MARTIN, B. M., MOHARRAM, R., GHERMAN, R. B. & MILLER, J. L. 2005. A newly discovered human alpha-globin gene. *Blood,* 106**,** 1466-72.

- GONG, Y. Z., NITITHAM, J., TAYLOR, K., MICELI-RICHARD, C., SORDET, C., WACHSMANN, D., BAHRAM, S., GEORGEL, P., CRISWELL, L. A., SIBILIA, J., MARIETTE, X., ALSALEH, G. & GOTTENBERG, J. E. 2014. Differentiation of follicular helper T cells by salivary gland epithelial cells in primary Sjögren's syndrome. *J Autoimmun,* 51**,** 57-66.

- GOTTENBERG, J. E., CAGNARD, N., LUCCHESI, C., LETOURNEUR, F., MISTOU, S., LAZURE, T., JACQUES, S., BA, N., ITTAH, M., LEPAJOLEC, C., LABETOULLE, M., ARDIZZONE, M., SIBILIA, J., FOURNIER, C., CHIOCCHIA, G. & MARIETTE, X. 2006. Activation of IFN pathways and plasmacytoid dendritic cell recruitment in target organs of primary Sjögren's syndrome. *Proc Natl Acad Sci U S A,* 103**,** 2770-5.

- GOTTENBERG, J. E., SEROR, R., MICELI-RICHARD, C., BENESSIANO, J., DEVAUCHELLE-PENSEC, V., DIEUDE, P., DUBOST, J. J., FAUCHAIS, A. L., GOEB, V., HACHULLA, E., HATRON, P. Y., LARROCHE, C., LE

GUERN, V., MOREL, J., PERDRIGER, A., PUECHAL, X., RIST, S., SARAUX, A., SENE, D., SIBILIA, J., VITTECOQ, O., NOCTURNE, G., RAVAUD, P. & MARIETTE, X. 2013. Serum levels of beta2-microglobulin and free light chains of immunoglobulins are associated with systemic disease activity in primary Sjögren's syndrome. Data at enrollment in the prospective ASSESS cohort. *PLoS One,* 8**,** e59868.

- GROOM, J., KALLED, S. L., CUTLER, A. H., OLSON, C., WOODCOCK, S. A., SCHNEIDER, P., TSCHOPP, J., CACHERO, T. G., BATTEN, M., WHEWAY, J., MAURI, D., CAVILL, D., GORDON, T. P., MACKAY, C. R. & MACKAY, F. 2002. Association of BAFF/BLyS overexpression and altered B cell differentiation with Sjögren's syndrome. *J Clin Invest,* 109**,** 59-68.

- GUAN, J., FANG, F. & SHAO, Z. 2015. Lymphoplasmacytic Lymphoma with IgA Paraproteinemia. *Clin Lab,* 61**,** 835-7.

- GUGGENBUHL, P., JEAN, S., JEGO, P., GROSBOIS, B., CHALES, G., SEMANA, G., LANCIEN, G., VEILLARD, E., PAWLOTSKY, Y. & PERDRIGER, A. 1998. Primary Sjögren's syndrome: role of the HLA-DRB1*0301-*1501 heterozygotes. *J Rheumatol,* 25**,** 900-5.

- GUNDERSON, K. L., KRUGLYAK, S., GRAIGE, M. S., GARCIA, F., KERMANI, B. G., ZHAO, C., CHE, D., DICKINSON, T., WICKHAM, E., BIERLE, J., DOUCET, D., MILEWSKI, M., YANG, R., SIEGMUND, C., HAAS, J., ZHOU, L., OLIPHANT, A., FAN, J. B., BARNARD, S. & CHEE, M. S. 2004. Decoding randomly ordered DNA arrays. *Genome Res,* 14**,** 870-7.

- HAGA, H. J., NADERI, Y., MORENO, A. M. & PEEN, E. 2012. A study of the prevalence of sicca symptoms and secondary Sjögren's syndrome in patients with rheumatoid arthritis, and its association to disease activity and treatment profile. *Int J Rheum Dis,* 15**,** 284-8.

- HALL, A. C. & KAHL, B. 2015. Splenic Lymphomas Presenting as Splenomegaly. *Oncology (Williston Park),* 29**,** 427-8, 441.

- HAMM-ALVAREZ, S. F., JANGA, S. R., EDMAN, M. C., MADRIGAL, S., SHAH, M., FROUSIAKIS, S. E., RENDUCHINTALA, K., ZHU, J., BRICEL, S., SILKA, K., BACH, D., HEUR, M., CHRISTIANAKIS, S., ARKFELD, D. G., IRVINE, J., MACK, W. J. & STOHL, W. 2014. Tear cathepsin S as a candidate biomarker for Sjögren's syndrome. *Arthritis Rheumatol,* 66**,** 1872-81.

- HANSEN, A., ODENDAHL, M., REITER, K., JACOBI, A. M., FEIST, E., SCHOLZE, J., BURMESTER, G. R., LIPSKY, P. E. & DORNER, T. 2002. Diminished peripheral blood memory B cells and accumulation of memory B cells in the salivary glands of patients with Sjögren's syndrome. *Arthritis Rheum,* 46**,** 2160-71.

- HATZIMANOLIS, A., MCGRATH, J. A., WANG, R., LI, T., WONG, P. C., NESTADT, G., WOLYNIEC, P. S., VALLE, D., PULVER, A. E. & AVRAMOPOULOS, D. 2013. Multiple variants aggregate in the neuregulin signaling pathway in a subset of schizophrenia patients. *Transl Psychiatry,* 3**,** e264.

- HEYEN, C. A., TAGLIABRACCI, V. S., ZHAI, L. & ROACH, P. J. 2009. Characterization of mouse UDP-glucose pyrophosphatase, a Nudix hydrolase encoded by the Nudt14 gene. *Biochem Biophys Res Commun,* 390**,** 1414-8.

- HILDEBRAND, J. M., LUO, Z., MANSKE, M. K., PRICE-TROSKA, T., ZIESMER, S. C., LIN, W., HOSTAGER, B. S., SLAGER, S. L., WITZIG, T. E., ANSELL, S. M., CERHAN, J. R., BISHOP, G. A. & NOVAK, A. J. 2010. A BAFF-R mutation associated with non-Hodgkin lymphoma alters TRAF recruitment and reveals new insights into BAFF-R signaling. *J Exp Med,* 207**,** 2569-79.

- HJELMERVIK, T. O., PETERSEN, K., JONASSEN, I., JONSSON, R. & BOLSTAD, A. I. 2005. Gene expression profiling of minor salivary glands clearly distinguishes primary Sjögren's syndrome patients from healthy control subjects. *Arthritis Rheum,* 52**,** 1534-44.

- HOLDER, J. L., JR., LOTZE, T. E., BACINO, C. & CHEUNG, S. W. 2012. A child with an inherited 0.31 Mb microdeletion of chromosome 14q32.33: further delineation of a critical region for the 14q32 deletion syndrome. *Am J Med Genet A,* 158a**,** 1962-6.

- HSU, T. C., CHANG, C. H., LIN, M. C., LIU, S. T., YEN, T. J. & TSAY, G. J. 2006. Anti-CENP-H antibodies in patients with Sjögren's syndrome. *Rheumatol Int,* 26**,** 298-303.

- HTTP://WWW.INGENUITY.COM 2015. Ingenuity Pathway analysis.

- HU, S., ARELLANO, M., BOONTHEUNG, P., WANG, J., ZHOU, H., JIANG, J., ELASHOFF, D., WEI, R., LOO, J. A. & WONG, D. T. 2008. Salivary Proteomics for Oral Cancer Biomarker Discovery. *Clinical cancer research :*

an official journal of the American Association for Cancer Research, 14**,** 6246-6252.

- HU, S., WANG, J., MEIJER, J., IEONG, S., XIE, Y., YU, T., ZHOU, H., HENRY, S., VISSINK, A., PIJPE, J., KALLENBERG, C., ELASHOFF, D., LOO, J. A. & WONG, D. T. 2007. Salivary proteomic and genomic biomarkers for primary Sjögren's syndrome. *Arthritis Rheum,* 56**,** 3588-600.

- HU, S., ZHOU, M., JIANG, J., WANG, J., ELASHOFF, D., GORR, S., MICHIE, S. A., SPIJKERVET, F. K., BOOTSMA, H., KALLENBERG, C. G., VISSINK, A., HORVATH, S. & WONG, D. T. 2009. Systems biology analysis of Sjögren's syndrome and mucosa-associated lymphoid tissue lymphoma in parotid glands. *Arthritis Rheum,* 60**,** 81-92.

- IKUTA, T. & KAWAJIRI, K. 2006. Zinc finger transcription factor Slug is a novel target gene of aryl hydrocarbon receptor. *Exp Cell Res,* 312**,** 3585-94.

- INGENUITYSYSTEMS. IPA. *Ingenuity Upstream Regulator*

- *Analysis in IPA®* [Online]. http://pages.ingenuity.com/rs/ingenuity/images/0812%20upstream_regulator_analysis_whitepaper.pdf: IPA.

- INOUE, H., MISHIMA, K., YAMAMOTO-YOSHIDA, S., USHIKOSHI-NAKAYAMA, R., NAKAGAWA, Y., YAMAMOTO, K., RYO, K., IDE, F. & SAITO, I. 2012. Aryl hydrocarbon receptor-mediated induction of EBV reactivation as a risk factor for Sjögren's syndrome. *J Immunol,* 188**,** 4654-62.

- IOANNIDIS, J. P., VASSILIOU, V. A. & MOUTSOPOULOS, H. M. 2002. Long-term risk of mortality and lymphoproliferative disease and predictive classification of primary Sjögren's syndrome. *Arthritis Rheum,* 46**,** 741-7.

- ISHIHARA, K. & HIRANO, T. 2002. IL-6 in autoimmune disease and chronic inflammatory proliferative disease. *Cytokine Growth Factor Rev,* 13**,** 357-68.

- IWATANI, K., TAKATA, K., SATO, Y., MIYATA-TAKATA, T., IWAKI, N., CUI, W., SAWADA-KITAMURA, S., SONOBE, H., TAMURA, M., SAITO, K., MIYATANI, K., YAMASAKI, R., YAMADORI, I., FUJII, N., TERASAKI, Y., MAEDA, Y., TANIMOTO, M., NAKAMURA, N. & YOSHINO, T. 2014. Low-grade B-cell lymphoma presenting primarily in the bone marrow. *Hum Pathol,* 45**,** 1379-87.

- IZIDORO-TOLEDO, T. C., BORGES, A. C., ARAUJO, D. D., MAZZI, D. P., NASCIMENTO JUNIOR, F. O., SOUSA, J. F., ALVES, C. P., PAIVA, A. P., TRINDADE, D. M., PATUSSI, E. V., PEIXOTO, P. M., KINNALLY, K. W. & ESPREAFICO, E. M. 2013. A myosin-Va tail fragment sequesters dynein light chains leading to apoptosis in melanoma cells. *Cell Death Dis,* 4**,** e547.

- JAKOBSSON, P. J., MANCINI, J. A., RIENDEAU, D. & FORD-HUTCHINSON, A. W. 1997. Identification and characterization of a novel microsomal enzyme with glutathione-dependent transferase and peroxidase activities. *J Biol Chem,* 272**,** 22934-9.

- JAMES, J. A., HARLEY, J. B. & SCOFIELD, R. H. 2001. Role of viruses in systemic lupus erythematosus and Sjogren syndrome. *Curr Opin Rheumatol,* 13**,** 370-6.

- JAYAPPA, K. D., AO, Z., WANG, X., MOULAND, A. J., SHEKHAR, S., YANG, X. & YAO, X. 2015. Human immunodeficiency virus type 1 employs the cellular dynein light chain 1 protein for reverse transcription through interaction with its integrase protein. *J Virol,* 89**,** 3497-511.

- JERUSS, J. S., MITTENDORF, E. A., TUCKER, S. L., GONZALEZ-ANGULO, A. M., BUCHHOLZ, T. A., SAHIN, A. A., CORMIER, J. N., BUZDAR, A. U., HORTOBAGYI, G. N. & HUNT, K. K. 2008. Staging of breast cancer in the neoadjuvant setting. *Cancer Res,* 68**,** 6477-81.

- JIMENEZ-MARíN, Á., COLLADO-ROMERO, M., RAMIREZ-BOO, M., ARCE, C. & GARRIDO, J. J. 2009. Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. *BMC Proceedings,* 3**,** 1-6.

- JIN, L., YU, D., LI, X., YU, N., LI, X., WANG, Y. & WANG, Y. 2014a. CD4+CXCR5+ follicular helper T cells in salivary gland promote B cells maturation in patients with primary Sjögren's syndrome. *Int J Clin Exp Pathol,* 7**,** 1988-96.

- JIN, L., ZUO, X. Y., SU, W. Y., ZHAO, X. L., YUAN, M. Q., HAN, L. Z., ZHAO, X., CHEN, Y. D. & RAO, S. Q. 2014b. Pathway-based analysis tools for complex diseases: a review. *Genomics Proteomics Bioinformatics,* 12**,** 210-20.

- JOHNSEN, S. J., BRUN, J. G., GORANSSON, L. G., SMASTUEN, M. C., JOHANNESEN, T. B., HALDORSEN, K., HARBOE, E., JONSSON, R., MEYER, P. A. & OMDAL, R. 2013. Risk of non-Hodgkin's lymphoma in

primary Sjögren's syndrome: a population-based study. *Arthritis Care Res (Hoboken),* 65**,** 816-21.

- JOHNSON, W. E., LI, C. & RABINOVIC, A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics,* 8**,** 118-27.

- JOHNSTONE, D., RIVEROS, C., HEIDARI, M., GRAHAM, R., TRINDER, D., BERRETTA, R., OLYNYK, J., SCOTT, R., MOSCATO, P. & MILWARD, E. 2013. Evaluation of Different Normalization and Analysis Procedures for Illumina Gene Expression Microarray Data Involving Small Changes. *Microarrays,* 2**,** 131.

- JONSSON, M. V., THEANDER, E. & JONSSON, R. 2012. Predictors for the development of non-Hodgkin lymphoma in primary Sjögren's syndrome. *Presse Med,* 41**,** e511-6.

- KAM, S. H., SINGH, A., HE, J. Q., RUAN, J., GAUVREAU, G. M., O'BYRNE, P. M., FITZGERALD, J. M. & TEBBUTT, S. J. 2012. Peripheral blood gene expression changes during allergen inhalation challenge in atopic asthmatic individuals. *J Asthma,* 49**,** 219-26.

- KANG, E. H., LEE, Y. J., HYON, J. Y., YUN, P. Y. & SONG, Y. W. 2011. Salivary cytokine profiles in primary Sjögren's syndrome differ from those in non-Sjogren sicca in terms of TNF-alpha levels and Th-1/Th-2 ratios. *Clin Exp Rheumatol,* 29**,** 970-6.

- KASSAN, S. S. & MOUTSOPOULOS, H. M. 2004. CLinical manifestations and early diagnosis of sjögren syndrome. *Archives of Internal Medicine,* 164**,** 1275-1284.

- KASSAN, S. S., THOMAS, T. L., MOUTSOPOULOS, H. M., HOOVER, R., KIMBERLY, R. P., BUDMAN, D. R., COSTA, J., DECKER, J. L. & CHUSED, T. M. 1978. Increased risk of lymphoma in sicca syndrome. *Ann Intern Med,* 89**,** 888-92.

- KATSIFIS, G. E., REKKA, S., MOUTSOPOULOS, N. M., PILLEMER, S. & WAHL, S. M. 2009. Systemic and local interleukin-17 and linked cytokines associated with Sjögren's syndrome immunopathogenesis. *Am J Pathol,* 175**,** 1167-77.

- KAUFFMANN, A., GENTLEMAN, R. & HUBER, W. 2009. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. *Bioinformatics,* 25**,** 415-6.

- KAUPPI, M., PUKKALA, E. & ISOMAKI, H. 1997. Elevated incidence of hematologic malignancies in patients with Sjögren's syndrome compared with patients with rheumatoid arthritis (Finland). *Cancer Causes Control,* 8**,** 201-4.

- KAUSAR, S., ASIF, M., BIBI, N. & RASHID, S. 2013. Comparative molecular docking analysis of cytoplasmic dynein light chain DYNLL1 with pilin to explore the molecular mechanism of pathogenesis caused by Pseudomonas aeruginosa PAO. *PLoS One,* 8**,** e76730.

- KELLY, C., BAIRD, G., FOSTER, H., HOSKER, H. & GRIFFITHS, I. 1991. Prognostic significance of paraproteinaemia in rheumatoid arthritis. *Ann Rheum Dis,* 50**,** 290-4.

- KIM, C. S., CHOI, Y. D., CHOI, J. S., BAE, E. H., MA, S. K. & KIM, S. W. 2012. EBV-positive diffuse large B-cell lymphoma in a patient with primary Sjögren's syndrome and membranous glomerulonephritis. *BMC Nephrol,* 13**,** 149.

- KIM, S. H., CLEARY, M. M., FOX, H. S., CHANTRY, D. & SARVETNICK, N. 2002. CCR4-bearing T cells participate in autoimmune diabetes. *J Clin Invest,* 110**,** 1675-86.

- KIMOTO, O., SAWADA, J., SHIMOYAMA, K., SUZUKI, D., NAKAMURA, S., HAYASHI, H. & OGAWA, N. 2011. Activation of the interferon pathway in peripheral blood of patients with Sjögren's syndrome. *J Rheumatol,* 38**,** 310-6.

- KONSTA, O. D., THABET, Y., LE DANTEC, C., BROOKS, W. H., TZIOUFAS, A. G., PERS, J. O. & RENAUDINEAU, Y. 2014. The contribution of epigenetics in Sjögren's Syndrome. *Front Genet,* 5**,** 71.

- KOUMAKIS, E., GIRAUD, M., DIEUDE, P., COHIGNAC, V., CUOMO, G., AIRO, P., HACHULLA, E., MATUCCI-CERINIC, M., DIOT, E., CARAMASCHI, P., MOUTHON, L., RICCIERI, V., CRACOWSKI, J. L., TIEV, K. P., FRANCES, C., AMOURA, Z., SIBILIA, J., COSNES, A., CARPENTIER, P., VALENTINI, G., MANETTI, M., GUIDUCCI, S., MEYER, O., KAHAN, A., BOILEAU, C., CHIOCCHIA, G. & ALLANORE, Y. 2012. Brief report: candidate gene study in systemic sclerosis identifies a

rare and functional variant of the TNFAIP3 locus as a risk factor for polyautoimmunity. *Arthritis Rheum,* 64**,** 2746-52.

- KRAMER, A., GREEN, J., POLLARD, J., JR. & TUGENDREICH, S. 2014. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics,* 30**,** 523-30.

- KRAMER, J. M., KLIMATCHEVA, E. & ROTHSTEIN, T. L. 2013. CXCL13 is elevated in Sjögren's syndrome in mice and humans and is implicated in disease pathogenesis. *J Leukoc Biol,* 94**,** 1079-89.

- KRIEF, G., DEUTSCH, O., GARIBA, S., ZAKS, B., AFRAMIAN, D. J. & PALMON, A. 2011. Improved visualization of low abundance oral fluid proteins after triple depletion of alpha amylase, albumin and IgG. *Oral Diseases,* 17**,** 45-52.

- KROCK, B. L., EISINGER-MATHASON, T. S., GIANNOUKOS, D. N., SHAY, J. E., GOHIL, M., LEE, D. S., NAKAZAWA, M. S., SESEN, J., SKULI, N. & SIMON, M. C. 2015. The aryl hydrocarbon receptor nuclear translocator is an essential regulator of murine hematopoietic stem cell viability. *Blood,* 125**,** 3263-72.

- KYRIAKIDIS, N. C., KAPSOGEORGOU, E. K., GOURZI, V. C., KONSTA, O. D., BALTATZIS, G. E. & TZIOUFAS, A. G. 2014. Toll-like receptor 3 stimulation promotes Ro52/TRIM21 synthesis and nuclear redistribution in salivary gland epithelial cells, partially via type I interferon pathway. *Clin Exp Immunol,* 178**,** 548-60.

- LABI, V. & ERLACHER, M. 2015. How cell death shapes cancer. *Cell Death Dis,* 6**,** e1675.

- LARSEN, M. J., THOMASSEN, M., TAN, Q., SORENSEN, K. P. & KRUSE, T. A. 2014. Microarray-based RNA profiling of breast cancer: batch effect removal improves cross-platform consistency. 2014**,** 651751.

- LAVOIE, T. N., STEWART, C. M., BERG, K. M., LI, Y. & NGUYEN, C. Q. 2011. Expression of interleukin-22 in Sjögren's syndrome: significant correlation with disease parameters. *Scand J Immunol,* 74**,** 377-82.

- LAZARUS, M. N., ROBINSON, D., MAK, V., MOLLER, H. & ISENBERG, D. A. 2006. Incidence of cancer in a cohort of patients with primary Sjögren's syndrome. *Rheumatology (Oxford),* 45**,** 1012-5.

- LEE, E. G., BOONE, D. L., CHAI, S., LIBBY, S. L., CHIEN, M., LODOLCE, J. P. & MA, A. 2000. Failure to regulate TNF-induced NF-kappaB and cell death responses in A20-deficient mice. *Science,* 289**,** 2350-4.

- LEE, S. J., LEE, J. S., SHIN, M. G., TANAKA, Y., PARK, D. J., KIM, T. J., PARK, Y. W. & LEE, S. S. 2012. Detection of HTLV-1 in the labial salivary glands of patients with Sjögren's syndrome: a distinct clinical subgroup? *J Rheumatol,* 39**,** 809-15.

- LESSARD, C. J., LI, H., ADRIANTO, I., ICE, J. A., RASMUSSEN, A., GRUNDAHL, K. M., KELLY, J. A., DOZMOROV, M. G., MICELI-RICHARD, C., BOWMAN, S., LESTER, S., ERIKSSON, P., ELORANTA, M. L., BRUN, J. G., GORANSSON, L. G., HARBOE, E., GUTHRIDGE, J. M., KAUFMAN, K. M., KVARNSTROM, M., JAZEBI, H., CUNNINGHAME GRAHAM, D. S., GRANDITS, M. E., NAZMUL-HOSSAIN, A. N., PATEL, K., ADLER, A. J., MAIER-MOORE, J. S., FARRIS, A. D., BRENNAN, M. T., LESSARD, J. A., CHODOSH, J., GOPALAKRISHNAN, R., HEFNER, K. S., HOUSTON, G. D., HUANG, A. J., HUGHES, P. J., LEWIS, D. M., RADFAR, L., ROHRER, M. D., STONE, D. U., WREN, J. D., VYSE, T. J., GAFFNEY, P. M., JAMES, J. A., OMDAL, R., WAHREN-HERLENIUS, M., ILLEI, G. G., WITTE, T., JONSSON, R., RISCHMUELLER, M., RONNBLOM, L., NORDMARK, G., NG, W. F., MARIETTE, X., ANAYA, J. M., RHODUS, N. L., SEGAL, B. M., SCOFIELD, R. H., MONTGOMERY, C. G., HARLEY, J. B. & SIVILS, K. L. 2013. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren's syndrome. *Nat Genet,* 45**,** 1284-92.

- LEVINGS, P. P. & BUNGERT, J. 2002. The human beta-globin locus control region. *Eur J Biochem,* 269**,** 1589-99.

- LI, C., MA, J., ZHAO, H., BLAGG, B. S. & DOBROWSKY, R. T. 2012. Induction of heat shock protein 70 (Hsp70) prevents neuregulin-induced demyelination by enhancing the proteasomal clearance of c-Jun. *ASN Neuro,* 4**,** e00102.

- LI, Y., ZHANG, K., CHEN, H., SUN, F., XU, J., WU, Z., LI, P., ZHANG, L., DU, Y., LUAN, H., LI, X., WU, L., LI, H., WU, H., LI, X., LI, X., ZHANG, X., GONG, L., DAI, L., SUN, L., ZUO, X., XU, J., GONG, H., LI, Z., TONG, S., WU, M., LI, X., XIAO, W., WANG, G., ZHU, P., SHEN, M., LIU, S., ZHAO, D., LIU, W., WANG, Y., HUANG, C., JIANG, Q., LIU, G., LIU, B., HU, S., ZHANG, W., ZHANG, Z., YOU, X., LI, M., HAO, W., ZHAO, C.,

LENG, X., BI, L., WANG, Y., ZHANG, F., SHI, Q., QI, W., ZHANG, X., JIA, Y., SU, J., LI, Q., HOU, Y., WU, Q., XU, D., ZHENG, W., ZHANG, M., WANG, Q., FEI, Y., ZHANG, X., LI, J., JIANG, Y., TIAN, X., ZHAO, L., WANG, L., ZHOU, B., LI, Y., ZHAO, Y., ZENG, X., OTT, J., WANG, J. & ZHANG, F. 2013. A genome-wide association study in Han Chinese identifies a susceptibility locus for primary Sjögren's syndrome at 7q11.23. *Nat Genet,* 45**,** 1361-1365.

- LIANG, Y., YANG, Z., QIN, B. & ZHONG, R. 2014. Primary Sjögren's syndrome and malignancy risk: a systematic review and meta-analysis. *Ann Rheum Dis,* 73**,** 1151-6.

- LIN, S. M., DU, P., HUBER, W. & KIBBE, W. A. 2008. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Research,* 36**,** e11-e11.

- LIN, X., RUI, K., DENG, J., TIAN, J., WANG, X., WANG, S., KO, K. H., JIAO, Z., CHAN, V. S., LAU, C. S., CAO, X. & LU, L. 2015. Th17 cells play a critical role in the development of experimental Sjögren's syndrome. *Ann Rheum Dis,* 74**,** 1302-10.

- LIN, X., TIAN, J., RUI, K., MA, K. Y., KO, K. H., WANG, S. & LU, L. 2014. The role of T helper 17 cell subsets in Sjögren's syndrome: similarities and differences between mouse model and humans. *Ann Rheum Dis,* 73**,** e43.

- LIU, B. & LI, Z. 2008. Endoplasmic reticulum HSP90b1 (gp96, grp94) optimizes B-cell function via chaperoning integrin and TLR but not immunoglobulin. *Blood,* 112**,** 1223-30.

- LIU, C. M., LIN, Y. M., YEH, K. T., CHEN, M. K., CHANG, J. H., CHEN, C. J., CHOU, M. Y., YANG, S. F. & CHIEN, M. H. 2012. Expression of carbonic anhydrases I/II and the correlation to clinical aspects of oral squamous cell carcinoma analyzed using tissue microarray. *J Oral Pathol Med,* 41**,** 533-9.

- LIU, J., WALTER, E., STENGER, D. & THACH, D. 2006. Effects of globin mRNA reduction methods on gene expression profiles from whole blood. *J Mol Diagn,* 8**,** 551-8.

- LUO, J., SCHUMACHER, M., SCHERER, A., SANOUDOU, D., MEGHERBI, D., DAVISON, T., SHI, T., TONG, W., SHI, L., HONG, H., ZHAO, C., ELLOUMI, F., SHI, W., THOMAS, R., LIN, S., TILLINGHAST, G., LIU, G., ZHOU, Y., HERMAN, D., LI, Y., DENG, Y., FANG, H.,

BUSHEL, P., WOODS, M. & ZHANG, J. 2010. A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data. *The Pharmacogenomics Journal,* 10**,** 278-291.

- MA, A. & MALYNN, B. A. 2012. A20: linking a complex regulator of ubiquitylation to immunity and human disease. *Nat Rev Immunol,* 12**,** 774-85.

- MAEHARA, T., MORIYAMA, M., HAYASHIDA, J. N., TANAKA, A., SHINOZAKI, S., KUBO, Y., MATSUMURA, K. & NAKAMURA, S. 2012. Selective localization of T helper subsets in labial salivary glands from primary Sjögren's syndrome patients. *Clin Exp Immunol,* 169**,** 89-99.

- MANOUSSAKIS, M. N., BOIU, S., KORKOLOPOULOU, P., KAPSOGEORGOU, E. K., KAVANTZAS, N., ZIAKAS, P., PATSOURIS, E. & MOUTSOPOULOS, H. M. 2007. Rates of infiltration by macrophages and dendritic cells and expression of interleukin-18 and interleukin-12 in the chronic inflammatory lesions of Sjögren's syndrome: correlation with certain features of immune hyperactivity and factors associated with high risk of lymphoma development. *Arthritis Rheum,* 56**,** 3977-88.

- MANTHORPE, R., PERMIN, H. & TAGE-JENSEN, U. 1979. Auto-antibodies in Sjögren's syndrome. With special reference to liver-cell membrane antibody (LMA). *Scand J Rheumatol,* 8**,** 168-72.

- MARIA, N. I., BRKIC, Z., WARIS, M., VAN HELDEN-MEEUWSEN, C. G., HEEZEN, K., VAN DE MERWE, J. P., VAN DAELE, P. L., DALM, V. A., DREXHAGE, H. A. & VERSNEL, M. A. 2014. MxA as a clinically applicable biomarker for identifying systemic interferon type I in primary Sjögren's syndrome. *Ann Rheum Dis,* 73**,** 1052-9.

- MARIETTE, X., SIBILIA, J., DELAFORGE, C., BENGOUFA, D., BROUET, J. C. & SOUSSI, T. 1999. Anti-p53 antibodies are rarely detected in serum of patients with rheumatoid arthritis and Sjögren's syndrome. *J Rheumatol,* 26**,** 1672-5.

- MARNEROS, A. G. 2013. BMS1 is mutated in aplasia cutis congenita. *PLoS Genet,* 9**,** e1003573.

- MARTEL, C., GONDRAN, G., LAUNAY, D., LALLOUE, F., PALAT, S., LAMBERT, M., LY, K., LOUSTAUD-RATTI, V., BEZANAHARY, H., HACHULLA, E., JAUBERTEAU, M. O., VIDAL, E., HATRON, P. Y. &

FAUCHAIS, A. L. 2011. Active immunological profile is associated with systemic Sjögren's syndrome. *J Clin Immunol,* 31**,** 840-7.

- MASAKI, Y., DONG, L., KUROSE, N., KITAGAWA, K., MORIKAWA, Y., YAMAMOTO, M., TAKAHASHI, H., SHINOMURA, Y., IMAI, K., SAEKI, T., AZUMI, A., NAKADA, S., SUGIYAMA, E., MATSUI, S., ORIGUCHI, T., NISHIYAMA, S., NISHIMORI, I., NOJIMA, T., YAMADA, K., KAWANO, M., ZEN, Y., KANEKO, M., MIYAZAKI, K., TSUBOTA, K., EGUCHI, K., TOMODA, K., SAWAKI, T., KAWANAMI, T., TANAKA, M., FUKUSHIMA, T., SUGAI, S. & UMEHARA, H. 2009. Proposal for a new clinical entity, IgG4-positive multiorgan lymphoproliferative syndrome: analysis of 64 cases of IgG4-related disorders. *Ann Rheum Dis,* 68**,** 1310-5.

- MASAKI, Y. & SUGAI, S. 2004. Lymphoproliferative disorders in Sjögren's syndrome. *Autoimmun Rev,* 3**,** 175-82.

- MASTROKOLIAS, A., DEN DUNNEN, J. T., VAN OMMEN, G. B., T HOEN, P. A. & VAN ROON-MOM, W. M. 2012. Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. *BMC Genomics,* 13**,** 28.

- MAVRAGANI, C. P. & KASSAN, S. S. 2012. Classic Immunosuppressive and Immunomodulatory Drugs. *In:* RAMOS-CASALS, M., STONE, J. H. & MOUTSOPOULOS, H. (eds.) *Sjögren's Syndrome: Diagnosis and Therapeutics.* London: Springer-Verlag London.

- MAVRAGANI, C. P. & MOUTSOPOULOS, H. M. 2010. The geoepidemiology of Sjögren's syndrome. *Autoimmun Rev,* 9**,** A305-10.

- MCNAMARA, N. A., GALLUP, M. & PORCO, T. C. 2014. Establishing PAX6 as a biomarker to detect early loss of ocular phenotype in human patients with Sjögren's syndrome. *Invest Ophthalmol Vis Sci,* 55**,** 7079-84.

- MICELI-RICHARD, C., COMETS, E., LOISEAU, P., PUECHAL, X., HACHULLA, E. & MARIETTE, X. 2007. Association of an IRF5 gene functional polymorphism with Sjögren's syndrome. *Arthritis Rheum,* 56**,** 3989-94.

- MILETIĆ, M., STOJANOVIĆ, R., PAJIĆ, O., BUGARSKI, D., MOJSILOVIĆ, S., ČOKIĆ, V. & MILENKOVIĆ, P. 2012. Serum interleukin-17 & nitric oxide levels in patients with primary Sjögren's syndrome. *The Indian Journal of Medical Research,* 135**,** 513-519.

- MITSIAS, D. I., TZIOUFAS, A. G., VEIOPOULOU, C., ZINTZARAS, E., TASSIOS, I. K., KOGOPOULOU, O., MOUTSOPOULOS, H. M. & THYPHRONITIS, G. 2002. The Th1/Th2 cytokine balance changes with the progress of the immunopathological lesion of Sjögren's syndrome. *Clin Exp Immunol,* 128**,** 562-8.

- MOGASS, M., YORK, T. P., LI, L., RUJIRABANJERD, S. & SHIANG, R. 2004. Genomewide analysis of gene expression associated with Tcof1 in mouse neuroblastoma. *Biochem Biophys Res Commun,* 325**,** 124-32.

- MOREY, J. S., RYAN, J. C. & VAN DOLAH, F. M. 2006. Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. *Biol Proced Online,* 8**,** 175-93.

- MORGENSTERN, R., ZHANG, J. & JOHANSSON, K. 2011. Microsomal glutathione transferase 1: mechanism and functional roles. *Drug Metab Rev,* 43**,** 300-6.

- NAKKEN, B., JONSSON, R., BROKSTAD, K. A., OMHOLT, K., NERLAND, A. H., HAGA, H. J. & HALSE, A. K. 2001. Associations of MHC class II alleles in Norwegian primary Sjögren's syndrome patients: implications for development of autoantibodies to the Ro52 autoantigen. *Scand J Immunol,* 54**,** 428-33.

- NARDI, N., BRITO-ZERON, P., RAMOS-CASALS, M., AGUILO, S., CERVERA, R., INGELMO, M. & FONT, J. 2006. Circulating auto-antibodies against nuclear and non-nuclear antigens in primary Sjögren's syndrome: prevalence and clinical significance in 335 patients. *Clin Rheumatol,* 25**,** 341-6.

- NEBERT, D. W. & VASILIOU, V. 2004. Analysis of the glutathione S-transferase (GST) gene family. *Hum Genomics,* 1**,** 460-4.

- NEZOS, A. & MAVRAGANI, C. P. 2015. Contribution of Genetic Factors to Sjögren's Syndrome and Sjögren's Syndrome Related Lymphomagenesis. *J Immunol Res,* 2015**,** 754825.

- NEZOS, A., PAPAGEORGIOU, A., FRAGOULIS, G., IOAKEIMIDIS, D., KOUTSILIERIS, M., TZIOUFAS, A. G., MOUTSOPOULOS, H. M., VOULGARELIS, M. & MAVRAGANI, C. P. 2014. B-cell activating factor genetic variants in lymphomagenesis associated with primary Sjögren's syndrome. *J Autoimmun,* 51**,** 89-98.

- NG, W. F., BOWMAN, S. J. & GRIFFITHS, B. 2011. United Kingdom Primary Sjögren's Syndrome Registry--a united effort to tackle an orphan rheumatic disease. *Rheumatology (Oxford),* 50**,** 32-9.

- NGO, V. N., YOUNG, R. M., SCHMITZ, R., JHAVAR, S., XIAO, W., LIM, K.-H., KOHLHAMMER, H., XU, W., YANG, Y., ZHAO, H., SHAFFER, A. L., ROMESSER, P., WRIGHT, G., POWELL, J., ROSENWALD, A., MULLER-HERMELINK, H. K., OTT, G., GASCOYNE, R. D., CONNORS, J. M., RIMSZA, L. M., CAMPO, E., JAFFE, E. S., DELABIE, J., SMELAND, E. B., FISHER, R. I., BRAZIEL, R. M., TUBBS, R. R., COOK, J. R., WEISENBURGER, D. D., CHAN, W. C. & STAUDT, L. M. 2011. Oncogenically active MYD88 mutations in human lymphoma. *Nature,* 470**,** 115-119.

- NGUYEN, C. Q., SHARMA, A., SHE, J. X., MCINDOE, R. A. & PECK, A. B. 2009. Differential gene expressions in the lacrimal gland during development and onset of keratoconjunctivitis sicca in Sjögren's syndrome (SJS)-like disease of the C57BL/6.NOD-Aec1Aec2 mouse. *Exp Eye Res,* 88**,** 398-409.

- NISHISHINYA, M. B., PEREDA, C. A., MUNOZ-FERNANDEZ, S., PEGO-REIGOSA, J. M., RUA-FIGUEROA, I., ANDREU, J. L., FERNANDEZ-CASTRO, M., ROSAS, J. & LOZA SANTAMARIA, E. 2015. Identification of lymphoma predictors in patients with primary Sjögren's syndrome: a systematic literature review and meta-analysis. *Rheumatol Int,* 35**,** 17-26.

- NOCTURNE, G., BOUDAOUD, S., MICELI-RICHARD, C., VIENGCHAREUN, S., LAZURE, T., NITITHAM, J., TAYLOR, K. E., MA, A., BUSATO, F., MELKI, J., LESSARD, C. J., SIVILS, K. L., DUBOST, J. J., HACHULLA, E., GOTTENBERG, J. E., LOMBES, M., TOST, J., CRISWELL, L. A. & MARIETTE, X. 2013. Germline and somatic genetic variations of TNFAIP3 in lymphoma complicating primary Sjögren's syndrome. *Blood,* 122**,** 4068-76.

- NOCTURNE, G. & MARIETTE, X. 2015. Sjogren Syndrome-associated lymphomas: an update on pathogenesis and management. *Br J Haematol,* 168**,** 317-27.

- NOCTURNE, G., SEROR, R., FOGEL, O., BELKHIR, R., BOUDAOUD, S., SARAUX, A., LARROCHE, C., LE GUERN, V., GOTTENBERG, J. E. & MARIETTE, X. 2015a. CXCL13 and CCL11 Serum Levels and Lymphoma

and Disease Activity in Primary Sjögren's Syndrome. *Arthritis Rheumatol,* 67**,** 3226-33.

- NOCTURNE, G., TARN, J., BOUDAOUD, S., LOCKE, J., MICELI-RICHARD, C., HACHULLA, E., DUBOST, J. J., BOWMAN, S., GOTTENBERG, J. E., CRISWELL, L. A., LESSARD, C. J., SIVILS, K. L., CARAPITO, R., BAHRAM, S., SEROR, R., NG, W. F. & MARIETTE, X. 2015b. Germline variation of TNFAIP3 in primary Sjögren's syndrome-associated lymphoma. *Ann Rheum Dis*. 75**,** 780-3.

- NOCTURNE, G., VIRONE, A., NG, W. F., LE GUERN, V., HACHULLA, E., CORNEC, D., DAIEN, C., VITTECOQ, O., BIENVENU, B., MARCELLI, C., WENDLING, D., AMOURA, Z., DHOTE, R., LAVIGNE, C., FIOR, R., GOTTENBERG, J. E., SEROR, R. & MARIETTE, X. 2015c. Rheumatoid factor and disease activity are independent predictors of lymphoma in primary Sjögren's Syndrome. *Arthritis Rheumatol*. a head of print.

- NORDMARK, G., WANG, C., VASAITIS, L., ERIKSSON, P., THEANDER, E., KVARNSTROM, M., FORSBLAD-D'ELIA, H., JAZEBI, H., SJOWALL, C., REKSTEN, T. R., BRUN, J. G., JONSSON, M. V., JOHNSEN, S. J., WAHREN-HERLENIUS, M., OMDAL, R., JONSSON, R., BOWMAN, S., NG, W. F., ELORANTA, M. L. & SYVANEN, A. C. 2013. Association of genes in the NF-kappaB pathway with antibody-positive primary Sjögren's syndrome. *Scand J Immunol,* 78**,** 447-54.

- ODA, T., MORIKAWA, N., SAITO, Y., MASUHO, Y. & MATSUMOTO, S. 2000. Molecular cloning and characterization of a novel type of histamine receptor preferentially expressed in leukocytes. *J Biol Chem,* 275**,** 36781-6.

- OGAWA, N. 2010. Unique gene expression pattern of peripheral blood in patients with Sjögren's syndrome. *Inflammation and Regeneration,* 30**,** 169-175.

- OGAWA, N., PING, L., ZHENJUN, L., TAKADA, Y. & SUGAI, S. 2002. Involvement of the interferon-gamma-induced T cell-attracting chemokines, interferon-gamma-inducible 10-kd protein (CXCL10) and monokine induced by interferon-gamma (CXCL9), in the salivary gland lesions of patients with Sjögren's syndrome. *Arthritis Rheum,* 46**,** 2730-41.

- OKE, V. & WAHREN-HERLENIUS, M. 2012. The immunobiology of Ro52 (TRIM21) in autoimmunity: a critical review. *J Autoimmun,* 39**,** 77-82.

- OXHOLM, P., DANIELS, T. E. & BENDTZEN, K. 1992. Cytokine expression in labial salivary glands from patients with primary Sjögren's syndrome. *Autoimmunity,* 12**,** 185-91.

- PAISANSINSUP, T., DESHMUKH, U. S., CHOWDHARY, V. R., LUTHRA, H. S., FU, S. M. & DAVID, C. S. 2002. HLA class II influences the immune response and antibody diversification to Ro60/Sjögren's syndrome-A: heightened antibody responses and epitope spreading in mice expressing HLA-DR molecules. *J Immunol,* 168**,** 5876-84.

- PALOMINO-MORALES, R. J., DIAZ-GALLO, L. M., WITTE, T., ANAYA, J. M. & MARTIN, J. 2010. Influence of STAT4 polymorphism in primary Sjögren's syndrome. *J Rheumatol,* 37**,** 1016-9.

- PAN, H. F., YE, D. Q., WANG, Q., LI, W. X., ZHANG, N., LI, X. P., XU, J. H. & DAI, H. 2008. Clinical and laboratory profiles of systemic lupus erythematosus associated with Sjögren syndrome in China: a study of 542 patients. *Clin Rheumatol,* 27**,** 339-43.

- PANNUCCI, C. J. & WILKINS, E. G. 2010. Identifying and Avoiding Bias in Research. *Plastic and reconstructive surgery,* 126**,** 619-625.

- PAPAGEORGIOU, A., MAVRAGANI, C. P., NEZOS, A., ZINTZARAS, E., QUARTUCCIO, L., DE VITA, S., KOUTSILIERIS, M., TZIOUFAS, A. G., MOUTSOPOULOS, H. M. & VOULGARELIS, M. 2015a. A BAFF receptor His159Tyr mutation in Sjögren's syndrome-related lymphoproliferation. *Arthritis Rheumatol,* 67**,** 2732-41.

- PAPAGEORGIOU, A., ZIOGAS, D. C., MAVRAGANI, C. P., ZINTZARAS, E., TZIOUFAS, A. G., MOUTSOPOULOS, H. M. & VOULGARELIS, M. 2015b. Predicting the Outcome of Sjögren's Syndrome-Associated Non-Hodgkin's Lymphoma Patients. *PLoS ONE,* 10**,** e0116189.

- PATEL, R. & SHAHANE, A. 2014. The epidemiology of Sjögren's syndrome. *Clin Epidemiol,* 6**,** 247-55.

- PAULEY, K. M., STEWART, C. M., GAUNA, A. E., DUPRE, L. C., KUKLANI, R., CHAN, A. L., PAULEY, B. A., REEVES, W. H., CHAN, E. K. & CHA, S. 2011. Altered miR-146a expression in Sjögren's syndrome and its functional role in innate immunity. *Eur J Immunol,* 41**,** 2029-39.

- PECK, A. B., SAYLOR, B. T., NGUYEN, L., SHARMA, A., SHE, J. X., NGUYEN, C. Q. & MCINDOE, R. A. 2011. Gene expression profiling of early-phase Sjögren's syndrome in C57BL/6.NOD-Aec1Aec2 mice identifies focal adhesion maturation associated with infiltrating leukocytes. *Invest Ophthalmol Vis Sci,* 52**,** 5647-55.

- PERTOVAARA, M., PUKKALA, E., LAIPPALA, P., MIETTINEN, A. & PASTERNACK, A. 2001. A longitudinal cohort study of Finnish patients with primary Sjögren's syndrome: clinical, immunological, and epidemiological aspects. *Ann Rheum Dis,* 60**,** 467-72.

- PETERS, J. E. & ISENBERG, D. A. 2012. Sjögren's Syndrome and Associations with Other Autoimmune and Rheumatic Diseases. *In:* RAMOS-CASALS, M. S., J.H. AND MOUTSOPOULOS, H.M. (ed.) *Sjögren's Syndrome, Diagnosis and Therapeutics.* London: Springer-Verlag.

- PILLEMER, S. R. 2006. Lymphoma and other malignancies in primary Sjögren's syndrome. *Ann Rheum Dis,* 65**,** 704-6.

- PON, J. R., WONG, J., SABERI, S., ALDER, O., MOKSA, M., GRACE CHENG, S. W., MORIN, G. B., HOODLESS, P. A. & HIRST, M. 2015. MEF2B mutations in non-Hodgkin lymphoma dysregulate cell migration by decreasing MEF2B target gene activation. 6**,** 7953.

- PONTARINI, E., FABRIS, M., QUARTUCCIO, L., CAPPELETTI, M., CALCATERRA, F., ROBERTO, A., CURCIO, F., MAVILIO, D., DELLA BELLA, S. & DE VITA, S. 2015. Treatment with belimumab restores B cell subsets and their expression of B cell activating factor receptor in patients with primary Sjögren's syndrome. *Rheumatology (Oxford),* 54**,** 1429-34.

- PUTHALAKATH, H., HUANG, D. C., O'REILLY, L. A., KING, S. M. & STRASSER, A. 1999. The proapoptotic activity of the Bcl-2 family member Bim is regulated by interaction with the dynein motor complex. *Mol Cell,* 3**,** 287-96.

- QIN, B., WANG, J., YANG, Z., YANG, M., MA, N., HUANG, F. & ZHONG, R. 2015. Epidemiology of primary Sjögren's syndrome: a systematic review and meta-analysis. *Ann Rheum Dis,* 74**,** 1983-9.

- QUARTUCCIO, L., FABRIS, M., SALVIN, S., MASET, M., DE MARCHI, G. & DE VITA, S. 2009. Controversies on rituximab therapy in sjogren syndrome-associated lymphoproliferation. *Int J Rheumatol,* 2009**,** 424935.

- QUARTUCCIO, L., ISOLA, M., BALDINI, C., PRIORI, R., BARTOLONI, E., CARUBBI, F., GREGORACI, G., GANDOLFO, S., SALVIN, S., LUCIANO, N., MINNITI, A., ALUNNO, A., GIACOMELLI, R., GERLI, R., VALESINI, G., BOMBARDIERI, S. & DE VITA, S. 2015. Clinical and biological differences between cryoglobulinaemic and hypergammaglobulinaemic purpura in primary Sjögren's syndrome: results of a large multicentre study. *Scand J Rheumatol,* 44**,** 36-41.

- QUARTUCCIO, L., SALVIN, S., FABRIS, M., MASET, M., PONTARINI, E., ISOLA, M. & DE VITA, S. 2013. BLyS upregulation in Sjögren's syndrome associated with lymphoproliferative disorders, higher ESSDAI score and B-cell clonal expansion in the salivary glands. *Rheumatology (Oxford),* 52**,** 276-81.

- QUINTANA, F. J., BASSO, A. S., IGLESIAS, A. H., KORN, T., FAREZ, M. F., BETTELLI, E., CACCAMO, M., OUKKA, M. & WEINER, H. L. 2008. Control of T(reg) and T(H)17 cell differentiation by the aryl hydrocarbon receptor. *Nature,* 453**,** 65-71.

- RAGHAVACHARI, N., XU, X., MUNSON, P. J. & GLADWIN, M. T. 2009. Characterization of whole blood gene expression profiles as a sequel to globin mRNA reduction in patients with sickle cell disease. *PLoS One,* 4**,** e6484.

- RAINEN, L., OELMUELLER, U., JURGENSEN, S., WYRICH, R., BALLAS, C., SCHRAM, J., HERDMAN, C., BANKAITIS-DAVIS, D., NICHOLLS, N., TROLLINGER, D. & TRYON, V. 2002. Stabilization of mRNA expression in whole blood samples. *Clin Chem,* 48**,** 1883-90.

- RAJEEVAN, M. S., VERNON, S. D., TAYSAVANG, N. & UNGER, E. R. 2001. Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. *J Mol Diagn,* 3**,** 26-31.

- RAMOS-CASALS, M., GARCIA-CARRASCO, M., BRITO ZERON, M. P., CERVERA, R. & FONT, J. 2002. Viral etiopathogenesis of Sjögren's syndrome: role of the hepatitis C virus. *Autoimmun Rev,* 1**,** 238-43.

- RAMOS-CASALS, M., STONE, J. H. & MOUTSOPOULOS, H. M. 2012. *Sjögren's Syndrome: Diagnosis and Therapeutics,* London, Springer-Verlag.

- RAWE, V. Y., PAYNE, C. & SCHATTEN, G. 2006. Profilin and actin-related proteins regulate microfilament dynamics during early mammalian embryogenesis. *Hum Reprod,* 21**,** 1143-53.

- ROESCHER, N., TAK, P. P. & ILLEI, G. G. 2009. Cytokines in Sjögren's syndrome. *Oral Dis,* 15**,** 519-26.

- RONNEKLEIV-KELLY, S. M., NUKAYA, M., DIAZ-DIAZ, C. J., MEGNA, B. W., CARNEY, P. R., GEIGER, P. G. & KENNEDY, G. D. 2015. Aryl hydrocarbon receptor-dependent apoptotic cell death induced by the flavonoid chrysin in human colorectal cancer cells. *Cancer Lett.* 370**,** 91-9.

- ROSE, N. R. & MACKAY, I. R. 1998. *The Autoimmune Diseases,* San Diego, California, USA, Academic Press.

- ROUTSIAS, J. G. & TZIOUFAS, A. G. 2010. Autoimmune response and target autoantigens in Sjögren's syndrome. *Eur J Clin Invest,* 40**,** 1026-36.

- ROYER, B., CAZALS-HATEM, D., SIBILIA, J., AGBALIKA, F., CAYUELA, J. M., SOUSSI, T., MALOISEL, F., CLAUVEL, J. P., BROUET, J. C. & MARIETTE, X. 1997. Lymphomas in patients with Sjögren's syndrome are marginal zone B-cell neoplasms, arise in diverse extranodal and nodal sites, and are not associated with viruses. *Blood,* 90**,** 766-75.

- SAFAVI, S., HANSSON, M., KARLSSON, K., BILOGLAV, A., JOHANSSON, B. & PAULSSON, K. 2015. Novel gene targets detected by genomic profiling in a consecutive series of 126 adults with acute lymphoblastic leukemia. *Haematologica,* 100**,** 55-61.

- SALOMONSSON, S., JONSSON, M. V., SKARSTEIN, K., BROKSTAD, K. A., HJELMSTROM, P., WAHREN-HERLENIUS, M. & JONSSON, R. 2003. Cellular basis of ectopic germinal center formation and autoantibody production in the target organ of patients with Sjögren's syndrome. *Arthritis Rheum,* 48**,** 3187-201.

- SCHLUESCHE, P., STELZER, G., PIAIA, E., LAMB, D. C. & MEISTERERNST, M. 2007. NC2 mobilizes TBP on core promoter TATA boxes. *Nat Struct Mol Biol,* 14**,** 1196-201.

- SCHMID, R., BAUM, P., ITTRICH, C., FUNDEL-CLEMENS, K., HUBER, W., BRORS, B., EILS, R., WEITH, A., MENNERICH, D. & QUAST, K. 2010. Comparison of normalization methods for Illumina BeadChip HumanHT-12 v3. *BMC Genomics,* 11**,** 349.

- SELLAR, G. C., LI, L., WATT, K. P., NELKIN, B. D., RABIASZ, G. J., STRONACH, E. A., MILLER, E. P., PORTEOUS, D. J., SMYTH, J. F. & GABRA, H. 2001. BARX2 induces cadherin 6 expression and is a functional suppressor of ovarian cancer progression. *Cancer Res,* 61**,** 6977-81.

- SEROR, R., RAVAUD, P., BOWMAN, S. J., BARON, G., TZIOUFAS, A., THEANDER, E., GOTTENBERG, J. E., BOOTSMA, H., MARIETTE, X. & VITALI, C. 2010. EULAR Sjögren's syndrome disease activity index: development of a consensus systemic disease activity index for primary Sjögren's syndrome. *Ann Rheum Dis,* 69**,** 1103-9.

- SEROR, R., SORDET, C., GUILLEVIN, L., HACHULLA, E., MASSON, C., ITTAH, M., CANDON, S., LE GUERN, V., AOUBA, A., SIBILIA, J., GOTTENBERG, J. E. & MARIETTE, X. 2007. Tolerance and efficacy of rituximab and changes in serum B cell biomarkers in patients with systemic complications of primary Sjögren's syndrome. *Ann Rheum Dis,* 66**,** 351-7.

- SHIBOSKI, S. C., SHIBOSKI, C. H., CRISWELL, L., BAER, A., CHALLACOMBE, S., LANFRANCHI, H., SCHIODT, M., UMEHARA, H., VIVINO, F., ZHAO, Y., DONG, Y., GREENSPAN, D., HEIDENREICH, A. M., HELIN, P., KIRKHAM, B., KITAGAWA, K., LARKIN, G., LI, M., LIETMAN, T., LINDEGAARD, J., MCNAMARA, N., SACK, K., SHIRLAW, P., SUGAI, S., VOLLENWEIDER, C., WHITCHER, J., WU, A., ZHANG, S., ZHANG, W., GREENSPAN, J. & DANIELS, T. 2012. American College of Rheumatology classification criteria for Sjögren's syndrome: a data-driven, expert consensus approach in the Sjögren's International Collaborative Clinical Alliance cohort. *Arthritis Care Res (Hoboken),* 64**,** 475-87.

- SHIN, H., SHANNON, C. P., FISHBANE, N., RUAN, J., ZHOU, M., BALSHAW, R., WILSON-MCMANUS, J. E., NG, R. T., MCMANUS, B. M. & TEBBUTT, S. J. 2014. Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. *PLoS One,* 9**,** e91041.

- SKOPOULI, F. N., DAFNI, U., IOANNIDIS, J. P. & MOUTSOPOULOS, H. M. 2000. Clinical evolution, and morbidity and mortality of primary Sjögren's syndrome. *Semin Arthritis Rheum,* 29**,** 296-304.

- SMEDBY, K. E., HJALGRIM, H., ASKLING, J., CHANG, E. T., GREGERSEN, H., PORWIT-MACDONALD, A., SUNDSTROM, C., AKERMAN, M., MELBYE, M., GLIMELIUS, B. & ADAMI, H. O. 2006.

Autoimmune and chronic inflammatory disorders and risk of non-Hodgkin lymphoma by subtype. *J Natl Cancer Inst,* 98**,** 51-60.

- SMITH, M. L., DUNNING, M. J., TAVARE, S. & LYNCH, A. G. 2010. Identification and correction of previously unreported spatial phenomena using raw Illumina BeadArray data. *BMC Bioinformatics,* 11**,** 208-208.

- SMYTH, G. K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol,* 3**,** Article3.

- SOLANS-LAQUE, R., LOPEZ-HERNANDEZ, A., BOSCH-GIL, J. A., PALACIOS, A., CAMPILLO, M. & VILARDELL-TARRES, M. 2011. Risk, predictors, and clinical characteristics of lymphoma development in primary Sjögren's syndrome. *Semin Arthritis Rheum,* 41**,** 415-23.

- SOMER, B. G., TSAI, D. E., DOWNS, L., WEINSTEIN, B. & SCHUSTER, S. J. 2003. Improvement in Sjögren's syndrome following therapy with rituximab for marginal zone lymphoma. *Arthritis Rheum,* 49**,** 394-8.

- SONG, E. Y., CHUNG, H. Y., JOO, S. Y., ROH, E. Y., SEONG, M. W., SHIN, Y. & PARK, M. H. 2012. Detection of HLA-DRB1 microchimerism using nested polymerase chain reaction and single-strand conformation polymorphism analysis. *Hum Immunol,* 73**,** 291-7.

- SORENSEN, K. D., WILD, P. J., MORTEZAVI, A., ADOLF, K., TORRING, N., HEEBOLL, S., ULHOI, B. P., OTTOSEN, P., SULSER, T., HERMANNS, T., MOCH, H., BORRE, M., ORNTOFT, T. F. & DYRSKJOT, L. 2009. Genetic and epigenetic SLC18A2 silencing in prostate cancer is an independent adverse predictor of biochemical recurrence after radical prostatectomy. *Clin Cancer Res,* 15**,** 1400-10.

- SPACHIDOU, M. P., BOURAZOPOULOU, E., MARATHEFTIS, C. I., KAPSOGEORGOU, E. K., MOUTSOPOULOS, H. M., TZIOUFAS, A. G. & MANOUSSAKIS, M. N. 2007. Expression of functional Toll-like receptors by salivary gland epithelial cells: increased mRNA expression in cells derived from patients with primary Sjögren's syndrome. *Clinical & Experimental Immunology,* 147**,** 497-503.

- STEGAEV, V., SILLAT, T., POROLA, P., HANNINEN, A., FALUS, A., MIELIAUSKAITE, D., BUZAS, E., ROTAR, Z., MACKIEWICZ, Z., STARK, H., CHAZOT, P. L. & KONTTINEN, Y. T. 2012. Brief report: first

identification of H(4) histamine receptor in healthy salivary glands and in focal sialadenitis in Sjögren's syndrome. *Arthritis Rheum,* 64**,** 2663-8.

- STEGAJEV, V., KOURI, V. P., SALEM, A., ROZOV, S., STARK, H., NORDSTROM, D. C. & KONTTINEN, Y. T. 2014. Activation of histamine H4 receptor inhibits TNFalpha/IMD-0354-induced apoptosis in human salivary NS-SV-AC cells. *Apoptosis,* 19**,** 1702-11.

- STEIN, C. K., QU, P., EPSTEIN, J., BUROS, A., ROSENTHAL, A., CROWLEY, J., MORGAN, G. & BARLOGIE, B. 2015. Removing batch effects from purified plasma cell gene expression microarrays with modified ComBat. *BMC Bioinformatics,* 16**,** 63.

- STEVENS, T. A., IACOVONI, J. S., EDELMAN, D. B. & MEECH, R. 2004. Identification of novel binding elements and gene targets for the homeodomain protein BARX2. *J Biol Chem,* 279**,** 14520-30.
- STORDEUR, P., ZHOU, L., BYL, B., BROHET, F., BURNY, W., DE GROOTE, D., VAN DER POLL, T. & GOLDMAN, M. 2003. Immune monitoring in whole blood using real-time PCR. *J Immunol Methods,* 276**,** 69-77.

- STRIMBU, K. & TAVEL, J. A. 2010. What are Biomarkers? *Current opinion in HIV and AIDS,* 5**,** 463-466.

- SUGAI, S., SAITO, I., MASAKI, Y., TAKESHITA, S., SHIMIZU, S., TACHIBANA, J. & MIYASAKA, N. 1994. Rearrangement of the rheumatoid factor-related germline gene Vg and bcl-2 expression in lymphoproliferative disorders in patients with Sjögren's syndrome. *Clin Immunol Immunopathol,* 72**,** 181-6.

- SULLIVAN, G. M. & FEINN, R. 2012. Using Effect Size—or Why the P Value Is Not Enough. *Journal of Graduate Medical Education,* 4**,** 279-282.

- SUPURAN, C. T. 2008. Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat Rev Drug Discov,* 7**,** 168-81.

- SZCZERBA, B. M., RYBAKOWSKA, P. D., DEY, P., PAYERHIN, K. M., PECK, A. B., BAGAVANT, H. & DESHMUKH, U. S. 2013. Type I interferon receptor deficiency prevents murine Sjögren's syndrome. *J Dent Res,* 92**,** 444-9.

- SZODORAY, P., ALEX, P., JONSSON, M. V., KNOWLTON, N., DOZMOROV, I., NAKKEN, B., DELALEU, N., JONSSON, R. & CENTOLA, M. 2005. Distinct profiles of Sjögren's syndrome patients with ectopic salivary gland germinal centers revealed by serum cytokines and BAFF. *Clin Immunol,* 117**,** 168-76.

- SZODORAY, P., JELLESTAD, S., ALEX, P., ZHOU, T., WILSON, P. C., CENTOLA, M., BRUN, J. G. & JONSSON, R. 2004. Programmed cell death of peripheral blood B cells determined by laser scanning cytometry in Sjögren's syndrome with a special emphasis on BAFF. *J Clin Immunol,* 24**,** 600-11.

- TAKABATAKE, N., SAYAMA, T., SHIDA, K., MATSUDA, M., NAKAMURA, H. & TOMOIKE, H. 1999. Lung adenocarcinoma in lymphocytic interstitial pneumonitis associated with primary Sjögren's syndrome. *Respirology,* 4**,** 181-4.

- TAKEMOTO, F., HOSHINO, J., SAWA, N., TAMURA, Y., TAGAMI, T., YOKOTA, M., KATORI, H., YOKOYAMA, K., UBARA, Y., HARA, S., TAKAICHI, K., YAMADA, A. & UCHIDA, S. 2005. Autoantibodies against carbonic anhydrase II are increased in renal tubular acidosis associated with Sjogren syndrome. *Am J Med,* 118**,** 181-4.

- TALAL, N. 1991. *Molecular Autoimmunity,* London, Academic Press.

- TALAL, N. & BUNIM, J. J. 1964. THE DEVELOPMENT OF MALIGNANT LYMPHOMA IN THE COURSE OF SJOEGREN'S SYNDROME. *Am J Med,* 36**,** 529-40.

- TANDON, M., GALLO, A., JANG, S. I., ILLEI, G. G. & ALEVIZOS, I. 2012. Deep sequencing of short RNAs reveals novel microRNAs in minor salivary glands of patients with Sjögren's syndrome. *Oral Dis,* 18**,** 127-31.

- TARCA, A. L., ROMERO, R. & DRAGHICI, S. 2006. Analysis of microarray experiments of gene expression profiling. *American journal of obstetrics and gynecology,* 195**,** 373-388.

- TERADA, K., KATAMINE, S., EGUCHI, K., MORIUCHI, R., KITA, M., SHIMADA, H., YAMASHITA, I., IWATA, K., TSUJI, Y., NAGATAKI, S. & ET AL. 1994. Prevalence of serum and salivary antibodies to HTLV-1 in Sjögren's syndrome. *Lancet,* 344**,** 1116-9.

- THACH, D. C., LIN, B., WALTER, E., KRUZELOCK, R., ROWLEY, R. K., TIBBETTS, C. & STENGER, D. A. 2003. Assessment of two methods for handling blood in collection tubes with RNA stabilizing agent for surveillance of gene expression profiles with high density microarrays. *Journal of Immunological Methods,* 283**,** 269-279.

- THEANDER, E., HENRIKSSON, G., LJUNGBERG, O., MANDL, T., MANTHORPE, R. & JACOBSSON, L. T. 2006. Lymphoma and other malignancies in primary Sjögren's syndrome: a cohort study on cancer incidence and lymphoma predictors. *Ann Rheum Dis,* 65**,** 796-803.

- THEANDER, E., VASAITIS, L., BAECKLUND, E., NORDMARK, G., WARFVINGE, G., LIEDHOLM, R., BROKSTAD, K., JONSSON, R. & JONSSON, M. V. 2011. Lymphoid organisation in labial salivary gland biopsies is a possible predictor for the development of malignant lymphoma in primary Sjögren's syndrome. *Ann Rheum Dis,* 70**,** 1363-8.

- TISHLER, M., YARON, I., SHIRAZI, I. & YARON, M. 1999. Hydroxychloroquine treatment for primary Sjögren's syndrome: its effect on salivary and serum inflammatory markers. *Ann Rheum Dis,* 58**,** 253-6.

- TOBON, G. J., SARAUX, A., GOTTENBERG, J. E., QUARTUCCIO, L., FABRIS, M., SEROR, R., DEVAUCHELLE-PENSEC, V., MOREL, J., RIST, S., MARIETTE, X., DE VITA, S., YOUINOU, P. & PERS, J. O. 2013. Role of Fms-like tyrosine kinase 3 ligand as a potential biologic marker of lymphoma in primary Sjögren's syndrome. *Arthritis Rheum,* 65**,** 3218-27.

- TOMI, A. L., BELKHIR, R., NOCTURNE, G., DESMOULINS, F., BERGE, E., PAVY, S., MICELI-RICHARD, C., MARIETTE, X. & SEROR, R. 2015. Monoclonal gammopathy and risk of lymphoma and multiple myeloma in patients with primary Sjögren's syndrome. *Arthritis Rheumatol*.

- TORO-DOMINGUEZ, D., CARMONA-SAEZ, P. & ALARCON-RIQUELME, M. E. 2014. Shared signatures between rheumatoid arthritis, systemic lupus erythematosus and Sjögren's syndrome uncovered through gene expression meta-analysis. *Arthritis Res Ther,* 16**,** 489.

- TROPPAN, K., WENZL, K., NEUMEISTER, P. & DEUTSCH, A. 2015. Molecular Pathogenesis of MALT Lymphoma. *Gastroenterol Res Pract,* 2015**,** 102656.

- TSIFETAKI, N., KITSOS, G., PASCHIDES, C. A., ALAMANOS, Y., EFTAXIAS, V., VOULGARI, P. V., PSILAS, K. & DROSOS, A. A. 2003. Oral pilocarpine for the treatment of ocular symptoms in patients with Sjögren's syndrome: a randomised 12 week controlled study. *Ann Rheum Dis,* 62**,** 1204-7.

- TSUJI, N., FUKUDA, K., NAGATA, Y., OKADA, H., HAGA, A., HATAKEYAMA, S., YOSHIDA, S., OKAMOTO, T., HOSAKA, M., SEKINE, K., OHTAKA, K., YAMAMOTO, S., OTAKA, M., GRAVE, E. & ITOH, H. 2014. The activation mechanism of the aryl hydrocarbon receptor (AhR) by molecular chaperone HSP90. *FEBS Open Bio,* 4**,** 796-803.

- TURTURRO, F. 2015. Constitutive NF- kappa B Activation Underlines Major Mechanism of Drug Resistance in Relapsed Refractory Diffuse Large B Cell Lymphoma. *Biomed Res Int,* 2015**,** 484537.

- TZIOUFAS, A. G., BOUMBA, D. S., SKOPOULI, F. N. & MOUTSOPOULOS, H. M. 1996. Mixed monoclonal cryoglobulinemia and monoclonal rheumatoid factor cross-reactive idiotypes as predictive factors for the development of lymphoma in primary Sjögren's syndrome. *Arthritis Rheum,* 39**,** 767-72.

- TZIOUFAS, A. G., KAPSOGEORGOU, E. K., MANOUSSAKIS, M. N. & MOUTSOPOULOS, H. M. 2012. Pathogenetic Aspects of Primary Sjögren's Syndrome *Sjögren's Syndrome Diagnosis and Therapeutics.* London: Springer-Verlag.

- TZIOUFAS, A. G., WASSMUTH, R., DAFNI, U. G., GUIALIS, A., HAGA, H. J., ISENBERG, D. A., JONSSON, R., KALDEN, J. R., KIENER, H., SAKARELLOS, C., SMOLEN, J. S., SUTCLIFFE, N., VITALI, C., YIANNAKI, E. & MOUTSOPOULOS, H. M. 2002. Clinical, immunological, and immunogenetic aspects of autoantibody production against Ro/SSA, La/SSB and their linear epitopes in primary Sjögren's syndrome (pSS): a European multicentre study. *Ann Rheum Dis,* 61**,** 398-404.

- UNO, Y., MURAYAMA, N., KUNORI, M. & YAMAZAKI, H. 2013. Characterization of microsomal glutathione S-transferases MGST1, MGST2, and MGST3 in cynomolgus macaque. *Drug Metab Dispos,* 41**,** 1621-5.

- VALDAGNI, R., RANCATI, T., GHILOTTI, M., COZZARINI, C., VAVASSORI, V., FELLIN, G., FIORINO, C., GIRELLI, G., BARRA, S., ZAFFARONI, N., PIEROTTI, M. A. & GARIBOLDI, M. 2009. To bleed or

not to bleed. A prediction based on individual gene profiling combined with dose-volume histogram shapes in prostate cancer patients undergoing three-dimensional conformal radiation therapy. *Int J Radiat Oncol Biol Phys,* 74**,** 1431-40.

- VALESINI, G., PRIORI, R., BAVOILLOT, D., OSBORN, J., DANIELI, M. G., DEL PAPA, N., GERLI, R., PIETROGRANDE, M., SABBADINI, M. G., SILVESTRIS, F. & VALSECCHI, L. 1997. Differential risk of non-Hodgkin's lymphoma in Italian patients with primary Sjögren's syndrome. *J Rheumatol,* 24**,** 2376-80.

- VAN WOERKOM, J. M., KRUIZE, A. A., WENTING-VAN WIJK, M. J., KNOL, E., BIHARI, I. C., JACOBS, J. W., BIJLSMA, J. W., LAFEBER, F. P. & VAN ROON, J. A. 2005. Salivary gland and peripheral blood T helper 1 and 2 cell activity in Sjögren's syndrome compared with non-Sjögren's sicca syndrome. *Ann Rheum Dis,* 64**,** 1474-9.

- VARTANIAN, K., SLOTTKE, R., JOHNSTONE, T., CASALE, A., PLANCK, S. R., CHOI, D., SMITH, J. R., ROSENBAUM, J. T. & HARRINGTON, C. A. 2009. Gene expression profiling of whole blood: comparison of target preparation methods for accurate and reproducible microarray analysis. *BMC Genomics,* 10**,** 2.

- VELDHOEN, M., HIROTA, K., WESTENDORF, A. M., BUER, J., DUMOUTIER, L., RENAULD, J. C. & STOCKINGER, B. 2008. The aryl hydrocarbon receptor links TH17-cell-mediated autoimmunity to environmental toxins. *Nature,* 453**,** 106-9.

- VENABLES, P. J. 2004. Sjögren's syndrome. *Best Pract Res Clin Rheumatol,* 18**,** 313-29.

- VENABLES, P. J. & RIGBY, S. P. 1997. Viruses in the etiopathogenesis of Sjögren's syndrome. *J Rheumatol Suppl,* 50**,** 3-5.

- VITALI, C., BOMBARDIERI, S., JONSSON, R., MOUTSOPOULOS, H., ALEXANDER, E., CARSONS, S., DANIELS, T., FOX, P., FOX, R., KASSAN, S., PILLEMER, S., TALAL, N. & WEISMAN, M. 2002. Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Annals of the Rheumatic Diseases,* 61**,** 554-558.

- VITALI, C., BOMBARDIERI, S., MOUTSOPOULOS, H. M., BALESTRIERI, G., BENCIVELLI, W., BERNSTEIN, R. M., BJERRUM, K. B., BRAGA, S., COLL, J., DE VITA, S. & ET AL. 1993. Preliminary criteria for the classification of Sjögren's syndrome. Results of a prospective concerted action supported by the European Community. *Arthritis Rheum,* 36**,** 340-7.

- VOSTERS, J. L., ROESCHER, N., POLLING, E. J., ILLEI, G. G. & TAK, P. P. 2012. The expression of APRIL in Sjögren's syndrome: aberrant expression of APRIL in the salivary gland. *Rheumatology (Oxford),* 51**,** 1557-62.

- VOULGARELIS, M., DAFNI, U. G., ISENBERG, D. A. & MOUTSOPOULOS, H. M. 1999. Malignant lymphoma in primary Sjögren's syndrome: a multicenter, retrospective, clinical study by the European Concerted Action on Sjögren's Syndrome. *Arthritis Rheum,* 42**,** 1765-72.

- VOULGARELIS, M., MAVRAGANI, C. P., XU, L., TREON, S. P. & MOUTSOPOULOS, H. M. 2014. Absence of somatic MYD88 L265P mutations in patients with primary Sjögren's syndrome. *Genes Immun,* 15**,** 54-6.

- VOULGARELIS, M. & MOUTSOPOULOS, H. M. 2001. Malignant lymphoma in primary Sjögren's syndrome. *Isr Med Assoc J,* 3**,** 761-6.

- VOULGARELIS, M., ZIAKAS, P. D., PAPAGEORGIOU, A., BAIMPA, E., TZIOUFAS, A. G. & MOUTSOPOULOS, H. M. 2012. Prognosis and outcome of non-Hodgkin lymphoma in primary Sjogren syndrome. *Medicine (Baltimore),* 91**,** 1-9.

- VUCICEVIC BORAS, V., BRAILO, V., LUKAC, J., KORDIC, D., PICEK, P. & BLAZIC-POTOCKI, Z. 2006. Salivary interleukin-6 and tumor necrosis factor alpha in patients with drug-induced xerostomia. *Oral Dis,* 12**,** 509-11.

- WANG, G., REN, G., CUI, X., LU, Z., MA, Y., QI, Y., HUANG, Y., LIU, Z., SUN, Z. & RUAN, Q. 2016. Human cytomegalovirus RL13 protein interacts with host NUDT14 protein affecting viral DNA replication. *Mol Med Rep*.

- WANG, J., ROBINSON, J. F., KHAN, H. M. R., CARTER, D. E., MCKINNEY, J., MISKIE, B. A. & HEGELE, R. A. 2004. Optimizing RNA extraction yield from whole blood for microarray gene expression analysis. *Clinical Biochemistry,* 37**,** 741-744.

- WANG, S. & KAUFMAN, R. J. 2012. The impact of the unfolded protein response on human disease. *J Cell Biol,* 197**,** 857-67.

- WANG, Y., SUN, N., LI, S., DU, Q., XU, Y., LIU, Z. & ZHANG, K. 2015. A Genetic Susceptibility Mechanism for Major Depression: Combinations of polymorphisms Defined the Risk of Major Depression and Subpopulations. *Medicine,* 94**,** e778.

- WENG, M. Y., HUANG, Y. T., LIU, M. F. & LU, T. H. 2012. Incidence of cancer in a nationwide population cohort of 7852 patients with primary Sjögren's syndrome in Taiwan. *Ann Rheum Dis,* 71**,** 524-7.

- WENZEL, J., HENZE, S., WORENKAMPER, E., BASNER-TSCHAKARJAN, E., SOKOLOWSKA-WOJDYLO, M., STEITZ, J., BIEBER, T. & TUTING, T. 2005. Role of the chemokine receptor CCR4 and its ligand thymus- and activation-regulated chemokine/CCL17 for lymphocyte recruitment in cutaneous lupus erythematosus. *J Invest Dermatol,* 124**,** 1241-8.

- WHITE, R. J., KHOO, B. C., INOSTROZA, J. A., REINBERG, D. & JACKSON, S. P. 1994. Differential regulation of RNA polymerases I, II, and III by the TBP-binding repressor Dr1. *Science,* 266**,** 448-50.

- WHITLEY P, G. J. A. G. M. 2007. Improved gene expression profiling with mouse blood samples. . *Ambion TechNotes,* 13**,** 27–28.

- WHITLEY P, M. S., SANTIAGO J, JOHNSON C AND SETTERQUIST R 2005. Improved microarray sensitivity using whole blood RNA samples. *Ambion TechNotes* 12**,** 20–23.

- WILDENBERG, M. E., VAN HELDEN-MEEUWSEN, C. G., VAN DE MERWE, J. P., DREXHAGE, H. A. & VERSNEL, M. A. 2008. Systemic increase in type I interferon activity in Sjögren's syndrome: A putative role for plasmacytoid dendritic cells. *European Journal of Immunology,* 38**,** 2024-2033.

- WILLEKE, P., SCHLUTER, B., SCHOTTE, H., DOMSCHKE, W., GAUBITZ, M. & BECKER, H. 2009. Interferon-gamma is increased in patients with primary Sjögren's syndrome and Raynaud's phenomenon. *Semin Arthritis Rheum,* 39**,** 197-202.

- WILLOUGHBY, C. E., BAKER, K., KAYE, S. B., CAREY, P., O'DONNELL, N., FIELD, A., LONGMAN, L., BUCKNALL, R. & HART, C. A. 2002.

Epstein-Barr virus (types 1 and 2) in the tear film in Sjögren's syndrome and HIV infection. *J Med Virol,* 68**,** 378-83.

- WINN, M. E., SHAW, M., APRIL, C., KLOTZLE, B., FAN, J. B., MURRAY, S. S. & SCHORK, N. J. 2011. Gene expression profiling of human whole blood samples with the Illumina WG-DASL assay. *BMC Genomics,* 12**,** 412.

- WINN, M. E., ZAPALA, M. A., HOVATTA, I., RISBROUGH, V. B., LILLIE, E. & SCHORK, N. J. 2010. The effects of globin on microarray-based gene expression analysis of mouse blood. *Mammalian Genome,* 21**,** 268-275.

- WINZER, M. & ARINGER, M. 2010. Use of methotrexate in patients with systemic lupus erythematosus and primary Sjögren's syndrome. *Clin Exp Rheumatol,* 28**,** S156-9.

- WITKE, W. 2004. The role of profilin complexes in cell motility and other cellular processes. *Trends Cell Biol,* 14**,** 461-9.

- WRIGHT, C., BERGSTROM, D., DAI, H., MARTON, M., MORRIS, M., TOKIWA, G., WANG, Y. & FARE, T. 2008. Characterization of globin RNA interference in gene expression profiling of whole-blood samples. *Clin Chem,* 54**,** 396-405.

- WU, J., FEI, P., GONG, Y., ZHAO, J., DONG, J. & TANG, F. 2000. [Apoptosis and related gene expression in lacrimal gland of cases with Sjögren's syndrome]. *Zhonghua Yan Ke Za Zhi,* 36**,** 255-8, 14.

- WU, Z. 2009. A Review of Statistical Methods for Preprocessing Oligonucleotide Microarrays. *Statistical methods in medical research,* 18**,** 533-541.

- WURMBACH, E., YUEN, T. & SEALFON, S. C. 2003. Focused microarray analysis. *Methods,* 31**,** 306-16.

- WWW.LIFETECHNOLOGIES.COM 2011. TaqMan® Array Plates,Fast 96-Well Plates, 96-Well Plates, Gene Signature Sets. 4391016 Rev. F.

- WWW.LIFETECHNOLOGIES.COM 2012. TaqMan® Gene Expression Assay solutions, Proven performance for fast, reliable results.

- YAGI, T., BAROJA-FERNáNDEZ, E., YAMAMOTO, R., MUñOZ, F. J., AKAZAWA, T., HONG, K. S. & POZUETA-ROMERO, J. 2003. Cloning, expression and characterization of a mammalian Nudix hydrolase-like enzyme that cleaves the pyrophosphate bond of UDP-glucose. *Biochemical Journal,* 370**,** 409-415.

- YAMAGAMI, S., EBIHARA, N. & AMANO, S. Y. 2005. Chemokine receptor gene expression in giant papillae of atopic keratoconjunctivitis. *Mol Vis,* 11**,** 192-200.

- YAN, M. X., YANG, J., SUN, Q., LIU, C. H., WANG, Y. G. & WANG, W. Q. 2013. Hepatocellular carcinoma that arose from primary Sjögren's syndrome. *Ann Hepatol,* 12**,** 824-9.

- YANG, K., SHI, H. X., LIU, X. Y., SHAN, Y. F., WEI, B., CHEN, S. & WANG, C. 2009. TRIM21 is essential to sustain IFN regulatory factor 3 activation during antiviral response. *J Immunol,* 182**,** 3782-92.

- YANG, X., LIU, D., MURRAY, T. J., MITCHELL, G. C., HESTERMAN, E. V., KARCHNER, S. I., MERSON, R. R., HAHN, M. E. & SHERR, D. H. 2005. The aryl hydrocarbon receptor constitutively represses c-myc transcription in human mammary tumor cells. *Oncogene,* 24**,** 7869-81.

- YAO, Y., LIU, Z., JALLAL, B., SHEN, N. & RONNBLOM, L. 2013. Type I interferons in Sjögren's syndrome. *Autoimmun Rev,* 12**,** 558-66.

- YI, P., HIGA, A., TAOUJI, S., BEXIGA, M. G., MARZA, E., ARMA, D., CASTAIN, C., LE BAIL, B., SIMPSON, J. C., ROSENBAUM, J., BALABAUD, C., BIOULAC-SAGE, P., BLANC, J. F. & CHEVET, E. 2012. Sorafenib-mediated targeting of the AAA(+) ATPase p97/VCP leads to disruption of the secretory pathway, endoplasmic reticulum stress, and hepatocellular cancer cell death. *Mol Cancer Ther,* 11**,** 2610-20.

- YOUINOU, P., DEVAUCHELLE-PENSEC, V. & PERS, J. O. 2010. Significance of B cells and B cell clonality in Sjögren's syndrome. *Arthritis Rheum,* 62**,** 2605-10.

- YOUNG, I. S. & WOODSIDE, J. V. 2001. Antioxidants in health and disease. *J Clin Pathol,* 54**,** 176-86.

- ZHANG, W., FENG, S., YAN, S., ZHAO, Y., LI, M., SUN, J., ZHANG, F. C., CUI, Q. & DONG, Y. 2010. Incidence of malignancy in primary Sjögren's syndrome in a Chinese cohort. *Rheumatology (Oxford),* 49**,** 571-7.

- ZHANG, Y., WEI, Z., LI, J. & LIU, P. 2015. Molecular pathogenesis of lymphomas of mucosa-associated lymphoid tissue-from (auto)antigen driven selection to the activation of NF-kappaB signaling. *Sci China Life Sci*. 58**,** 1246-55.

- ZHAO, N., LIU, Y., CHANG, Z., LI, K., ZHANG, R., ZHOU, Y., QIU, F., HAN, X. & XU, Y. 2015. Identification of Biomarker and Co-Regulatory Motifs in Lung Adenocarcinoma Based on Differential Interactions. *PLoS One,* 10**,** e0139165.

- ZHENG, L., ZHANG, Z., YU, C., TU, L., ZHONG, L. & YANG, C. 2009. Association between IFN-alpha and primary Sjögren's syndrome. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod,* 107**,** e12-8.

- ZINTZARAS, E., VOULGARELIS, M. & MOUTSOPOULOS, H. M. 2005. The risk of lymphoma development in autoimmune diseases: a meta-analysis. *Arch Intern Med,* 165**,** 2337-44.

- ZUCKERMAN, N. S., HAZANOV, H., BARAK, M., EDELMAN, H., HESS, S., SHCOLNIK, H., DUNN-WALTERS, D. & MEHR, R. 2010. Somatic hypermutation and antigen-driven selection of B cells are altered in autoimmune diseases. *J Autoimmun,* 35**,** 325-35.

- ZURGIL, N., BAKIMER, R., MOUTSOPOULOS, H. M., TZIOUFAS, A. G., YOUINOU, P., ISENBERG, D. A., SCHEINBERG, M., KVEDER, T., ROZMAN, B., LUDERSCHMIDT, C. & ET AL. 1992. Antimitochondrial (pyruvate dehydrogenase) autoantibodies in autoimmune rheumatic diseases. *J Clin Immunol,* 12**,** 201-9.

Supplementary data

Conferences and publications:

Poster presentations:

1. Al-Ali S, Cockell S, Tarn J, James K, Young D, UKPSSR study group, Griffiths B, Bowman S and Ng WF. 2013. "Effects of globin mRNA on the whole blood gene expression signature of primary Sjögren's syndrome". The 12th international Sjögren's syndrome symposium (New Era on Sjögren's Syndrome), Kyoto, Japan, a young travel award winning abstract**.**
2. Al-Ali S, Skelton A, Cockell S, James K, Tarn J , Young D, UKPSSR study group, Griffiths B, Bowman S and Ng WF. 2015. "Identification of whole blood gene expression signature of primary Sjögren's syndrome associated lymphoma". The 13th International Symposium on Sjögren's Syndrome (ISSS), Bergen, Norway, Scandinavian Journal of Immunology, 81(5), 273.

Oral presentations:

1. Al-Ali S, Skelton A, Cockell S, James K, Tarn J , Young D, UKPSSR study group, Griffiths B, Bowman S and Ng WF. 2014. "Identification of whole blood gene expression signature in primary Sjögren's syndrome-associated lymphoma". Oral presentation at the ACR/ARHP Annual meeting (American College of Rheumatology) in Boston, USA. 2014 ACR/ARHP Annual Meeting Abstract Supplement, Arthritis & Rheumatology, 66 (S10), S1-S1402.
2. Al-Ali S, Skelton A, Cockell S, James K, Tarn J , Young D, UKPSSR study group, Griffiths B, Bowman S and Ng WF. 2014. "Identification of whole blood gene expression signature in primary Sjögren's syndrome-associated lymphoma". Oral presentation at The Annual Northern and Yorkshire Rheumatology Meeting at York, UK.

Publications:

1. JAMES, K., AL-ALI, S., TARN, J., COCKELL, S. J., GILLESPIE, C. S., HINDMARSH, V., LOCKE, J., MITCHELL, S., LENDREM, D., BOWMAN, S., PRICE, E., PEASE, C. T., EMERY, P., LANYON, P., HUNTER, J. A., GUPTA, M., BOMBARDIERI, M., SUTCLIFFE, N., PITZALIS, C., MCLAREN, J., COOPER, A., REGAN, M., GILES, I., ISENBERG, D., SARAVANAN, V., COADY, D., DASGUPTA, B., MCHUGH, N., YOUNG-MIN, S., MOOTS, R., GENDI, N., AKIL, M., GRIFFITHS, B., WIPAT, A., NEWTON, J., JONES, D. E., ISAACS, J., HALLINAN, J. & NG, W. F. 2015. a transcriptional signature of fatigue derived from patients with primary sjogren's syndrome. *plos one,* 10**,** e0143970.

Supplementary Tables

S1 R scripts for microarray data analysis of the Globin mRNA depletion effects on pSS gene expression profiling

```
source("http://bioconductor.org/biocLite.R")
biocLite(c("lumi", "gplots", "ggplot2", "limma", "annotate", "lumiHumanAll.db", "limma", "sva",
"lumiHumanIDMapping"))
install.packages(c("scales", "reshape2"))
library(stringr)
library(sva)
library(lumi)
library(gplots)
library(ggplot2)
library(annotate)
library(lumiHumanAll.db)
library(limma)
library(scales)
library(reshape2)
library(lumiHumanIDMapping)
##'-----------------------------------------------------------------------------------------------------------#
filename            <- "raw_data_SA.txt"
raw_data            <- lumiR(filename)
pheno_table         <- read.table("pheno.txt",
                          header=T,
                          sep="\t",
                          row.names=1,
                          stringsAsFactors=F)
pData(raw_data)  <- pheno_table
##'-----------------------------------------------------------------------------------------------------------#
det <- melt(detection(raw_data))
dtp <- ggplot(data=det, aes(x=Var2, y=value)) +
        geom_boxplot(outlier.size=0.5,
       size=0.2) +
       scale_x_discrete(name="") +
       scale_y_continuous(name="Amplitude") +
       theme_bw() +
       theme(axis.text.x=element_text(angle=90, vjust=0.5,
       size=6),
       axis.text.y=element_text(size=6),
       axis.title.y=element_text(size=6))
image_deploy(dtp, "DetectionPval_")
##'-----------------------------------------------------------------------------------------------------------#
raw_data_det        <- raw_data
raw_data_det        <- raw_data[, grep("Globin_Clear", pData(raw_data)$class)]
# raw_data_det      <- raw_data[, grep("Full_RNA",    pData(raw_data)$class)]
##'-----------------------------------------------------------------------------------------------------------#
vst_data            <- lumiT(raw_data_det, method='vst')
rsn_data            <- lumiN(vst_data, method = "rsn")
lumi.Q              <- lumiQ(rsn_data)
exprs_data          <- exprs(lumi.Q)
present_count       <- detectionCall(lumi.Q)
normalised_data  <- exprs_data[present_count > 0, ]
```

```
##'-------------------------------------------------------------------------------------#
probe_list      <- rownames(normalised_data)
nuIDs           <- probeID2nuID(probe_list)[, "nuID"]
symbol          <- getSYMBOL(nuIDs, "lumiHumanAll.db")
name            <- unlist(lookUp(nuIDs, "lumiHumanAll.db", "GENENAME"))
anno_df         <- data.frame(ID = nuIDs, probe_list, symbol, name)
entrez_map      <- data.frame(nuID=as.vector(anno_df$ID),
                       EntrezID=nuID2EntrezID(as.vector(anno_df$ID),
                      "lumiHumanIDMapping"))
##'-------------------------------------------------------------------------------------#
design                  <- model.matrix(~0 + factor(pData(raw_data_det)$treatment,
                           levels=c("SS", "control")))
colnames(design)        <- c("SS", "control")
num_parameters          <- ncol(design)
fit                     <- lmFit(normalised_data, design)
cont_mat                <- makeContrasts(SS-control, levels=c("SS", "control"))
fit2                    <- contrasts.fit(fit, contrasts=cont_mat)
fit2                    <- eBayes(fit2)
fit2$genes              <- anno_df
##'-------------------------------------------------------------------------------------#
comparisons <- c("SS - control")
p_cut_off     <- 0.05
fold_change <- 1.2
i               <- 1
gene_list_unfiltered <- topTable(fit2,
                            coef="SS - control",
                            number=Inf,
                            adjust.method="BH")
gene_list               <- topTable(fit2,
                           coef="SS - control",
                           p.value=p_cut_off,
                           lfc=log2(fold_change),
                           number=Inf,
                           adjust.method="BH")
##'-------------------------------------------------------------------------------------#
pca         <- prcomp(t(normalised_data))
d           <- as.data.frame(pca$x)
d           <- cbind(d, pData(raw_data_det))
d$pairs     <- as.factor(d$pairs)

gg1         <- ggplot(d, aes(x=PC1, y=PC2, shape=class)) +
               geom_point(size=3) +
               geom_text(label=d$pairs, size=4, vjust=1.2, hjust=-0.2) +
            theme_bw() +
            theme(axis.title.x = element_text(size=20),
           axis.title.y = element_text(size=20))

gg2       <- ggplot(d, aes(x=PC1, y=PC2, shape=class)) +
              geom_point(aes(colour=treatment), size=3) +
              geom_text(label=d$pairs, size=4, vjust=1.2, hjust=-0.2) +
             scale_colour_grey(start = 0, end = .9) +
             theme_bw() +
             theme(axis.title.x = element_text(size=20),
```

```
            axis.title.y = element_text(size=20))
##'-----------------------------------------------------------------------------------------#
library(VennDiagram)
unfiltered_fullrna    <- topTable(fit2, coef="SS - control", number=Inf,
                          adjust.method="BH")
unfiltered_fullrna_in <- unfiltered_fullrna
unfiltered_globin     <- topTable(fit2, coef="SS - control", number=Inf,
                          adjust.method="BH")
unfiltered_globin_in  <- unfiltered_globin

afc  <- 1.2
pval <- 0.05

unfiltered_fullrna_in$pass  <- unfiltered_fullrna_in$adj.P.Val < pval &
abs(unfiltered_fullrna_in$logFC) > log2(afc)
unfiltered_fullrna_in$class <- "Full_RNA"
unfiltered_globin_in$pass   <- unfiltered_globin_in$adj.P.Val < pval &
abs(unfiltered_globin_in$logFC) > log2(afc)
unfiltered_globin_in$class  <- "Globin_Clear"
df                          <- rbind(unfiltered_fullrna_in, unfiltered_globin_in)

ggplot(df, aes(x=logFC, y=-log(adj.P.Val, 10))) +
  geom_point(aes(colour=pass, shape=pass), show_guide=F) +
  scale_colour_manual(values=c(alpha('grey', 0.5), 'black')) +
  geom_hline(yintercept=-log(pval, 10), colour="black", linetype=2) +
  geom_vline(xintercept=-log2(afc), colour="black", linetype=2) +
  geom_vline(xintercept=log2(afc), colour="black", linetype=2) +
  facet_grid(. ~ class) +
  theme_bw()

ggplot(unfiltered_fullrna_in, aes(x=logFC, y=-log(adj.P.Val, 10))) +
  geom_point(aes(colour=pass, shape=pass), show_guide=F) +
  scale_colour_manual(values=c(alpha('grey', 0.5), 'black')) +
  geom_hline(yintercept=-log(pval, 10), colour="black", linetype=2) +
  geom_vline(xintercept=-log2(afc), colour="black", linetype=2) +
  geom_vline(xintercept=log2(afc), colour="black", linetype=2) +
  ggtitle("Full RNA Differential Expression: PSS - Control") +
  theme_bw()

ggplot(unfiltered_globin_in, aes(x=logFC, y=-log(adj.P.Val, 10))) +
  geom_point(aes(colour=pass, shape=pass), show_guide=F) +
  scale_colour_manual(values=c(alpha('grey', 0.5), 'black')) +
  geom_hline(yintercept=-log(pval, 10), colour="black", linetype=2) +
  geom_vline(xintercept=-log2(afc), colour="black", linetype=2) +
  geom_vline(xintercept=log2(afc), colour="black", linetype=2) +
  ggtitle("Globin Clear Differential Expression: PSS - Control") +
  theme_bw()
##'-----------------------------------------------------------------------------------------#
```

S2 R scripts for microarray data analysis of the whole blood gene expression signature in pSS-associated lymphoma

```
filename            <- "Fai_Sample_Probe_Profile.txt"
raw_data            <- lumiR(filename)
##'Gender Discrepancies
raw_data            <- raw_data[, -c(2, 6)]
pheno_table          <- read.table("sample_info.txt",
                          header=T,
                          sep="\t",
                          stringsAsFactors=F)
rownames(pheno_table) <- pheno_table$Sentrix_ID
pData(raw_data)       <- pheno_table
##'-------------------------------------------------------------------------------------------------------#

##"'IPS-004-1", - Detected Outlier
removal      <- c("IPS-004-1", "IPS-002-1", "NCL-130-0", "SWI-084-0", "LEE-060-0", "NCL-113-
0", "TOR-006-1")
removal_pos <- match(rownames(pData(raw_data)[pData(raw_data)$SampleID %in% removal,]),
colnames(raw_data))
raw_data_in <- raw_data[,-removal_pos]

##'RIN SCORE EXCLUSION
##'Arrays Less than 7
Seven      <-      c("LEE-060-0","BIR-033-1","NCL-097-0","NCL-007-1","TOR-006-1","BIR-039-
1","LEE-034-1",
         "LEE-062-1","TOR-007-0","NOT-036-1","FIF-027-1","BIR-011-1","SWI-006-1","BIR-
030-1",
         "IPS-002-1","DER-019-1","FIF-014-1","WIN-016-1","LEE-016-1","LEE-012-1","GLA-
019-1",
         "BIR-041-1","SWI-031-1","NCL-053-1")
removal_pos      <-      match(rownames(pData(raw_data_in)[pData(raw_data_in)$SampleID      %in%
Seven,]), colnames(raw_data_in))
raw_data_in <- raw_data_in[,-removal_pos]

##'Arrays Less than 7 with lymphoma
SevenL <- c("LEE-060-0", "BIR-033-1", "NCL-097-0", "NCL-007-1", "TOR-006-1", "BIR-039-1",
"LEE-034-1",
          "LEE-062-1", "TOR-007-0", "BIR-011-1", "SWI-006-1", "BIR-030-1",
          "IPS-002-1",  "DER-019-1",  "FIF-014-1",  "WIN-016-1",  "LEE-016-1",  "LEE-012-1",
"GLA-019-1",
          "BIR-041-1", "SWI-031-1", "NCL-053-1")
removal_pos      <-      match(rownames(pData(raw_data_in)[pData(raw_data_in)$SampleID      %in%
SevenL,]), colnames(raw_data_in))
raw_data_in <- raw_data_in[,-removal_pos]


##'Arrays Less than 5
Five <- c("NCL-097-0", "BIR-011-1", "SWI-006-1", "BIR-041-1", "SWI-031-1", "NCL-053-1")
removal_pos      <-      match(rownames(pData(raw_data_in)[pData(raw_data_in)$SampleID      %in%
Five,]), colnames(raw_data_in))
raw_data_in <- raw_data_in[,-removal_pos]

##'BATCH EXCLUSION
raw_data_in <- raw_data_in[,-grep(1, pData(raw_data_in)$Batch)]
pData(raw_data_in)$Batch <- pData(raw_data_in)$Batch - 1
```

```
##'------------------------------------------------------------------------------------------------------------#

det <- melt(detection(raw_data))
dtp <- ggplot(data=det, aes(x=Var2, y=value)) +
  geom_boxplot(outlier.size=0.5,
               size=0.2) +
  scale_x_discrete(name="") +
  scale_y_continuous(name="Amplitude") +
  theme_bw() +
  theme(axis.text.x=element_text(angle=90,
                                 vjust=0.5,
                                 size=6),
        axis.text.y=element_text(size=6),
        axis.title.y=element_text(size=6))

image_deploy(dtp, "DetectionPval_")
##'------------------------------------------------------------------------------------------------------------#
vst_data         <- lumiT(raw_data_in, method='vst')
rsn_data         <- lumiN(vst_data, method = "rsn")
lumi.Q           <- lumiQ(rsn_data)

##'Detection Threshold Filtering
exprs_data       <- exprs(lumi.Q)
present_count    <- detectionCall(lumi.Q)
normalised_data  <- exprs_data[present_count > 0, ]
##'------------------------------------------------------------------------------------------------------------#
batches            <- pData(raw_data_in)$Batch
pheno              <- data.frame(sample=c(1:ncol(normalised_data)),
                                 outcome=as.factor(pData(lumi.Q)$Group),
                                 batch=batches)
rownames(pheno)    <- colnames(normalised_data)
batch              <- pheno$batch
mod                <- model.matrix(~as.factor(outcome), data = pheno)
batchCorrected_data <- ComBat(dat=normalised_data,
                              batch=batch,
                              mod=mod,
                              par.prior=T,
                              prior.plots=F)
##'------------------------------------------------------------------------------------------------------------#
probe_list       <- rownames(batchCorrected_data)
nuIDs            <- probeID2nuID(probe_list)[, "nuID"]
symbol           <- getSYMBOL(nuIDs, "lumiHumanAll.db")
name             <- unlist(lookUp(nuIDs, "lumiHumanAll.db", "GENENAME"))
anno_df          <- data.frame(ID = nuIDs, probe_list, symbol, name)
entrez_map       <- data.frame(nuID=as.vector(anno_df$ID),
                               EntrezID=nuID2EntrezID(as.vector(anno_df$ID),
                                                      "lumiHumanIDMapping"))
##'------------------------------------------------------------------------------------------------------------#



treatments          <- unique(pData(raw_data_in)$Group)
```

```
treatment_arrays    <- pData(raw_data_in)$Group
batchCorrected_data <- data.matrix(batchCorrected_data)
design              <- model.matrix(~0 + factor(treatment_arrays, levels = treatments))
colnames(design)    <- treatments
num_parameters      <- ncol(design)
fit                 <- lmFit(batchCorrected_data, design)

cont_mat            <- makeContrasts(SS-Control,
                                     Lymphoma-Control,
                                     Cancer-Control,
                                     PreMalignancy-Control,
                                     SS-Lymphoma,
                                     SS-Cancer,
                                     SS-PreMalignancy,
                                     Lymphoma-Cancer,
                                     Lymphoma-PreMalignancy,
                                     Cancer-PreMalignancy,
                                     levels=treatments)
fit2                <- contrasts.fit(fit, contrasts=cont_mat)
fit2                <- eBayes(fit2)
##'-----------------------------------------------------------------------------------------#
comparisons <- c("SS - Control", "Lymphoma - Control", "Cancer - Control",
                 "PreMalignancy - Control", "SS - Lymphoma", "SS - Cancer",
                 "SS - PreMalignancy", "Lymphoma - Cancer", "Lymphoma - PreMalignancy",
                 "Cancer - PreMalignancy")
p_cut_off   <- 0.05
fold_change <- 1.2
i           <- 5

gene_list_unfiltered <- topTable(fit2,
                                 coef=comparisons[i],
                                 number=Inf,
                                 adjust.method="BH")

gene_list            <- topTable(fit2,
                                 coef=comparisons[i],
                                 p.value=p_cut_off,
                                 lfc=log2(fold_change),
                                 number=Inf,
                                 adjust.method="BH")
##'-----------------------------------------------------------------------------------------#
```

S3 The qRT-PCR data for the remaining validated genes tested in the validation cohort*. The table includes p values, fold changes and regula**tion in pSS-associated lymphoma.** Fc was calculated according to the housekeeping gene normalisation*

Gene symbol	Gene name	p value	Fold change and regulation
CNPY3	Canopy FGF signaling regulator 3	0.163	1.09 ↓
CYFIP2	Cytoplasmic FMR1 interacting protein 2	0.1949	1.11 ↑
ESYT1	Extended synaptotagmin-like protein 1	0.1098	1.26 ↑
HNRNPUL1	Heterogeneous nuclear ribonucleoprotein U-like 1	0.3141	1.07 ↑
MAGED1	Melanoma antigen family D, 1	0.0617	1.30 ↑
POM121C	POM121 transmembrane nucleoporin C	0.2661	1.00 ↑
PRPF8	Pre-mRNA processing factor 8	0.0515	1.19 ↑
RBP7	Retinol binding protein 7, cellular	0.6217	1.24 ↓
SEC61G	Sec61 gamma subunit	0.0248	1.31 ↓
SF3A1	Splicing factor 3a, subunit 1, 120kDa	0.0005	1.39 ↑
SRP14	Signal recognition particle 14kDa (homologous Alu RNA binding protein)	0.0185	1.35 ↓
UBXN11	UBX domain protein 11	0.4145	1.10 ↓
VCP	Valosin containing protein	0.037	1.20 ↑

S4 The qRT-PCR data for the remaining validated genes tested in the untreated pSS-associated lymphoma*. The table includes p values, fold changes and regulation in pSS-associated lymphoma. Fc was calculated according to the housekeeping gene normalisation*

Gene symbol	Gene name	p value	Fold change and regulation
BMS1	BMS1 ribosome biogenesis factor	0.1695	1.17 ↓
C10orf32	Chromosome 10 open reading frame 32	0.8647	1.06 ↓
CBLL1	Cbl proto-oncogene-like 1, E3 ubiquitin protein ligase	0.0884	1.41 ↑
CNPY3	Canopy FGF signaling regulator 3	0.9406	1.10 ↓
CYFIP2	Cytoplasmic FMR1 interacting protein 2	0.5159	1.08 ↑
ESYT1	Extended synaptotagmin-like protein 1	0.5652	1.13 ↑
HNRNPUL1	Heterogeneous nuclear ribonucleoprotein U-like 1	0.2773	1.08 ↑
LEF1	Lymphoid enhancer-binding factor 1	0.1359	1.21 ↓
LGALS1	Lectin, galactoside-binding, soluble, 1	0.101	1.30 ↑
MAGED1	Melanoma antigen family D, 1	0.6701	1.03 ↑
MGST3	Microsomal glutathione S-transferase 3	0.0526	1.43 ↑
OAF	OAF homolog (Drosophila)	0.0864	1.31 ↑
POM121C	POM121 transmembrane nucleoporin C	0.2371	1.31 ↑
PRPF8	Pre-mRNA processing factor 8	0.9067	1.12 ↓
SEC61G	Sec61 gamma subunit	0.628	1.05 ↑
SGK223	Homolog of rat pragma of Rnd2	0.5298	1.32 ↑
SRP14	Signal recognition particle 14kDa (homologous Alu RNA binding protein)	1.0000	1.22 ↑
UBXN11	UBX domain protein 11	0.8899	1.08 ↓

S5 The canonical pathways in pSS-associated lymphoma. The table represents the canonical pathways in pSS vs pSS-associated lymphoma

Ingenuity Canonical Pathways	-log(p-value)	Ratio	z-score	Overlapped downregulated genes	Overlapped upregulated genes	Molecules
Aryl Hydrocarbon Receptor Signaling	3.99E+00	3.70E-02	NaN	1/135 (1%)	4/135 (3%)	*MYC,ALDH9A1,RBL2,HSP90B1,MGST3*
Histamine Degradation	3.12E+00	1.67E-01	NaN	1/12 (8%)	1/12 (8%)	*HNMT,ALDH9A1*
Unfolded protein response	3.09E+00	5.66E-02	NaN	0/53 (0%)	3/53 (6%)	*HSPA9,HSP90B1,VCP*
Neuregulin Signaling	2.50E+00	3.53E-02	NaN	0/85 (0%)	3/85 (4%)	*MYC,HSP90B1,PRKCQ*
T Cell Receptor Signaling	2.38E+00	3.19E-02	NaN	0/94 (0%)	3/94 (3%)	*PRKCQ,ITK,RASGRP1*
Telomerase Signaling	2.35E+00	3.12E-02	NaN	0/96 (0%)	3/96 (3%)	*MYC,HSP90B1,ETS1*
iCOS-iCOSL Signaling in T Helper Cells	2.33E+00	3.06E-02	NaN	0/98 (0%)	3/98 (3%)	*HLA-DRB1,PRKCQ,ITK*
CD28 Signaling in T Helper Cells	2.21E+00	2.78E-02	NaN	0/108 (0%)	3/108 (3%)	*HLA-DRB1,PRKCQ,ITK*
Molecular Mechanisms of Cancer	2.10E+00	1.39E-02	NaN	0/359 (0%)	5/359 (1%)	*NCSTN,MYC,PRKCQ,LEF1,RASGRP1*
Thyroid Cancer Signaling	2.10E+00	5.13E-02	NaN	0/39 (0%)	2/39 (5%)	*MYC,LEF1*
L-carnitine Biosynthesis	1.99E+00	3.33E-01	NaN	0/3 (0%)	1/3 (33%)	*ALDH9A1*
eNOS Signaling	1.94E+00	2.22E-02	NaN	0/135 (0%)	3/135 (2%)	*HSPA9,HSP90B1,PRKCQ*
Xenobiotic Metabolism Signaling	1.93E+00	1.56E-02	NaN	1/256 (0%)	3/256 (1%)	*ALDH9A1,HSP90B1,PRKCQ,MGST3*
Retinoate Biosynthesis II	1.86E+00	2.50E-01	NaN	1/4 (25%)	0/4 (0%)	*RBP7*
Endometrial Cancer Signaling	1.86E+00	3.85E-02	NaN	0/52 (0%)	2/52 (4%)	*MYC,LEF1*
Calcium-induced T Lymphocyte Apoptosis	1.84E+00	3.77E-02	NaN	0/53 (0%)	2/53 (4%)	*HLA-DRB1,PRKCQ*
Aldosterone Signaling in Epithelial Cells	1.81E+00	1.99E-02	NaN	0/151 (0%)	3/151 (2%)	*HSPA9,HSP90B1,PRKCQ*
Thrombopoietin Signaling	1.81E+00	3.64E-02	NaN	0/55 (0%)	2/55 (4%)	*MYC,PRKCQ*
ErbB4 Signaling	1.77E+00	3.45E-02	NaN	0/58 (0%)	2/58 (3%)	*NCSTN,PRKCQ*
Role of NFAT in Regulation of the Immune Response	1.75E+00	1.87E-02	NaN	0/160 (0%)	3/160 (2%)	*HLA-DRB1,PRKCQ,ITK*
Cell Cycle: G1/S Checkpoint Regulation	1.70E+00	3.17E-02	NaN	0/63 (0%)	2/63 (3%)	*MYC,RBL2*
NRF2-mediated Oxidative Stress Response	1.63E+00	1.69E-02	NaN	1/177 (1%)	2/177 (1%)	*PRKCQ,VCP,MGST3*
Regulation of the Epithelial-Mesenchymal Transition Pathway	1.61E+00	1.66E-02	NaN	0/181 (0%)	3/181 (2%)	*NCSTN,LEF1,ETS1*
Prolactin Signaling	1.58E+00	2.74E-02	NaN	0/73 (0%)	2/73 (3%)	*MYC,PRKCQ*
RAR Activation	1.57E+00	1.60E-02	NaN	1/187 (1%)	2/187 (1%)	*RBP7,PRKCQ,SMARCA2*
Acute Myeloid Leukemia Signaling	1.54E+00	2.60E-02	NaN	0/77 (0%)	2/77 (3%)	*MYC,LEF1*
Leukocyte Extravasation Signaling	1.54E+00	1.55E-02	NaN	0/193 (0%)	3/193 (2%)	*PRKCQ,ITK,RASGRP1*
Role of BRCA1 in DNA Damage Response	1.53E+00	2.56E-02	NaN	0/78 (0%)	2/78 (3%)	*RBL2,SMARCA2*
Prostate Cancer Signaling	1.51E+00	2.50E-02	NaN	0/80 (0%)	2/80 (3%)	*HSP90B1,LEF1*

Factors Promoting Cardiogenesis in Vertebrates	1.42E+00	2.25E-02	NaN	0/89 (0%)	2/89 (2%)	*PRKCQ,LEF1*
Chronic Myeloid Leukemia Signaling	1.39E+00	2.17E-02	NaN	0/92 (0%)	2/92 (2%)	*MYC,RBL2*
Assembly of RNA Polymerase III Complex	1.39E+00	8.33E-02	NaN	0/12 (0%)	1/12 (8%)	*SF3A1*
Glioma Signaling	1.38E+00	2.13E-02	NaN	0/94 (0%)	2/94 (2%)	*RBL2,PRKCQ*
Mouse Embryonic Stem Cell Pluripotency	1.38E+00	2.13E-02	NaN	0/94 (0%)	2/94 (2%)	*MYC,LEF1*
Nitric Oxide Signaling in the Cardiovascular System	1.37E+00	2.11E-02	NaN	0/95 (0%)	2/95 (2%)	*HSP90B1,PRKCQ*
Huntington's Disease Signaling	1.36E+00	1.33E-02	NaN	0/226 (0%)	3/226 (1%)	*HSPA9,PRKCQ,SDHA*
HGF Signaling	1.30E+00	1.92E-02	NaN	0/104 (0%)	2/104 (2%)	*PRKCQ,ETS1*
The Visual Cycle	1.30E+00	6.67E-02	NaN	1/15 (7%)	0/15 (0%)	*RBP7*
Fatty Acid ¿±-oxidation	1.30E+00	6.67E-02	NaN	0/15 (0%)	1/15 (7%)	*ALDH9A1*
Oxidative Ethanol Degradation III	1.30E+00	6.67E-02	NaN	0/15 (0%)	1/15 (7%)	*ALDH9A1*
PKC¿¸ Signaling in T Lymphocytes	1.27E+00	1.85E-02	NaN	0/108 (0%)	2/108 (2%)	*HLA-DRB1,PRKCQ*
Putrescine Degradation III	1.27E+00	6.25E-02	NaN	0/16 (0%)	1/16 (6%)	*ALDH9A1*
Tryptophan Degradation X (Mammalian, via Tryptamine)	1.24E+00	5.88E-02	NaN	0/17 (0%)	1/17 (6%)	*ALDH9A1*
Protein Ubiquitination Pathway	1.24E+00	1.18E-02	NaN	1/254 (0%)	2/254 (1%)	*HSPA9,HSP90B1,PSMC1*
Glutathione Redox Reactions I	1.22E+00	5.56E-02	NaN	1/18 (6%)	0/18 (0%)	*MGST3*
P2Y Purigenic Receptor Signaling Pathway	1.20E+00	1.69E-02	NaN	0/118 (0%)	2/118 (2%)	*MYC,PRKCQ*
phagosome maturation	1.20E+00	1.69E-02	NaN	1/118 (1%)	1/118 (1%)	*HLA-DRB1,DYNLL1*
Ethanol Degradation IV	1.20E+00	5.26E-02	NaN	0/19 (0%)	1/19 (5%)	*ALDH9A1*
D-myo-inositol (1,4,5,6)-Tetrakisphosphate Biosynthesis	1.18E+00	1.64E-02	NaN	1/122 (1%)	1/122 (1%)	*ATP1A1,NUDT14*
D-myo-inositol (3,4,5,6)-tetrakisphosphate Biosynthesis	1.18E+00	1.64E-02	NaN	1/122 (1%)	1/122 (1%)	*ATP1A1,NUDT14*
Glucocorticoid Receptor Signaling	1.18E+00	1.11E-02	NaN	0/270 (0%)	3/270 (1%)	*HSPA9,HSP90B1,SMARCA2*
Cdc42 Signaling	1.17E+00	1.63E-02	NaN	0/123 (0%)	2/123 (2%)	*HLA-DRB1,ITK*
Dopamine Degradation	1.15E+00	4.76E-02	NaN	0/21 (0%)	1/21 (5%)	*ALDH9A1*
Endoplasmic Reticulum Stress Pathway	1.15E+00	4.76E-02	NaN	0/21 (0%)	1/21 (5%)	*HSP90B1*
Polyamine Regulation in Colon Cancer	1.13E+00	4.55E-02	NaN	0/22 (0%)	1/22 (5%)	*MYC*
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	1.12E+00	1.06E-02	NaN	0/284 (0%)	3/284 (1%)	*MYC,PRKCQ,LEF1*
B Cell Development	1.12E+00	4.35E-02	NaN	0/23 (0%)	1/23 (4%)	*HLA-DRB1*
TCA Cycle II (Eukaryotic)	1.12E+00	4.35E-02	NaN	0/23 (0%)	1/23 (4%)	*SDHA*
Estrogen-mediated S-phase Entry	1.10E+00	4.17E-02	NaN	0/24 (0%)	1/24 (4%)	*MYC*

Glutathione-mediated Detoxification	1.10E+00	4.17E-02	NaN	1/24 (4%)	0/24 (0%)	*MGST3*
D-myo-inositol-5-phosphate Metabolism	1.08E+00	1.45E-02	NaN	1/138 (1%)	1/138 (1%)	*ATP1A1,NUDT14*
3-phosphoinositide Degradation	1.08E+00	1.44E-02	NaN	1/139 (1%)	1/139 (1%)	*ATP1A1,NUDT14*
Cell Cycle Control of Chromosomal Replication	1.05E+00	3.70E-02	NaN	0/27 (0%)	1/27 (4%)	*RPA2*
Glioblastoma Multiforme Signaling	1.05E+00	1.38E-02	NaN	0/145 (0%)	2/145 (1%)	*MYC,LEF1*
3-phosphoinositide Biosynthesis	1.03E+00	1.34E-02	NaN	1/149 (1%)	1/149 (1%)	*ATP1A1,NUDT14*
Tec Kinase Signaling	1.02E+00	1.33E-02	NaN	0/150 (0%)	2/150 (1%)	*PRKCQ,ITK*
Role of p14/p19ARF in Tumor Suppression	1.02E+00	3.45E-02	NaN	0/29 (0%)	1/29 (3%)	*SF3A1*
Mitochondrial Dysfunction	1.01E+00	1.32E-02	NaN	0/152 (0%)	2/152 (1%)	*NCSTN,SDHA*
Retinoate Biosynthesis I	1.01E+00	3.33E-02	NaN	1/30 (3%)	0/30 (0%)	*RBP7*
Retinol Biosynthesis	1.01E+00	3.33E-02	NaN	1/30 (3%)	0/30 (0%)	*RBP7*
Ethanol Degradation II	1.01E+00	3.33E-02	NaN	0/30 (0%)	1/30 (3%)	*ALDH9A1*
Noradrenaline and Adrenaline Degradation	9.79E-01	3.12E-02	NaN	0/32 (0%)	1/32 (3%)	*ALDH9A1*
Autoimmune Thyroid Disease Signaling	9.54E-01	2.94E-02	NaN	0/34 (0%)	1/34 (3%)	*HLA-DRB1*
Wnt/²ç-catenin Signaling	9.49E-01	1.20E-02	NaN	0/166 (0%)	2/166 (1%)	*MYC,LEF1*
Endothelin-1 Signaling	9.45E-01	1.20E-02	NaN	0/167 (0%)	2/167 (1%)	*MYC,PRKCQ*
B Cell Receptor Signaling	9.45E-01	1.20E-02	NaN	0/167 (0%)	2/167 (1%)	*PRKCQ,ETS1*
Nucleotide Excision Repair Pathway	9.42E-01	2.86E-02	NaN	0/35 (0%)	1/35 (3%)	*RPA2*
Antigen Presentation Pathway	9.31E-01	2.78E-02	NaN	0/36 (0%)	1/36 (3%)	*HLA-DRB1*
autophagy	9.31E-01	2.78E-02	NaN	1/36 (3%)	0/36 (0%)	*ATG12*
Notch Signaling	9.20E-01	2.70E-02	NaN	0/37 (0%)	1/37 (3%)	*NCSTN*
Allograft Rejection Signaling	8.98E-01	2.56E-02	NaN	0/39 (0%)	1/39 (3%)	*HLA-DRB1*
Graft-versus-Host Disease Signaling	8.98E-01	2.56E-02	NaN	0/39 (0%)	1/39 (3%)	*HLA-DRB1*
ILK Signaling	8.88E-01	1.10E-02	NaN	0/181 (0%)	2/181 (1%)	*MYC,LEF1*
Mechanisms of Viral Exit from Host Cells	8.78E-01	2.44E-02	NaN	0/41 (0%)	1/41 (2%)	*PRKCQ*
ERK/MAPK Signaling	8.73E-01	1.08E-02	NaN	0/185 (0%)	2/185 (1%)	*MYC,ETS1*
Superpathway of Inositol Phosphate Compounds	8.69E-01	1.08E-02	NaN	1/186 (1%)	1/186 (1%)	*ATP1A1,NUDT14*
UVC-Induced MAPK Signaling	8.68E-01	2.38E-02	NaN	0/42 (0%)	1/42 (2%)	*PRKCQ*
Nur77 Signaling in T Lymphocytes	8.32E-01	2.17E-02	NaN	0/46 (0%)	1/46 (2%)	*HLA-DRB1*
nNOS Signaling in Neurons	8.32E-01	2.17E-02	NaN	0/46 (0%)	1/46 (2%)	*PRKCQ*
OX40 Signaling Pathway	8.15E-01	2.08E-02	NaN	0/48 (0%)	1/48 (2%)	*HLA-DRB1*
Amyloid Processing	7.98E-01	2.00E-02	NaN	0/50 (0%)	1/50 (2%)	*NCSTN*
LPS/IL-1 Mediated Inhibition of RXR Function	7.92E-01	9.62E-03	NaN	1/208 (0%)	1/208 (0%)	*ALDH9A1,MGST3*

UVB-Induced MAPK Signaling	7.75E-01	1.89E-02	NaN	0/53 (0%)	1/53 (2%)	*PRKCQ*
Serotonin Degradation	7.61E-01	1.82E-02	NaN	0/55 (0%)	1/55 (2%)	*ALDH9A1*
ErbB2-ErbB3 Signaling	7.54E-01	1.79E-02	NaN	0/56 (0%)	1/56 (2%)	*MYC*
Phospholipase C Signaling	7.51E-01	9.05E-03	NaN	0/221 (0%)	2/221 (1%)	*PRKCQ,ITK*
Myc Mediated Apoptosis Signaling	7.40E-01	1.72E-02	NaN	0/58 (0%)	1/58 (2%)	*MYC*
Colorectal Cancer Metastasis Signaling	7.22E-01	8.66E-03	NaN	0/231 (0%)	2/231 (1%)	*MYC,LEF1*
CCR5 Signaling in Macrophages	7.14E-01	1.61E-02	NaN	0/62 (0%)	1/62 (2%)	*PRKCQ*
T Helper Cell Differentiation	7.14E-01	1.61E-02	NaN	0/62 (0%)	1/62 (2%)	*HLA-DRB1*
ERK5 Signaling	7.14E-01	1.61E-02	NaN	0/62 (0%)	1/62 (2%)	*MYC*
GM-CSF Signaling	7.14E-01	1.61E-02	NaN	0/62 (0%)	1/62 (2%)	*ETS1*
Mitotic Roles of Polo-Like Kinase	7.07E-01	1.59E-02	NaN	0/63 (0%)	1/63 (2%)	*HSP90B1*
Pyridoxal 5'-phosphate Salvage Pathway	7.07E-01	1.59E-02	NaN	0/63 (0%)	1/63 (2%)	*PRKCQ*
Hypoxia Signaling in the Cardiovascular System	7.07E-01	1.59E-02	NaN	0/63 (0%)	1/63 (2%)	*HSP90B1*
Remodeling of Epithelial Adherens Junctions	6.89E-01	1.52E-02	NaN	0/66 (0%)	1/66 (2%)	*CBLL1*
GABA Receptor Signaling	6.89E-01	1.52E-02	NaN	0/66 (0%)	1/66 (2%)	*ALDH9A1*
Erythropoietin Signaling	6.83E-01	1.49E-02	NaN	0/67 (0%)	1/67 (1%)	*PRKCQ*
Melatonin Signaling	6.78E-01	1.47E-02	NaN	0/68 (0%)	1/68 (1%)	*PRKCQ*
Macropinocytosis Signaling	6.78E-01	1.47E-02	NaN	0/68 (0%)	1/68 (1%)	*PRKCQ*
Growth Hormone Signaling	6.72E-01	1.45E-02	NaN	0/69 (0%)	1/69 (1%)	*PRKCQ*
Renal Cell Carcinoma Signaling	6.72E-01	1.45E-02	NaN	0/69 (0%)	1/69 (1%)	*ETS1*
Basal Cell Carcinoma Signaling	6.72E-01	1.45E-02	NaN	0/69 (0%)	1/69 (1%)	*LEF1*
IL-3 Signaling	6.61E-01	1.41E-02	NaN	0/71 (0%)	1/71 (1%)	*PRKCQ*
Small Cell Lung Cancer Signaling	6.61E-01	1.41E-02	NaN	0/71 (0%)	1/71 (1%)	*MYC*
IL-4 Signaling	6.61E-01	1.41E-02	NaN	0/71 (0%)	1/71 (1%)	*HLA-DRB1*
Role of Wnt/GSK-3²خ Signaling in the Pathogenesis of Influenza	6.56E-01	1.39E-02	NaN	0/72 (0%)	1/72 (1%)	*LEF1*
LPS-stimulated MAPK Signaling	6.51E-01	1.37E-02	NaN	0/73 (0%)	1/73 (1%)	*PRKCQ*
NF-ؽخB Activation by Viruses	6.51E-01	1.37E-02	NaN	0/73 (0%)	1/73 (1%)	*PRKCQ*
STAT3 Pathway	6.51E-01	1.37E-02	NaN	0/73 (0%)	1/73 (1%)	*MYC*
Communication between Innate and Adaptive Immune Cells	6.45E-01	1.35E-02	NaN	0/74 (0%)	1/74 (1%)	*HLA-DRB1*
BMP signaling pathway	6.45E-01	1.35E-02	NaN	0/74 (0%)	1/74 (1%)	*MAGED1*
VEGF Family Ligand-Receptor Interactions	6.40E-01	1.33E-02	NaN	0/75 (0%)	1/75 (1%)	*PRKCQ*
HER-2 Signaling in Breast Cancer	6.35E-01	1.32E-02	NaN	0/76 (0%)	1/76 (1%)	*PRKCQ*
Altered T Cell and B Cell Signaling in	6.35E-01	1.32E-02	NaN	0/76 (0%)	1/76 (1%)	*HLA-DRB1*

Rheumatoid Arthritis						
VDR/RXR Activation	6.30E-01	1.30E-02	NaN	0/77 (0%)	1/77 (1%)	*PRKCQ*
PDGF Signaling	6.30E-01	1.30E-02	NaN	0/77 (0%)	1/77 (1%)	*MYC*
GPCR-Mediated Nutrient Sensing in Enteroendocrine Cells	6.02E-01	1.20E-02	NaN	0/83 (0%)	1/83 (1%)	*PRKCQ*
Adrenergic Signaling-±ż	5.93E-01	1.18E-02	NaN	0/85 (0%)	1/85 (1%)	*PRKCQ*
ErbB Signaling	5.93E-01	1.18E-02	NaN	0/85 (0%)	1/85 (1%)	*PRKCQ*
Bladder Cancer Signaling	5.89E-01	1.16E-02	NaN	0/86 (0%)	1/86 (1%)	*MYC*
G Beta Gamma Signaling	5.80E-01	1.14E-02	NaN	0/88 (0%)	1/88 (1%)	*PRKCQ*
Apoptosis Signaling	5.80E-01	1.14E-02	NaN	0/88 (0%)	1/88 (1%)	*PRKCQ*
Virus Entry via Endocytic Pathways	5.76E-01	1.12E-02	NaN	0/89 (0%)	1/89 (1%)	*PRKCQ*
Crosstalk between Dendritic Cells and Natural Killer Cells	5.76E-01	1.12E-02	NaN	0/89 (0%)	1/89 (1%)	*HLA-DRB1*
PPAR Signaling	5.72E-01	1.11E-02	NaN	0/90 (0%)	1/90 (1%)	*HSP90B1*
Salvage Pathways of Pyrimidine Ribonucleotides	5.68E-01	1.10E-02	NaN	0/91 (0%)	1/91 (1%)	*PRKCQ*
Fc ³żReceptor-mediated Phagocytosis in Macrophages and Monocytes	5.63E-01	1.09E-02	NaN	0/92 (0%)	1/92 (1%)	*PRKCQ*
Oxidative Phosphorylation	5.63E-01	1.09E-02	NaN	0/92 (0%)	1/92 (1%)	*SDHA*
Neuropathic Pain Signaling In Dorsal Horn Neurons	5.37E-01	1.01E-02	NaN	0/99 (0%)	1/99 (1%)	*PRKCQ*
phagosome formation	5.37E-01	1.01E-02	NaN	0/99 (0%)	1/99 (1%)	*PRKCQ*
Cholecystokinin/Gastrin-mediated Signaling	5.33E-01	1.00E-02	NaN	0/100 (0%)	1/100 (1%)	*PRKCQ*
Type I Diabetes Mellitus Signaling	5.29E-01	9.90E-03	NaN	0/101 (0%)	1/101 (1%)	*HLA-DRB1*
Rac Signaling	5.22E-01	9.71E-03	NaN	0/103 (0%)	1/103 (1%)	*CYFIP2*
Corticotropin Releasing Hormone Signaling	5.12E-01	9.43E-03	NaN	0/106 (0%)	1/106 (1%)	*PRKCQ*
Fc Epsilon RI Signaling	5.08E-01	9.35E-03	NaN	0/107 (0%)	1/107 (1%)	*PRKCQ*
fMLP Signaling in Neutrophils	5.08E-01	9.35E-03	NaN	0/107 (0%)	1/107 (1%)	*PRKCQ*
Natural Killer Cell Signaling	5.05E-01	9.26E-03	NaN	0/108 (0%)	1/108 (1%)	*PRKCQ*
Renin-Angiotensin Signaling	5.05E-01	9.26E-03	NaN	0/108 (0%)	1/108 (1%)	*PRKCQ*
Androgen Signaling	4.98E-01	9.09E-03	NaN	0/110 (0%)	1/110 (1%)	*PRKCQ*
CCR3 Signaling in Eosinophils	4.92E-01	8.93E-03	NaN	0/112 (0%)	1/112 (1%)	*PRKCQ*
Sperm Motility	4.89E-01	8.85E-03	NaN	0/113 (0%)	1/113 (1%)	*PRKCQ*
Type II Diabetes Mellitus Signaling	4.86E-01	8.77E-03	NaN	0/114 (0%)	1/114 (1%)	*PRKCQ*
p38 MAPK Signaling	4.83E-01	8.70E-03	NaN	0/115 (0%)	1/115 (1%)	*MYC*
Synaptic Long Term Potentiation	4.80E-01	8.62E-03	NaN	0/116 (0%)	1/116 (1%)	*PRKCQ*

14-3-3-mediated Signaling	4.80E-01	8.62E-03	NaN	0/116 (0%)	1/116 (1%)	*PRKCQ*
p70S6K Signaling	4.74E-01	8.47E-03	NaN	0/118 (0%)	1/118 (1%)	*PRKCQ*
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	4.71E-01	8.40E-03	NaN	0/119 (0%)	1/119 (1%)	*PRKCQ*
PI3K/AKT Signaling	4.68E-01	8.33E-03	NaN	0/120 (0%)	1/120 (1%)	*HSP90B1*
Hereditary Breast Cancer Signaling	4.51E-01	7.94E-03	NaN	0/126 (0%)	1/126 (1%)	*SMARCA2*
GNRH Signaling	4.48E-01	7.87E-03	NaN	0/127 (0%)	1/127 (1%)	*PRKCQ*
Ovarian Cancer Signaling	4.42E-01	7.75E-03	NaN	0/129 (0%)	1/129 (1%)	*LEF1*
IL-12 Signaling and Production in Macrophages	4.37E-01	7.63E-03	NaN	0/131 (0%)	1/131 (1%)	*PRKCQ*
Human Embryonic Stem Cell Pluripotency	4.37E-01	7.63E-03	NaN	0/131 (0%)	1/131 (1%)	*LEF1*
Protein Kinase A Signaling	4.36E-01	5.41E-03	NaN	0/370 (0%)	2/370 (1%)	*PRKCQ,LEF1*
Synaptic Long Term Depression	4.27E-01	7.41E-03	NaN	0/135 (0%)	1/135 (1%)	*PRKCQ*
Epithelial Adherens Junction Signaling	4.07E-01	6.99E-03	NaN	0/143 (0%)	1/143 (1%)	*LEF1*
Gζ±q Signaling	4.03E-01	6.90E-03	NaN	0/145 (0%)	1/145 (1%)	*PRKCQ*
CXCR4 Signaling	3.89E-01	6.62E-03	NaN	0/151 (0%)	1/151 (1%)	*PRKCQ*
Gap Junction Signaling	3.89E-01	6.62E-03	NaN	0/151 (0%)	1/151 (1%)	*PRKCQ*
Dopamine-DARPP32 Feedback in cAMP Signaling	3.76E-01	6.37E-03	NaN	0/157 (0%)	1/157 (1%)	*PRKCQ*
Hepatic Cholestasis	3.74E-01	6.33E-03	NaN	0/158 (0%)	1/158 (1%)	*PRKCQ*
Dendritic Cell Maturation	3.68E-01	6.21E-03	NaN	0/161 (0%)	1/161 (1%)	*HLA-DRB1*
NF-ΚB Signaling	3.62E-01	6.10E-03	NaN	0/164 (0%)	1/164 (1%)	*PRKCQ*
PPARζ±/RXRζ± Activation	3.60E-01	6.06E-03	NaN	0/165 (0%)	1/165 (1%)	*HSP90B1*
Acute Phase Response Signaling	3.56E-01	5.99E-03	NaN	1/167 (1%)	0/167 (0%)	*RBP7*
CREB Signaling in Neurons	3.52E-01	5.92E-03	NaN	0/169 (0%)	1/169 (1%)	*PRKCQ*
Role of NFAT in Cardiac Hypertrophy	3.41E-01	5.71E-03	NaN	0/175 (0%)	1/175 (1%)	*PRKCQ*
AMPK Signaling	3.39E-01	5.68E-03	NaN	0/176 (0%)	1/176 (1%)	*SMARCA2*
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	3.34E-01	5.59E-03	NaN	0/179 (0%)	1/179 (1%)	*PRKCQ*
mTOR Signaling	3.29E-01	5.49E-03	NaN	0/182 (0%)	1/182 (1%)	*PRKCQ*
IL-8 Signaling	3.27E-01	5.46E-03	NaN	0/183 (0%)	1/183 (1%)	*PRKCQ*
Thrombin Signaling	3.20E-01	5.35E-03	NaN	0/187 (0%)	1/187 (1%)	*PRKCQ*
Systemic Lupus Erythematosus Signaling	3.19E-01	5.32E-03	NaN	0/188 (0%)	1/188 (1%)	*PRPF8*
Breast Cancer Regulation by Stathmin1	3.15E-01	5.26E-03	NaN	0/190 (0%)	1/190 (1%)	*PRKCQ*
Actin Cytoskeleton Signaling	2.85E-01	4.76E-03	NaN	0/210 (0%)	1/210 (0%)	*CYFIP2*
Role of Osteoblasts, Osteoclasts and	2.80E-01	4.67E-03	NaN	0/214 (0%)	1/214 (0%)	*LEF1*

Chondrocytes in Rheumatoid Arthritis						
G-Protein Coupled Receptor Signaling	2.31E-01	3.94E-03	NaN	0/254 (0%)	1/254 (0%)	*RASGRP1*

S6 DEGs list in pSS vs. healthy control. *The (-) sign in FC column represents the downregulated genes in pSS-associated lymphoma (p<0.02, FC cut off=1.2). NAs excluded*

Gene symbol	Adjusted p value	FC	Gene symbol	Adjusted p value	FC
IFI27	6.1203E-07	11.51	*IRF7*	1.1052E-07	1.99
IFI44L	9.8518E-11	8.75	*OAS1*	0.00286509	1.98
ISG15	2.1509E-10	5.63	*LAP3*	1.5486E-06	1.96
IFIT1	9.8518E-11	5.60	*PARP9*	2.3007E-08	1.96
RSAD2	1.6184E-10	5.39	*OAS2*	2.485E-07	1.93
IFI44	1.0004E-09	5.23	*USP18*	1.027E-05	1.92
OAS3	8.2797E-10	4.35	*PARP14*	8.0137E-09	1.92
EPSTI1	9.8518E-11	4.29	*OAS2*	1.2484E-07	1.91
IFIT3	9.8518E-11	4.28	*HERC6*	2.1619E-06	1.89
IFITM3	6.0553E-09	4.09	*HELZ2*	1.8633E-07	1.87
HERC5	2.1509E-10	4.04	*IFIH1*	1.1189E-06	1.86
LY6E	1.6081E-08	3.98	*TIMM10*	0.00141889	1.86
IFI6	1.564E-10	3.64	*PARP12*	5.9388E-07	1.83
OAS1	1.1626E-09	3.61	*SPATS2L*	3.8514E-06	1.80
MX1	1.3295E-09	3.56	*STAT1*	4.7985E-08	1.79
OASL	1.564E-10	3.45	*SCO2*	1.8633E-07	1.79
OAS1	2.278E-09	3.33	*MT1A*	1.0666E-05	1.79
XAF1	1.1351E-10	3.19	*DDX60*	1.8622E-06	1.78
IFIT3	1.1052E-07	3.05	*HLA-DRB4*	0.00011162	1.78
XAF1	8.9622E-09	3.01	*STAT2*	2.8561E-08	1.77
OAS2	1.1833E-07	2.95	*RTP4*	7.1145E-07	1.77
IFIT3	3.2097E-08	2.84	*TNFAIP6*	0.00337936	1.76
IFIT2	1.1833E-07	2.70	*PARP9*	3.9313E-08	1.75
OTOF	0.00182888	2.53	*OASL*	6.2852E-08	1.75
EIF2AK2	6.4838E-09	2.43	*BATF2*	1.5139E-06	1.74
SERPING1	6.3208E-08	2.37	*CEACAM1*	3.7958E-07	1.72
SAMD9L	8.2597E-10	2.31	*STAT1*	1.0171E-08	1.71
GBP1	1.7593E-07	2.28	*STAT1*	1.5486E-06	1.69
HLA-DRB6	0.01348724	2.26	*UBE2L6*	3.2965E-07	1.68
IRF7	6.2852E-08	2.19	*DHX58*	1.2289E-06	1.67
ZBP1	4.7985E-08	2.12	*GBP1P1*	5.1765E-07	1.63
IFI44L	2.2601E-06	2.11	*FBXO6*	3.3811E-07	1.63
TRIM22	3.3867E-08	2.10	*IFITM1*	1.1052E-07	1.63
HES4	2.7065E-05	2.09	*IFI16*	3.3811E-07	1.60
IFI6	3.9313E-08	2.07	*MX2*	1.815E-05	1.60
OAS1	7.1042E-07	2.06	*BST2*	8.0137E-09	1.60
IFI35	9.3684E-09	2.04	*CEACAM1*	8.0804E-07	1.58
GBP1	1.1189E-06	2.03	*TNFSF10*	4.2476E-05	1.56
MT2A	1.9606E-06	2.03	*FCGR1A*	0.00599652	1.55
GBP5	8.0916E-06	1.99	*LAMP3*	2.5802E-07	1.54

SHISA5	9.1194E-06	1.54	*TOR1B*	0.00017337	1.36
DHRS9	0.00172583	1.54	*APOBEC3A*	0.02503283	1.36
FCGR1B	0.01085872	1.53	*HSH2D*	0.00067639	1.36
TNFSF13B	0.00019361	1.53	*SP110*	0.00056916	1.35
OAS3	6.4975E-07	1.53	*TYMP*	0.00218665	1.35
GBP4	0.00020481	1.53	*CXCL10*	0.02057559	1.35
CHMP5	0.01816947	1.53	*C3AR1*	4.2476E-05	1.35
TMEM140	0.00032783	1.52	*TRIM69*	9.3418E-06	1.34
PARP10	2.9039E-06	1.52	*CASP1*	0.00590912	1.34
SAMD9	0.00668398	1.52	*LGALS9*	9.1817E-05	1.34
DDX60L	0.00016309	1.52	*CMTR1*	2.3946E-06	1.33
PLSCR1	3.5562E-05	1.51	*PIK3AP1*	0.02021398	1.33
REC8	1.5429E-08	1.50	*C1QB*	0.04448298	1.33
DHRS9	0.01251414	1.50	*GPR1*	0.02591635	1.33
DDX58	2.2505E-06	1.49	*CASP1*	0.01348724	1.33
IFIT5	0.00056916	1.48	*ISG20*	0.00029951	1.32
ANKRD22	0.03561342	1.47	*DRAP1*	0.00037743	1.32
TAP1	0.00025813	1.47	*GBP2*	0.00681183	1.31
CCR1	0.00062612	1.46	*NT5C3A*	0.03924027	1.31
IRF9	4.1905E-08	1.46	*GCH1*	2.9457E-05	1.31
TMEM123	0.03561342	1.46	*NUB1*	0.000285	1.31
PSME2	8.4344E-07	1.45	*RHBDF2*	0.00110442	1.30
TDRD7	6.1203E-07	1.45	*FGD2*	0.02977371	1.30
TNFSF13B	0.00069558	1.45	*IFI30*	0.03504812	1.30
FCGR1C	0.00484405	1.45	*MAFB*	0.02364289	1.30
CMPK2	2.6863E-06	1.44	*LGALS9*	0.00011464	1.30
UBE2L6	0.00019071	1.44	*SMCO4*	0.00608333	1.30
PHF11	2.8262E-06	1.43	*GPBAR1*	0.04494512	1.30
SP110	9.9778E-05	1.43	*GALM*	0.0086087	1.29
LGALS3BP	0.00038232	1.42	*SP140*	0.00221738	1.29
TAP2	0.00013868	1.42	*SP110*	0.01097291	1.29
LAG3	0.01843014	1.42	*CTSL*	0.01365326	1.29
BLVRA	3.1112E-05	1.41	*TRIM38*	0.00510626	1.28
FAM46A	2.2154E-05	1.40	*C19orf66*	4.4689E-05	1.28
PHF11	0.00015576	1.39	*ADAR*	0.00399478	1.28
UNC93B1	8.7056E-05	1.39	*SAT1*	0.00892901	1.28
SP110	0.00043463	1.39	*IFIT1*	0.00267798	1.28
MOV10	8.5572E-06	1.38	*PSMB9*	0.02475062	1.28
TRIM5	0.00013588	1.38	*OAS2*	0.00012809	1.27
VAMP5	0.00127331	1.38	*SCARB2*	0.00056916	1.27
CEACAM1	4.2476E-05	1.36	*LOC100128274*	0.00587583	1.27
IL1RN	0.02859237	1.36	*ANKFY1*	0.00011731	1.27
TYMP	0.00290922	1.36	*SEPT4*	0.01100047	1.27
ADAR	3.3401E-05	1.36	*RHBDF2*	0.00243183	1.26

SP100	0.00011464	1.26	*IMPA2*	0.00518347	-1.39
TRAFD1	0.00267487	1.26	*GPR162*	0.02652862	-1.37
CTSL	0.00854154	1.26	*PYGL*	0.03946276	-1.35
TRIM21	0.00217263	1.26	*FAM212B*	0.00969497	-1.32
HLA-A	0.0165758	1.26	*PPM1F*	0.0006649	-1.31
HLA-F	0.03282891	1.25	*RPL3*	0.00202947	-1.30
PSMB9	0.00165278	1.25	*MIR181A2HG*	0.03875588	-1.29
C1orf85	0.02995833	1.25	*RPS5*	0.01348724	-1.28
TRIM25	0.01463022	1.25	*TBL1X*	0.01551496	-1.27
MT1G	0.00087699	1.25	*RXRA*	0.02023698	-1.27
TYMP	0.02107649	1.24	*WLS*	0.04310778	-1.27
ETV7	0.00012809	1.24	*EIF3L*	0.00186693	-1.26
TRIM6	0.00056916	1.24	*EIF3L*	0.00026111	-1.26
BTN3A1	0.02976317	1.24	*IMPDH1*	0.011807	-1.26
LYSMD2	0.02394556	1.24	*FAM101B*	0.00902316	-1.26
LMO2	0.02672869	1.24	*IRS2*	0.04448298	-1.26
TYMP	0.03490352	1.24	*EEF1G*	0.04786736	-1.25
ZC3HAV1	0.00280296	1.24	*TBC1D14*	0.00376538	-1.25
PCK2	0.00025813	1.23	*RPS3*	0.0101884	-1.25
LHFPL2	0.04327692	1.23	*ICAM3*	0.00881026	-1.25
REC8	1.0666E-05	1.23	*CCNY*	0.00027901	-1.25
RNF213	0.04533332	1.23	*RPS14*	0.03669829	-1.25
DUSP5	0.00235267	1.23	*EEF2*	0.00188502	-1.24
ZC3HAV1	0.00787496	1.22	*RPS6KA5*	0.01089737	-1.24
ACOT9	0.00620092	1.22	*ZNF746*	0.00280527	-1.24
MT1IP	0.00362303	1.22	*EIF4B*	0.00071312	-1.23
TRIM5	7.4935E-05	1.21	*FEZ1*	0.00890452	-1.23
SOCS1	0.00255775	1.21	*RASSF2*	0.02995833	-1.23
ANKFY1	0.00045049	1.21	*RPS14*	0.03715746	-1.23
CD38	0.03927095	1.21	*RPS8*	0.04823418	-1.22
MYD88	0.01463022	1.21	*CCNJL*	0.01884647	-1.22
HLA-DRA	0.03421476	1.21	*SCAP*	0.02756616	-1.21
TRIM56	0.00404656	1.21	*ARRB1*	0.03875588	-1.21
GADD45B	0.00532618	1.21	*RPL13A*	0.02685321	-1.21
CCL2	0.00349719	1.21	*PTOV1*	0.00205049	-1.21
BTN3A3	0.00510626	1.21	*TP53INP2*	0.00344213	-1.21
SP100	0.00100736	1.20	*TPRG1L*	0.00188502	-1.20
MYOF	0.03504812	1.20	*DSC1*	0.00023024	-1.20
ODF3B	0.00058825	1.20			
EPB41L3	0.01796162	1.20			
ATF3	0.00029951	1.20			
HLA-DRB4	0.02191445	-2.24			
MYOM2	0.00632069	-1.84			
TXNDC12	0.00356056	-1.46			

S7 DEGs list in pSS-paraproteinemia vs. healthy control. *The (-) sign in FC column represents the downregulated genes in pSS-associated lymphoma (p<0.02, FC cut off =1.2). NAs excluded*

Gene symbol	Adjusted p value	FC	Gene symbol	Adjusted p value	FC
IFI27	0.00078165	9.26	*PARP14*	0.0000452	1.82
IFI44L	0.00000159	7.62	*USP18*	0.00393644	1.82
IFIT1	0.00000159	5.09	*MT1A*	0.00082371	1.80
ISG15	0.00000481	4.88	*TNFAIP6*	0.02834293	1.80
RSAD2	0.00000481	4.61	*OAS2*	0.00069217	1.80
IFI44	0.0000683	4.05	*TIMM10*	0.04089623	1.79
IFIT3	0.00000159	3.82	*LAP3*	0.00450693	1.76
IFITM3	0.000018	3.79	*GBP1*	0.02173078	1.76
OAS3	0.000015	3.76	*SCO2*	0.00012054	1.76
HERC5	0.00000444	3.66	*OAS2*	0.00092249	1.73
IFI6	0.00000147	3.66	*MX2*	0.0000952	1.73
EPSTI1	0.00000458	3.63	*TRIM22*	0.00502449	1.71
LY6E	0.0000627	3.59	*PARP9*	0.00142196	1.69
MX1	0.000005	3.39	*OASL*	0.00014469	1.67
OAS1	0.00000938	3.29	*STAT2*	0.00017389	1.67
OASL	0.00000159	3.27	*SPATS2L*	0.00469705	1.66
OAS1	0.00000798	3.15	*BATF2*	0.00208978	1.63
XAF1	0.00000798	3.00	*DHX58*	0.00062508	1.63
XAF1	0.00000274	2.93	*STAT1*	0.00117505	1.61
OAS2	0.00017389	2.76	*GBP1*	0.04749386	1.60
IFIT3	0.00165209	2.48	*DDX60*	0.0086334	1.59
EIF2AK2	0.00000444	2.48	*BST2*	0.00000798	1.59
IFIT3	0.0007451	2.37	*TNFSF10*	0.00142196	1.59
IFIT2	0.00196135	2.24	*RTP4*	0.00471405	1.59
HES4	0.00070573	2.19	*PARP10*	0.0000795	1.58
OAS1	0.01079063	2.15	*TNFSF13B*	0.00225167	1.58
IRF7	0.0000931	2.10	*IFITM1*	0.00023134	1.56
IRF7	0.00000938	2.09	*STAT1*	0.00047068	1.56
SERPING1	0.00068254	2.09	*PARP9*	0.00215663	1.54
SAMD9L	0.0000261	2.08	*IFIH1*	0.03874027	1.53
IFI44L	0.0012519	2.00	*CEACAM1*	0.00803466	1.51
OAS1	0.00057187	1.97	*FBXO6*	0.00185527	1.51
ZBP1	0.00019114	1.97	*UBE2L6*	0.00502449	1.50
IFI6	0.0000927	1.97	*SHISA5*	0.00293771	1.49
MT2A	0.00076998	1.96	*IFI16*	0.00211597	1.48
IFI35	0.0000348	1.95	*OAS3*	0.00069217	1.48
HELZ2	0.0000703	1.87	*DDX58*	0.0005829	1.48
HERC6	0.00066256	1.85	*TNFSF13B*	0.00890402	1.47
PARP12	0.00012054	1.84	*CEACAM1*	0.00421342	1.46

REC8	0.0000823	1.44	*ANKFY1*	0.00443378	1.27
IRF9	0.0000627	1.44	*PSMB9*	0.01873972	1.26
FAM46A	0.00060689	1.43	*KIAA0319L*	0.01143196	1.26
BLVRA	0.00097532	1.43	*SP100*	0.00125743	1.26
GBP1P1	0.01832553	1.42	*ODF3B*	0.00068254	1.26
ISG20	0.00024796	1.42	*TRIM21*	0.04027824	1.25
UNC93B1	0.00153611	1.42	*MS4A14*	0.03229019	1.25
TYMP	0.00661229	1.41	*GSTO1*	0.04427576	1.24
TYMP	0.01229455	1.41	*TRIM69*	0.04802239	1.24
VAMP5	0.01037169	1.41	*ACOT9*	0.03438561	1.23
RNF213	0.04487247	1.41	*ZC3HAV1*	0.04241708	1.23
FGD2	0.02784943	1.40	*TTC21A*	0.00048808	1.23
SP110	0.01037169	1.39	*TRIM56*	0.02426733	1.23
DRAP1	0.0007451	1.39	*PCK2*	0.01079063	1.23
VCAN	0.04857966	1.39	*TRIM14*	0.0007451	1.21
KLHDC8B	0.03177238	1.39	*CD86*	0.03479376	1.21
PLSCR1	0.03471629	1.38	*NELL2*	0.0016615	-1.51
LAMP3	0.00940818	1.38	*TXNDC12*	0.04806856	-1.45
LGALS9	0.00091277	1.38	*BCL11B*	0.03711457	-1.32
CASP1	0.03962541	1.37	*EIF3L*	0.00087021	-1.31
MOV10	0.00171684	1.36	*EIF4B*	0.00364144	-1.27
ADAR	0.00197242	1.36	*TBC1D14*	0.02947428	-1.27
LGALS9	0.00044535	1.36	*EEF2*	0.01107357	-1.27
PHF11	0.01650733	1.36	*SERTAD2*	0.03359516	-1.26
TOR1B	0.00695755	1.36	*FEZ1*	0.03100429	-1.26
SCARB2	0.00082371	1.34	*ALDOC*	0.02060631	-1.26
CMPK2	0.01107357	1.34	*GTF2IP1*	0.00417508	-1.26
SAT1	0.02060631	1.33	*LRRC26*	0.00242618	-1.26
TYMP	0.01930483	1.33	*EIF3L*	0.03677415	-1.25
TDRD7	0.00974749	1.33	*CTDSP2*	0.04539763	-1.25
MT1G	0.00061761	1.33	*ANXA7*	0.00171491	-1.23
PHF11	0.01722046	1.32	*PAQR8*	0.03493695	-1.22
PSME2	0.01372558	1.32	*SIAH1*	0.01998406	-1.22
TRIM38	0.02173078	1.32	*ESYT2*	0.03403468	-1.22
CMTR1	0.00153534	1.30	*NMT2*	0.03644079	-1.22
TYMP	0.03229019	1.30	*SMARCA2*	0.03358119	-1.22
HK3	0.02598473	1.30	*SIN3A*	0.01180687	-1.22
CEACAM1	0.02868831	1.29	*CDKN1B*	0.03711457	-1.22
REC8	0.0000266	1.29	*SESN1*	0.01079063	-1.21
RRAS	0.04052875	1.28	*MBP*	0.02598473	-1.21
MS4A4A	0.00082371	1.28	*MID2*	0.0000795	-1.21
SP100	0.00185527	1.28	*LINC00623*	0.03677967	-1.21
MT1IP	0.0032513	1.28			
C19orf66	0.00400343	1.27			

S8 DEGs list in pSS-other cancers vs. healthy control. *The (-) sign in FC column represents the downregulated genes in pSS-associated lymphoma (p<0.02, FC cut off =1.2). NAs excluded*

Gene symbol	Adjusted p value	FC	Gene symbol	Adjusted p value	FC
IFI44L	8.90E-05	7.13	*GBP1*	0.045682442	1.81
IFI27	0.03815265	6.60	*PARP12*	0.001679555	1.80
IFIT1	8.90E-05	4.90	*HERC6*	0.007581113	1.80
ISG15	8.90E-05	4.79	*OAS2*	0.003663257	1.77
RSAD2	0.000102181	4.43	*OAS2*	0.006587646	1.77
IFIT3	5.16E-05	4.00	*USP18*	0.034708812	1.77
IFI44	0.00085592	3.95	*STAT2*	0.000233021	1.76
OAS3	0.000166919	3.77	*MT1A*	0.01245208	1.74
EPSTI1	8.90E-05	3.62	*MX2*	0.000929208	1.72
LY6E	0.000653176	3.56	*TRIM22*	0.024572779	1.72
HERC5	8.90E-05	3.55	*STAT1*	0.00167034	1.69
MX1	8.90E-05	3.53	*OASL*	0.000880267	1.69
OAS1	8.90E-05	3.42	*STAT1*	0.000233021	1.67
IFI6	8.90E-05	3.28	*PARP9*	0.015134363	1.65
OAS1	0.000102181	3.16	*BATF2*	0.010675641	1.64
IFITM3	0.003696931	3.07	*BST2*	8.90E-05	1.61
OASL	8.90E-05	3.06	*DHX58*	0.007337187	1.60
XAF1	7.27E-05	3.01	*UBE2L6*	0.006303719	1.58
XAF1	0.00017948	2.90	*TNFSF13B*	0.014490376	1.57
IFIT3	0.003414857	2.68	*CEACAM1*	0.013311905	1.57
OAS2	0.002156565	2.68	*SP110*	0.000880267	1.56
IFIT3	0.000852243	2.63	*RTP4*	0.03815265	1.56
EIF2AK2	8.90E-05	2.43	*PARP10*	0.001909718	1.54
HES4	0.00121236	2.34	*TNFSF10*	0.03188489	1.52
IFIT2	0.010804285	2.25	*PARP9*	0.022472172	1.51
IRF7	0.000866371	2.10	*FAM46A*	0.000451667	1.50
SAMD9L	0.000236626	2.09	*TNFSF13B*	0.024845865	1.50
LAP3	0.000774552	2.06	*REC8*	0.00014025	1.50
ZBP1	0.001019367	2.00	*UNC93B1*	0.00109746	1.49
IFI35	0.00017948	1.99	*IRF9*	0.000102181	1.49
SERPING1	0.012098879	1.97	*FGD2*	0.029604923	1.46
MT2A	0.005060379	1.96	*IFITM1*	0.016946873	1.46
IFI6	0.000880267	1.96	*IFI16*	0.026198366	1.45
IRF7	0.000710192	1.96	*SHISA5*	0.03786659	1.45
OAS1	0.005093636	1.94	*OAS3*	0.011555212	1.44
PARP14	0.000102181	1.91	*CEACAM1*	0.042650839	1.43
SCO2	0.000166919	1.87	*LAMP3*	0.018542689	1.42
HELZ2	0.000880267	1.85	*PSME2*	0.005060379	1.41
IFI44L	0.034708812	1.84	*BLVRA*	0.01109232	1.41

SP110	0.042650839	1.39			
DDX58	0.034037138	1.39			
PHF11	0.013636559	1.38			
MOV10	0.008061982	1.37			
ADAR	0.008240337	1.37			
LGALS9	0.006757455	1.37			
SP110	0.037659257	1.37			
ISG20	0.008951273	1.37			
CMPK2	0.023505453	1.37			
LGALS9	0.003663257	1.35			
FOXP1-IT1	0.042817254	1.34			
TDRD7	0.042650839	1.33			
CMTR1	0.003817828	1.32			
ANKFY1	0.003459426	1.32			
DRAP1	0.04332153	1.32			
NUB1	0.034037138	1.31			
SP100	0.005751943	1.30			
SYAP1	0.045683609	1.28			
DUSP5	0.016001267	1.28			
MT1G	0.028621277	1.27			
SP100	0.006759859	1.26			
TRIM56	0.039988487	1.25			
ATF3	0.003663257	1.25			
REC8	0.010675641	1.22			
ODF3B	0.03728581	1.21			
NEXN	0.042650839	1.21			
C9orf91	0.045680309	1.20			
NELL2	0.013636559	-1.49			
ANXA7	0.005060379	-1.25			
ESYT2	0.042817254	-1.25			
SMARCA2	0.042817254	-1.24			
MID2	8.90E-05	-1.24			
USP9X	0.008951273	-1.24			
HNRNPA0	0.048907002	-1.22			
NR3C2	0.002069335	-1.21			
DSC1	0.034199174	-1.20			

***S9 DEGs list of pSS-associated lymphoma vs. healthy control**. The (-) sign in FC column represent the downregulated genes in pSS-associated lymphoma (p<0.02, FC=1.2)*

Gene symbol	Adjusted p value	FC	Gene symbol	Adjusted p value	FC
IFI27	1.5E-05	16.86	*OAS2*	1.41E-06	2.19
IFI44L	7.5E-09	12.44	*TNFAIP6*	0.000917	2.18
IFIT1	7.5E-09	7.66	*PARP12*	1.98E-06	2.15
RSAD2	9.09E-09	7.33	*GBP1*	0.000172	2.10
ISG15	2.5E-08	7.23	*HERC6*	6.21E-05	2.05
IFI44	1.09E-07	6.51	*RTP4*	2.72E-06	2.05
LY6E	1.25E-07	5.43	*OAS1*	0.018468	2.04
EPSTI1	7.5E-09	5.39	*OAS2*	2.84E-05	2.04
IFIT3	7.5E-09	5.34	*MT1A*	6.21E-05	2.01
OAS3	1.2E-07	5.19	*PARP9*	1.94E-05	1.99
HERC5	2.3E-08	5.03	*USP18*	0.000917	1.96
IFITM3	7.27E-07	4.84	*TNFSF10*	2.05E-06	1.95
IFI6	7.5E-09	4.78	*HELZ2*	4.17E-05	1.94
OAS1	3.1E-08	4.61	*PARP14*	1.08E-05	1.93
MX1	2.06E-07	4.14	*GBP5*	0.003049	1.90
OAS1	6.55E-08	4.13	*RNASE2*	0.012504	1.90
OASL	4.11E-08	3.94	*IFIH1*	0.000238	1.90
IFIT3	4.79E-06	3.59	*STAT2*	4.38E-06	1.88
XAF1	3.37E-08	3.59	*SPATS2L*	0.00029	1.88
IFIT3	2.1E-06	3.28	*DDX60*	0.000119	1.88
OAS2	3.93E-05	3.09	*TNFSF13B*	3.37E-05	1.82
XAF1	1.1E-05	3.06	*DHX58*	2.18E-05	1.82
EIF2AK2	7.35E-08	2.92	*OASL*	1.49E-05	1.81
IFIT2	2.39E-05	2.89	*STAT1*	3.76E-05	1.81
HES4	8.14E-06	2.76	*BATF2*	0.000172	1.80
SAMD9L	2.5E-08	2.70	*HLA-DRB4*	0.003986	1.80
OTOF	0.016387	2.61	*IFITM1*	1.43E-06	1.80
IRF7	2.5E-08	2.59	*PARP9*	2.6E-05	1.77
IFI44L	5.58E-06	2.59	*STAT1*	5.61E-06	1.76
ZBP1	1E-06	2.45	*SAMD9*	0.002968	1.75
IRF7	3.9E-06	2.44	*FCGR1A*	0.004621	1.75
SERPING1	4.89E-05	2.39	*MX2*	0.000144	1.74
OAS1	7.28E-06	2.38	*UBE2L6*	4.51E-05	1.74
MT2A	1.96E-05	2.32	*ANKRD22*	0.009801	1.73
IFI35	3.18E-07	2.30	*IL1RN*	0.000405	1.73
SCO2	2.25E-08	2.30	*FBXO6*	1.94E-05	1.73
LAP3	8.42E-06	2.29	*CHMP5*	0.012217	1.72
GBP1	0.000202	2.23	*SHISA5*	2.6E-05	1.71
IFI6	5.61E-06	2.23	*TNFSF13B*	8.86E-05	1.70
TRIM22	8.54E-06	2.21	*IFIT5*	0.000218	1.70

BLVRA	5.31E-07	1.70	*LGALS9*	6.71E-06	1.47
DRAP1	5.47E-08	1.68	*TYMP*	0.000474	1.47
FCGR1B	0.01376	1.67	*CASP1*	0.003753	1.46
GUSBP1	0.015554	1.65	*PSMB9*	8.42E-06	1.46
CEACAM1	0.000714	1.65	*NT5C3A*	0.012504	1.46
PARP10	2.72E-05	1.65	*IFI30*	0.008201	1.46
IL1RN	0.001121	1.65	*TOR1B*	0.000429	1.46
BST2	4.32E-06	1.64	*CHMP5*	0.018223	1.45
DDX58	8.4E-06	1.64	*CCR1*	0.016069	1.45
IRF9	5.93E-08	1.64	*HSH2D*	0.001736	1.44
STAT1	0.001501	1.62	*PHF11*	0.002192	1.44
VAMP5	0.000105	1.61	*NT5C3A*	0.044945	1.43
GBP1P1	0.000404	1.61	*TMEM140*	0.045044	1.41
TMEM123	0.031091	1.59	*ADAR*	0.000572	1.41
CEACAM1	0.000405	1.57	*TAP2*	0.007716	1.40
UNC93B1	2.84E-05	1.56	*MS4A6A*	0.036008	1.40
OAS3	0.000122	1.56	*DYNLT1*	0.007837	1.40
PSME2	9.01E-06	1.56	*SEPT4*	0.002296	1.40
IFI16	0.00059	1.55	*GBP4*	0.047205	1.40
FAM46A	1.94E-05	1.55	*ISG20*	0.000723	1.40
ANKRD22	0.020882	1.55	*TRIM5*	0.003753	1.40
LAMP3	0.00015	1.55	*SAT1*	0.003753	1.40
DDX60L	0.00514	1.53	*KLHDC8B*	0.02115	1.40
LGALS9	6.71E-06	1.53	*MS4A7*	0.006492	1.39
CMPK2	4.34E-05	1.53	*MS4A6A*	0.017423	1.39
PHF11	4.31E-05	1.52	*HK3*	0.001602	1.39
TYMP	0.000563	1.52	*CD36*	0.031481	1.39
GPBAR1	0.002153	1.52	*SP110*	0.010107	1.39
CXCL10	0.00656	1.52	*IFIT1*	0.001648	1.38
REC8	1.33E-05	1.51	*PSMB9*	0.012791	1.38
TYMP	0.001319	1.51	*TAP1*	0.038526	1.38
UBE2L6	0.001394	1.51	*TRIM21*	0.00059	1.38
CASP1	0.0023	1.51	*CST3*	0.030373	1.38
FFAR2	0.012071	1.49	*LYSMD2*	0.004199	1.37
GLRX	0.044792	1.49	*TRIM38*	0.005591	1.37
SMA4	0.03066	1.49	*GSTO1*	0.000651	1.36
PLSCR1	0.003909	1.49	*HSPA7*	0.011826	1.36
FCGR1C	0.020263	1.49	*MT1G*	0.00025	1.36
GPBAR1	0.006059	1.49	*ODF3B*	3.43E-06	1.36
FGD2	0.004199	1.49	*LGALS3BP*	0.036504	1.35
TDRD7	7.65E-05	1.49	*HSPA6*	0.02616	1.35
SP110	0.001172	1.49	*CMTR1*	0.00038	1.34
MOV10	4.16E-05	1.49	*SCARB2*	0.001003	1.34
LGALS1	0.001067	1.48	*PLAC8*	0.024225	1.34

LOC100128274	0.007811	1.34	*TRAFD1*	0.023378	1.27
TYMP	0.010652	1.33	*ADAR*	0.046774	1.27
RRAS	0.008912	1.33	*SP140*	0.043972	1.26
NBN	0.003653	1.33	*KIAA0319L*	0.008183	1.26
C19orf66	0.000457	1.33	*RAB24*	0.030636	1.26
CREB1	0.001062	1.33	*BRSK1*	0.046237	1.26
C1GALT1	0.010189	1.32	*MS4A14*	0.015701	1.26
TRIM69	0.00282	1.32	*FAM13A*	0.009829	1.26
WSB1	0.001292	1.32	*POMP*	0.000218	1.26
LMO2	0.018223	1.32	*CTSH*	0.043277	1.26
C3AR1	0.008294	1.31	*NMI*	0.042268	1.26
AP5B1	0.046669	1.31	*ANKFY1*	0.008562	1.25
GCH1	0.002263	1.31	*OAF*	0.008962	1.25
GSTO1	0.006376	1.31	*ACOT9*	0.015421	1.25
TRIM56	0.001003	1.31	*RNF7*	0.006271	1.25
KYNU	0.049183	1.30	*PRO0628*	0.033767	1.25
SOCS1	0.000954	1.30	*ZDHHC19*	0.049994	1.25
ACER3	0.008612	1.30	*NAPA*	0.001067	1.25
PRDX4	0.005904	1.30	*LOC100128288*	0.03597	1.24
DUSP19	0.000463	1.30	*GADD45B*	0.012504	1.24
LHFPL2	0.032922	1.30	*REC8*	0.000689	1.24
SLIRP	0.018223	1.30	*ATP5J2*	0.005904	1.24
LOC284837	0.006376	1.29	*PSMA3*	0.030059	1.24
CAMK1	0.006759	1.29	*DYNLL1*	0.003715	1.24
NEXN	0.000172	1.29	*MGST3*	0.036168	1.24
MT1F	0.023286	1.29	*NUDT14*	0.002457	1.24
SP110	0.046669	1.29	*IL15*	0.016102	1.24
MYD88	0.007856	1.29	*ZC3HAV1*	0.028613	1.24
SP100	0.002126	1.29	*TRIM6*	0.015483	1.24
CARD16	0.013794	1.29	*PSMB8*	0.003134	1.24
RAB24	0.013469	1.28	*CYSLTR1*	0.018349	1.23
KLHL28	0.028917	1.28	*GNG5*	0.007716	1.23
SP100	0.000318	1.28	*ACOT9*	0.011321	1.23
MT1IP	0.003419	1.28	*RBM43*	0.007909	1.23
ADAP2	0.016826	1.28	*ZC3HAV1*	0.03379	1.23
HINT3	0.036666	1.28	*PTTG1*	0.006343	1.23
NCOA7	0.036373	1.28	*UNC93B1*	0.003849	1.23
CEACAM1	0.031829	1.27	*TMOD2*	0.014744	1.23
MAD2L1BP	0.001509	1.27	*GALNT4*	0.021994	1.23
CCL2	0.003057	1.27	*HCFC1R1*	0.004443	1.23
OAS2	0.00667	1.27	*TRIM14*	0.000378	1.23
KCTD12	0.040119	1.27	*MS4A4A*	0.011826	1.23
CCDC53	0.000346	1.27	*LRRFIP1*	0.020685	1.23
IGFBP7	0.019533	1.27	*PCK2*	0.010186	1.23

SP100	0.006946	1.23	*FAM102A*	0.000786	-1.42
HNMT	0.025315	1.22	*SEC14L1*	0.017125	-1.41
MED28	0.005891	1.22	*ETS1*	0.006399	-1.41
THOC7	0.018952	1.22	*SERTAD2*	0.00029	-1.40
ETV7	0.012464	1.22	*ALDOC*	0.000109	-1.40
TRIM5	0.003057	1.22	*CTDSP2*	0.000346	-1.40
SFT2D1	0.046894	1.22	*MAGED1*	0.002516	-1.39
SCIMP	0.014539	1.22	*POM121C*	0.002162	-1.38
C4orf3	0.028728	1.22	*HIST1H1C*	0.038916	-1.38
PSMB8	0.013067	1.22	*SCAP*	0.000884	-1.38
SEC61G	0.020633	1.22	*SIRPA*	0.019085	-1.38
TMEM219	0.005891	1.21	*HNRNPUL1*	0.001521	-1.37
RAB37	0.015716	1.21	*IMPA2*	0.049109	-1.37
GNB4	0.002286	1.21	*HNRNPA1P10*	0.001159	-1.37
TMSB10	0.007643	1.21	*SF3A1*	0.000713	-1.37
DUSP5	0.035707	1.21	*TBC1D14*	0.001101	-1.36
SRBD1	0.00656	1.21	*EIF4B*	6.21E-05	-1.36
COMMD1	0.003134	1.21	*JAK1*	0.027337	-1.36
RNF31	0.038024	1.21	*ZNF746*	0.000446	-1.36
SSB	0.038673	1.20	*TBL1X*	0.012504	-1.36
FAM13A	0.044387	1.20	*SUN2*	0.003786	-1.35
PNPT1	0.00028	1.20	*CYFIP2*	0.008696	-1.35
PSME1	0.03379	1.20	*CD247*	0.023958	-1.35
FRMD3	0.006751	1.20	*BAG3*	0.000368	-1.35
HLA-DRB1	0.019579	-4.75	*RPL23AP5*	0.008183	-1.35
MYH9	0.001545	-1.59	*ALDH9A1*	0.00042	-1.34
SPOCK2	0.000774	-1.55	*ADCK3*	0.002508	-1.34
SGK223	0.000418	-1.55	*FAM102A*	0.002693	-1.34
LEF1	0.008441	-1.52	*SMARCA2*	0.00026	-1.34
TXNDC12	0.012538	-1.52	*ESYT1*	0.008661	-1.33
NELL2	0.002296	-1.51	*ATP6V1A*	0.014592	-1.33
ABLIM1	0.009953	-1.49	*RPL10A*	0.018349	-1.33
BCL11B	0.000501	-1.48	*ARRB1*	0.007317	-1.33
PIK3IP1	0.002026	-1.48	*DDB1*	0.000495	-1.33
IL2RB	0.023723	-1.47	*FNBP1*	0.010107	-1.32
ETS1	0.008164	-1.47	*CD247*	0.049653	-1.32
IGF2R	0.004967	-1.47	*APMAP*	0.030463	-1.32
SYTL2	0.009112	-1.45	*HNRNPDL*	0.000445	-1.32
RNA28S5	0.021678	-1.45	*SIN3A*	8.86E-05	-1.32
SORL1	0.019438	-1.44	*LAMP1*	0.007708	-1.32
CACNA1I	0.001334	-1.43	*PRKDC*	0.000214	-1.32
TKT	0.015824	-1.42	*VCP*	0.000704	-1.32
PRPF8	0.0004	-1.42	*VEGFB*	0.020685	-1.31
LEF1	0.042038	-1.42	*PARP1*	0.038153	-1.31

GTF2IP1	0.000448	-1.31	*CIRBP*	0.013827	-1.27
UBE2H	0.004967	-1.31	*WAC*	0.004259	-1.26
MEF2D	0.000956	-1.30	*BACH2*	0.010319	-1.26
ITK	0.03597	-1.30	*HNRNPUL1*	0.01454	-1.26
TPRG1L	0.000346	-1.30	*VPS51*	0.012045	-1.26
SCYL1	0.004443	-1.30	*ANKRD11*	0.004535	-1.26
CD96	0.011884	-1.30	*AXIN2*	0.012045	-1.26
MYH9	0.045193	-1.30	*EPHX2*	0.041367	-1.26
MBP	0.000619	-1.30	*ULK1*	0.023685	-1.26
OSBPL10	0.012791	-1.30	*RPS6KA5*	0.039348	-1.26
GLS	0.000451	-1.30	*PAF1*	0.006384	-1.26
MAP7D1	0.02294	-1.30	*NCL*	0.02294	-1.26
PRKCH	0.028917	-1.30	*EEF2*	0.014462	-1.25
RGCC	0.048765	-1.30	*LOC283070*	0.047081	-1.25
MSN	0.020569	-1.30	*RPS4X*	0.029797	-1.25
RBL2	0.001591	-1.29	*ABLIM1*	0.04298	-1.25
EIF3L	0.008456	-1.29	*PEBP1*	0.023532	-1.25
ULK1	0.011854	-1.29	*FOXJ2*	0.002608	-1.25
ID3	0.010653	-1.29	*PHRF1*	0.034796	-1.25
KIAA1147	0.017676	-1.29	*LRFN3*	0.02487	-1.25
KLF13	0.028613	-1.29	*KHDRBS1*	0.017423	-1.25
NMT2	0.0029	-1.28	*BTG1*	0.012504	-1.25
MED16	0.044467	-1.28	*MID2*	8.54E-06	-1.25
RNF44	0.006524	-1.28	*COBLL1*	0.023532	-1.25
CDR2	0.005421	-1.28	*TP53INP2*	0.007716	-1.25
RPS4X	0.02473	-1.28	*SLC12A9*	0.018349	-1.25
MYC	0.038922	-1.28	*HDAC1*	0.007716	-1.25
ACOX1	0.038153	-1.28	*PRKCQ*	0.021994	-1.25
FAM53C	0.03421	-1.28	*GMEB2*	0.00018	-1.24
TGOLN2	0.010088	-1.28	*SAFB*	0.004935	-1.24
SMAP2	0.041808	-1.28	*BANP*	0.005891	-1.24
CRKL	0.001067	-1.27	*FOXO1*	0.004005	-1.24
ARID1A	0.016338	-1.27	*FRMD8*	0.002608	-1.24
ICAM3	0.028873	-1.27	*CBLL1*	0.006399	-1.24
ARHGEF18	0.013838	-1.27	*EPHA1*	0.018349	-1.24
ANXA7	0.000346	-1.27	*HNRNPA0*	0.005641	-1.24
ZSCAN18	0.021678	-1.27	*RPRD2*	1.96E-05	-1.24
PEX5	0.000122	-1.27	*EIF3B*	0.023886	-1.24
RPA2	0.000321	-1.27	*PI4KA*	0.048768	-1.24
RPS3	0.032879	-1.27	*DOCK2*	0.049261	-1.24
HSP90B1	0.018107	-1.27	*FOXJ3*	0.003753	-1.24
NOTCH1	0.030178	-1.27	*DHRS3*	0.018349	-1.24
LRRC26	0.002057	-1.27	*RAB22A*	0.000364	-1.24
EIF3L	0.006489	-1.27	*TAF4*	0.005587	-1.24

SLC7A1	0.000346	-1.24	*ACTG1*	0.008675	-1.21
TUBB	0.008534	-1.23	*NR3C2*	0.000446	-1.21
ST13P4	0.046641	-1.23	*COCH*	0.000349	-1.21
GEMIN4	0.023723	-1.23	*EIF3D*	0.005577	-1.21
KAT8	0.002078	-1.23	*BMS1*	0.005252	-1.21
INTS1	0.007234	-1.23	*RAPGEF6*	0.031091	-1.21
U2AF1	0.008294	-1.23	*TNFRSF13B*	0.023908	-1.21
RPA1	0.002187	-1.23	*THOC5*	0.032389	-1.21
QRICH1	0.002162	-1.23	*PAFAH1B1*	0.007971	-1.21
RPN1	0.000218	-1.23	*KLHL3*	0.042037	-1.21
ASNS	0.030636	-1.23	*MAGED1*	0.003599	-1.21
PDPK1	0.03597	-1.23	*STXBP5*	0.014454	-1.21
LRIG1	0.014497	-1.23	*EIF4A3*	0.043043	-1.21
AP1M1	0.044312	-1.23	*ENO2*	0.024421	-1.21
VPS35	0.021994	-1.23	*PRPF19*	0.012045	-1.21
ALKBH5	0.009412	-1.23	*FAM120B*	0.016568	-1.21
DSC1	0.002296	-1.23	*LRRC47*	0.006532	-1.21
MAN1C1	0.040336	-1.23	*RNF216*	0.00652	-1.21
DNAJA3	0.00667	-1.23	*USP7*	0.002126	-1.21
ARCN1	0.03678	-1.23	*FHL1*	0.021284	-1.21
CS	0.028841	-1.23	*SRRM1*	0.003582	-1.20
RRN3	0.009783	-1.23	*IMPDH2*	0.048273	-1.20
SYTL2	0.024498	-1.22	*SNRPN*	0.040929	-1.20
SLC16A10	0.00038	-1.22	*MDC1*	0.036666	-1.20
USP9X	0.003986	-1.22	*HSPA9*	0.019311	-1.20
NDRG3	0.0023	-1.22	*TBC1D9*	0.048261	-1.20
CALM3	0.03379	-1.22	*ARF1*	0.028643	-1.20
DANCR	0.00042	-1.22	*DEXI*	0.021246	-1.20
PRKCSH	0.028917	-1.22	*ASF1B*	0.046464	-1.20
CDK19	0.008365	-1.22	*NIPSNAP1*	0.001486	-1.20
SPECC1L	0.011998	-1.22	*BCL2L13*	0.010699	-1.20
SMARCC1	0.001067	-1.22	*FAM174A*	0.029365	-1.20
DSC1	0.002665	-1.22			
EPHB4	0.015716	-1.22			
XRN2	0.007856	-1.22			
GTF2IP1	0.02167	-1.22			
HSPBAP1	0.006979	-1.22			
CYB561D1	0.000767	-1.22			
BCAP31	0.01671	-1.22			
ATP1A1	0.006399	-1.22			
EXTL3	0.030463	-1.22			
AKR1B1	0.046774	-1.22			
AKTIP	0.039388	-1.22			
PTBP1	0.023723	-1.22			

S10 The canonical pathways identified in pSS-associated lymphoma vs healthy controls *The table contains only the pathways that have a significant z-score (2 < z-score > -2)*

Ingenuity Canonical Pathways	-log(p-value)	Ratio	z-score	Molecules
Interferon Signaling	1.74E+01	4.44E-01	2.84	*OAS1,IRF9,IFITM1,IFIT3,STAT2,JAK1,IFI6,IFITM3, TAP1,IFIT1,STAT1,ISG15,MX1,IFI35,SOCS1,PSMB8*
Dendritic Cell Maturation	2.95E+00	6.21E-02	2.333	*FCGR1B,IL15,HLA-DRB1,IL1RN,MYD88,CREB1, STAT1,FCGR1A,STAT2,HLA-DRB4,FCER1G*
RhoGDI Signaling	1.18E+00	4.05E-02	2.236	*GNB4,GNG5,ACTG1,MSN,PI4KA,ARHGEF18,FNBP1*
Huntington's Disease Signaling	4.65E+00	6.99E-02	-2	*GNB4,CREB1,PRKCQ,NAPA,CASP1,PSME1,PSME2, HSPA6,TAF4,HSPA9,PDPK1,GNG5,SIN3A,PRKCH,HDAC1,GLS*
CD28 Signaling in T Helper Cells	1.96E+00	5.93E-02	-2	*HLA-DRB1,PDPK1,CALM1 (includes others),PRKCQ,ITK,FCER1G,CD247*
RhoA Signaling	9.71E-01	4.10E-02	-2	*RAPGEF6,EPHA1,ACTG1,MSN,PI4KA*
Calcium-induced T Lymphocyte Apoptosis	4.32E+00	1.25E-01	-2.236	*HLA-DRB1,CALM1 (includes others),PRKCQ,FCER1G,PRKCH,MEF2D,CD247,HDAC1*
iCOS-iCOSL Signaling in T Helper Cells	2.76E+00	7.41E-02	-2.236	*HLA-DRB1,PDPK1,CALM1 (includes others),PRKCQ,ITK,IL2RB, FCER1G,CD247*
Aldosterone Signaling in Epithelial Cells	1.87E+00	5.26E-02	-2.236	*HSPA9,PDPK1,HSP90B1,PRKCQ,NR3C2, PRKCH,HSPA6,PI4KA*
Nitric Oxide Signaling in the Cardiovascular System	1.77E+00	6.00E-02	-2.236	*CALM1 (includes others),SLC7A1,HSP90B1,PRKCQ,PRKCH, VEGFB*
Signaling by Rho Family GTPases	6.84E-01	2.99E-02	-2.236	*GNB4,GNG5,ACTG1,MSN,PI4KA,ARHGEF18,FNBP1*
Leukocyte Extravasation Signaling	9.41E-01	3.54E-02	-2.646	*ACTG1,CRKL,PRKCQ,ICAM3,ITK,MSN,PRKCH*

S11 The downstream analysis in pSS-associated lymphoma vs healthy controls *The table contains only the diseases and functions that have a significant z-score (2 < z-score > -2)*

Categories	Diseases or Functions Annotation	p-Value	Predicted Activation State	Activation z-score	Molecules
Cell-To-Cell Signaling and Interaction, Cellular Function and Maintenance, Inflammatory Response	phagocytosis of blood cells	1.25E-03	Increased	2.012	*CD36,CXCL10,FCGR1A,GLRX,IL15,ISG15,MYD88,RGCC,SIRPA*
Cell Death and Survival	cell death of leukocyte cell lines	4.56E-08	Increased	2.026	*CD247,CREB1,DNAJA3,EIF2AK2,EIF4B,GADD45B,GLS,HLADRB4,IFIH1,IL15, LGALS1,MYC,NBN,NOTCH1,PNPT1,PRKCQ,RRAS,SOCS1,TNFSF10,TUBB*
Cell Death and Survival	apoptosis of hematopoietic cell lines	7.96E-05	Increased	2.175	*CREB1,DNAJA3,EIF2AK2,EIF4B,ICAM3,IFIH1,IL15,MYC,NBN,NOTCH1,PARP1, PNPT1,PRKCQ,SOCS1,TNFSF10*
Cell-To-Cell Signaling and Interaction, Inflammatory Response	immune response of T lymphocytes	3.88E-04	Increased	2.232	*C3AR1,CASP1,ETS1,FOXO1,IL15,ITK,LGALS9,MYD88,PSME2*
Cell-To-Cell Signaling and Interaction, Hair and Skin Development and Function	response of epithelial cell lines	8.45E-07	Increased	2.377	*DDX58,IFIH1,MID2,TRIM14,TRIM21,TRIM38,TRIM5,TRIM56,TRIM6*
Cell-To-Cell Signaling and Interaction, Embryonic Development	response of embryonic cell lines	1.00E-06	Increased	2.377	*DDX58,IFIH1,MID2,TRIM14,TRIM21,TRIM38,TRIM5,TRIM56,TRIM6*
Cell-To-Cell Signaling and Interaction, Renal and Urological System Development and Function	response of kidney cell lines	1.19E-06	Increased	2.377	*DDX58,IFIH1,MID2,TRIM14,TRIM21,TRIM38,TRIM5,TRIM56,TRIM6*
Neurological Disease, Skeletal and Muscular Disorders	neuromuscular disease	1.16E-10	Increased	2.391	*ANKRD11,ARRB1,ATP6V1A,BANP,BCL11B,CASP1,CCL2,CIRBP,COCH,CREB1, CST3,CYFIP2,DUSP5,DYNLT1,ENO2,EPHB4,EPHX2,EPSTI1,FCGR1A,FCGR1B, GBP1,GCH1,HNRNPDL,HSP90B1,ID3,IFIT1,IL1RN,IMPDH2,IRF7,ISG15,LAMP1 ,*

					LY6E,MBP,MEF2D,MT1G,MT2A,MX1,NAPA,NDRG3,OAS1,OAS3,PARP1,PEBP1, PSMB8,PSMB9,PSME1,RPS4X,RPS6KA5,RSAD2,SAT1,SCARB2,SERPING1, SIN3A,SLIRP,SORL1,VCP
Inflammatory Response	immune response of cells	3.56E-10	Increased	2.403	*C3AR1,CASP1,CCL2,CD247,CD36,CEACAM1,CXCL10,DDX58,DOCK2,ETS1,FC ER1G,FCGR1A,FOXO1,GLRX,IFIH1,IL15,IRF7,ISG15,ITK,JAK1,LGALS9,MID2, MYC, MYD88,MYH9,PARP1,PSME2,RGCC,SF3A1,SIRPA,SOCS1,STAT1,TRIM14,TRIM2 1, TRIM38,TRIM5,TRIM56,TRIM6,UBE2L6*
Cell-To-Cell Signaling and Interaction, Inflammatory Response	immune response of leukocytes	6.21E-06	Increased	2.479	*C3AR1,CASP1,CCL2,CD36,CXCL10,DDX58,ETS1,FCER1G,FOXO1,GLRX,IL15, ISG15,ITK,LGALS9,MYD88,PSME2,RGCC,SIRPA,UBE2L6*
Cell Death and Survival	apoptosis of leukocyte cell lines	6.71E-04	Increased	2.751	*CREB1,DNAJA3,EIF2AK2,EIF4B,IFIH1,MYC,NBN,NOTCH1,PNPT1,PRKCQ,SOC S1,TNFSF10*
Neurological Disease	progressive motor neuropathy	1.66E-06	Increased	2.891	*ARRB1,CALM1(includesothers),CASP1,CCL2,CD36,CST3,ENO2,EPSTI1,FCGR1A, FCGR1B,FOXO1,GCH1,HNRNPDL,IFIT1,IMPDH2,IRF7,ISG15,LAMP1,LY6E, MAGED1,MBP,MEF2D,MX1,OAS1,OAS3,PEBP1,RPS4X,RSAD2,SCARB2,SERPIN G1*
Cell Signaling	I-kappaB kinase/NF-kappaB cascade	5.73E-07	Increased	3.067	*BST2,CASP1,CD36,DNAJA3,HDAC1,LGALS1,LGALS9,MID2,MYD88,PDPK1,RNF 31,SHISA5,STAT1,TRIM22,TRIM38,TRIM5,ZC3HAV1*
Inflammatory Disease, Neurological Disease, Skeletal and Muscular Disorders	Multiple Sclerosis	8.20E-08	Increased	3.148	*ARRB1,CASP1,CCL2,CST3,EPSTI1,FCGR1A,FCGR1B,IFIT1,IMPDH2,IRF7,ISG15 , LY6E,MBP,MX1,OAS1,OAS3,RSAD2,SERPING1*
Inflammatory Disease, Neurological Disease, Skeletal and Muscular Disorders	relapsing-remitting multiple sclerosis	1.54E-07	Increased	3.148	*EPSTI1,IFIT1,IRF7,ISG15,LY6E,MX1,OAS1,OAS3,RSAD2,SERPING1*
Cell Signaling	protein kinase cascade	1.74E-06	Increased	3.313	*BST2,CASP1,CCL2,CD36,CEACAM1,CRKL,DNAJA3,DUSP19,EIF2AK2,GADD45 B, HDAC1,LGALS1,LGALS9,MID2,MYC,MYD88,NMI,PAFAH1B1,PDPK1,PEBP1,RN F31,SHISA5,SOCS1,STAT1,STAT2,TRIM22,TRIM38,TRIM5,ZC3HAV1*
Infectious Diseases	replication of virus	1.41E-23	Decreased	-4.852	*ADAR,AKTIP,AP1M1,ARCN1,ATP6V1A,BST2,CBLL1,CCL2,CCR1,CEACAM1,CRE B1,CST3,CXCL10,DDX58,EIF2AK2,EIF3L,EIF4A3,FCGR1A,GBP1,IFIH1,IFIT1,IF*

					ITM1, *IFITM3,IL15,IRF9,ISG15,ISG20,JAK1,LGALS9,LY6E,MED28,MX1,MX2,MYC,MYD88,NCL,NMI,OAS1,OASL,PARP12,PI4KA,PRPF8,RSAD2,SAFB,SF3A1,SOCS1,SP110,STAT1,STAT2,TAP1,TNFSF10,TRIM14,TRIM21,TRIM38,TRIM5,TUBB,UBE2L6,USP7,VCP,ZC3HAV1*
Infectious Diseases	Viral Infection	4.46E-27	Decreased	-4.474	*ADAR,AKTIP,AP1M1,ARCN1,ARF1,ARID1A,ARRB1,ATP6V1A,BST2,C3AR1,CAMK1,CARD16,CBLL1,CCL2,CCR1,CD247,CD36,CEACAM1,CHMP5,CREB1,CST3,CXCL10,DDX58,DDX60L,DNAJA3,EIF2AK2,EIF3L,EIF4A3,FCGR1A,FCGR1B,FOXJ2,GBP1,GSTO1,HDAC1,HIST1H1C,HNRNPDL,HSPA6,HSPA9,IFI35,IFIH,IFIT1,IFIT2,IFIT3,* *IFITM1,IFITM3,IGF2R,IL15,IL1RN,IL2RB,IMPA2,IMPDH2,IRF9,ISG15,ISG20,ITK,* *JAK1,KHDRBS1,LEF1,LGALS1,LGALS9,LY6E,MED16,MED28,MOV10,MS4A4A,MT2A,MX1,MX2,MYC,MYD88,NCL,NMI,OAS1,OASL,PAF1,PARP1,PARP12,PARP9,* *PI4KA,PRKCH,PRPF8,PSMA3,PSME2,PTTG1,RNASE2,RNF216,RPL10A,RSAD2,SAFB,SAMD9,SCARB2,SEC14L1,SEC61G,SF3A1,SFT2D1,SMARCA2,SOCS1,SP100,* *SP110,SPATS2L,SSB,STAT1,STAT2,TAP1,TKT,TNFSF10,TRAFD1,TRIM14,TRIM21,* *TRIM22,TRIM38,TRIM5,TRIM56,TUBB,UBE2H,UBE2L6,UNC93B1,USP7,VCP,ZC3HAV1*
Cell Signaling	replication of viral replicon	2.63E-18	Decreased	-4.3	*ADAR,BST2,EIF2AK2,IFI16,IFIT1,IFITM1,IFITM3,ISG15,ISG20,MX1,OAS1,OAS3,OASL,PARP10,PI4KA,PLSCR1,RSAD2,TRIM6,ZC3HAV1*
Infectious Diseases	replication of RNA virus	1.21E-21	Decreased	-4.225	*ADAR,AKTIP,AP1M1,ARCN1,ATP6V1A,BST2,CBLL1,CCR1,CEACAM1,CREB1,CST3,CXCL10,DDX58,EIF2AK2,EIF3L,EIF4A3,FCGR1A,GBP1,IFIH1,IFIT1,IFITM1,IFITM3,IRF9,ISG15,ISG20,JAK1,LGALS9,LY6E,MED28,MX1,MYC,MYD88,NCL,NMI,OAS1,OASL,PARP12,PI4KA,PRPF8,RSAD2,SAFB,SF3A1,SOCS1,SP110,STAT1,STAT2,TAP1,TNFSF10,TRIM14,TRIM21,TRIM38,TRIM5,TUBB,UBE2L6,ZC3HAV1*
Protein Synthesis	metabolism of protein	1.49E-04	Decreased	-3.372	*ARRB1,BAG3,BANP,CASP1,CIRBP,CREB1,CST3,CTSH,CYFIP2,DNAJA3,EEF2,EIF2AK2,EIF3B,EIF3D,EIF3L,EIF4A3,EIF4B,FBXO6,FOXO1,IFI30,IL1RN,LAMP1,LAP3,MYC,MYD88,MYH9,NOTCH1,PARP12,PRKCQ,PTBP1,RPS4X,SAMD9L,SAT1,SLC7A1,SORL1,SSB,TBL1X,TNFSF10,TNFSF13B,TP53INP2,UBE2H,USP18,USP7,VCP*
Infectious Diseases	replication of Herpesviridae	1.58E-13	Decreased	-3.189	*ADAR,CST3,CXCL10,DDX58,IFIH1,IRF9,ISG20,MX2,MYD88,OAS1,OASL,PARP12,* *RSAD2,STAT1,STAT2*

Infectious Diseases	infection of kidney cell lines	1.78E-06	Decreased	-3.038	*ADAR,ARID1A,EIF2AK2,FOXJ2,HDAC1,HNRNPDL,IFITM1,IFITM3,ISG20,KHDRBS1,MT2A,PARP9,PRKCH,PRPF8,PSMA3,RNF216,RPL10A,RSAD2,SF3A1,TRIM56,* *UBE2H,ZC3HAV1*
Infectious Diseases, Organismal Injury and Abnormalities	infection of embryonic cell lines	1.32E-06	Decreased	-3.022	*ADAR,ARID1A,EIF2AK2,FOXJ2,HDAC1,HNRNPDL,IFITM1,IFITM3,ISG20,KHDRBS1,MT2A,PARP9,PRKCH,PRPF8,PSMA3,RNF216,RPL10A,RSAD2,SF3A1,TRIM56,UBE2H,ZC3HAV1*
Infectious Diseases	infection of epithelial cell lines	1.32E-06	Decreased	-3.022	*ADAR,ARID1A,EIF2AK2,FOXJ2,HDAC1,HNRNPDL,IFITM1,IFITM3,ISG20,KHDRBS1,MT2A,PARP9,PRKCH,PRPF8,PSMA3,RNF216,RPL10A,RSAD2,SF3A1,TRIM56,UBE2H,ZC3HAV1*
Infectious Diseases	replication of Flaviviridae	4.97E-07	Decreased	-2.931	*CXCL10,EIF2AK2,IFIT1,IFITM1,IFITM3,ISG15,OASL,PI4KA,RSAD2,UBE2L6*
Infectious Diseases	replication of Murine herpesvirus 4	1.68E-10	Decreased	-2.828	*ADAR,CXCL10,DDX58,IFIH1,ISG20,MX2,OAS1,PARP12*
Infectious Diseases	infection by DNA virus	5.62E-08	Decreased	-2.669	*ADAR,EIF2AK2,FCGR1A,FCGR1B,IGF2R,IL15,IL1RN,LGALS1,MYD88,SAMD9,STAT1,TRIM21,UNC93B1*
Infectious Diseases	infection of mammalia	4.72E-06	Decreased	-2.572	*CASP1,CD36,DDX58,EIF2AK2,IFI30,IFIH1,IL15,ISG15,ITK,MYD88,STAT1,UNC93B1,USP18*
Cell Cycle	interphase	6.05E-07	Decreased	-2.476	*ARRB1,BAG3,BRSK1,BTG1,CAMK1,CD247,EIF2AK2,ETS1,FOXO1,GADD45B,HDAC1,ID3,IGFBP7,IL15,LEF1,LGALS1,MT1A,MX2,MYC,NBN,NOTCH1,PAF1,PARP1,PLAC8,PLSCR1,PNPT1,PRKCH,PRKDC,PRPF19,PRPF8,PTTG1,RBL2,RGCC,RNF31,RPA1,SMARCA2,STAT1,SUN2,TRIM21*
Infectious Diseases	infection by Herpesviridae	8.89E-06	Decreased	-2.463	*FCGR1A,FCGR1B,IGF2R,IL15,IL1RN,LGALS1,MYD88,STAT1,UNC93B1*
Cell Cycle	G1 phase	6.93E-06	Decreased	-2.393	*ARRB1,BAG3,CAMK1,CD247,EIF2AK2,ETS1,FOXO1,GADD45B,ID3,IGFBP7,LEF1,LGALS1,MT1A,MX2,MYC,NOTCH1,PARP1,PLSCR1,PNPT1,PRKCH,PRKDC,PRPF8,* *RBL2,RGCC,SMARCA2*
Infectious Diseases	Bacterial Infections	3.31E-07	Decreased	-2.342	*CASP1,CCL2,CCR1,CD36,CXCL10,CYSLTR1,FCER1G,FCGR1A,FCGR1B,IFI30,IL15,IL1RN,IL2RB,MYD88,PARP1,RNASE2,RPS6KA5,SIRPA,SOCS1,STAT1,TRAFD1,* *UNC93B1,USP18*
Infectious Diseases	replication of vesicular stomatitis	1.47E-12	Decreased	-2.328	*DDX58,FCGR1A,IFIH1,IFITM3,IRF9,ISG20,LY6E,OAS1,OASL,PARP12,SOCS1,SP110,STAT2,TAP1,TRIM38*

	virus				
Infectious Diseases	infection of cells	1.18E-10	Decreased	-2.265	*ADAR,ARCN1,ARF1,ARID1A,ARRB1,CAMK1,CARD16,CASP1,CCR1,DDX58,DDX60L,EIF2AK2,FOXJ2,HDAC1,HNRNPDL,HSPA9,IFI35,IFITM1,IFITM3,IGF2R,IL15,* *ISG20,JAK1,KHDRBS1,LGALS9,MED28,MT2A,PAF1,PARP9,PI4KA,PRKCH,PRPF8,PSMA3,PSME2,RNF216,RPL10A,RSAD2,SAMD9,SEC14L1,SEC61G,SF3A1,SFT2D1,* *SP110,SPATS2L,SSB,STAT1,STAT2,TAP1,TRAFD1,TRIM21,TRIM5,TRIM56,UBE2H,* *UNC93B1,ZC3HAV1*
Infectious Diseases	replication of Influenza A virus	8.05E-09	Decreased	-2.257	*ADAR,AKTIP,ARCN1,ATP6V1A,CBLL1,CREB1,DDX58,EIF2AK2,EIF3L,EIF4A3,GBP1,IFITM1,IFITM3,ISG15,JAK1,MX1,MYC,NMI,PRPF8,RSAD2,SAFB,SF3A1,STAT1,TRIM14,TRIM21,TUBB*
Inflammatory Response	inflammation of eye	3.81E-04	Decreased	-2.236	*CST3,IL1RN,IRF9,LGALS1,SOCS1,STAT1*
Infectious Diseases	replication of Hepatitis C virus	5.35E-04	Decreased	-2.201	*EIF2AK2,IFIT1,IFITM1,ISG15,PI4KA,UBE2L6*
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	mammary tumor	3.76E-04	Decreased	-2.2	*ACOT9,AKR1B1,ALDOC,ARID1A,ARRB1,ASF1B,ATP5J2,AXIN2,BLVRA,* *CALM(includesothers),CCL2,CMPK2,CXCL10,EIF2AK2,EIF3B,EIF4B,EPHX2,ETS1,* *GADD45B,GSTO1,HDAC1,HIST1H1C,HNRNPUL1,HSP90B1,IFIT1,LGALS1,LMO2,* *MT1F,MT1G,MT2A,MYC,MYH9,NBN,NCL,NOTCH1,PARP1,PDPK1,PRKCQ,PTBP1,PTTG1,RNF7,RPS3,RPS4X,SIN3A,SLIRP,STAT1,TBC1D9,TNFSF10,TUBB,TYMP,U2AF1/U2AF1L5,USP9X*
Infectious Diseases	replication of Influenza virus	2.08E-09	Decreased	-2.135	*ADAR,AKTIP,ARCN1,ATP6V1A,CBLL1,CREB1,DDX58,EIF2AK2,EIF3L,EIF4A3,GBP1,IFITM1,IFITM3,ISG15,JAK1,MX1,MYC,NMI,PRPF8,RSAD2,SAFB,SF3A1,STAT1,TNFSF10,TRIM14,TRIM21,TUBB*
Infectious Diseases	infection by RNA virus	7.27E-10	Decreased	-2.077	*ARCN1,ARF1,ARID1A,ARRB1,CAMK1,CARD16,CD36,CEACAM1,DDX58,DDX60L,* *EIF2AK2,FCGR1A,FCGR1B,FOXJ2,HDAC1,HNRNPDL,HSPA9,IFI35,IFIH1,IFITM3,* *IL2RB,IMPDH2,ISG15,ISG20,JAK1,KHDRBS1,LGALS9,MED28,MT2A,MYD88,PAF1,PARP9,PI4KA,PRKCH,PRPF8,PSMA3,PSME2,RNF216,RPL10A,RSAD2,SEC14L1,* *SEC61G,SF3A1,SFT2D1,SP110,SPATS2L,SSB,STAT1,STAT2,TAP1,TNFSF10,TRAF*

					D1,TRIM5,TRIM56,UBE2H,ZC3HAV1

S12 The upstream regulators in pSS-associated lymphoma vs healthy controls *The table contains only the upstream regulators that have a significant z-score (2 < z-score > -2)*

Upstream Regulator	**Molecule Type**	**Predicted Activation State**	**Activation z-score**	**p-value**	***Target molecules in dataset***
IFNL1	cytokine	Activated	6.664	9.18E-66	*BST2,C19orf66,CXCL10,DDX58,DDX60,EIF2AK2,GBP1,HERC5,HERC6,IFI27,IFI35,IFI44,IFI44L,IFI6,IFIH1,IFIT1,IFIT2,IFIT3,IFIT5,IFITM1,IFITM3,IRF9,ISG15,ISG20,LGALS3BP,MX1,OAS1,OAS2,OAS3,OASL,PHF11,PLSCR1,PSMB9,RSAD2,SAMD9,SP100,SP110,STAT1,TDRD7,TMEM140,TRIM14,TRIM22,UBE2L6,USP18,ZC3HAV1*
IFNA2	cytokine	Activated	6.539	6.77E-55	*BST2,C19orf66,CXCL10,DDX58,DDX60,EIF2AK2,GBP1,HERC5,HERC6,HSH2D,IFI27,IFI35,IFI44,IFI44L,IFI6,IFIH1,IFIT1,IFIT2,IFIT3,IFIT5,IFITM1,IFITM3,IRF7,IRF9,ISG15,ISG20,LGALS3BP,LY6E,MX1,OAS1,OAS2,OAS3,PARP12,PLSCR1,RSAD2,SOCS1,SP100,SP110,STAT1,TDRD7,TNFSF10,TRIM14,UBE2L6,USP18,ZC3HAV1*
PRL	cytokine	Activated	6.291	2.15E-54	*ADAR,BST2,C19orf66,CAMK1,CMPK2,CST3,CTSH,CXCL10,DDX58,DDX60L,DHX58,EIF2AK2,EPSTI1,HELZ2,HERC5,HERC6,ID3,IFI35,IFI44,IFI44L,IFI6,IFIH1,IFIT1,IFIT3,IFIT5,IFITM1,IRF7,IRF9,ISG15,LAMP3,LY6E,MAGED1,MSN,MX2,MYC,OAS1,OAS2,OAS3,PARP10,PARP12,PARP14,PLSCR1,PNPT1,PSME1,PSME2,REC8,RSAD2,SAMD9,SAMD9L,SHISA5,SOCS1,SP100,SP110,STAT2,TDRD7,TMEM140,TRIM14,USP18,XAF1*
IRF7	transcription regulator	Activated	5.798	4.99E-40	*CMPK2,CXCL10,DDX58,DHX58,FCGR1A,GBP5,HELZ2,IFI16,IFIH1,IFIT2,IFIT3,IFITM3,IL15,IRF7,ISG15,ISG20,NT5C3A,OAS1,OAS2,OAS3,OASL,PARP12,PARP14,PHF11,PLAC8,RSAD2,SAMD9L,STAT1,STAT2,TAP1,TDRD7,TNFSF10,UBE2L6,USP18,ZBP1*
IRF3	transcription regulator	Activated	5.944	1.46E-38	*CCL2,CMPK2,CXCL10,DDX58,DHX58,FCGR1A,GBP5,HELZ2,IFI16,IFI6,IFIH1,IFIT1,IFIT2,IFIT3,IFITM3,IL15,IRF7,ISG15,ISG20,NT5C3A,OAS1,OAS2,OAS3,OASL,PARP12,PARP14,PHF11,PLAC8,RSAD2,SAMD9L,STAT1,STAT2,TAP1,TDRD7,UBE2L6,USP18,ZBP1*
Ifnar	group	Activated	5.728	5.59E-37	*CCL2,CXCL10,DDX58,EIF2AK2,FCER1G,IFI16,IFI35,IFIH1,IFIT2,IFIT3,IFITM3,IL15,IRF7,IRF9,ISG20,MYD88,OAS1,OAS2,OASL,PNPT1,PSMB8,PSMB9,RSAD2,STAT1,STAT2,TAP1,TAP2,TNFSF10,TRIM21,UBE2L6,UNC93B1,USP18,XAF1,ZBP1*
IFNG	cytokine	Activated	7.442	1.91E-35	*ATP1A1,BATF2,BTG1,CASP1,CCL2,CEACAM1,CMPK2,CTSH,CXCL10,DDB1,DDX58,EIF2AK2,FCER1G,FCGR1A,FCGR1B,GBP1,GBP4,GBP5,GCH1,GNB4,HERC6,IFI16,IFI27,IFI44,IFI44L,IFI6,IFIH1,IFIT1,IFIT2,IFIT3,IFIT5,IL15,IL1RN,IRF7,IRF9,ISG15,ISG20,KCTD12,LAMP3,LGALS3BP,LGALS9,LY6E,MX1,MX2,MYC,MYD88,NMI,NOTCH1,OAS*

					1,OAS2,OAS3,OASL,PRKCQ,PRPF8,PSMB8,PSMB9,PSME1,PSME2,RNF31,RSAD2,RTP4,SAMD9,SF3A1,SOCS1,SP100,SP110,STAT1,STAT2,TAP1,TAP2,TNFSF10,TNFSF13B,TRIM22,TYMP,USP18
STAT1	transcription regulator	Activated	5.378	4.25E-31	*BATF2,CASP1,CCL2,CEACAM1,CMPK2,CXCL10,EIF2AK2,FCER1G,GBP1,GBP5,HERC6,IFI16,IFI27,IFI6,IFIT1,IFIT2,IFIT3,IFITM1,IL15,IRF7,IRF9,ISG15,KCTD12,MX1,MYC,OAS2,OASL,PARP9,PSMB8,PSMB9,PSME1,PSME2,RSAD2,SOCS1,SP110,STAT1,STAT2,TAP1,TNFSF10,TRAFD1*
MAVS	other	Activated	4.516	6.85E-26	*CMPK2,CXCL10,DDX58,DHX58,IFIT1,IFIT2,IFIT3,IFITM3,IRF7,ISG15,ISG20,NT5C3A,OAS1,OAS2,OASL,PARP12,RSAD2,SOCS1,STAT1,STAT2,UBE2L6*
IFNB1	cytokine	Activated	5.093	2.56E-25	*BST2,CASP1,CCL2,CMPK2,CXCL10,DDX58,EIF2AK2,GBP4,GBP5,GNB4,IFI16,IFI27,IFIH1,IFIT1,IFIT2,IFIT3,IRF7,IRF9,ISG15,ISG20,MX1,MYC,NMI,NOTCH1,NT5C3A,OAS1,OAS2,RSAD2,SOCS1,STAT1,STAT2,TNFSF10,TRIM21,USP18*
IRF5	transcription regulator	Activated	4.383	1.95E-24	*CMPK2,CXCL10,DDX58,DHX58,IFIH1,IFIT2,IFIT3,IFITM3,IRF7,ISG15,ISG20,NT5C3A,OAS1,OAS2,OASL,PARP12,RSAD2,STAT1,STAT2,UBE2L6*
EIF2AK2	kinase	Activated	4.856	1.52E-23	*DDX58,EIF2AK2,IFI27,IFI35,IFI6,IFIT1,IFIT5,IFITM1,ISG15,ISG20,LGALS3BP,MYC,NMI,OAS1,OAS3,PARP12,PARP9,PLSCR1,REC8,SP140,STAT1,UBE2L6,USP18,ZC3HAV1*
TGM2	enzyme	Activated	4.906	2.13E-23	*CCL2,CD36,CEACAM1,CXCL10,DDX60,DDX60L,FCER1G,FFAR2,HK3,ICAM3,IFI35,IFI6,IFIT1,IFIT2,IFIT3,IFIT5,IRF9,LGALS9,LY6E,MT2A,MYC,OAS1,OAS2,OAS3,OASL,PARP14,PARP9,PLSCR1,SAMD9L,SIRPA,SP110,STAT1,TAP1,TOR1B,TRIM22,XAF1*
TLR3	transmembrane receptor	Activated	4.042	5.14E-22	*CMPK2,CXCL10,DDX58,DHX58,EIF2AK2,GADD45B,GBP4,GCH1,HERC5,IFI16,IFI44,IFI44L,IFI6,IFIH1,IFIT1,IFIT2,IFIT3,IL15,IRF7,ISG15,ISG20,MX1,MX2,MYD88,OAS1,OASL,RSAD2,STAT1,TNFSF10,TNFSF13B,TRIM38,USP18,ZC3HAV1*
DDX58	enzyme	Activated	3.615	1.07E-19	*CXCL10,DDX58,EIF2AK2,IFI35,IFI44,IFIH1,IFIT1,IFIT2,IFIT3,IRF7,ISG15,ISG20,NMI,OAS1,RSAD2,SOCS1,STAT1,STAT2*
IFNAR1	transmembrane receptor	Activated	3.665	1.53E-19	*CMPK2,CXCL10,EIF2AK2,IFI16,IFI44,IFIH1,IFIT2,IFIT3,IL15,IRF7,ISG15,MYC,OAS1,OAS2,OAS3,OASL,PARP12,RNASE2,RSAD2,RTP4,SOCS1,STAT1,TNFSF13B,USP18*
Interferon alpha	group	Activated	4.666	5.96E-17	*BCL2L13,CASP1,CXCL10,DDX58,EIF2AK2,GBP1,IFI16,IFI27,IFIT1,IFIT2,IFIT3,IL15,IRF7,ISG15,ISG20,MYD88,OAS1,OAS2,PHF11,RNF31,SF3A1,SOCS1,TNFSF10,TNFSF13B,USP18*
TLR7	transmembrane receptor	Activated	4.534	6.61E-16	*CCL2,CXCL10,IFI35,IFI44L,IFIT1,IFIT3,IFITM1,IRF7,IRF9,ISG15,ISG20,MX1,MX2,MYD88,OAS2,OAS3,RSAD2,SOCS1,STAT1,STAT2,TRIM38*
TLR9	transmembrane receptor	Activated	4.357	2.27E-15	*CXCL10,GADD45B,GCH1,IFI16,IFI35,IFI44L,IFIT1,IFIT2,IFIT3,IFITM1,IRF7,IRF9,ISG15,ISG20,MX1,MX2,OAS2,OAS3,RSAD2,SOCS1,STAT1,STAT2,TNFRSF13B,USP18*
SASH1	other	Activated	3.638	1.39E-14	*CMPK2,CXCL10,HDAC1,HELZ2,IFIT2,IFIT3,IL15,IRF7,ISG15,ISG20,NMI,OASL,RSAD2,SOCS1,STAT1,STAT2,TRIM21*
SAMSN1	other	Activated	3.3	2.34E-14	*CMPK2,CXCL10,HDAC1,HELZ2,IFIT2,IFIT3,IL15,IRF7,ISG15,ISG20,NMI,OASL,RGCC,RSAD2,SOCS1,STAT1,STAT2,TRIM21*

PAF1	other	Activated	3.742	1.10E-13	*DDX58,FAM46A,HELZ2,HERC5,IFI44,IFI44L,IFIT3,IFITM3,ISG15,ISG20,OAS2,OAS3, OASL,ZC3HAV1*
DOCK8	other	Activated	3.5	1.13E-13	*CMPK2,CXCL10,HDAC1,HELZ2,IFIT2,IFIT3,IL15,IRF7,ISG15,ISG20,NMI,RSAD2,SOC S1,STAT1,STAT2,TRIM21*
TLR4	transmembran e receptor	Activated	3.517	1.12E-12	*CCL2,CMPK2,CXCL10,HDAC1,HELZ2,IFI16,IFIT2,IFIT3,IFITM3,IL15,IL2RB,IRF7,ISG 15,ISG20,MX1,NMI,OASL,RGCC,RSAD2,SOCS1,STAT1,STAT2,TNFRSF13B,TNFSF10,T RIM21,TRIM38*
IFNA1/IFN A13	cytokine	Activated	2.736	2.32E-12	*EIF2AK2,IFI6,IFIT2,ISG15,MX1,OAS1,OAS2,RSAD2*
FADD	other	Activated	2.851	2.90E-12	*CXCL10,DDX58,DHX58,EIF2AK2,GADD45B,IFIH1,IFIT2,IRF7,LY6E,MYC,PSMB8,SOC S1,STAT1,STAT2,TRAFD1*
NFATC2	transcription regulator	Activated	3.638	6.41E-12	*CMPK2,CXCL10,HDAC1,HELZ2,IFIT2,IFIT3,IL15,IRF7,ISG15,ISG20,MYC,NMI,OASL,R SAD2,SOCS1,STAT1,STAT2*
IRF1	transcription regulator	Activated	2.569	7.88E-12	*CASP1,CCL2,CEACAM1,CXCL10,IFI44L,IL15,MYC,PSMB8,PSMB9,PSME1,PSME2,SO CS1,TAP1,TAP2,TNFSF10,TRIM22*
IFNAR2	transmembran e receptor	Activated	2.646	8.45E-11	*HERC5,ISG15,OAS1,PSMB8,PSMB9,PSME2,UBE2L6,USP18*
TNF	cytokine	Activated	4.241	6.62E-10	*ATP1A1,BST2,CASP1,CCL2,CCR1,CXCL10,DUSP5,ETS1,GBP1,GCH1,GNB4,HK3,HLA-DRB4,IFI16,IFIT3,IL15,IL1RN,ISG15,JAK1,KYNU,LAMP3,MYC,OAS1,OAS2,OASL,PLSC R1,PSMB8,PSMB9,PSME2,RNASE2,RNF31,RPS3,SAT1,SCO2,SLC7A1,SOCS1,STAT1,TA P1,TDRD7,TNFSF10,TNFSF13B,TYMP,UBE2H*
TICAM1	other	Activated	3.921	1.10E-09	*CCL2,CMPK2,CXCL10,DDX58,IFI16,IFIT1,IFIT2,IFIT3,IL15,IRF7,ISG15,ISG20,OASL,R SAD2,SOCS1,TNFSF10*
TNK1	kinase	Activated	2.646	1.32E-08	*IFI16,IFIH1,IFIT2,IRF7,ISG20,OAS2,TNFSF10*
IFN alpha/beta	group	Activated	3.132	1.17E-07	*CCL2,IFI16,IFIT3,LY6E,RSAD2,SOCS1,STAT1,STAT2,TNFSF10,TRIM21*
FZD9	g-protein coupled receptor	Activated	2.433	4.45E-07	*CXCL10,IFI16,IFI44,IRF7,ISG15,STAT1*
IFN Beta	group	Activated	2	1.73E-06	*IFI6,IFIT1,IRF9,MX1,OAS2,SOCS1,STAT1,USP18*
MAPKAP1	other	Activated	2	8.15E-06	*IFI16,IFIT2,OAS2,PHF11*
MYD88	other	Activated	2.845	1.99E-05	*CCL2,CMPK2,CXCL10,IFIT2,IL15,IL1RN,IRF7,ISG15,OASL,RSAD2,SOCS1,TNFRSF13B ,TNFSF13B,USP18*
MAP2K6	kinase	Activated	2.224	2.83E-05	*CXCL10,IRF9,ISG15,STAT1,TNFSF10*
Map3k7	kinase	Activated	2.63	1.05E-04	*CMPK2,CXCL10,IFIT2,IL15,ISG15,ISG20,RSAD2*

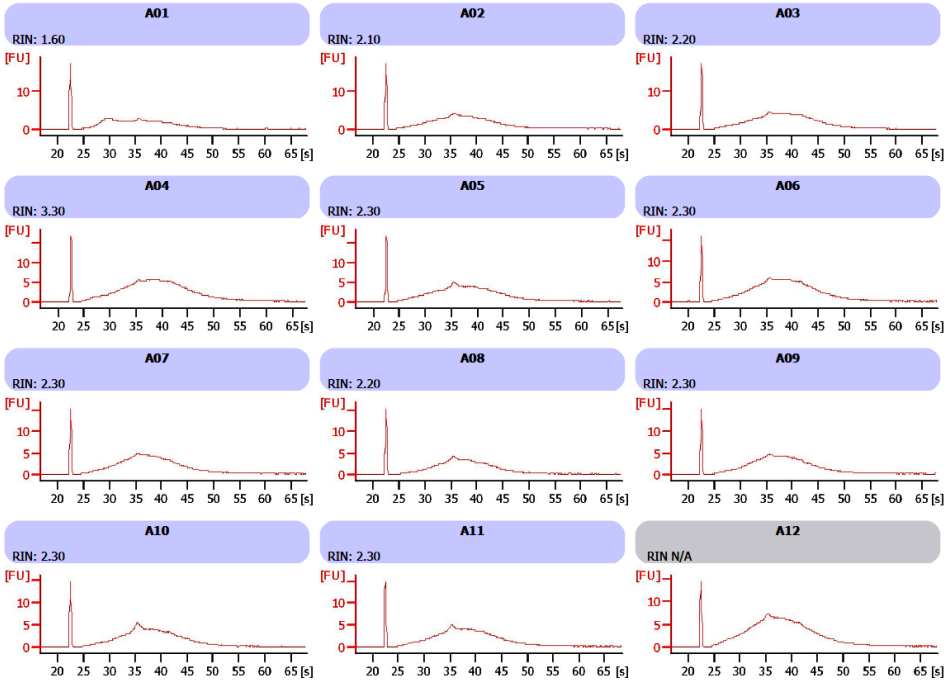
ARHGAP21	other	Activated	2.449	1.45E-04	*CXCL10,IFIT3,IL15,ISG20,NMI,TRIM21*
JAK1	kinase	Activated	2.219	1.59E-04	*CCL2,EIF2AK2,IFIT2,MX1,STAT1*
IFIH1	enzyme	Activated	2.2	1.59E-04	*CXCL10,IRF7,ISG15,OAS1,OAS2*
IL12B	cytokine	Activated	2.393	2.12E-04	*CXCL10,DUSP5,IRF9,SOCS1,STAT1,STAT2*
SMARCB1	transcription regulator	Activated	2.449	2.35E-04	*BAG3,BTG1,CEACAM1,EIF2AK2,IFITM1,MX1,OAS1,OAS3*
MAP2K3	kinase	Activated	2.213	3.05E-04	*CXCL10,IRF9,ISG15,STAT1,TNFSF10*
JUN	transcription regulator	Activated	2.184	4.50E-04	*ASNS,C3AR1,CCL2,CXCL10,IFI16,IGF2R,MBP,MT2A,MYC,RNF7,SNRPN,STAT1*
TMEM173	other	Activated	2.236	4.70E-04	*GBP5,IFI16,ISG15,OAS1,OASL*
IL12A	cytokine	Activated	2.183	6.94E-04	*CXCL10,IRF9,SOCS1,STAT1,STAT2*
PLK4	kinase	Activated	2	1.14E-03	*CMPK2,CXCL10,IFIT2,IL15*
IKBKB	kinase	Activated	3.124	1.68E-03	*CCL2,CCR1,CXCL10,ETS1,GCH1,IFI16,IL1RN,ISG15,MBP,MYC,SOCS1*
IL27	cytokine	Activated	2.412	2.89E-03	*CCL2,CXCL10,IL1RN,SOCS1,STAT1,TAP1*
CD14	transmembrane receptor	Activated	2	3.67E-03	*CCL2,CXCL10,IFIT1,SOCS1*
P38 MAPK	group	Activated	2.071	3.99E-03	*BATF2,CCL2,CXCL10,ETS1,GBP1,IRF7,MYC,STAT1,TNFSF10,ULK1*
NFkB (complex)	complex	Activated	2.775	4.49E-03	*CCL2,CXCL10,DUSP5,IL15,IRF7,MYC,NOTCH1,PSMB9,RSAD2,SOCS1,TAP1,TAP2,TNFSF10,TRIM38*
CD40LG	cytokine	Activated	2	1.27E-02	*ANXA7,CCL2,GCH1,HLA-DRB4,LAMP3,STAT1*
IL1B	cytokine	Activated	2.629	1.96E-02	*AKR1B1,ATP1A1,CCL2,CXCL10,FOXO1,GCH1,IL15,IL1RN,ISG15,MT2A,NMI,PLSCR1,TNFAIP6*
IKBKG	kinase	Activated	2.417	2.17E-02	*CCR1,GCH1,IFI16,IL1RN,IRF7,ISG15*
BTNL2	transmembrane receptor	Activated	2	5.30E-02	*GBP4,IFITM3,NCOA7,TNFSF10*
MAPK1	kinase	Inhibited	-7.59	1.81E-46	*ADAR,BST2,CDK19,DDX58,EIF2AK2,GBP1,GBP5,GLS,HERC5,IFI16,IFI27,IFI35,IFI44,IFI6,IFIH1,IFIT1,IFIT2,IFIT3,IFIT5,IFITM1,IFITM3,IGFBP7,IRF7,IRF9,ISG15,ISG20,LAMP3,LAP3,LGALS1,LGALS3BP,MX2,NMI,OAS1,OAS2,OAS3,OASL,PARP12,PHF11,PLSCR1,PSMB8,PSMB9,PSME2,SMARCC1,SP100,SP110,STAT1,STAT2,SUN2,TAP1,TDRD7,TNFSF10,TRIM14,TRIM21,TRIM22,TRIM38,TRIM5,UBE2L6,USP18,ZC3HAV1*
TRIM24	transcription regulator	Inhibited	-5.874	2.04E-37	*CMPK2,CXCL10,DDX58,DDX60,DHX58,EPSTI1,GBP4,HERC6,IFI35,IFI44,IFIH1,IFIT2,IFIT3,IRF7,IRF9,ISG15,LGALS3BP,MOV10,NMI,OAS1,OASL,PARP12,PHF11,PLAC8,P*

					RKCQ,PSMB8,PSMB9,RTP4,SAMD9L,SHISA5,SOCS1,STAT1,STAT2,TAP1,TRAFD1,USP18
CNOT7	transcription regulator	Inhibited	-2.621	1.11E-31	*BST2,CMPK2,HERC6,IFI27,IFI35,IFI44L,IFI6,IFIT5,IFITM1,ISG15,LGALS3BP,OAS1,OAS2,OAS3,PARP12,PLSCR1,PSMB8,SP110,STAT1,TAP1,TAP2,UBE2L6*
IL1RN	cytokine	Inhibited	-5.385	6.14E-29	*DDX58,GBP1,HERC6,IFI27,IFI44,IFI44L,IFI6,IFIH1,IFIT3,IFIT5,IRF7,IRF9,ISG20,LAMP3,LGALS9,MX1,MX2,OAS1,OAS2,OAS3,OASL,RSAD2,RTP4,SAMD9,SP100,STAT2,TNFSF10,TRIM22,USP18*
ACKR2	g-protein coupled receptor	Inhibited	-4.583	3.51E-27	*ADAR,CXCL10,DDX58,DDX60,DHX58,EIF2AK2,IFI16,IFI44,IFIT2,IFIT3,IRF7,ISG15,ISG20,OAS1,OAS2,OAS3,OASL,RSAD2,STAT1,STAT2,USP18*
SOCS1	other	Inhibited	-3.674	3.61E-14	*CXCL10,DDX58,GBP5,IFI16,IFI44,IFIH1,IFIT1,IFIT2,IFIT3,IL2RB,IRF7,ISG15,ISG20,MX1,OAS1,OAS2,SOCS1,STAT1*
USP18	peptidase	Inhibited	-3.104	1.92E-13	*CXCL10,IFI6,IFITM3,IRF7,IRF9,ISG15,MX1,OAS1,SOCS1,TNFSF10*
PTGER4	g-protein coupled receptor	Inhibited	-4.013	9.02E-13	*CCL2,CMPK2,CXCL10,DDX58,GBP4,HERC6,IFI16,IFI35,IFIH1,IFIT2,IRF7,ISG20,LHFPL2,PARP14,RSAD2,RTP4,TNFSF10,TRIM21,USP18,XAF1*
DNASE2	enzyme	Inhibited	-2.606	3.66E-12	*CXCL10,DHX58,IFIT3,IRF7,ISG15,OAS1,OAS3,RSAD2,RTP4,USP18,ZBP1*
BTK	kinase	Inhibited	-3.101	5.55E-12	*CXCL10,ETS1,IFI35,IFI44L,IFIT1,IFIT3,IFITM1,IRF9,ISG15,ISG20,MX1,MX2,OAS2,OAS3,STAT1*
TAB1	enzyme	Inhibited	-2.985	1.76E-10	*CXCL10,GBP1,GCH1,IFIH1,IFIT1,IRF7,TNFSF10,TNFSF13B,XAF1*
GAPDH	enzyme	Inhibited	-2.534	2.24E-10	*CCL2,FCER1G,IFI6,IFIT2,IFITM1,OAS1,OAS2,OAS3,STAT1,UBE2L6*
Irgm1	other	Inhibited	-2.53	4.31E-09	*CXCL10,ID3,IFI16,IFIT2,IFIT3,IRF7,OAS2,OASL,RSAD2,USP18*
MAP3K7	kinase	Inhibited	-2.985	7.04E-09	*CXCL10,GBP1,GCH1,IFIH1,IFIT1,IRF7,TNFSF10,TNFSF13B,XAF1*
ISG15	other	Inhibited	-2.219	1.36E-08	*DDX58,IFI6,IFITM3,MX1,OAS1*
IRF4	transcription regulator	Inhibited	-2.377	1.52E-08	*CXCL10,GBP1,IL1RN,IRF7,IRF9,PHF11,PLSCR1,PRKCQ,STAT1,STAT2,TNFSF10,TNFSF13B*
mir-21	microrna	Inhibited	-2.892	9.25E-07	*CXCL10,DHX58,FCGR1A,GBP5,IFI16,ITK,MYD88,OAS2,OAS3,PSME2,STAT1,STAT2,TAP1*
SOCS3	phosphatase	Inhibited	-2.804	1.73E-06	*CXCL10,IFIT1,IFIT2,ISG20,MX1,OAS1,OAS2,SOCS1*
mir-146	microrna	Inhibited	-2.207	8.49E-06	*CCL2,CXCL10,MBP,MYD88,SOCS1*
IL10RA	transmembrane receptor	Inhibited	-3	1.34E-04	*BATF2,CD36,CYFIP2,FHL1,GBP5,HK3,IFI16,IL1RN,IRF7,PSMB8,PSMB9,RSAD2,STAT1,* *TAP1,TAP2,ZBP1*

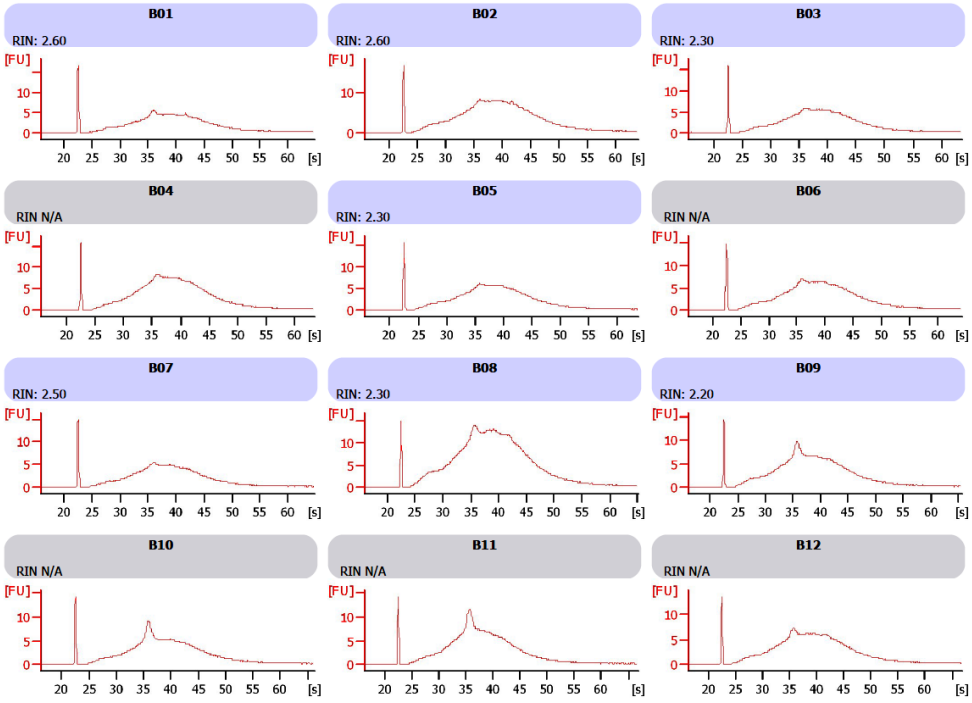
CD3	complex	Inhibited	-2.411	1.75E-03	*BTG1,CD247,ETS1,HSP90B1,IL2RB,IRF9,LAMP1,LEF1,STAT1,TNFSF10*
RPSA	translation regulator	Inhibited	-2	2.02E-03	*GBP4,ISG15,STAT1,TAP1*
SUMO3	other	Inhibited	-2	3.29E-03	*CEACAM1,ISG15,LAMP3,LY6E*
SUMO2	enzyme	Inhibited	-2	5.00E-03	*CEACAM1,ISG15,LAMP3,LY6E*
ERK1/2	group	Inhibited	-2.18	5.84E-03	*ARRB1,CCL2,FOXO1,IFIT1,MYC,PSMB8,PSMB9,TAP1,TAP2*
MKNK1	kinase	Inhibited	-2.236	3.64E-02	*BAG3,MYC,NELL2,PAFAH1B1,PTBP1*
ADORA2A	g-protein coupled receptor	Inhibited	-2.236	6.64E-02	*ALDOC,CXCL10,EEF2,IFITM3,NAPA*
CLDN7	other	Inhibited	-2	6.66E-02	*GLS,LYSMD2,MT1A,MT2A*
IGF1	growth factor	Inhibited	-2.207	1.63E-01	*MBP,PSMB8,PSMB9,TAP1,TAP2*

Supplementary Figures

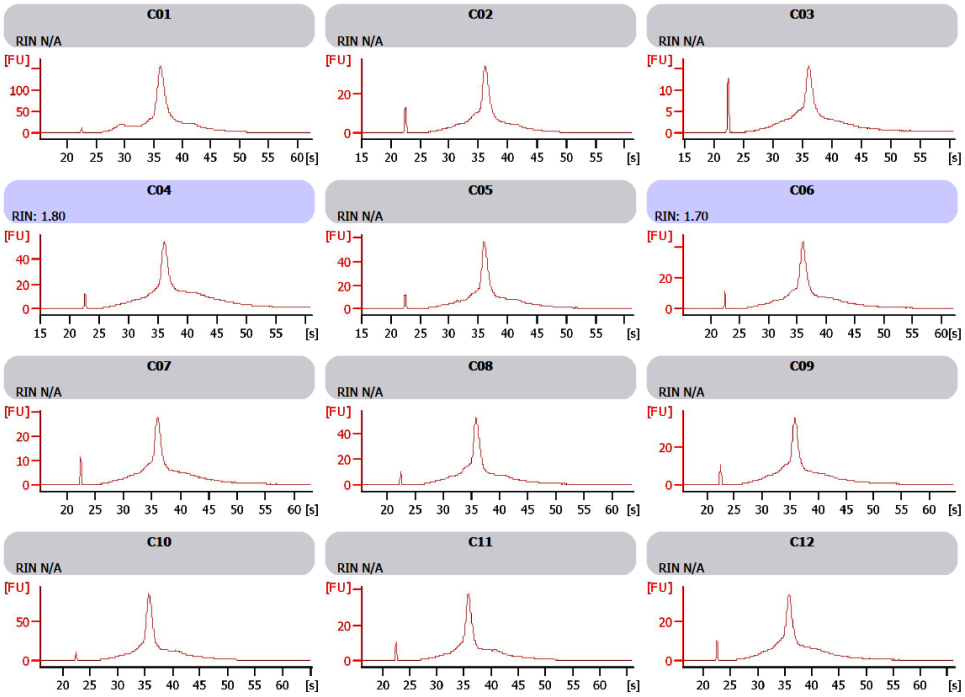
SF1 Electropherograms of the amplified cRNA of the first 12 G-depleted samples. *A01-A12 represent the samples which distributed randomly on the Bioanalyzer chip*



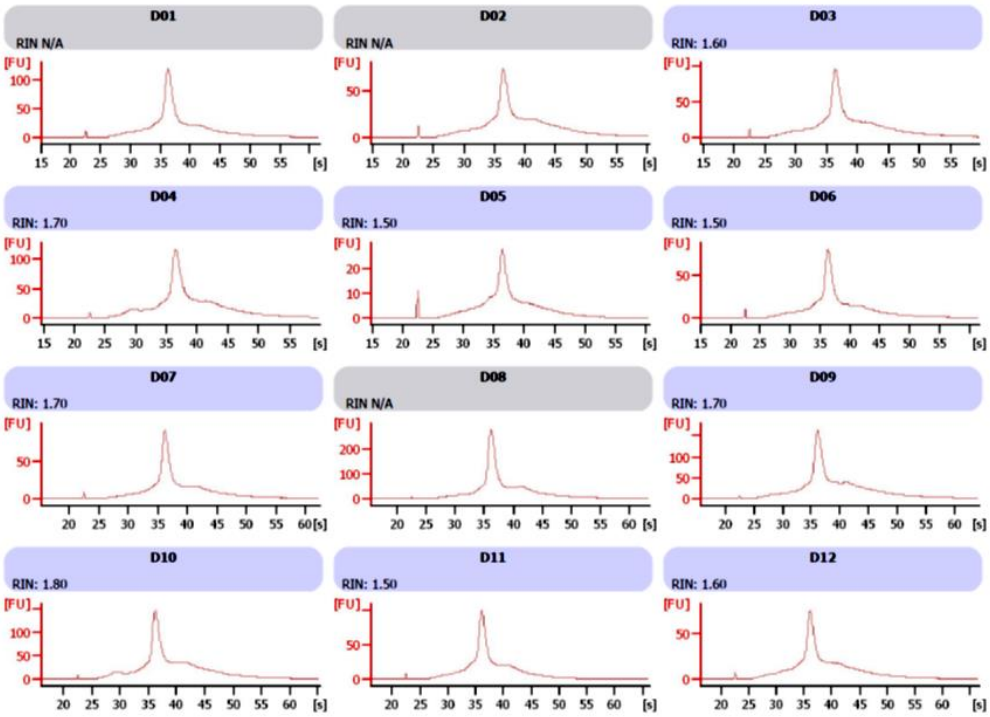
SF2 Electropherograms of the amplified cRNA of the second 12 G-depleted samples. *B01-B12 represent the samples which distributed randomly on the Bioanalyzer chip*



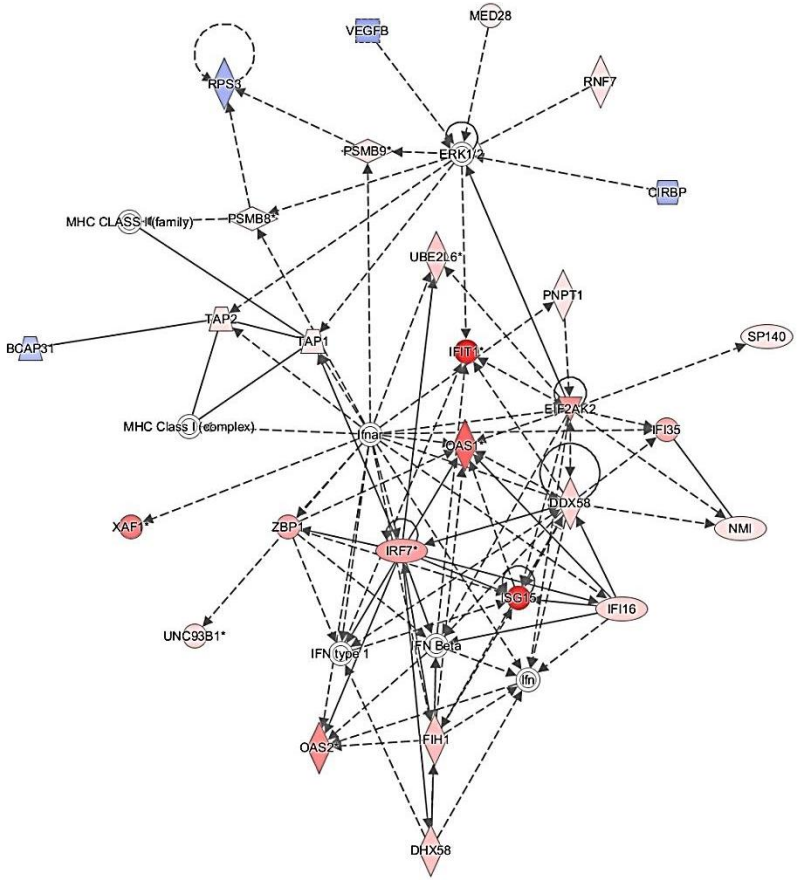
SF3 Electropherograms of the amplified cRNA of the first 12 G-non depleted samples. *C01-C12 represent the samples which distributed randomly on the Bioanalyzer chip. The sharp peaks represents the globin mRNA within the samples.*



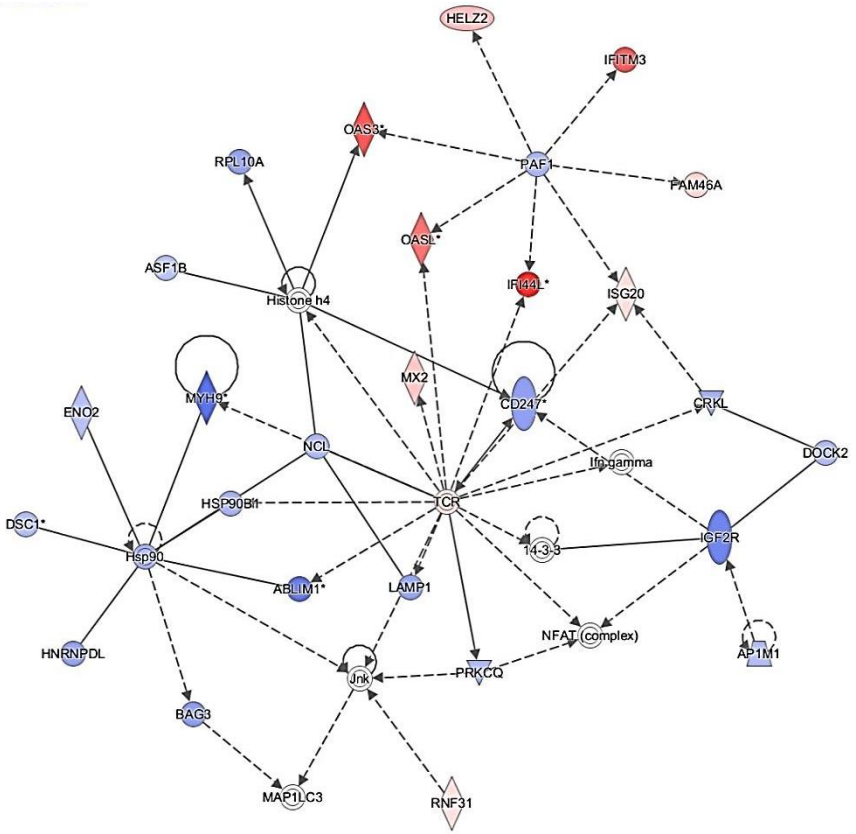
SF4 Electropherograms of the amplified cRNA of the second 12 G-non depleted samples. *D01-D12 represent the samples which distributed randomly on the Bioanalyzer chip. The sharp peaks represents the globin mRNA within the samples.*



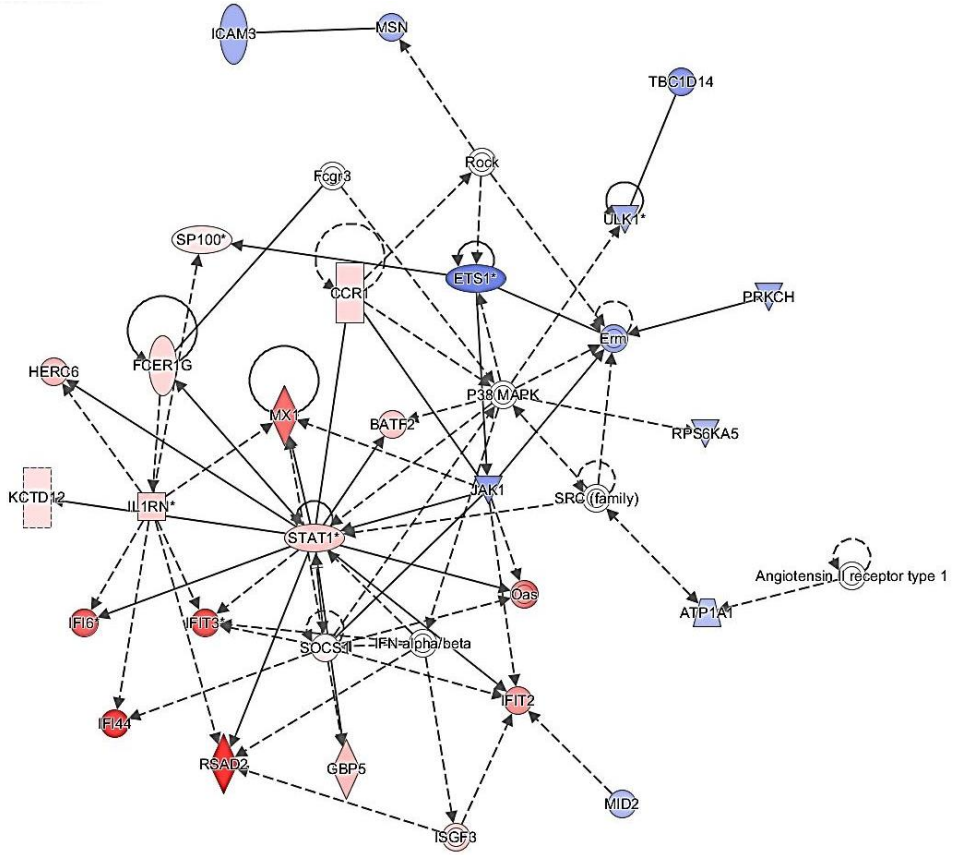
SF5 The molecular networks analysis in pSS-associated lymphoma vs healthy controls *Only 18 networks from the analysis are shown as 3 networks have only 2 genes*



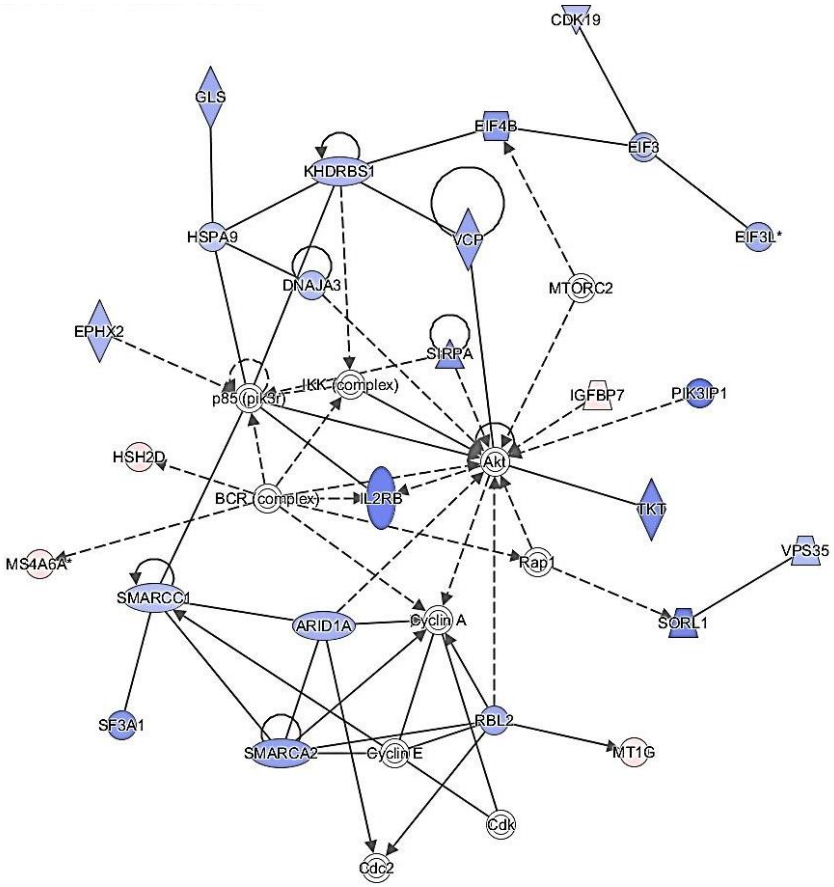
Network 2



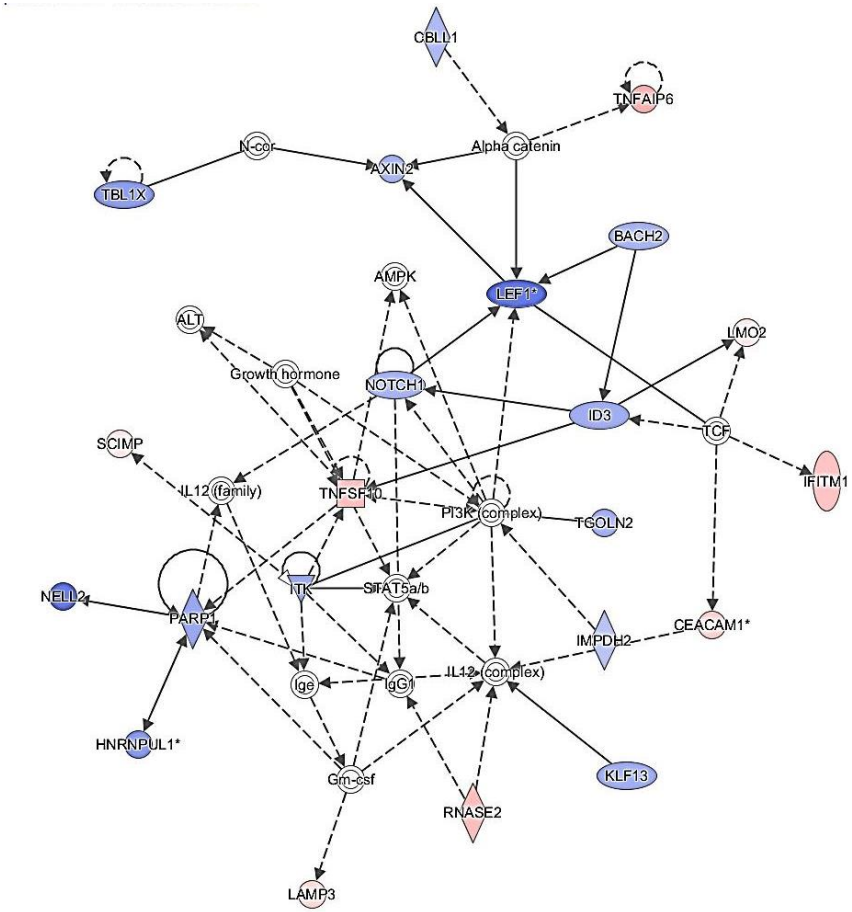
Network 3



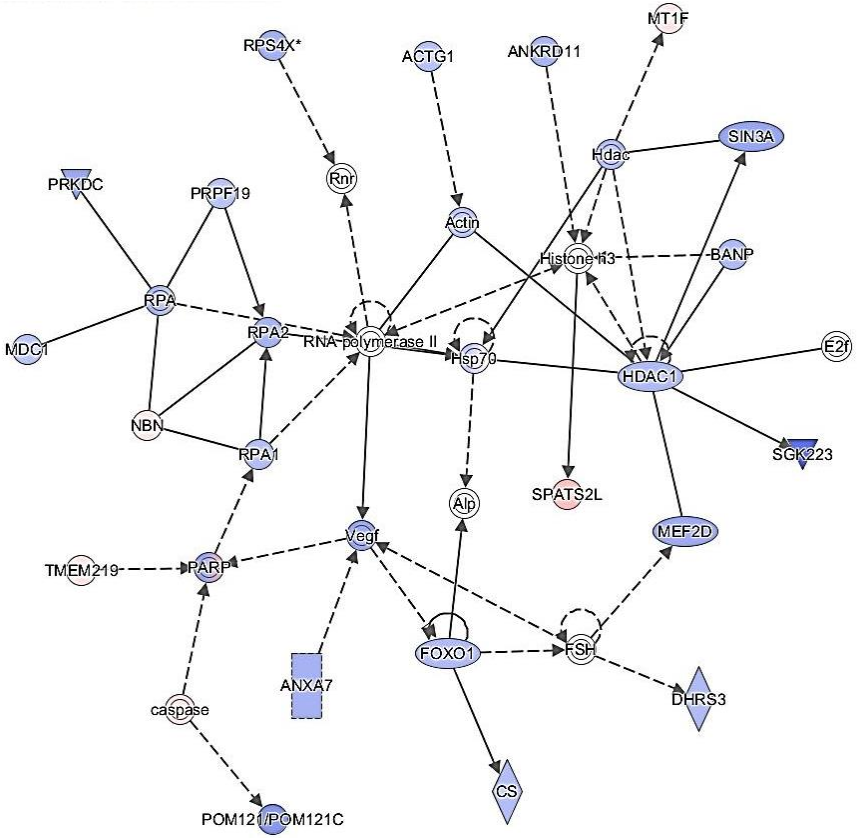
Network 4



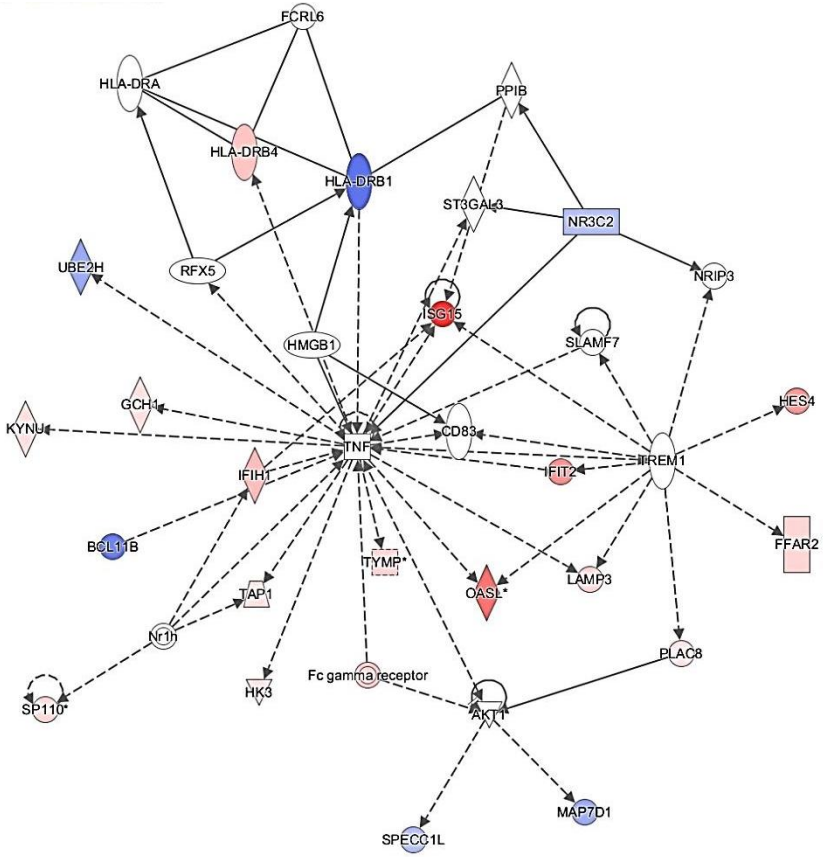
Network 6



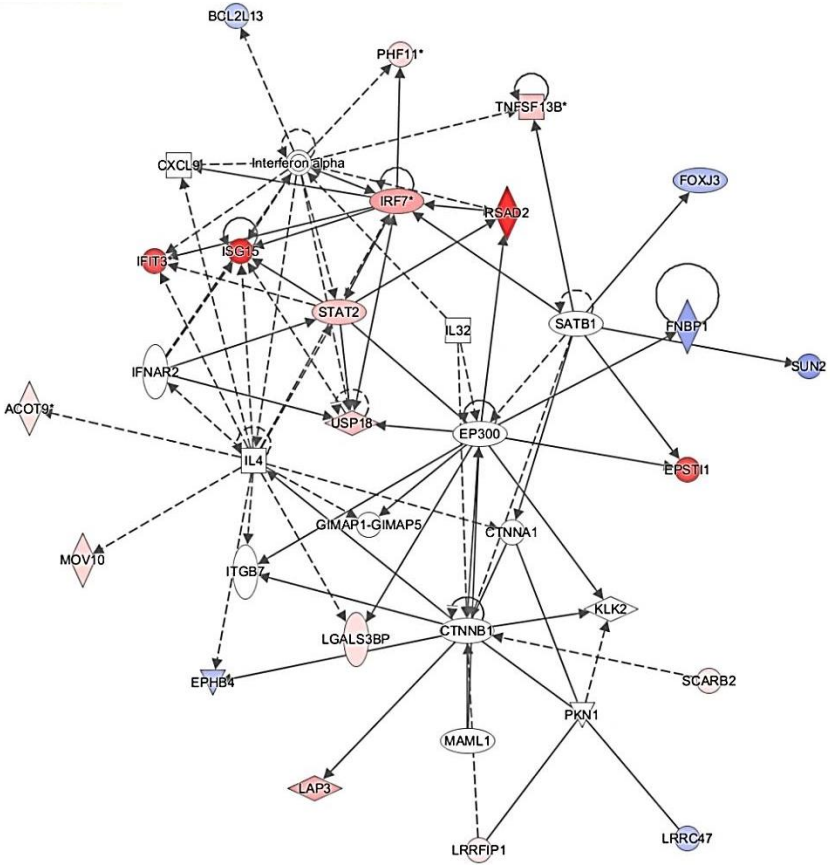
Network 7



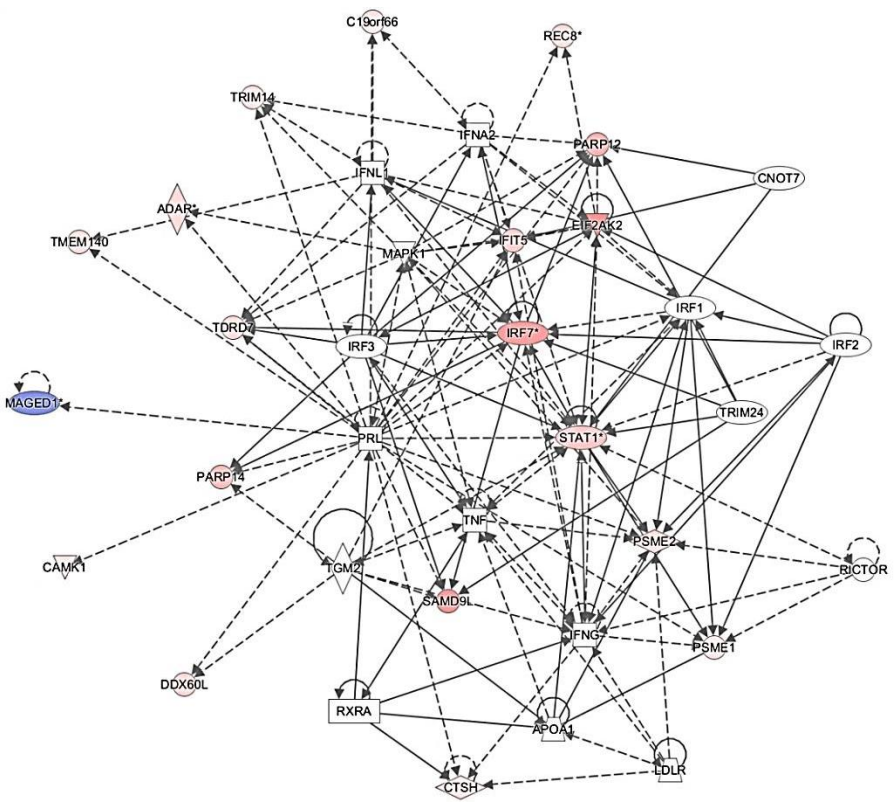
Network 8



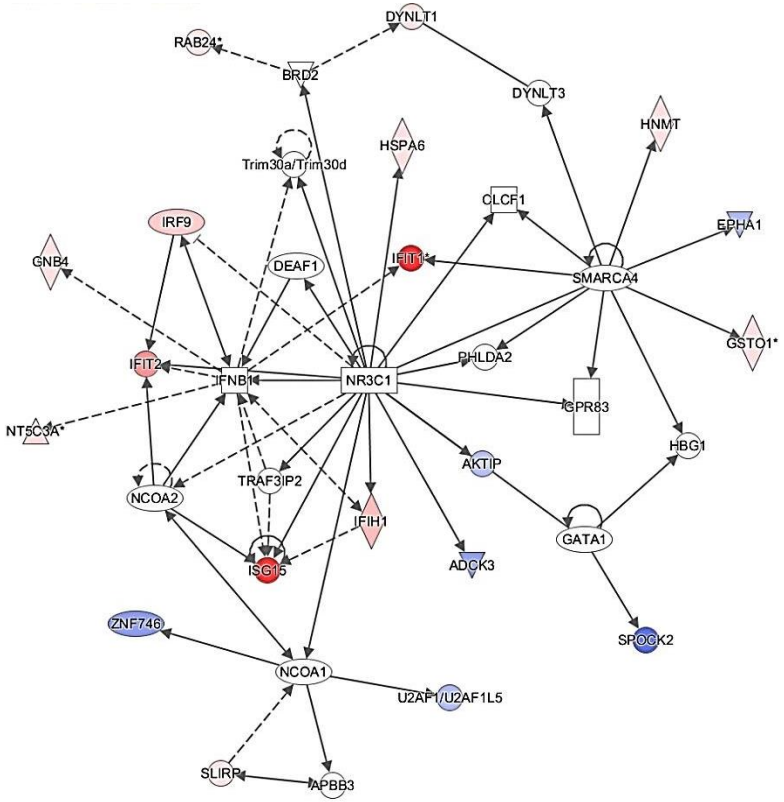
Network 9



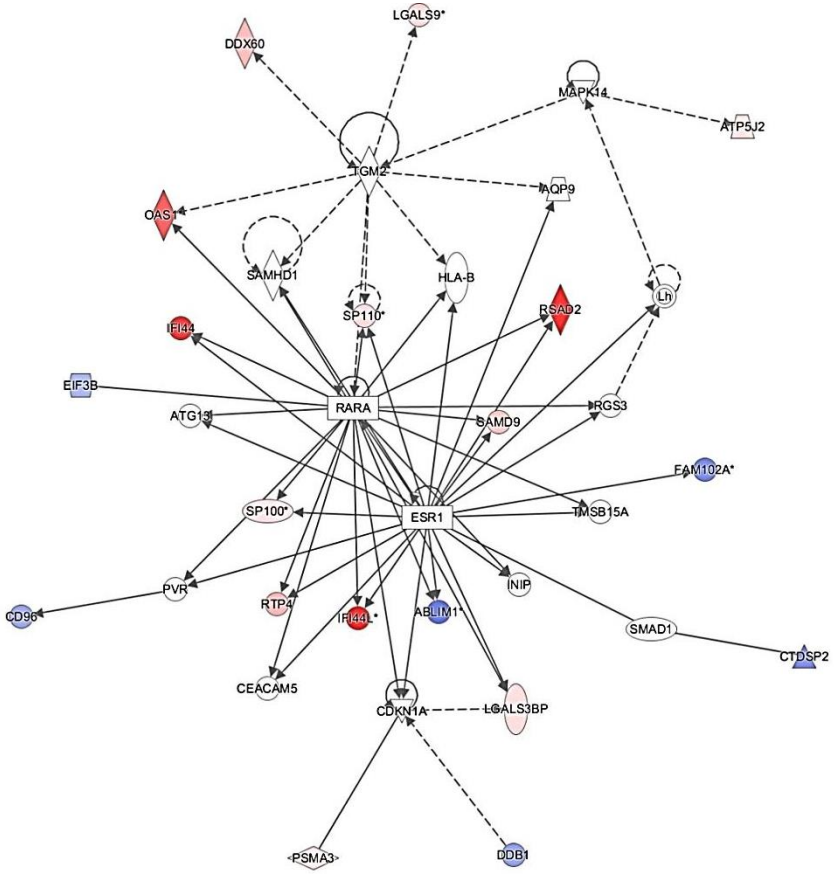
Network 10



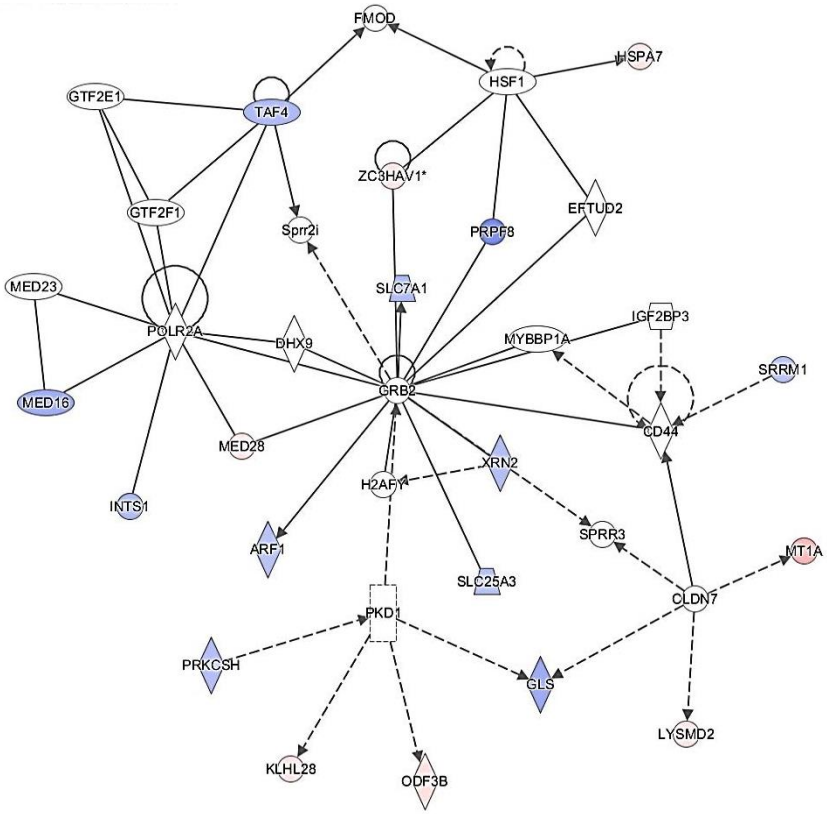
Network 11



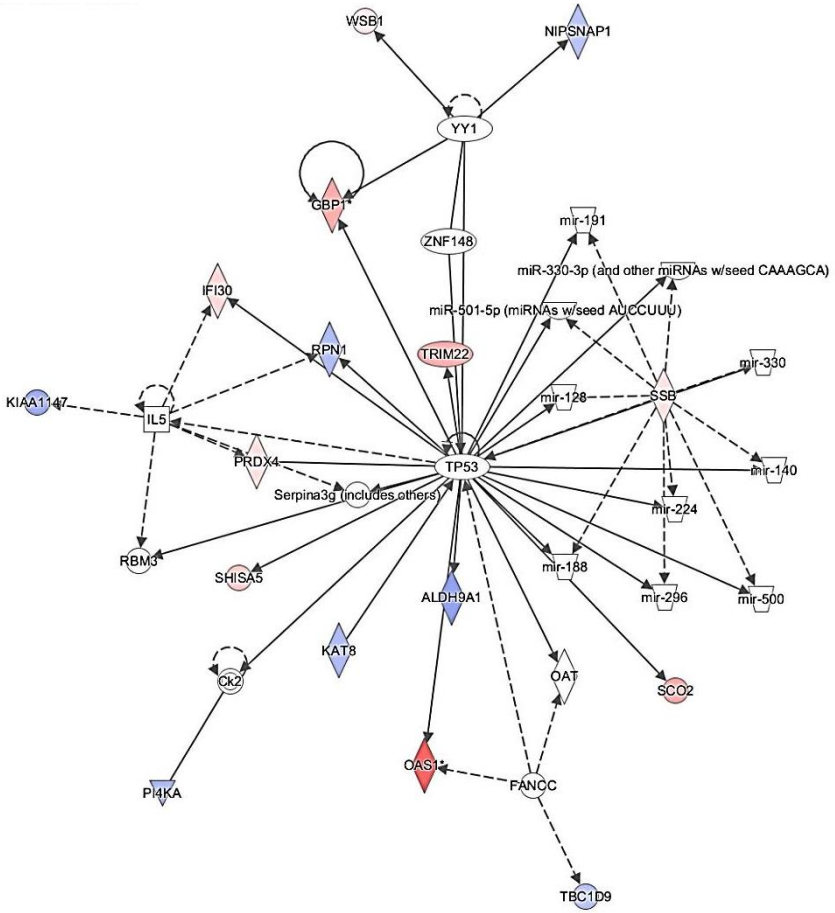
Network 12



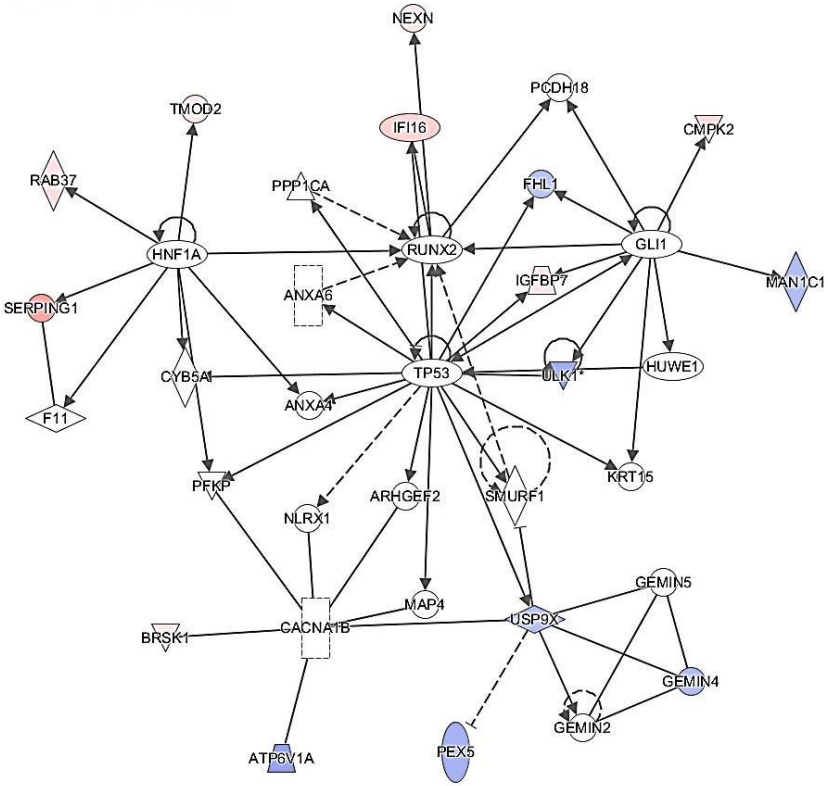
Network 13



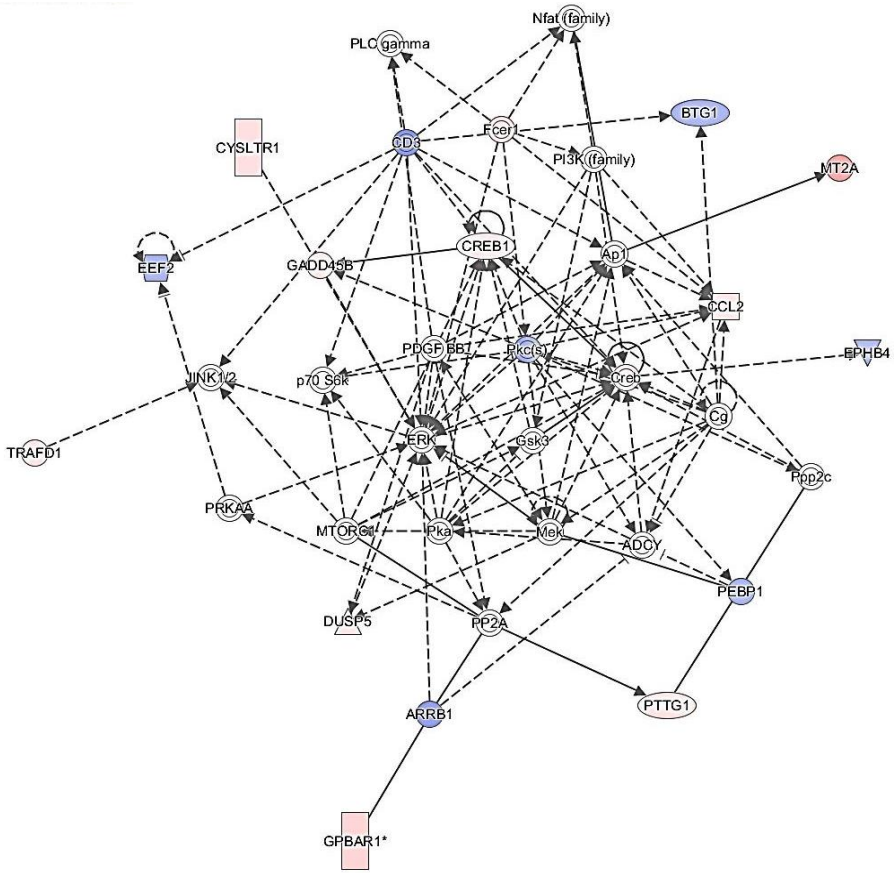
Network 14



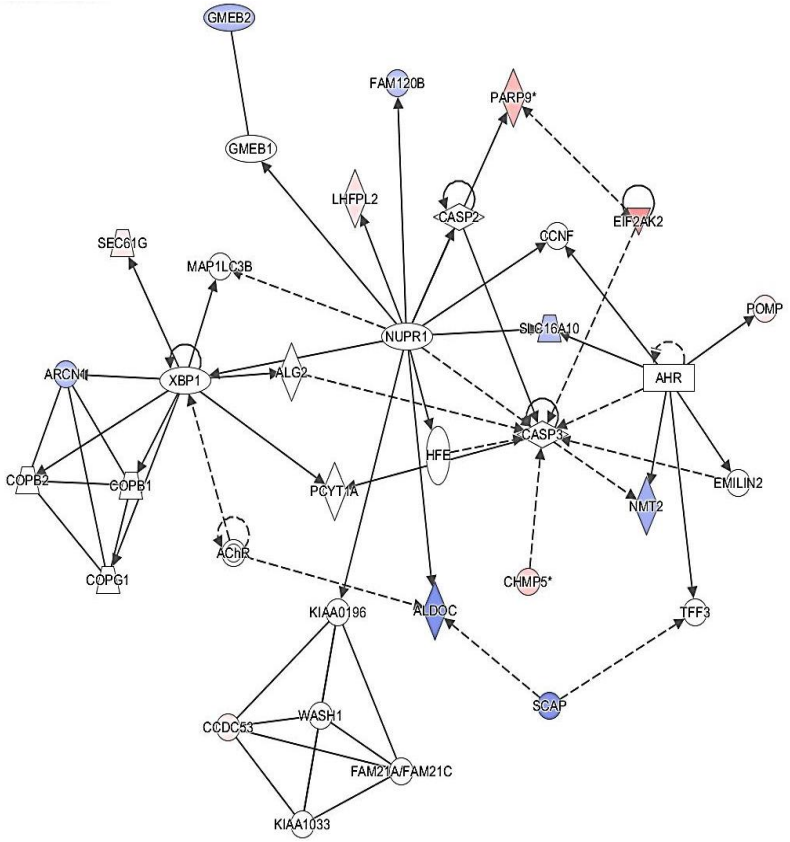
Network 15



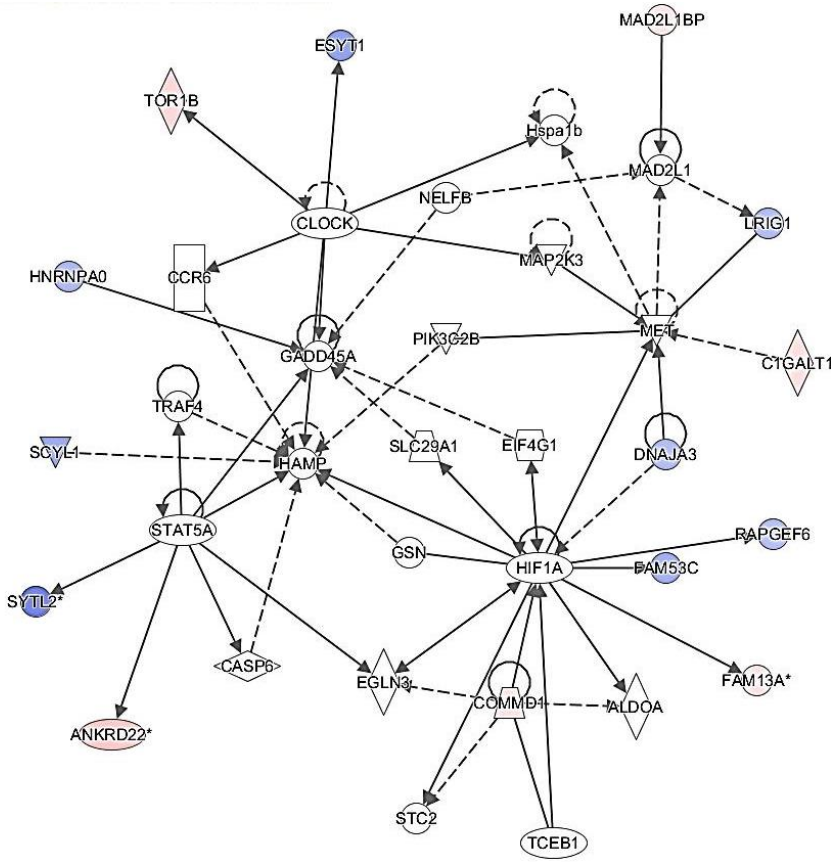
Network 17



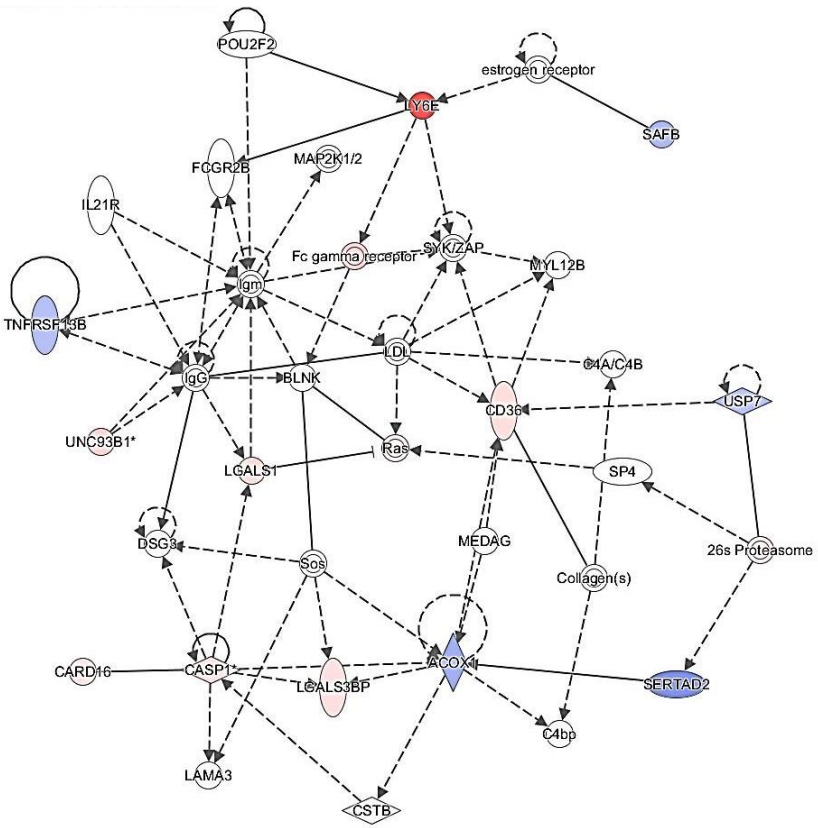
Network 18



Network 19



Network 20



Network 21